Block Permutation Principles for the Change Analysis of Dependent Data*

Claudia Kirch†

2007

Abstract

We study an AMOC-model with an abrupt change in the mean and dependent errors that form a linear process. Different kinds of statistics are considered, such as maximum-type statistics (particularly different CUSUM procedures) or sum-type statistics. Approximations of the critical values for change-point tests are obtained through permutation methods. The theoretical results show that the original test statistics and their corresponding block permutation counterparts follow the same distributional asymptotics. The main step in the proof is to obtain limit theorems for the corresponding rank statistics and then use laws of large numbers to obtain the permutation asymptotics conditionally on the given data.

Some simulation studies illustrate that the permutation tests usually behave better than the original tests if performance is measured by the α- and β-errors respectively.

Keywords: Permutation principle, change in mean, rank statistic, dependent observations, linear process

1. Introduction

A series of papers has been published on the use of permutation principles for obtaining reasonable approximations to the critical values of change-point tests. This approach was first suggested by Antoch and Hušková [1] and later pursued by other authors (cf. Hušková [9] for a recent survey). So far, it has mostly been dealt with independent observations. In many situations dependent observations are much more realistic. Kirch and Steinebach [12] considered change-point tests for possibly dependent processes under strong invariance. In that situation we were close enough to the independent case for the usual procedure to work.

---

*This is a preprint of an article published in *J. Statist. Plann. Inference*,137:2453-2474, 2007, Copyright 2007 by Elsevier.

†Technical University Kaiserslautern, Departement of Mathematics, Erwin-Schrödinger-Str., 67663 Kaiserslautern, ckirch@mathematik.uni-kl.de

1. Introduction

In this paper we examine a block permutation method (to take account of the dependency) for the most commonly used test statistics for an abrupt change in the mean (AMOC location model). A discussion of the behavior of these test statistics under the null hypothesis for dependent observations can be found in Antoch et al. [2] as well as Horváth [8].

We consider the following AMOC location model

$$X(i) = \mu + d\,1_{\{i>m^*\}} + e(i), \quad 1 \leqslant i \leqslant n, \tag{1.1}$$

where the errors $\{e(i), 1 \leqslant i \leqslant n\}$ are given by the linear process

$$e(i) = \sum_{j \geqslant 0} w_j\, \epsilon(i-j)$$

and $m^* = m^*(n)$, $d = d(n)$ may depend on n. We are interested in testing the null hypothesis of "no change"

$$H_0: \quad m^* = n$$

against the alternative of a change in the mean

$$H_1: \quad 1 \leqslant m^* < n \text{ and } d \neq 0.$$

Moreover we assume that the innovations $\{\epsilon(i) : -\infty < i < \infty\}$ are i.i.d. random variables with

$$\mathrm{E}\,\epsilon(i) = 0, \quad 0 < \sigma^2 = \mathrm{E}\,\epsilon(i)^2 < \infty, \quad \mathrm{E}\,|\epsilon(i)|^\nu < \infty \text{ for some } \nu > 2. \tag{1.2}$$

We suppose that the weights $\{w_j : j \geqslant 0\}$ satisfy

$$\sum_{j \geqslant 0} w_j \neq 0, \quad \sum_{j \geqslant 0} \sqrt{j}\,|w_j| < \infty. \tag{1.3}$$

In the case where the errors $\{e(i) : 1 \leqslant i \leqslant n\}$ are a sequence of i.i.d. random variables (with mean zero and finite variance) one often uses statistics based on the partial sums $S_m = \sum_{i=1}^m (X_i - \overline{X}_n)$ where $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.
It turns out that these statistics work also very well in the present setup (confer Antoch et al. [2] and Horváth [8]).
Typical test statistics are

$$T_n^{(1)} = \max_{1 \leqslant m < n} \left(\sqrt{\frac{n}{m(n-m)}}\, |S_m| \right),$$

$$T_n^{(2)}(G) = \max_{G < m \leqslant n} \frac{1}{\sqrt{G}}\, |S_m - S_{m-G}|,$$

$$T_n^{(3)}(q) = \max_{1 \leqslant m < n} \left(\frac{1}{\sqrt{n}\, q(\frac{m}{n})}\, |S_m| \right),$$

$$T_n^{(4)}(r) = \frac{1}{n} \sum_{m=1}^{n-1} \frac{1}{r(m/n)} \left(\frac{1}{\sqrt{n}} S_m \right)^2,$$

where $q(\cdot)$ and $r(\cdot)$ are weight functions defined on $(0,1)$ and $G < n$.
We assume that the weight function q belongs to the class

$$Q_{0,1} = \{q \; : \; q \text{ is non-decreasing in a neighborhood of zero, non-increasing in a neighborhood of one and } \inf_{\eta \leqslant t \leqslant 1-\eta} q(t) > 0 \text{ for all } 0 < \eta < 1/2\}.$$

The following integral plays a crucial role for the convergence of statistics based on a weight function q

$$I^*(q,c) = \int_0^1 \frac{1}{t(1-t)} \exp\left\{\frac{-cq^2(t)}{t(1-t)}\right\} dt.$$

For details and further references confer Csörgő and Horváth [4], Chapter 4.
We assume that the weight function r fulfills for all $x \in (0,1)$

$$r(x) > 0 \quad \text{and} \quad \int_0^1 \frac{t(1-t)}{r(t)}\, dt < \infty. \tag{1.4}$$

For more details and further references confer Csörgő and Horváth [5], Chapter 2.

The following theorem gives the asymptotic distribution under H_0 for the above statistics.

Theorem 1.1. *Assume that (1.1) - (1.3) and H_0 holds. Let $\alpha(x) = \sqrt{2\log x}$ and $\beta(x) = 2\log x + \frac{1}{2}\log\log x - \frac{1}{2}\log \pi$.*

a) Then we have for all $x \in \mathbb{R}$

$$P\left(\alpha(\log n)\frac{T_n^{(1)}}{\widehat{\tau}} - \beta(\log n) \leqslant x\right) \longrightarrow \exp(-2e^{-x}) \qquad \text{as } n \to \infty,$$

where $\widehat{\tau} - \tau = o_P((\log\log n)^{-1})$, $\tau^2 := \sigma^2 \left(\sum_{s\geqslant 0} w_s\right)^2 (> 0)$.

b) If $G = G(n) \to \infty$, $\frac{G}{n} \to 0$ and $G^{-1}n^{2/\nu}\log n \to 0$ as $n \to \infty$, then we have for all $x \in \mathbb{R}$

$$P\left(\alpha(n/G)\frac{T_n^{(2)}(G)}{\widehat{\tau}} - \beta(n/G) \leqslant x\right) \longrightarrow \exp(-2e^{-x}) \quad \text{as } n \to \infty,$$

where $\widehat{\tau} - \tau = o_P(\log(n/G)^{-1})$.

c) If $q \in Q_{0,1}$ and $I^*(q,c) < \infty$ for some $c > 0$, then

$$\frac{1}{\widehat{\tau}}\, T_n^{(3)}(q) \xrightarrow{\mathcal{D}} \sup_{0<t<1} \frac{|B(t)|}{q(t)} \qquad \text{as } n \to \infty,$$

where $\{B(t) : 0 \leqslant t \leqslant 1\}$ is a Brownian bridge and $\widehat{\tau} - \tau = o_P(1)$.

d) If r fulfills condition (1.4), then

$$\frac{1}{\widehat{\tau}^2}\, T_n^{(4)}(r) \xrightarrow{\mathcal{D}} \int_0^1 \frac{B^2(t)}{r(t)}\, dt \qquad \text{as } n \to \infty,$$

where $\{B(t) : 0 \leqslant t \leqslant 1\}$ is a Brownian bridge and $\widehat{\tau} - \tau = o_P(1)$.

Proof. Confer Theorem 2.1 in Antoch et al. [2], there (1.3) is replaced by $\sum_{j\geqslant 0} j\,|w_j| < \infty$. Yet this assumption is only needed to obtain the Beveridge-Nelson decomposition, which also holds under (1.3), confer Phillips and Solo [14], equation (15) et sqq. ■

Remark 1.1. Antoch et al. [2], Theorem 2.3 and Remark 3, propose the following Bartlett estimator:

$$\widetilde{\tau}_n^2(\Lambda) = \widehat{R}(0) + 2\sum_{k=1}^{\Lambda}\left(1 - \frac{k}{\Lambda}\right)\widehat{R}(k), \tag{1.5}$$

where

$$\widehat{R}(k) = \frac{1}{n}\left(\sum_{t=1}^{\widehat{m}-k}(X_t - \overline{X}_{\widehat{m}})(X_{t+k} - \overline{X}_{\widehat{m}}) + \sum_{t=\widehat{m}+1}^{n-k}(X_t - \overline{X}^*_{\widehat{m}})(X_{t+k} - \overline{X}^*_{\widehat{m}})\right),$$

$\widehat{m} = \min\{\arg\max\{|S_k| : k = 1, \ldots, n\}\}$ and $\overline{X}^*_{\widehat{m}} = \frac{1}{n-\widehat{m}}\sum_{i=\widehat{m}+1}^{n} X_i$. Λ should be chosen such that $\Lambda^2/\log(n) \to \infty$ and $n^{-1}\Lambda^2 \log\Lambda = o(\log(n))$, then the above theorem holds true with $\widehat{\tau} = \widetilde{\tau}_n(\Lambda)$.

Remark 1.2. Csörgő and Horváth [5], Theorem 4.1.2, prove that the asymptotic for the q-weighted CUSUM statistic remains true for somewhat more general error sequences. The resampling methods then also hold if the necessary strong laws of large numbers are fulfilled, i.e. one only has to prove equations (6.1) - (6.5).

The above theorem allows to construct asymptotic tests where the quantiles of the limit distributions are chosen as critical values. Yet, it is well known that the rate of convergence of these statistics (especially the extreme value statistics) can be very slow (cf. Berkes et al. [3]). Furthermore the limit distributions of $T_n^{(3)}(q)$ and $T_n^{(4)}(r)$ are explicitly known only for $q, r \equiv 1$.
So alternative methods to obtain critical values for the test are of interest. In this article we study a version of the permutation method by Hušková [9], which is based on blocks of the observations.

We assume that we split the observation sequence of length n into L sequences of length K (i.e. $n = KL$). K and L depend on n and converge to infinity with n. Instead of permuting the observations $X(i)$, we permute the blocks $X(Kl+1), \ldots, X(K(l+1))$, $l = 0, \ldots, L-1$, and compute the statistics using the permuted blocks (there are no changes in the order of $X(\cdot)$ within the blocks).

The idea is that the block contains enough information about the dependency structure so that the estimate is close to the null hypothesis.

We assume in the following that $L \to \infty$ and $K = K(L)$, $n = n(L) = KL$, $K/L = O(1)$.

Remark 1.3. It is also possible to look at $n = K(L-1) + K^*$, $0 < K^* \leqslant K$. Then we still have L blocks altogether, but only $L-1$ are of length K and one is of length smaller or equal to K. The proofs remain the same, yet one always has to take care of the shorter block, which makes notations much more complicated.

The paper is organized as follows: In Section 2 we give the rank asymptotics of the statistics of interest. This is needed to obtain the main results of Section 3, which show the validity of the above approach. Results of a simulation study are presented in Section 4, while the proofs are given in Sections 5 respectively 6. An appendix summarizes some results for strong mixing sequences.

