Resampling Methods for the Change Analysis of Dependent Data

I n a u g u r a l – D i s s e r t a t i o n

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Claudia Kirch

aus Saarbrücken

Köln, März 2006

1. Berichterstatter:	Prof. Dr. Josef Steinebach, Universität zu Köln
2. Berichterstatter:	Prof. Dr. Wolfgang Wefelmeyer, Universität zu Köln
3. Berichterstatterin:	Prof. Dr. Marie Hušková, Karls-Universität, Prag

Tag der mündlichen Prüfung: 14.06.2006

Change alone is eternal, perpetual, immortal.

Arthur Schopenhauer (1788-1860)

Abstract

The fundamental question in change-point analysis is whether an observed stochastic process follows one model or whether the underlying model changes at least once during the observational period. Thus the field can essentially be divided into two subfields, hypothesis testing and point estimating. This thesis deals with the first problem.

Most of the older works discuss independent observations, yet from a practical point of view cases of dependent data have become more and more important. We develop testing procedures for dependent models.

In change-point analysis critical values for testing procedures are usually obtained by distributional asymptotics. These critical values, however, do not sufficiently reflect dependency. Moreover it is a well-known fact that convergence rates especially for extreme-value statistics are very slow. Using resampling methods we obtain better approximations, which take possible dependency structures more efficiently into account. We prove that the original statistics and their resampling counterparts follow the same distributional asymptotics. First we obtain limit theorems for the corresponding rank statistics, which then combined with laws of large numbers imply the resampling asymptotics conditionally on the given data.

In a first part we consider abrupt and gradual changes in models of possibly dependent observations satisfying a strong invariance principle.

The main part of this thesis studies a location model with dependent errors that form a linear process. Different types of statistics are considered, such as maximum-type statistics (particularly different CUSUM procedures) or sum-type statistics. The resampling-methods have to be adapted to allow for dependent errors. Thus, we analyze a block bootstrap as well as a bootstrap in the frequency domain.

Finally, some simulation studies illustrate that the permutation tests usually behave better than the original tests if performance is measured by the $\alpha-$ and $\beta-$errors, respectively.

Zusammenfassung

Die Changepoint Analyse beschäftigt sich mit der Fragestellung, ob ein beobachteter stochastischer Prozess einem festen Modell folgt oder ob sich das zu Grunde liegende Modell einmal oder mehrmals während des Beobachtungszeitraums ändert. Im Wesentlichen ergeben sich hieraus zwei Teilgebiete, wovon eines sich mit der Entwicklung statistischer Tests das andere mit dem Schätzen der Stelle, an der der Strukturbruch stattgefunden hat, beschäftigt. Diese Arbeit beschäftigt sich mit dem ersten Problem.

Eine große Mehrheit der älteren Arbeiten behandelt unabhängige Beobachtungen. Der Fall abhängiger Daten ist jedoch für praktische Zwecke immer wichtiger geworden. Wir entwickeln Test-Verfahren in abhängigen Modellen.

Die Festlegung kritischer Werte für Testverfahren zur Aufdeckung von Strukturbrüchen erfolgt häufig auf der Basis von Verteilungsasymptotiken. Die so gewonnenen kritischen

Werte beziehen jedoch mögliche Abhängigkeitsstrukturen nicht in ausreichendem Maße ein. Außerdem sind die Konvergenzraten vor allem bei Extremwertasymptotiken bekanntermaßen sehr langsam. Durch den Einsatz von Resampling-Methoden können bessere Approximationen erzielt werden, die auch mögliche Abhängigkeitsstrukturen besser abbilden. Wir zeigen, dass die ursprünglichen Teststatistiken und die dazugehörigen Resampling-Statistiken der gleichen Verteilungsasymptotik folgen. Hierzu werden erst entsprechende Resultate für die dazugehörigen Rangstatistiken bewiesen. Aus diesen lassen sich dann unter Zuhilfenahme von Gesetzen der Großen Zahlen entsprechende Grenzwertsätze für die auf den Beobachtungen bedingten Resampling-Statistiken herleiten.