2. Rank Asymptotics

In this section we prove the asymptotics for the corresponding rank statistics. We can then deduce the validity of the block resampling methods in the next section by choosing multiples of $X(i)$ as scores and proving that they fulfill almost surely conditions (2.1) and (2.2).

Let $\boldsymbol{\pi} = (\pi(1), \dots, \pi(L))$ be a random permutation of $(1, \dots, L)$ chosen with probability $\frac{1}{L!}$ and $a_n(1), \dots, a_n(n)$ scores satisfying

$$\frac{1}{L}\sum_{l=0}^{L-1} \max_{k=0,\dots,K-1} \left| \frac{1}{\sqrt{K}} \sum_{j=k+1}^{K} (a_n(Kl+j) - \overline{a}_n) \right|^{\nu} \leqslant D_1 \tag{2.1}$$

for some $\nu > 2$ and

$$\tau_n^2(\mathbf{a}) := \frac{1}{L}\sum_{l=0}^{L-1} \left[\frac{1}{\sqrt{K}} \sum_{k=1}^{K} (a_n(Kl+k) - \overline{a}_n) \right]^2 \geqslant D_2, \tag{2.2}$$

where $\overline{a}_n := \frac{1}{n}\sum_{i=1}^{n} a_n(i)$ and $D_1, D_2 > 0$ are some constants. The rank statistics of our interest are based on partial sums

$$S_{L,K}^{\mathbf{a},\boldsymbol{\pi}}(l,k) := \sum_{i=1}^{l-1}\sum_{j=1}^{K}(a_n[K(\pi(i)-1)+j] - \overline{a}_n) + \sum_{j=1}^{k}(a_n[K(\pi(l)-1)+j] - \overline{a}_n).$$

Precisely we are interested in:

$$T_{L,K}^{(1)}(\boldsymbol{\pi},\mathbf{a}) := \max_{2\leqslant l\leqslant L-1} \max_{1\leqslant k\leqslant K} \sqrt{\frac{LK}{(K(l-1)+k)(LK - K(l-1)-k)}} \left| S_{L,K}^{\mathbf{a},\boldsymbol{\pi}}(l,k) \right|,$$

$$T_{L,K}^{(2)}(G,\boldsymbol{\pi},\mathbf{a}) := \frac{1}{\sqrt{G}} \max_{\substack{1\leqslant l\leqslant L, 1\leqslant k\leqslant K\\ K(l-1)+k>G}} \left| S_{L,K}^{\mathbf{a},\boldsymbol{\pi}}(l,k) - S_{L,K}^{\mathbf{a},\boldsymbol{\pi}}(l_G,k_G) \right|,$$

$$T_{L,K}^{(3)}(q,\boldsymbol{\pi},\mathbf{a}) := \max_{\substack{1\leqslant l\leqslant L, 1\leqslant k\leqslant K\\ (l,k)\neq(L,K)}} \frac{1}{\sqrt{KL}\, q\left(\frac{K(l-1)+k}{KL}\right)} \left| S_{L,K}^{\mathbf{a},\boldsymbol{\pi}}(l,k) \right|,$$

$$T_{L,K}^{(4)}(r,\boldsymbol{\pi},\mathbf{a}) := \frac{1}{(KL)^2} \sum_{\substack{1\leqslant l\leqslant L, 1\leqslant k\leqslant K\\ (l,k)\neq(L,K)}} \frac{1}{r\left(\frac{K(l-1)+k}{KL}\right)} \left(S_{L,K}^{\mathbf{a},\boldsymbol{\pi}}(l,k) \right)^2,$$

where $K(l_G-1)+k_G = K(l-1)+k-G$, i.e. $l_G - 1 = \lfloor \frac{K(l-1)+k-G}{K} \rfloor$, $k_G = (K(l-1)+k-G) \mod K$.

The following theorem is the main tool to derive the validity of the block permutation method, yet it might also be of independent interest.

Theorem 2.1. *Let $\boldsymbol{\pi} = (\pi(1), \dots, \pi(L))$ be a random permutation of $(1, \dots, L)$ as above. Moreover let $a_n(1), \dots, a_n(n)$ be scores satisfying (2.1) and (2.2). $\alpha(x)$, $\beta(x)$ are as in Theorem 1.1.*

a) If $K = O((\log n)^\gamma)$ for some $\gamma > 0$, we have for all $x \in \mathbb{R}$

$$P\left(\alpha(\log n)\, \frac{T^{(1)}_{L,K}(\boldsymbol{\pi}, \mathbf{a})}{\tau_n(\mathbf{a})} - \beta(\log n) \leqslant x \right) \to \exp(-2e^{-x}) \quad \text{as } L \to \infty.$$

b) If, as $L \to \infty$, $G = G(n) \to \infty$, $G/n \to 0$, and

$$G^{-1} L^{-2\mu} n \log(n/G) = o(1) \qquad \text{for some} \quad 0 \leqslant \mu < \min\left(\frac{\kappa - 2}{2\kappa}, \frac{1}{4} \right), \tag{2.3}$$

then we have for all $x \in \mathbb{R}$

$$P\left(\alpha(n/G)\, \frac{T^{(2)}_{L,K}(G, \boldsymbol{\pi}, \mathbf{a})}{\tau_n(\mathbf{a})} - \beta(n/G) \leqslant x \right) \to \exp(-2e^{-x}) \quad \text{as } L \to \infty.$$

c) If $q \in Q_{0,1}$, $I^*(q,c) < \infty$ for some $c > 0$, and as $L \to \infty$

$$\frac{1}{L\, q^2\left(\frac{1}{KL}\right)} \to 0, \qquad \frac{1}{L\, q^2\left(1 - \frac{1}{KL}\right)} \to 0, \tag{2.4}$$

then

$$\frac{T^{(3)}_{L,K}(q, \boldsymbol{\pi}, \mathbf{a})}{\tau_n(\mathbf{a})} \xrightarrow{\mathcal{D}} \sup_{0<t<1} \frac{|B(t)|}{q(t)} \quad \text{as } L \to \infty,$$

where $\{B(t) : 0 \leqslant t \leqslant 1\}$ denotes a Brownian bridge.

d) If r fulfills condition (1.4) and as $L \to \infty$

$$\frac{1}{L^2 K} \sum_{k=1}^{K} \frac{1}{r\left(\frac{k}{KL}\right)} \to 0, \qquad \frac{1}{L^2 K} \sum_{k=1}^{K-1} \frac{1}{r\left(1 - \frac{k}{KL}\right)} \to 0, \tag{2.5}$$

then

$$\frac{T^{(4)}_{L,K}(r, \boldsymbol{\pi}, \mathbf{a})}{\tau_n^2(\mathbf{a})} \xrightarrow{\mathcal{D}} \int_0^1 \frac{B^2(t)}{r(t)}\, dt,$$

where $\{B(t) : 0 \leqslant t \leqslant 1\}$ denotes a Brownian bridge.

Remark 2.1. Concerning the weighted CUSUM-statistic $\widetilde{T}^{(1)}_{L,K}(\boldsymbol{\pi}, \mathbf{a})$ with the maximum over the complete range $1 \leqslant K(l-1) + k < n$ (instead of $K \leqslant K(l-1) + k \leqslant n - K$), the assertion remains true, if

$$\max_{1 \leqslant k \leqslant K} \left| \frac{1}{\sqrt{k}} \sum_{j=1}^{k} (a_n(K(\pi(1) - 1) + j) - \overline{a}_n) \right| = o_P\left(\sqrt{\log \log n} \right)$$

resp.

$$\max_{1 \leqslant k \leqslant K} \left| \frac{1}{\sqrt{k}} \sum_{j=K-k+1}^{K} (a_n(K(\pi(1) - 1) + j) - \overline{a}_n) \right| = o_P\left(\sqrt{\log \log n} \right).$$

3. Main Results: Block Permutation Method

In this section we prove that the block permutation method gives asymptotically correct critical values. Precisely we show that the quantiles from the permutation statistics given our observed data approximate the critical values corresponding to the null distribution. This holds not only when our observations follow the null hypothesis but even when they follow an alternative.
Main tool in the proof are the rank asymptotics developed in the previous section. Furthermore we need strong laws of large numbers for the blocks and even for the maximum of partial sums to prove that the conditions on the scores from the previous section are almost surely fulfilled. Such laws hold e.g. under $(\mathcal{A})$ below.

Let $\boldsymbol{\pi} = (\pi(1), \ldots, \pi(L))$ be a random permutation of $(1, \ldots, L)$ chosen with probability $\frac{1}{L!}$ independent of $X(\cdot)$. Then the permutation statistics $T^{(1)}_{L,K}(\boldsymbol{\pi}, X)$, $T^{(2)}_{L,K}(G, \boldsymbol{\pi}, X)$, $T^{(3)}_{L,K}(q, \boldsymbol{\pi}, X)$ respectively $T^{(4)}_{L,K}(r, \boldsymbol{\pi}, X)$ are given by $T^{(1)}_{L,K}(\boldsymbol{\pi}, \mathbf{a}), \ldots$, where the scores $a_n(i)$ are replaced by $X(i)$.

For the permutation result to hold true we standardize using the variance of the block rank statistic. Lemma 3.1 shows that this variance converges in probability to $\tau^2 := \sigma^2 \left(\sum w_j\right)^2$ under H_0, if we replace the scores by the observations. The convergence is even sufficiently fast for Theorem 1.1 under suitable conditions. That way we obtain the following estimator ($\widehat{\tau}_{LK} > 0$)

$$\widehat{\tau}^2_{LK} := \frac{1}{KL} \sum_{l=0}^{L-1} \left[\sum_{k=1}^{K} (X(Kl+k) - \overline{X}_n) \right]^2 . \tag{3.1}$$

Note that it does not depend on the permutations, thus the outcome of the permutation test is in fact independent of the actual value of that estimator. The estimator in Remark 1.1 has in general a different asymptotic behavior under alternatives and can thus not be used for the block permutation statistics.

Lemma 3.1. *Under (1.1) - (1.3) and* H_0

$$\frac{1}{KL} \sum_{l=0}^{L-1} \left[\sum_{k=1}^{K} (X(Kl+k) - \overline{X}_n) \right]^2 = \sigma^2 \left(\sum_{j \geqslant 0} w_j \right)^2 + O_P\left(\sqrt{\frac{1}{K}} + \sqrt{\frac{1}{L}} + \frac{\log\log n}{L} + n^{-\frac{\mu-2}{\mu}} \right),$$

where $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X(i)$ *and* $\mu < \max(\nu, 4)$ *with* $\mathrm{E}\,|X(i)|^\nu < \infty$, $i = 1, 2, \ldots$.

Now we prove that the above statistics conditioned on the observations have the same asymptotic behavior as the statistics under the null hypothesis (cf. Theorem 1.1). It does not matter whether our observations follow the null hypothesis or an alternative. This is true under the following assumption on the error sequence $\{e(i) : i \geqslant 1\}$ for certain δ, Δ.

$(\mathcal{A})$ Let $\{Z_{i,n} : 1 \leqslant i \leqslant n, n \geqslant 1\}$ be a strictly stationary sequence with $\mathrm{E}\, Z_{i,n} = 0$, $i \in \mathbb{Z}$. Assume there are $\delta, \Delta > 0$ with

$$\mathrm{E}\,|Z_{i,n}|^{2+\delta+\Delta} \leqslant D_1 \quad \text{for all } 1 \leqslant i \leqslant n, n \in \mathbb{N}$$

and there is a sequence $\alpha(k)$ with $\alpha_n(k) \leqslant \alpha(k)$, $k \in \mathbb{N}$, and

$$\sum_{k=0}^{\infty}(k+1)^{\delta/2}\alpha(k)^{\Delta/(2+\delta+\Delta)} \leqslant D_2, \tag{3.2}$$

where α_n is the corresponding strong mixing coefficient, i.e.

$$\alpha_n(k) = \sup_{A,B} |P(A \cap B) - P(A)P(B)|,$$

where A and B vary over the σ-fields $\mathcal{A}(Z_{0,n}, Z_{-1,n}, \ldots)$ respectively $\mathcal{A}(Z_{k,n}, Z_{k+1,n}, \ldots)$.

Under this assumption we get moment inequalities, which in turn yield strong laws of large numbers even for triangular arrays (cf. the appendix). This is why we introduce $\tilde{\alpha}(\cdot)$.

Corollary 4 in Withers [17] gives conditions under which linear sequences are strong mixing and even provides the mixing coefficients. This can be used to check the above condition. Causal ARMA sequences with appropriate error sequences, for example, fulfill it for any δ, Δ, if the $(2+\delta+\Delta)$-moment of the innovations exists.

Now we state our main theorem which shows that the block permutation method works in our setting.

Theorem 3.1. *Assume that* $\{X(i) : 1 \leqslant i \leqslant n\}$ *fulfills (1.1) - (1.3) with* $\nu > 4$. *Let* $0 < \tilde{\delta} < (\nu-4)/2$ *and let the sequence* $\{e(i) : i \geqslant 1\}$ *fulfill assumption* $(\mathcal{A})$ *for some* $\delta^{(j)}, \Delta^{(j)}$, $j = 1, 2$, *with* $2 + 2\tilde{\delta} < \delta^{(1)} < \nu - 2$, $\Delta^{(1)} := \nu - 2 - \delta^{(1)}$ *respectively* $0 < \delta^{(2)} < \frac{2+\delta^{(1)}}{2+\tilde{\delta}} - 2$ *and* $\Delta^{(2)} := \frac{2+\delta^{(1)}}{2+\tilde{\delta}} - 2 - \delta^{(2)}$.

Under the alternative let either

(i) $K^{(2+\tilde{\delta})/2}|d|^{2+\tilde{\delta}}\min(\frac{m^*}{n}, \frac{n-m^*}{n}) = O(1)$ *and* $\frac{d^2K}{L} = o(1)$ *or*

(ii) $\min(\frac{m^*}{n}, \frac{n-m^*}{n}) \geqslant \epsilon > 0$ *(no restriction on* $d = d_n$ *necessary).*

Let $\alpha(x)$, $\beta(x)$ *be as in Theorem 1.1 and* $K/L = O(1)$. *If* K *is bounded, we also need* $\operatorname{var}\left(\sum_{k=1}^{K} e(k)\right) \geqslant c > 0$ *as* $L \to \infty$. *The estimator* $\widehat{\tau}^2_{LK}$ *is as in (3.1).*

a) If $K = O((\log n)^{\gamma})$ *for some* $\gamma > 0$, *then we have for all* $x \in \mathbb{R}$ *as* $L \to \infty$

$$P\left(\alpha(\log n)\frac{T^{(1)}_{L,K}(\boldsymbol{\pi}, X)}{\widehat{\tau}_{LK}} - \beta(\log n) \leqslant x \middle| X_1, \ldots, X_n\right) \to \exp(-2e^{-x}) \quad a.s.$$

b) If $G = G(n) \to \infty$, $G/n \to 0$, *and (2.3), then we have for all* $x \in \mathbb{R}$ *as* $L \to \infty$

$$P\left(\alpha(n/G)\,\frac{T^{(2)}_{L,K}(G, \boldsymbol{\pi}, X)}{\widehat{\tau}_{LK}} - \beta(n/G) \leqslant x \middle| X_1, \ldots, X_n\right) \to \exp(-2e^{-x}) \quad a.s.$$

c) If $q \in Q_{0,1}$, $I^*(q,c) < \infty$ for some $c > 0$ and (2.4), then we have for all $x \in \mathbb{R}$ as $L \to \infty$

$$P\left(\frac{T^{(3)}_{L,K}(q, \boldsymbol{\pi}, X)}{\widehat{\tau}_{LK}} \leqslant x \middle| X_1, \ldots, X_n \right) \to P\left(\sup_{0<t<1} \frac{|B(t)|}{q(t)} \leqslant x \right) \quad a.s.,$$

where $\{B(t) : 0 \leqslant t \leqslant 1\}$ denotes a Brownian bridge.

d) If r fulfills conditions (1.4) and (2.5), then we have for all $x \in \mathbb{R}$ as $L \to \infty$

$$P\left(\frac{T^{(4)}_{L,K}(r, \boldsymbol{\pi}, X)}{\widehat{\tau}^2_{LK}} \leqslant x \middle| X_1, \ldots, X_n \right) \to P\left(\int_0^1 \frac{B^2(t)}{r(t)}\, dt \leqslant x \right) \quad a.s.,$$

where $\{B(t) : 0 \leqslant t \leqslant 1\}$ denotes a Brownian bridge.

Remark 3.1. If $2 < \nu \leqslant 4$ one gets the above results in a P-stochastic sense instead of almost surely. We then need $0 < \tilde{\delta} < \nu - 2$ and (3.2) for $\tilde{\delta} < \delta^{(1)} < \nu - 2$, $\Delta^{(1)} = \nu - 2 - \delta^{(1)}$ and $\delta^{(2)} > 2 + 2\tilde{\delta}$, $\Delta^{(2)} > 0$. For details confer Kirch [10], Remark 3.5.4.

Remark 3.2. Under the above conditions $\widetilde{T}_n^{(1)}$ also has the correct asymptotic behavior, if we take the maximum over the whole range $1 \leqslant K(l-1) + k < n$. The reason is that the conditions in Remark 2.1 are fulfilled $a.s.$ for logarithmic K. For details confer Kirch [10], Remark 3.5.5 b).

Remark 3.3. It is also possible to use a moving blocks bootstrap with replacement. A block-length of K then gives $(n - K)$ different blocks. The first K and the last K observations, however, are underrepresented in the bootstrap sample leading to some bias. This is why Politis and Romano [15] proposed a circular procedure, where a circular periodic extension of the data sequence is used. This has the advantage that the bootstrap is automatically centered around the sample mean. Then we have the following blocks $\{(X(l+1), \ldots, X(l+K)),\ l = 0, \ldots, n-1\}$, $X(i) = X(i-n)$, $i > n$. The proof uses the same methods, for details confer Kirch [10], Section 3.6.

4. Simulations

The purpose of this simulation study is to illustrate how well the asymptotic test as well as the block permutation test work for small sample sizes n.

Due to limitation of space we only provide a small part of the results for $T_n(q_1)$, $q_1 \equiv 1$. The complete results including Tables providing the critical values, the results for $q_2(t) = (t(1-t))^{1/4}$ as well as for the other three statistics can be found in Kirch [11], confer also Section 6.2 of Kirch [10]. There, one also finds a discussion of the variance estimator $\widehat{\tau}^2_{LK}$ introduced in this paper.

The simulations show that the results for the statistics except for the MOSUM-statistic $T^{(2)}(G)$ are very similar. The permutation statistic does behave better than the asymptotic method for an appropriately chosen block-length K.
However, the asymptotic method does perform better with statistic $T^{(2)}(G)$. The permutational method is not very well suited here. The reason is that we are looking at a generally very small (e.g. $G = 0.05\,n$ or $G = 0.1\,n$) window of data. So K has to

be small compared to G, otherwise the maximum is taken from the same numbers for many different permutations. For larger n and larger G the permutation method behaves somewhat better, but it is still not as good as the asymptotic method, which is working fairly well. Only in the case of i.i.d. error sequences, where we can choose $K = 1$, the above problem does not occur and the permutation method actually works comparably well, maybe even somewhat better than the asymptotic one.

As indicated in Remark 1.1 we use the Bartlett estimator (1.5) in the asymptotic case. We choose $\Lambda = 0.1\,n$, because it is the best choice according to the simulation study conducted by Antoch et al. [2].
Yet for the permutation method we need to use estimator (3.1) in order to have the correct asymptotic behavior given the observations. Since it is invariant under permutations, the original as well as permutation statistic are divided by the same value so that the actual value is irrelevant. This is an advantage, because the performance of the test does not depend on the performance of the estimator.

In the simulation study we use the model of Section 1, where $\{e(i) : i \geqslant 1\}$ forms an AR(1) sequence with autoregressive coefficient $\rho \in \{-0.5, -0.3 : 0.3, 0.5, 0.7\}$ and $\{\epsilon(j) : -\infty < j < \infty\}$ are i.i.d. $N(0,1)$, hence $\tau = \frac{1}{1-\rho}$. Sample sizes are $80, 120, 210$, the change-points under the alternative are at $\frac{n}{4}, \frac{n}{2}, \frac{3}{4}n$, we choose the block-length K approximately as $1, \frac{\log n}{2}, \log n, \frac{(\log n)^2}{2}$ and $d \in \{0.25, 0.5, 1, 2, 4\}$.

A simulation of the distribution of the statistic under the null hypothesis shows that the quantiles depend strongly on the correlation ρ between the observations.

Next we simulate the quantiles of the block permutation statistic given an observation sequence. The observations for different choices of parameters are based on the same underlying error sequence. We also use the same 10 000 permutations to calculate the permutation statistic.
The results show that the quantiles also depend on the correlation ρ.
Under positive correlation they are quite stable for different alternatives. However under negative correlation the quantiles get smaller, the more obvious the change. As we will see later, this phenomenon contributes to a better power of the test. For greater n they clearly stabilize.

There is a good match of the quantiles under the null hypothesis and the permutation quantiles for an appropriately chosen value of K. For stronger correlated errors the match is better the longer the block-length K is. This is not very surprising, since in this case the dependency structure is much better preserved. In the independent case, we have a better match for a shorter block-length.

To get a better impression how well the block permutation distribution of one specific realization matches the distribution of the statistic under H_0 we create QQ-plots.

1) *Exact distribution:* Determine the empirical distribution function of the statistic (under H_0) based on 10 000 samples of length n.

2) *Simulate observations:* Simulate one specific realization of the model for particular parameters of H_0 or H_1.

3) *Block permutation distribution:* Determine the empirical distribution function of the block permutation statistic based on 10 000 permutations given the realization from the previous step.

4) Draw a *QQ-plot* of the null distribution from step 1) against the permutation distributions from step 3).

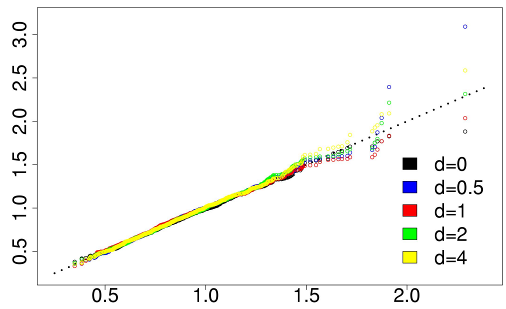
Figures 4.1(i) to 4.1(ii) show some typical results.
For strongly correlated errors we get best results for large values of K. But the closer to independence the errors are, the better the plots become for small K. For independent errors the plot for $K = 1$ is as good as the one for $K = 2$ or $K = 4$.