In einem ersten Teil betrachten wir abrupte sowie graduelle Strukturbrüche in Modellen von möglicherweise abhängigen Beobachtungen, die ein starkes Invarianzprinzip erfüllen.

Der Hauptteil dieser Arbeit beschäftigt sich mit einem Lokationsmodell mit abhängigen Fehlern, die einen linearen Prozess bilden. Hierbei benutzen wir verschiedene Statistiken wie zum Beispiel Maximum-Statistiken (insbesondere verschiedene CUSUM Verfahren) oder auch Summen-Statistiken. Die Resampling-Methoden müssen hierbei an den abhängigen Fall angepasst werden. Hierzu betrachten wir einerseits einen Block Bootstrap, andererseits einen Bootstrap im Frequenzbereich.

Schließlich zeigen Simulationsstudien, dass obige Resampling-Tests im Gegensatz zu den ursprünglichen Tests das Niveau sehr genau einhalten - und gleichzeitig über eine verhältnismäßig bessere Güte verfügen.

Contents

Part I.

Theoretical Results

1. Introduction

1.1. Change-Point Analysis

Change alone is eternal, perpetual, immortal.[1]

Indeed the world is filled with changes. We encounter them – or the possibility of them – everyday in such diverse fields as economics, finance, medicine, geology, physics and so on. Therefore the detection, location and investigation of changes is of particular human interest.

Change-point analysis provides statistical tools to decide whether a given (ordered) data set remains stable over time or whether it follows a certain model up to an unknown time-point and a different model afterwards. Usually this means that some parameters in a given model are subject to change.

The earliest change-point studies go back to the 1950s (cf. Page [67, 68, 69]), where they arose in the context of quality control. There, one usually observes the output of a production line and assumes that a certain characteristic varies around a certain *in-control* constant. Sometimes this characteristic suddenly starts to vary around another *out-of-control* constant, for example due to a failure of the production device. One then wants to know if and when such a change occurred to take appropriate measures. It is important that in the above setting the location of the change is unknown, otherwise the situation reduces to well-known two-sample problems.

Since then many articles have been published and the list of applications is becoming longer and longer. Many of them cover the topic of a single change-point in the mean (at-most-one-change or AMOC location model) for an independent sequence of random variables. Of course in many situations dependent observations are much more realistic, hence we will focus on that case in this thesis.

Recently, there has been an increased interest in the statistical analysis of change-point detection and estimation. This is probably due to the fact that with the growing field of information technology more and more data is collected, which can be (needs to be) analyzed.

There are essentially two aspects of change-point analysis. The first one is to detect whether there is a change at all. The second one is to estimate how many changes have occurred and where. We will concentrate on the first problem.

Statistical procedures in change-point analysis can further be divided into two main categories. In the *a-posteriori* analysis we have already observed past data and now

[1] accredited to the German philosopher Arthur Schopenhauer (1788-1860)

want to know whether it contains a change. As a contrast the *sequential* approach regularly takes new data and after each observation applies a new test. This way we hope to be warned as fast as possible after the change occurred.

Sequential analysis has recently received a lot of attention since it is more realistic in many situations again especially in view of the growing field of information technology. Most of the older papers, however, are dealing with a-posteriori methods which remain an active field of scientific research and which have proven to be very useful in the past.

Here, we will review a-posteriori methods, apply them for dependent data and improve the existing tests by applying resampling methods to determine the critical values.

A typical very general mathematical formulation of the change problem can be expressed as follows: Let $X(1), \ldots, X(n)$ be the observed data. We now would like to decide whether the null hypothesis of no change

$$H_0 : X(1) \overset{\mathcal{D}}{=} X(2) \overset{\mathcal{D}}{=} \ldots \overset{\mathcal{D}}{=} X(n)$$

holds or whether the data follows the alternative

$$H_1 : \text{There exists } 1 \leqslant i < j \leqslant n : X(i) \overset{\mathcal{D}}{\neq} X(j),$$

i.e. there is some kind of change in distribution.

Many problems can be transformed into a form that matches the above description. Let us assume for example we have observed a financial time series with a linear drift, i.e. $Y(i) = a\frac{i}{n} + e(i), i = 1, \ldots, n$, $Y(0) = 0$, for some stationary error sequence $\{e(\cdot)\}$. We are interested in the question whether the linear drift changes over time, i.e. whether it holds $Y(i) = a\frac{m}{n} + a^*\frac{i-m}{n} + e(i)$, $a \neq a^*$, for $i > m$. Here, $X(i) = Y(i) - Y(i-1)$, $i = 1, \ldots, n$, matches the above description. This is essentially the kind of problem we encounter in Chapter 2.

As already mentioned an important submodel – probably the most common one – is the AMOC-location model, which is given by

$$X(j) = \begin{cases} \mu + e(j), & 1 \leqslant j \leqslant m, \\ \mu + d + e(j), & m < j \leqslant n, \end{cases}$$

where $m = m(n)$ is the unknown change-point, $d = d_n$ the mean change, $\{e(\cdot)\}$ a centered error sequence. This error sequence is usually assumed to be a sequence of i.i.d. random variables with $\mathrm{E}\,|e(1)|^\nu < \infty$ for some $\nu > 2$. We are now interested in testing the null hypothesis of no change, i.e.

$$H_0 : m = n,$$

versus the alternative of a change in the mean, i.e.

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

Chapters 3 and 4 deal with this problem. There we drop the assumption of independence and only assume that the error sequence follows a linear process.

The above setting deals with an abrupt change, i.e. the mean jumps all of a sudden from μ to $\mu + d$ at time-point m. In some situations gradual changes, where the mean changes slowly from μ to $\mu + d$, are more realistic. One of the models in Chapter 2 deals with this kind of changes.

A detailed discussion of the field of change-point analysis can be found in the book of Chen and Gupta [16], which specializes in parametric models, and more importantly in the book of Csörgő and Horváth [19] who also take non-parametric models into account. Another very good introduction into the field is the paper by Antoch et al. [3]. The important models and tools are introduced and relevant references given.

1.2. Resampling Methods in Change-Point Analysis

One of the major problems in hypothesis testing is finding good approximations to the critical values. Assume we have found a reasonable test statistic to one of our change-point problems above. This can for example be done by using *pseudo maximum-likelihood* or *pseudo Bayes* methods (confer Section 3.2). However, in practice the distributions of these statistics – even for normal errors – are very complex, so that they can be computed explicitly only for small sample sizes. Another possibility is to use the Bonferroni inequality to derive an upper bound for the quantiles. But again this approach only gives satisfactory results for small samples. For details confer Hawkins [43] or Antoch et al. [3].

Critical values are thus often obtained by distributional asymptotics under the null hypothesis. One problem with that approach, however, is that the limit distribution sometimes depends on unknown parameters or its quantiles are not theoretically known so that one needs to use simulations. Another severe problem is that the convergence is often very slow, especially for extreme value statistics. This means that the asymptotic critical values are only good approximations for large sample sizes, otherwise they fail. Thus the test is usually not exact and might not even hold the chosen level. This is also confirmed by our simulation studies in Part II.

Therefore Antoch and Hušková [2] proposed resampling methods to approximate the critical values of an AMOC location model. More precisely they proposed a permutation test. Since then the method has been applied to all kinds of problems in change-point analysis such as kernel type statistics, U-statistics or linear regression models. For a recent survey see Hušková [47]. A thorough introduction to resampling methods can be found in the books of Good [35, 36].

How Do Permutation Tests Work?

Let $X(1), \dots, X(n)$ be a sequence of random variables. Moreover let us assume $T_n = T_n(X(1), \dots, X(n))$ is our test statistic. We reject the null hypothesis for large values thereof. Consider the statistic of the permuted observations

$$T_n^{\mathbf{R}} := T_n(X(R_1), \dots, X(R_n)).$$

$\mathbf{R} = (R_1, \dots, R_n)$ is a random permutation of $(1, \dots, n)$ independent of $\{X(\cdot)\}$ such that $P(\mathbf{R} = r) = \frac{1}{n!}$ for all permutations $r = (r_1, \dots, r_n)$.