For $K = 10$ and $n = 80$ the greatest simulated values show some kind of "step behavior", i.e. many different permutations give the same value of the permutation statistic. This is due to the fact that there are only 8 blocks to permute. So if there is one block that gives (e.g. if put on the first place) the "maximally" obtained value, there is a good chance that several permutations will put this block at the same place. However, this phenomenon disappears for greater n (cf. e.g. Figures 4.1(ii), which gives the QQ-Plots for $n = 210$ and $K = 15$). Even for small n, this phenomenon does not seem to influence the performance of the test, as the size-power-curves will show. The test behaves very well with large values of K under H_0 as well as H_1. Apparently the number of such "outliers" is too small to influence the quantiles.

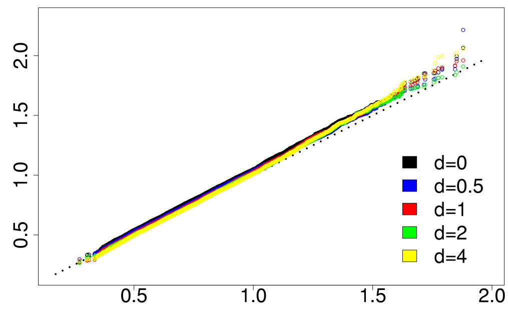
The plots displayed below have a change at $n/2$, but we get essentially the same picture for different choices of m^*. Moreover the distribution seems to be independent of d (i.e. the mean difference before and after the change) for positively correlated observations. There is again only the exception of $K = 10$, but fortunately the values become smaller if there is a more obvious change. This only means that we are more likely to reject the null hypothesis under alternatives, which improves the power of the test. For negatively correlated observations the values get also smaller for more obvious changes.

The permutation test rejects the null hypothesis if the value of the statistic $T_n(q_1)$ of an observation sequence is greater than the $(1 - \alpha)$-quantile of the block permutation statistic given the same sequence of observations. To get a better idea of how well the test performs we create size-power-curves of both methods under the null hypothesis and under various alternatives. Size-power-curves are plots of the empirical distribution function of the p-values of the statistic $T_n(q_1)$ for the null hypothesis respectively given alternatives with respect to the distribution used to determine the critical values of the test.
What we get is a plot that shows the actual α–errors resp. $1-(\beta$–errors) on the y-axis for the chosen quantiles on the x-axis. So, the graph for the null hypothesis should be close to the diagonal (which is given by the dotted line) and for the alternatives should be as steep as possible.
The results are presented in Figure 4.2. The asymptotic method does not perform too well, except in the case of $\rho = -0.5$. The α–errors in all other cases are too high, e.g. for $\rho = 0.5$ we even have an actual α–error of 20% for a nominal one of 10%.

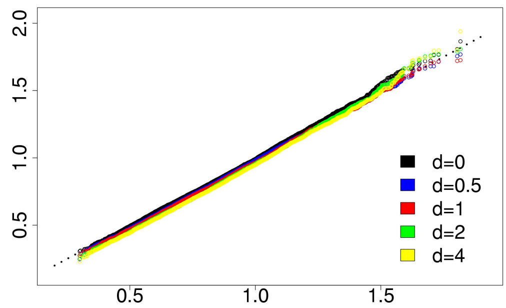
For the permutation method it is apparently very important to choose an appropriate K. For an independent error-sequence the α–error is good for all choices of K, however, the β–errors increase (i.e. the y-values decrease under alternatives) somewhat for higher



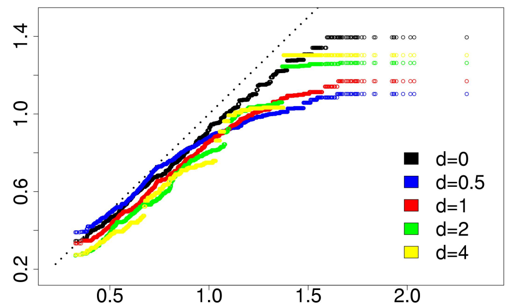
(1) $\rho = -0.5,\ K = 10$



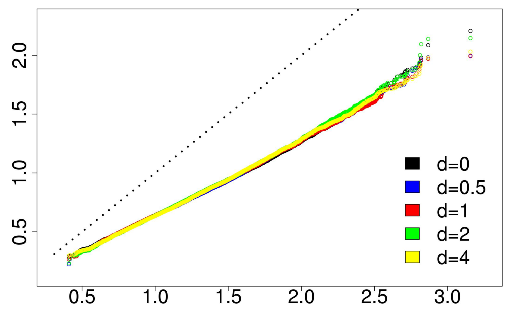
(2) $\rho = -0.1,\ K = 2$



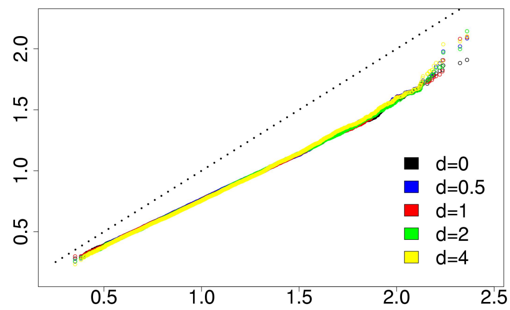
(3) $\rho = -0.1,\ K = 4$



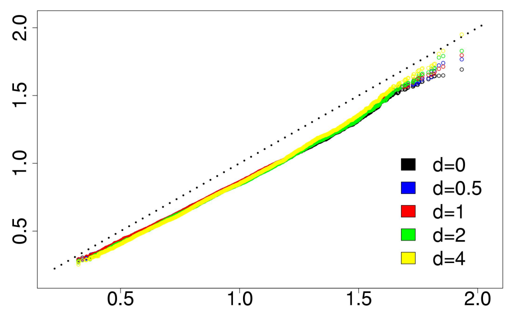
(4) $\rho = -0.1,\ K = 10$



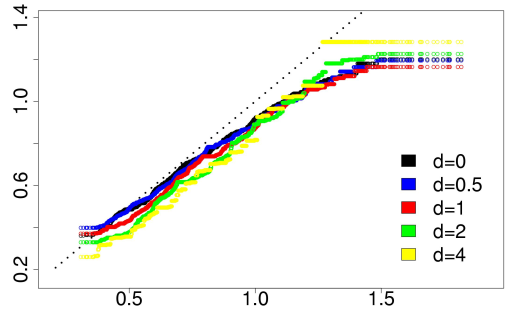
(5) $\rho = 0.5,\ K = 1$



(6) $\rho = 0.5,\ K = 2$

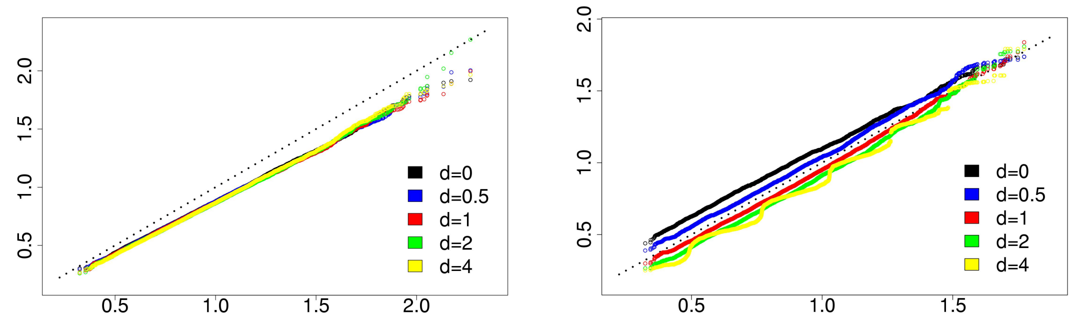


(7) $\rho = 0.5,\ K = 4$



(8) $\rho = 0.5,\ K = 10$

Figure 4.1(i): QQ-Plots of $T_n^{(3)}(q_1)/\widehat{\tau}_{LK}$ (under H_0) against $T_{L,K}^{(3)}(q_1, \boldsymbol{\pi}, X)/\widehat{\tau}_{LK}$ for $n = 80$, $m^* = 40$ and different values for ρ and K



(9) $\rho = 0.5$, $K = 5$ (10) $\rho = 0.5$, $K = 15$

Figure 4.1(ii): QQ-Plots of $T_n^{(3)}(q_1)/\widehat{\tau}_{LK}$ (under H_0) against $T_{L,K}^{(3)}(q_1, \boldsymbol{\pi}, X)/\widehat{\tau}_{LK}$ for $n = 210$, $m^* = 105$ and different values for ρ and K

K. The best choice here would indeed be $K = 1$. On the other hand a choice of $K = 10$ is best under strong dependence.

The figures also show that the power of the test is better for negatively correlated error sequences. This is due to the fact that the permutation quantiles slightly decrease for greater mean differences.

The more dependent we suspect the data to be the greater we should choose K, since a choice of too large a K does not negatively influence the α–errors. It does, however, increase the β–errors but only slightly. This also means that – if in doubt – it is always better to choose a larger K.

A comparison of both methods (for an appropriately chosen K) yields that the α–errors of the permutation method are always better than those of the asymptotic method (with the single exception of $\rho = -0.5$). Concerning the β–errors we realize that – depending on the data – they are comparable (if taken into account that a larger α-error goes usually along with a smaller β-error) for both methods.

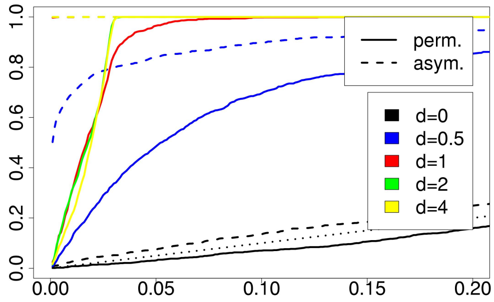
5. Proofs of Section 2

Throughout the proofs we use the notation $a_n \ll b_n$ for $a_n = O(b_n)$.

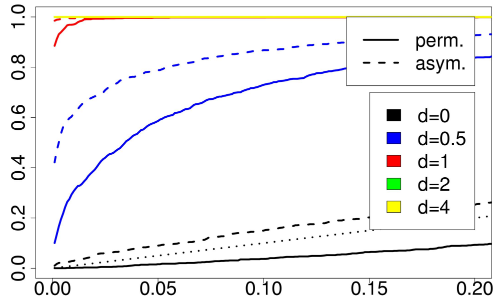
Before we prove Theorem 2.1 we need two lemmas that deal with the increments of the rank statistics respectively Brownian bridges.

Lemma 5.1. *Let $\pi_{\mathbf{L}} = (\pi(1), \ldots, \pi(L))$ be a random permutation of $(1, \ldots, L)$ chosen with probability $\frac{1}{L!}$. Moreover let $a_n(1), \ldots, a_n(n)$ be scores satisfying (2.1). Then it holds for all $\mu < \min\left(\frac{\nu-2}{2\nu}, \frac{1}{4}\right)$*

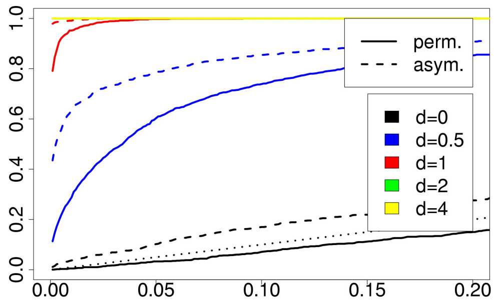
$$\max_{\substack{1\leqslant l\leqslant L-1\\ 1\leqslant k\leqslant K}} \left(\frac{l(L-l)}{L}\right)^{\mu} \frac{L}{\sqrt{l(L-l)}} \left|\frac{1}{\sqrt{LK}} \sum_{j=k+1}^{K} (a_n[K(\pi(l)-1)+j] - \overline{a}_n)\right| = O_P(1).$$



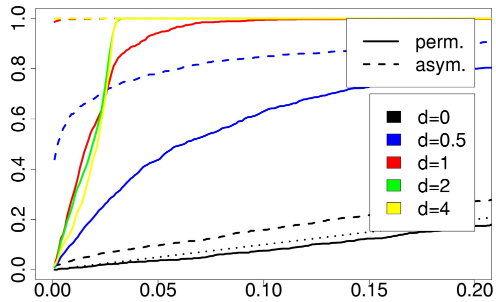
(11) $\rho = -0.5,\ K = 10$



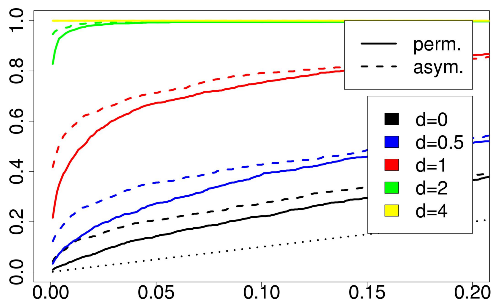
(12) $\rho = -0.3,\ K = 2$



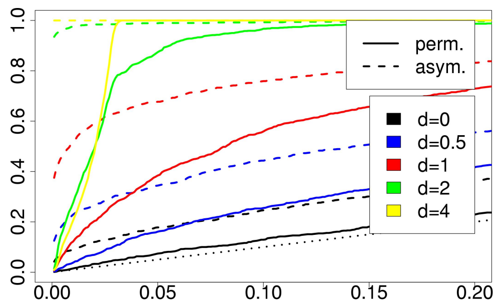
(13) $\rho = -0.3,\ K = 4$



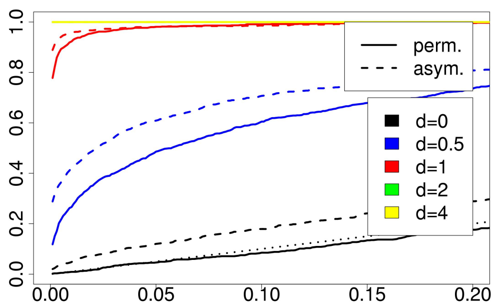
(14) $\rho = -0.3,\ K = 10$



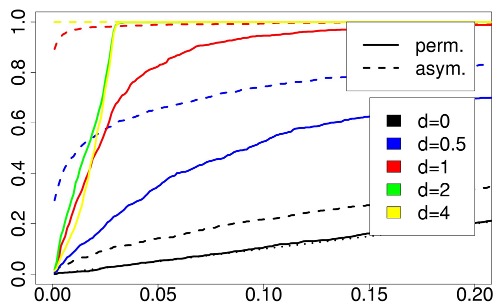
(15) $\rho = 0.5,\ K = 4$



(16) $\rho = 0.5,\ K = 10$



(17) $\rho = 0,\ K = 1$



(18) $\rho = 0,\ K = 10$

Figure 4.2: Size-power-curves of $T_n^{(3)}(q_1)/\widetilde{\tau}_n(\Lambda, X)$ with respect to the asymptotic distribution and of $T_{L,K}^{(3)}(q_1, \boldsymbol{\pi}, X)/\widehat{\tau}_{LK}$ with respect to the permutation distribution for $n = 80$, $m^* = 40$ and different values for ρ and K

Proof. For every $\epsilon > 0$ we find a C big enough such that

$$P\left(\max_{1\leqslant l\leqslant L-1}\max_{1\leqslant k\leqslant K}\left(\frac{L}{l(L-l)}\right)^{1/2-\mu}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K}(a_n[K(\pi(l)-1)+j]-\overline{a}_n)\right|\geqslant C\right)$$
$$\leqslant\sum_{l=1}^{L-1}P\left(\max_{1\leqslant k\leqslant K}\left(\frac{L}{l(L-l)}\right)^{1/2-\mu}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K}(a_n[K(\pi(l)-1)+j]-\overline{a}_n)\right|\geqslant C\right)$$
$$\leqslant\sum_{l=1}^{L-1}\frac{1}{C^\nu}\left(\frac{L}{l(L-l)}\right)^{(1/2-\mu)\nu}\mathrm{E}\left(\max_{1\leqslant k\leqslant K}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K}(a_n[K(\pi(l)-1)+j]-\overline{a}_n)\right|^\nu\right)$$
$$\leqslant\epsilon.$$

Note that $(1/2-\mu)\nu > 1$ and that for every $s > 1$ we have

$$\sum_{l=1}^{L-1}\left(\frac{L}{l(L-l)}\right)^s\leqslant 2\sum_{l=1}^{\lceil L/2\rceil}\left(\frac{L}{l(L-l)}\right)^s\leqslant 2L^s\int_1^{L/2}(x(L-x))^{-s}\,dx+O(1)$$
$$\leqslant 2^{s+1}\int_1^{L/2}x^{-s}\,dx+O(1)=O(1).$$

■

The next lemma corresponds to the one above dealing with the increments of Brownian bridges instead of the increments of rank statistics. It is based on results of Csörgő and Révész [6].

Lemma 5.2. *Let $\{B(t): 0\leqslant t\leqslant 1\}$ be a Brownian bridge. Then it holds:*

a)

$$\max_{\substack{1\leqslant l\leqslant L\\1\leqslant k\leqslant K}}\sqrt{L}\left|B\left(\frac{l}{L}\right)-B\left(\frac{K(l-1)+k}{KL}\right)\right|=O\left(\sqrt{\log L}\right)\quad a.s.$$

b)

$$\max_{\substack{(\log L)^s\leqslant l\leqslant L-(\log L)^s\\1\leqslant k\leqslant K}}\frac{L}{\sqrt{l(L-l)}}\left|B\left(\frac{l}{L}\right)-B\left(\frac{K(l-1)+k}{KL}\right)\right|=O((\log L)^{(1-s)/2})\quad a.s.,$$

where $s\geqslant 0$.

c)

$$\max_{\substack{1\leqslant l<L\\1\leqslant k\leqslant K}}\frac{L}{\sqrt{l(L-l)}}\left|B\left(\frac{l}{L}\right)-B\left(\frac{K(l-1)+k}{KL}\right)\right|=O_P(1).$$

Proof. a) follows immediately from Theorem 1.4.1 (p. 42) in Csörgő and Révész [6], and implies b).

Note that b) implies c) for $\log L \leqslant l \leqslant L - \log L$.
It holds

$$\left\{\sqrt{L}\left|B\left(\frac{l}{L}\right) - B\left(\frac{K(l-1)+k}{KL}\right)\right| : 1 \leqslant l \leqslant L, 1 \leqslant k \leqslant K\right\}$$
$$\stackrel{\mathcal{D}}{=} \left\{\left|W(l) - W\left(l - \frac{K-k}{K}\right) - \frac{K-k}{LK}W(L)\right| : 1 \leqslant l \leqslant L, 1 \leqslant k \leqslant K\right\},$$

where $\{W(t) : t \geqslant 0\}$ is a Wiener process. The law of iterated logarithm gives the assertion for the term $(K-k)/(LK)W(L)$ so that it suffices to investigate the maximum of $|W(l) - W(l-(K-k)/K)|$. The assertion for $l = 1$ also follows immediately. For $2 \leqslant l < \log L$ consider $l_j := \max(2, 2^{-j}\log L)$, then $[2, \log L) = \sum_{j=0}^{M_L}[l_{j+1}, l_j)$, where $M_L = \left\lceil \frac{\log\log L}{\log 2}\right\rceil - 2$.

Theorem 1.2.1 (p. 30) of Csörgő, Révész [6] gives now

$$\max_{0 \leqslant j \leqslant M_L} \max_{l_{j+1} \leqslant l < l_j} \max_{1 \leqslant k \leqslant K} \sqrt{\frac{L}{l(L-l)}}\left|W(l) - W\left(l - \frac{K-k}{K}\right)\right|$$
$$\ll \max_{0 \leqslant j \leqslant M_L} \max_{l_{j+1} \leqslant l < l_j} \sqrt{\frac{\log l_j}{l}} \leqslant \max_{0 \leqslant j \leqslant M_L} \sqrt{\frac{\log l_j}{l_{j+1}}} = O(1) \qquad \text{a.s.},$$

since

$$\max_{0 \leqslant j \leqslant M_L} \sqrt{\frac{\log l_j}{l_{j+1}}} \leqslant \max_{0 \leqslant j \leqslant M_L} \sqrt{2\frac{\log(2^{-j}\log L)}{2^{-j}\log L}} \leqslant \sup_{x \geqslant 1}\sqrt{2\frac{\log x}{x}} = O(1).$$

In view of $\{B(t)\} \stackrel{\mathcal{D}}{=} \{B(1-t)\}$, we get analogously

$$\max_{L - \log L \leqslant l < L, 1 \leqslant k \leqslant K} \frac{L}{\sqrt{l(L-l)}}\left|B\left(\frac{l}{L}\right) - B\left(\frac{K(l-1)+k}{KL}\right)\right| = O_P(1).$$

Putting everything together we arrive at the assertion. ■

Proof of Theorem 2.1. The idea of the proof is the following: Lemma 5.1 controls the error when replacing $S_{L,K}^{\mathbf{a},\boldsymbol{\pi}}(l,k)$, $k = 1, \ldots, K$, in the statistics by $S_{L,K}^{\mathbf{a},\boldsymbol{\pi}}(l,K)$. Theorem 1 by Einmahl and Mason [7] allows us to replace those partial sums by a Brownian bridge at time l/L. We finally need Lemma 5.2 to replace $B(l/L)$ by $B([K(l-1)+k]/[LK])$. Then we obtain the assertion from the one for i.i.d. observations.

Since it is well known how to get results of the type above using an embedding as in Theorem 1 of Einmahl and Mason [7], we will only give the proof for the q-weighted CUSUM-statistic. For details of the other proofs confer Kirch [10].

Let

$$D_{l,\mu} := \left(\frac{l(L-l)}{L}\right)^{\mu} \frac{L}{\sqrt{l(L-l)}}\left|\frac{1}{\sqrt{L}}\widetilde{\Pi}(l) - B\left(\frac{l}{L}\right)\right|,$$

where

$$\{\widetilde{\Pi}(l) : 1 \leqslant l \leqslant L\} \stackrel{\mathcal{D}}{=} \left\{\frac{1}{\sqrt{K\tau_n^2(\mathbf{a})}}\sum_{i=1}^{l}\sum_{j=1}^{K}(a_n[K(\pi_L(i)-1)+j] - \overline{a}_n) : 1 \leqslant l \leqslant L\right\}.$$

and $\{B(t) : 0 \leqslant t \leqslant 1\}$ is the Brownian bridge of Theorem 1 in Einmahl and Mason [7]. That Theorem with $X_L(i) := \frac{1}{\sqrt{K\,\tau_n^2(\mathbf{a})}} \sum_{k=1}^{K} a_n(K(i-1)+k)$ yields

$$\max_{1\leqslant l\leqslant L-1} D_{l,\mu} = O_P(1) \tag{5.1}$$

for all $\mu < \min\left(\frac{\nu-2}{2\nu}, \frac{1}{4}\right)$. As indicated in Remark 1 of Einmahl and Mason [7] we replaced the exponent 4 in assumption (2.3) of [7] by μ above, details can be found in Kirch and Steinebach [12], Theorem 8, or Kirch [10], Appendix D.

We will first verify that as $L \to \infty$

$$\max_{\substack{1\leqslant l\leqslant L, 1\leqslant k\leqslant K\\ (l,k)\neq(L,K)}} \frac{1}{q\left(\frac{K(l-1)+k}{KL}\right)} \left|\frac{1}{\sqrt{L}}\widetilde{\Pi}_L(l) - B\left(\frac{l}{L}\right)\right| = o_P(1) \quad \text{as } L\to\infty. \tag{5.2}$$

Since $\max_{\eta<t<1-\eta} q(t) \geqslant C(\eta) > 0$ for $0<\eta<1/2$, we obtain by (5.1) for the maximum over $\eta L \leqslant l \leqslant (1-\eta)L$

$$\begin{aligned}
&\max_{\eta L\leqslant l\leqslant(1-\eta)L}\ \max_{1\leqslant k\leqslant K} \frac{1}{q\left(\frac{K(l-1)+k}{KL}\right)} \left|\frac{1}{\sqrt{L}}\widetilde{\Pi}(l) - B\left(\frac{l}{L}\right)\right| \\
&\leqslant \max_{\eta L\leqslant l\leqslant(1-\eta)L} \frac{1}{C(\eta)} D_{L,\mu}\left[\frac{l}{L}\,\frac{L-l}{L}\right]^{1/2-\mu} L^{-\mu} = O_P(L^{-\mu}) = o_P(1).
\end{aligned}$$

To handle the maximum over $1<l<\eta L$ and $(1-\eta)L<l<L$ recall that Csörgő and Horváth [4], Chapter 4, Theorem 2.4, gives

$$\lim_{t\to 0}\frac{\sqrt{t}}{q(t)} = 0 = \lim_{t\to 1}\frac{\sqrt{1-t}}{q(t)}. \tag{5.3}$$

Combining these facts with $\inf_{\eta\leqslant t\leqslant 1-\eta} q(t) > 0$ we obtain

$$\sup_{0<t<1}\frac{\min(t,1-t)}{q^2(t)} = O(1). \tag{5.4}$$

Now by $\max_{2\leqslant l\leqslant L-1}\ \max_{1\leqslant k\leqslant K} \frac{l(L-l)}{(l-1+\frac{k}{K})(L-(l-1)-\frac{k}{K})} = O(1)$ we get uniformly in L

$$\begin{aligned}
&\max_{\substack{1<l<\eta L, (1-\eta)L<l<L\\ 1\leqslant k\leqslant K}} \frac{1}{q\left(\frac{K(l-1)+k}{KL}\right)} \left|\frac{1}{\sqrt{L}}\widetilde{\Pi}(l) - B\left(\frac{l}{L}\right)\right| \\
&\ll \max_{1\leqslant l<L} D_{l,0} \max_{\substack{1/L<l/L<\eta, (1-\eta)<l/L<1\\ 1\leqslant k\leqslant K}} \frac{\min\left[\sqrt{\frac{l-1+k/K}{L}}, \sqrt{1-\frac{l-1+k/K}{L}}\right]}{q\left(\frac{l-1+k/K}{L}\right)} \\
&= O_P(1)\, o(1) = o_P(1) \quad \text{as } \eta\to 0.
\end{aligned}$$

The assertion for $l=L$ is trivial and for $l=1$ (5.1), $\mu=0$, gives as $L\to\infty$

$$\max_{1\leqslant k\leqslant K} \frac{1}{q\left(\frac{k}{KL}\right)} \left|\frac{1}{\sqrt{L}}\widetilde{\Pi}_L(1) - B(1/L)\right| = \frac{1}{\sqrt{L}\, q\left(\frac{1}{KL}\right)} O_P(1) = o_P(1),$$

since $\frac{1}{Lq^2(1/KL)} \to 0$ and q is non-decreasing in a neighborhood of 0. By choosing $\eta>0$ small enough and then $L = L(\eta)$ sufficiently large we obtain (5.2).

Lemma 5.1 and assumptions (2.1) respectively (2.2) yield analogously as $L \to \infty$

$$\max_{\substack{1\leqslant l\leqslant L, 1\leqslant k\leqslant K \\ (l,k)\neq(L,K)}} \frac{1}{\sqrt{LK\tau_n^2(\mathbf{a})}} \frac{1}{q\left(\frac{K(l-1)+k}{KL}\right)} \left| \sum_{j=k+1}^{K} (a_n[K(\pi(l)-1)+j]-\overline{a}_n) \right| = o_P(1), \tag{5.5}$$

where an application of the Markov inequality yields the assertion for $l = L$. This shows that the error when replacing $S_{L,K}^{\mathbf{a},\boldsymbol{\pi}}(l,k)$ in the statistic $T_{L,K}^{(3)}(q,\boldsymbol{\pi},\mathbf{a})$ by $S_{L,K}^{\mathbf{a},\boldsymbol{\pi}}(l,K)$ is $o_P(1)$. Hence (5.2) and (5.5) yield

$$\begin{aligned}
\frac{T_{L,K}^{(3)}(q,\boldsymbol{\pi},\mathbf{a})}{\tau_n(\mathbf{a})} &= \max_{\substack{1\leqslant l\leqslant L, 1\leqslant k\leqslant K \\ (l,k)\neq(L,K)}} \frac{1}{\sqrt{KL}\, q\left(\frac{K(l-1)+k}{KL}\right)\tau_n(\mathbf{a})} \left| S_{L,K}^{\mathbf{a},\boldsymbol{\pi}}(l,k)\right| \\
&= \max_{\substack{1\leqslant l\leqslant L, 1\leqslant k\leqslant K \\ (l,k)\neq(L,K)}} \frac{1}{\sqrt{KL}\, q\left(\frac{K(l-1)+k}{KL}\right)\tau_n(\mathbf{a})} \left| S_{L,K}^{\mathbf{a},\boldsymbol{\pi}}(l,K)\right| + o_P(1) \\
&\overset{\mathcal{D}}{=} \max_{\substack{1\leqslant l\leqslant L, 1\leqslant k\leqslant K \\ (l,k)\neq(L,K)}} \frac{1}{q\left(\frac{K(l-1)+k}{KL}\right)} \left| \frac{1}{\sqrt{L}} \widetilde{\Pi}(l)\right| + o_P(1) \\
&= \max_{\substack{1\leqslant l\leqslant L, 1\leqslant k\leqslant K \\ (l,k)\neq(L,K)}} \frac{\left|B\left(\frac{l}{L}\right)\right|}{q\left(\frac{K(l-1)+k}{KL}\right)} + o_P(1).
\end{aligned}$$

We will now complete the proof by showing that

$$\max_{\substack{1\leqslant l\leqslant L, 1\leqslant k\leqslant K \\ (l,k)\neq(L,K)}} \frac{\left|B\left(\frac{l}{L}\right)\right|}{q\left(\frac{K(l-1)+k}{KL}\right)} = \max_{\substack{1\leqslant l\leqslant L, 1\leqslant k\leqslant K \\ (l,k)\neq(L,K)}} \frac{\left|B\left(\frac{K(l-1)+k}{KL}\right)\right|}{q\left(\frac{K(l-1)+k}{KL}\right)} + o_P(1). \tag{5.6}$$

This means that the block permutation statistic has the same asymptotic behavior as the original statistic for n i.i.d. standard normally distributed random variables. Thus an application of Theorem 1.1 yields the assertion. First Lemma 5.2 b) and equation (5.4) show for the maximum over $\log^2 L \leqslant l \leqslant L - \log^2 L$

$$\begin{aligned}
&\max_{\substack{\log^2 L\leqslant l\leqslant L-\log^2 L \\ 1\leqslant k\leqslant K}} \frac{1}{q\left(\frac{K(l-1)+k}{KL}\right)} \left| B\left(\frac{l}{L}\right) - B\left(\frac{l}{L} - \frac{K-k}{LK}\right)\right| \\
&= O\left(\sqrt{\frac{1}{\log L}} \max_{\substack{\log^2 L\leqslant l\leqslant L-\log^2 L \\ 1\leqslant k\leqslant K}} \sqrt{\frac{\min(l, L-l)}{L\, q^2\left(\frac{K(l-1)+k}{KL}\right)}} \right) = o(1) \quad \text{a.s.}
\end{aligned}$$

To handle the maximum over $l < \log^2 L$ and $L - \log^2 L < l < L$ note that assumption (2.4) respectively equation (5.3) give

$$\max_{\substack{1\leqslant l<\log^2 L,\ L-\log^2 L<l<L \\ 1\leqslant k\leqslant K}} \frac{\min(l, L-l)}{L\, q^2\left(\frac{K(l-1)+k}{KL}\right)} = o(1).$$

By Lemma 5.2 c) we conclude

$$\max_{\substack{1\leqslant l<\log^2 L,\ L-\log^2 L<l<L\\ 1\leqslant k\leqslant K}} \frac{1}{q\left(\frac{K(l-1)+k}{KL}\right)}\left|B\left(\frac{l}{L}\right)-B\left(\frac{l}{L}-\frac{K-k}{LK}\right)\right|$$
$$=O_P\left(\max_{\substack{1\leqslant l<\log^2 L,\ L-\log^2 L<l<L\\ 1\leqslant k\leqslant K}}\sqrt{\frac{\min(l,L-l)}{L\,q^2\left(\frac{K(l-1)+k}{KL}\right)}}\right)=o_P(1).$$

(2.4) gives the assertion for $l=L$ and we have proven (5.6). ■

6. Proofs of Section 3

Proof of Lemma 3.1. The proof follows the techniques outlined in Phillips and Solo [14]. The method uses an algebraic decomposition of the linear filter, which allows us to prove the limit theorem by means of a corresponding theorem for an independent sequence of random variables additionally to some limit theorems for the telescoping sums of another linear sequence.

First of all the following decomposition holds:

$$\frac{1}{KL}\sum_{l=0}^{L-1}\left[\sum_{k=1}^{K}(X(Kl+k)-\overline{X}_n)\right]^2=\frac{1}{KL}\sum_{l=0}^{L-1}\left(\sum_{k=1}^{K}e(Kl+k)\right)^2-K\,\overline{e}_n^2.$$

The law of iterated logarithm for linear processes (confer e.g. Theorem 3.3 of Philipps and Solo [14]) gives for the second term

$$K\,\overline{e}_n^2=O\left(\frac{\log\log n}{L}\right)\quad \text{a.s.}$$

The Beveridge-Nelson decomposition (confer Phillips and Solo [14], equation (15) et sqq.) states $e(j)=\epsilon(j)\left(\sum_{s\geqslant 0}w_s\right)+\widetilde{e}(j-1)-\widetilde{e}(j)$ where $\widetilde{e}(\cdot)$ is another stationary linear process with existing second moment. We thus get for the first term

$$\frac{1}{KL}\sum_{l=0}^{L-1}\left(\sum_{k=1}^{K}e(Kl+k)\right)^2$$
$$=\left(\sum_{s\geqslant 0}w_s\right)^2\frac{1}{KL}\sum_{l=0}^{L-1}\left(\sum_{k=1}^{K}\epsilon(Kl+k)\right)^2+\frac{1}{KL}\sum_{l=0}^{L-1}\left(\widetilde{e}(Kl)-\widetilde{e}(K(l+1))\right)^2$$
$$+\left(\sum_{s\geqslant 0}w_s\right)\frac{2}{KL}\sum_{l=0}^{L-1}\Bigl(\sum_{k=1}^{K}\epsilon(Kl+k)\Bigr)\left(\widetilde{e}(Kl)-\widetilde{e}(K(l+1))\right)$$
$$=:D_1(L,K)+D_2(L,K)+D_3(L,K).$$

Concerning $D_1(L,K)$, the law of Marcinkiewicz–Zygmund (cf. e.g. Loève [13], p. 254, Moments Lemma 4°) gives

$$\frac{1}{KL}\sum_{l=0}^{L-1}\left(\sum_{k=1}^{K}\epsilon(Kl+k)\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\epsilon^2(i) + \frac{1}{KL}\sum_{l=0}^{L-1}\sum_{\substack{k_1\neq k_2\\1}}^{K}\epsilon(Kl+k_1)\epsilon(Kl+k_2)$$
$$= \sigma^2 + O_P\left(n^{-\frac{\mu-2}{\mu}}\right) + O_P\left(\sqrt{\frac{1}{L}}\right),$$

where the last line follows from the Markov inequality, since

$$\mathrm{var}\Big(\frac{1}{KL}\sum_{l=0}^{L-1}\sum_{k_1\neq\ k_2}\epsilon(Kl+k_1)\epsilon(Kl+k_2)\Big) \ll \sigma^4\frac{1}{L}$$

as some computations show. Since

$$\mathrm{E}\left(\frac{1}{L}\sum_{l=0}^{L-1}(\widetilde{e}(Kl)-\widetilde{e}(K(l+1)))^2\right) \leqslant 4\,\mathrm{E}(\widetilde{e}(0)^2) < \infty,$$

the Markov inequality gives $D_2(L,K) = O_P\left(\frac{1}{K}\right)$. Concerning $D_3(L,K)$ we get by the Cauchy-Schwartz-Inequality

$$\mathrm{E}\left|\frac{1}{KL}\sum_{l=0}^{L-1}\left(\sum_{k=1}^{K}\epsilon(Kl+k)\right)(\widetilde{e}(Kl)-\widetilde{e}(K(l+1)))\right|$$
$$\leqslant \frac{1}{KL}\sum_{l=0}^{L-1}\left(\mathrm{var}\left(\sum_{k=1}^{K}\epsilon(Kl+k)\right)\mathrm{var}(\widetilde{e}(Kl)-\widetilde{e}(K(l+1)))\right)^{1/2} \ll \frac{1}{\sqrt{K}}.$$

Again the Markov inequality gives $D_3(L,K) = O_P\left(\sqrt{\frac{1}{K}}\right)$. Putting everything together we arrive at the assertion. ∎

To prove Theorem 3.1 we first need a small auxiliary lemma:

Lemma 6.1. *Under conditions (1.2) (the existence of the second moment suffices) and (1.3) we have as $n\to\infty$*

$$\mathrm{var}\left(\frac{1}{\sqrt{n}}\sum_{l=1}^{n}e(l)\right) \to \left(\sum_{s\geqslant 0}w_s\right)^2\sigma^2 > 0.$$

Proof. A straightforward calculation gives the assertion. For details confer Kirch [10], Lemma 3.5.2. ∎

Proof of Theorem 3.1. The idea of the proof is to apply Theorem 2.1 with special scores, more precisely under (i) we choose $a_n(i) := X(i)$ and under (ii) $a_n(i) = X(i)/\sqrt{d^2K}$. In a first step we prove that assumptions (2.1) and (2.2) are fulfilled almost surely for the underlying error sequence $\{e(\cdot)\}$, in a second step we conclude that they are also fulfilled almost surely for $\{X(\cdot)\}$.

In the following we will repeatedly use D as a constant. It may be different in every inequality.

We will now derive some strong laws of large numbers for the underlying linear process, which we need to obtain (2.1) and (2.2) for $\{X(\cdot)\}$. The Minkovski inequality and the monotone convergence theorem give $\mathrm{E}\,|e(i)|^\nu \leqslant D\left(\sum_{j\geqslant 0}|w_j|\right)^\nu$ for all $i \geqslant 0$. Thus Theorem A.2 b) yields as $L \to \infty$

$$\frac{1}{n}\sum_{j=1}^{n} e(j) = O\left(\sqrt{\frac{\log n}{n}}\right) \quad a.s. \tag{6.1}$$

To prove (2.2) under alternatives we need the following two laws for triangular arrays. Let $l^* := \lceil m^*/K \rceil$ $(m^* \neq n)$. Since $\mathrm{E}\left(\max_{k=1,\ldots,K}\left|\sum_{j=1}^{k} e(Kl^*+j)\right|^{2+\delta^{(1)}}\right) = \mathrm{E}\left(\max_{k=1,\ldots,K}\left|\sum_{j=1}^{k} e(j)\right|^{2+\delta^{(1)}}\right)$, Theorem A.2 and Remark A.1 give

$$\frac{1}{K}\sum_{k=1}^{K} e(Kl^*+k) = O\left(\sqrt{\frac{\log K}{K}}\right) \quad a.s.,$$

as $K \to \infty$. For K bounded the Markov inequality yields

$$P\left(\frac{1}{\sqrt{L}}\left|\sum_{k=1}^{K} e(Kl^*+k)\right| \geqslant \epsilon\right) \ll \frac{K^\nu}{\epsilon^\nu} L^{-\nu/2}\,\mathrm{E}\,|e(0)|^\nu \ll \frac{1}{\epsilon^\nu} L^{-\nu/2}.$$

Because $\sum_L L^{-\nu/2} < \infty$, it holds as $L \to \infty$

$$\frac{1}{\sqrt{L}}\left|\sum_{k=1}^{K} e(Kl^*+k)\right| = O\left(\sqrt{\log K}\right) \quad a.s. \tag{6.2}$$

for $K \to \infty$ as well as K bounded. Similarly we deduce

$$\frac{1}{\sqrt{L(n-m^*)}}\left|\sum_{j=Kl^*+1}^{n} e(j)\right| = o(1) \quad a.s. \tag{6.3}$$

The next two strong laws imply (2.1) and (2.2) for the error sequence $\{e(\cdot)\}$. Since the mixing coefficient of $\{\frac{1}{K}\sum_{k=1}^{K} e(Kl+k) : l \geqslant 0\}$ is smaller than the one of $\{e(i) : i \geqslant 1\}$ for all K, assumption (3.2) is uniformly fulfilled in K. Also this sequence remains stationary for stationary $\{e(\cdot)\}$. Consequently Theorem A.2 a) shows that there is a $D > 0$ such that

$$\mu_K(2+\delta^{(1)}) := \mathrm{E}\max_{k=0,\ldots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K} e(Kl+k)\right|^{2+\delta^{(1)}} < D,$$

uniformly in $l \geqslant 0$ and K. Note that $2(2+\delta^{(2)}+\Delta^{(2)}) < (2+\tilde{\delta})(2+\delta^{(2)}+\Delta^{(2)}) = 2+\delta^{(1)}$. From Corollary A.1 we conclude as $L \to \infty$

$$\begin{aligned}
&\frac{1}{L}\sum_{l=0}^{L-1}\Big(\frac{1}{\sqrt{K}}\sum_{k=1}^{K} e(Kl+k)\Big)^2 \\
&= \frac{1}{L}\sum_{l=0}^{L-1}\left[\Big(\frac{1}{\sqrt{K}}\sum_{k=1}^{K} e(Kl+k)\Big)^2 - \mathrm{var}\Big(\frac{1}{\sqrt{K}}\sum_{k=1}^{K} e(k)\Big)\right] \\
&\quad + \mathrm{var}\Big(\frac{1}{\sqrt{K}}\sum_{k=1}^{K} e(k)\Big) \;\to\; C > 0 \quad a.s.,
\end{aligned} \tag{6.4}$$

where we either use Lemma 6.1 (if $K \to \infty$) or the fact that var $\left(\frac{1}{\sqrt{K}} \sum_{k=1}^{K} e(k) \right) \geqslant c > 0$, if K is bounded. Moreover we get

$$\frac{1}{L} \sum_{l=0}^{L-1} \left[\max_{k=0,\dots,K-1} \left| \frac{1}{\sqrt{K}} \sum_{j=k+1}^{K} e(Kl+j) \right|^{2+\tilde{\delta}} - \mu_K(2+\tilde{\delta}) \right] + \mu_K(2+\tilde{\delta}) \leqslant D \quad a.s. \tag{6.5}$$

Now we are ready to use Theorem 2.1 to arrive at the assertion. First consider (i). Note that the condition includes the case of the null hypothesis (where $d = 0$). Choose the scores $a_n(i) := X(i)$. Without loss of generality assume $\mu = 0$. We will now prove that (2.1) and (2.2) are almost surely fulfilled for these scores. We begin with (2.2). Note that

$$\frac{1}{L} \sum_{l=0}^{L-1} \left(\frac{1}{\sqrt{K}} \sum_{k=1}^{K} \left(X(Kl+k) - \overline{X}_n \right) \right)^2 = \frac{1}{KL} \sum_{l=0}^{L-1} \left(\sum_{k=1}^{K} X(Kl+k) \right)^2 - K\overline{X}_n^2.$$

Moreover it holds $\overline{X}_n = d\,\frac{n-m^*}{n} + \overline{e}_n$ so that equation (6.1) yields

$$\begin{aligned} K\overline{X}_n^2 &= Kd^2 \left(\frac{n-m^*}{n} \right)^2 + 2\sqrt{K}d\frac{n-m^*}{n}\sqrt{K}\,\overline{e}_n + K\,\overline{e}_n^2 \\ &= Kd^2 \left(\frac{n-m^*}{n} \right)^2 + o\left(\sqrt{K}d\frac{n-m^*}{n} \right) + o(1) \quad a.s. \end{aligned}$$

Furthermore equation (6.1) gives as $L \to \infty$

$$\begin{aligned} &\frac{1}{KL} \sum_{l=0}^{L-1} \left(\sum_{k=1}^{K} \left(d1_{\{Kl+k>m^*\}} + e(Kl+k) \right) \right)^2 \\ &= \frac{1}{KL} \sum_{l=0}^{L-1} \left(\sum_{k=1}^{K} e(Kl+k) \right)^2 + \frac{1}{KL} \sum_{l=0}^{L-1} \left(\sum_{k=1}^{K} d1_{\{Kl+k>m^*\}} \right)^2 \\ &\quad + \frac{2}{KL} \sum_{l=0}^{L-1} \left(\sum_{j=1}^{K} d1_{\{Kl+j>m^*\}} \right) \left(\sum_{k=1}^{K} e(Kl+k) \right) \\ &= \frac{1}{KL} \sum_{l=0}^{L-1} \left(\sum_{k=1}^{K} e(Kl+k) \right)^2 + Kd^2\frac{n-m^*}{n} + o\left(\sqrt{|d|^2 K\frac{n-m^*}{n}} \right) + o(1) \quad a.s., \end{aligned}$$

since $\frac{d^2 K}{L} = o(1)$ so that

$$\frac{1}{KL} \sum_{l=0}^{L-1} \left(\sum_{k=1}^{K} d1_{\{Kl+k>m^*\}} \right)^2 = \frac{1}{KL}\, d^2\, K^2\, \frac{n-m^*}{K} + o(1).$$

Equations (6.2) and (6.3) now imply keeping in mind condition (i)

$$
\begin{aligned}
&\frac{2}{KL}\sum_{l=0}^{L-1}\left(\sum_{j=1}^{K} d1_{\{Kl+j>m^*\}}\right)\left(\sum_{k=1}^{K} e(Kl+k)\right)\\
&\ll |d|K\frac{1}{n}\left|\sum_{j=Kl^*+1}^{n} e(j)\right| + K|d|\frac{1}{n}\left|\sum_{k=1}^{K} e(Kl^*+k)\right|\\
&\ll |d|\sqrt{K\frac{n-m^*}{n}}\frac{1}{\sqrt{L(n-m^*)}}\left|\sum_{j=Kl^*+1}^{n} e(j)\right|\\
&\quad+\sqrt{K}|d|\min\left(\frac{n-m^*}{n},\frac{m^*}{n}\right)^{\frac{1}{(2+\tilde{\delta})}}\frac{1}{n^{\frac{\tilde{\delta}}{2(2+\tilde{\delta})}}}\frac{1}{\sqrt{L}}\left|\sum_{k=1}^{K} e(Kl^*+k)\right|\\
&= o\left(\sqrt{|d|^2K\frac{n-m^*}{n}}\right)+o(1) \quad a.s.
\end{aligned}
$$

Putting everything together equation (6.4) gives as $L\to\infty$

$$
\begin{aligned}
&\frac{1}{L}\sum_{l=0}^{L-1}\left(\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\left(X(Kl+k)-\overline{X}_n\right)\right)^2\\
&= C + Kd^2(n-m^*)/n - Kd^2((n-m^*)/n)^2 + o\left(\sqrt{|d|^2K\frac{n-m^*}{n}}\right)+o(1)\\
&= C + Kd^2[(n-m^*)m^*/n^2] + o\left(\sqrt{|d|^2K\frac{n-m^*}{n}}\right)+o(1) \quad a.s.
\end{aligned}
$$

We now use the representation $X(i)=(\mu+d)-d1_{\{i\leqslant m^*\}}+e(i)$ instead of (1.1), thus an analogous calculation yields as $L\to\infty$

$$
\begin{aligned}
&\frac{1}{L}\sum_{l=0}^{L-1}\left(\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\left(X(Kl+k)-\overline{X}_n\right)\right)^2\\
&= C + Kd^2m^*/n - Kd^2(m^*/n)^2 + o\left(\sqrt{|d|^2K\frac{m^*}{n}}\right)+o(1)\\
&= C + Kd^2[(n-m^*)m^*/n^2] + o\left(\sqrt{|d|^2K\frac{m^*}{n}}\right)+o(1) \quad a.s.
\end{aligned}
$$

Together this gives

$$
\begin{aligned}
&\frac{1}{L}\sum_{l=0}^{L-1}\left(\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\left(X(Kl+k)-\overline{X}_n\right)\right)^2\\
&= C + Kd^2\left[(n-m^*)m^*/n^2\right] + o\left(\sqrt{|d|^2K\min\left(\frac{n-m^*}{n},\frac{m^*}{n}\right)}\right)+o(1)\\
&\geqslant C + o(1) \quad a.s.
\end{aligned}
$$

Now we prove that (2.1) holds almost surely for $a_n(i) = X(i)$. Equations (6.1) and (6.5) imply

$$\begin{aligned}
&\frac{1}{L}\sum_{l=0}^{L-1} \max_{k=0,\ldots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K}\left(X(Kl+j)-\overline{X}_n\right)\right|^{2+\tilde{\delta}} \\
&\ll \frac{1}{L}\sum_{l=0}^{L-1} \max_{k=0,\ldots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K} e(Kl+j)\right|^{2+\tilde{\delta}} + |\sqrt{K}\overline{e}_n|^{2+\tilde{\delta}} \qquad (6.6)\\
&\qquad + |\sqrt{K}d|^{2+\tilde{\delta}}((n-m^*)/n)^{2+\tilde{\delta}} + |\sqrt{K}d|^{2+\tilde{\delta}}(n-m^*)/n \\
&\ll 1 + |\sqrt{K}d|^{2+\tilde{\delta}}(n-m^*)/n \quad a.s.
\end{aligned}$$

Analogously we get using the representation $X(i) = (\mu + d) - d1_{\{i\leqslant m^*\}} + e(i)$

$$\frac{1}{L}\sum_{l=0}^{L-1} \max_{k=0,\ldots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K}\left(X(Kl+j)-\overline{X}_n\right)\right|^{2+\tilde{\delta}} \ll 1 + |\sqrt{K}d|^{2+\tilde{\delta}}m^*/n \quad a.s.,$$

which gives

$$\begin{aligned}
&\frac{1}{L}\sum_{l=0}^{L-1} \max_{k=0,\ldots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K}\left(X(Kl+j)-\overline{X}_n\right)\right|^{2+\tilde{\delta}} \\
&\ll 1 + |\sqrt{K}d|^{2+\tilde{\delta}} \min\left(\frac{m^*}{n}, \frac{n-m^*}{n}\right) \ll 1 \quad a.s.
\end{aligned}$$

For the proof of (ii) we have to distinguish two main cases, i.e. $Kd^2 = O(1)$ and $1/(Kd^2) = O(1)$. The first one is included in (i), so we can assume that $1/(Kd^2) = O(1)$. Choosing $a_n(i) := X(i)/\sqrt{Kd^2}$ as scores the proof is analogous to the one above and therefore omitted. For details confer Kirch [10], proof of Theorem 3.5.1. In the case where $d = d(n)$ is such that it follows neither of the above possibilities, we have infinitely many n with $Kd^2 \leqslant 1$ and also infinitely many with $Kd^2 > 1$. Then just choose as scores

$$a_n(i) = \begin{cases} X(i) & Kd^2 \leqslant 1 \\ X(i)/\sqrt{Kd^2} & Kd^2 > 1. \end{cases}$$

As above (2.1) and (2.2) hold almost surely for both subsequences, hence also for the complete sequence. The assertion now follows from Theorem 2.1. ■

A. Some results for strong mixing random sequences

In this appendix we summarize some results on moment inequalities for dependent (strong mixing) random sequences, which also yield a strong law of large numbers for triangular arrays of dependent random variables.

Theorem A.1 (Yokoyama [18], Theorem 1)**.** *Let assumption ($\mathcal{A}$) be fulfilled. Then we have*

$$\mathrm{E}\left|\sum_{i=1}^{n} Z_i\right|^{2+\delta} \leqslant \Gamma(D_1, \tilde{\alpha}, \delta, \Delta)\; n^{(2+\delta)/2},$$

where $\Gamma(D_1, \tilde{\alpha}, \delta, \Delta)$ is a constant just depending on D_1, $\tilde{\alpha}(\cdot)$, δ, and Δ.

Theorem A.2 (Serfling [16], Lemma B, Theorem 3.1). *Under assumption $(\mathcal{A})$ it holds:*

a)

$$\mathrm{E}\left(\max_{l=1,\ldots,n}\left|\sum_{j=1}^{l} Z_j\right|^{2+\delta}\right) \leqslant D\, n^{(2+\delta)/2},$$

where D only depends on δ and the joint distribution of the Z_i.

b)

$$\frac{1}{n}\left|\sum_{j=1}^{n} Z_j\right| = O\left(\frac{(\log n)^{1/(2+\delta)}(\log\log n)^{2/(2+\delta)}}{n^{1/2}}\right) \quad a.s.$$

Remark A.1. Assertion b) of Theorem A.2 remains true for a triangular array that fulfills uniformly the assertion in a).

Corollary A.1. *Let $\{Z_{i,n} : 1 \leqslant i \leqslant n, n \geqslant 1\}$ be a triangular array, which fulfills uniformly condition $(\mathcal{A})$. Then we have as $n \to \infty$*

$$\frac{1}{n}\sum_{i=1}^{n} Z_{i,n} \to 0 \quad a.s.$$

Proof. The assertion is an easy consequence of the Markov inequality.

∎

Acknowledgement

The author is grateful to Professor Marie Hušková from Charles University, Prague, and Professor Josef Steinebach from University of Cologne for many productive discussions and suggestions.

References

[1] Antoch, J. and Hušková, M. Permutation tests for change point analysis. *Statist. Probab. Lett.*, 53:37–46, 2001.

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

[3] Berkes, I., Horváth, L., Hušková, M., and Steinebach, J. Applications of permutations to the simulations of critical values. *J. Nonparametr. Stat.*, 16:197–216, 2003.

[4] Csörgő, M., and Horváth, L. *Weighted Approximations in Probability and Statistics.* Wiley, Chichester, 1993.

[5] Csörgő, M., and Horváth, L. *Limit Theorems in Change-Point Analysis.* Wiley, Chichester, 1997.

[6] Csörgő, M., and Révész, P. *Strong Approximations in Probability and Statistics.* Academic Press, New York, 1981.

[7] Einmahl, U., and Mason, D.M. Approximations to permutation and exchangeable processes. *J. Theoret. Probab.*, 5:101–126, 1992.

[8] Horváth, L. Detection of changes in linear sequences. *Ann. Inst. Statist. Math.*, 49:271–283, 1997.

[9] Hušková, M. Permutation principle and bootstrap in change point analysis. *Fields Inst. Commun.*, 44:273–291, 2004.

[10] Kirch, C. *Resampling Methods for the Change Analysis of Dependent Data.* PhD thesis, University of Cologne, Cologne, 2006. http://kups.ub.uni-koeln.de/volltexte/2006/1795/.

[11] Kirch, C. Simulation study of the block permutation method as well as frequency permutation method for the change analysis of dependent data. http://www.mi.uni-koeln.de/~jost/simulations.pdf, 2006. University of Cologne, 109 p.

[12] Kirch, C., and Steinebach, J. Permutation principles for the change analysis of stochastic processes under strong invariance. *J. Comput. Appl. Math.*, 186:64–88, 2006.

[13] Loève, M. *Probability Theory I.* Springer, New York, 1963.

[14] Phillips, P.C.B., and Solo, V. Asymptotics for linear processes. *Ann. Statist.*, 20:971–1001, 1992.

[15] Politis, D.N., and Romano, J.P. A circular block-resampling procedure for stationary data. In LePage, R. and Billard, L., editors, *Exploring the Limits of Bootstrap*, pages 263–270, New York, 1992. Wiley.

[16] Serfling, R.J. Convergence properties of s_n under moment restrictions. *Ann. Math. Statist.*, 41:1235–1248, 1970.

[17] Withers, C.S. Conditions for linear processes to be strong-mixing. *Z. Wahrsch. verw. Geb.*, 57:477–480, 1981.

[18] Yokoyama, R. Moment bounds for stationary mixing sequences. *Z. Wahrsch. verw. Geb.*, 52:45–57, 1980.