Note that we can easily calculate (or rather simulate for computational matters) the exact distribution of $T_n^{\mathbf{R}}$ conditioned on the observed data sequence $X(1), \ldots, X(n)$. There the randomness only comes from the permutation. So we reject the null hypothesis if $T_n(X(1), \ldots, X(n))$ is larger than the α-quantile of $T_n^{\mathbf{R}|X}$, i.e. $T_n^{\mathbf{R}}$ conditioned on $X(1), \ldots, X(n)$.

The computational costs for exactly calculating $T_n^{\mathbf{R}}$ are too high, so for practical purposes we will use its empirical distribution function based on N random permutations.

For clarification we will state the algorithm here:

1) Calculate the value of the statistic for the given observations: $T := T_n(X(1), \ldots, X(n))$.
2) Calculate $T_n^{\mathbf{R}^{(1)}} = T_n(X(R_1^{(1)}), \ldots, X(R_n^{(1)}))$ for some random permutation $\mathbf{R}^{(1)} = (R_1^{(1)}, \ldots, R_n^{(1)})$.
3) Repeat step 2) N times, e.g. $N = 10\,000$.
4) Calculate the α-quantile of $T_n^{\mathbf{R}^{(1)}}, \ldots, T_n^{\mathbf{R}^{(N)}}$ from Step 3), i.e. choose c minimal such that
$$\frac{1}{N} \sum_{j=1}^{N} 1_{\{T_n^{\mathbf{R}^{(j)}} \leqslant c\}} \geqslant 1 - \alpha.$$
5) Reject the null hypothesis for $T > c$.

The above test is in the non-randomized form, in the randomized form we reject the null hypothesis also for $T = c$ with an appropriately chosen probability γ. As we will see below, we then get an exact test under certain conditions on the error sequence. Yet, we only make a small mistake using the non-randomized form (same as with the empirical distribution function) and both versions are asymptotically equivalent in our examples, since the limiting distribution of T_n is continuous. This is why we will use the above version for practical purposes.

Formally the permutational test with level α is defined as follows:

$$\varphi^{\mathbf{R}}(X(1), \ldots, X(n)) = \begin{cases} 1, & > \\ \gamma, \qquad T_n(X(1), \ldots, X(n)) & = \quad c(X(\cdot)), \\ 0, & < \end{cases}$$

where $c(X_{(\cdot)}), \gamma$ are chosen such that

$$P\left(T_n^{\mathbf{R}} > c(X(\cdot)) \,|\, X(1), \ldots, X(n)\right) + \gamma P\left(T_n^{\mathbf{R}} = c(X(\cdot)) \,|\, X(1), \ldots, X(n)\right) = \alpha.$$

The next lemma shows that the above algorithm really corresponds to this test.

Lemma 1.2.1. *If* $\mathbf{R} = (R_1, \ldots, R_n)$ *and* $\{X(\cdot)\}$ *are independent, we have for any* $x \in \mathbb{R}$

$$P\left(T_n^{\mathbf{R}} \leqslant x \,|\, X(1), \ldots, X(n)\right) = \frac{1}{n!} \sum_{r \in \mathcal{R}_n} I_{\{T_n^r \leqslant x\}},$$

where $\mathcal{R}_n$ *is the set of all permutations of* $(1, \ldots, n)$.

Proof.

$$
\begin{aligned}
&P\left(T_n^{\mathbf{R}} \leqslant x \mid X(1),\ldots,X(n)\right)\\
&= \sum_{r\in\mathcal{R}_n} \mathrm{E}\left(1_{\{T_n(X(r_1),\ldots,X(r_n))\leqslant x\}}\, 1_{\{\mathbf{R}=r\}} \mid X(1),\ldots,X(n)\right)\\
&= \sum_{r\in\mathcal{R}_n} 1_{\{T_n(X(r_1),\ldots,X(r_n))\leqslant x\}}\, \mathrm{E}\left(1_{\{\mathbf{R}=r\}} \mid X(1),\ldots,X(n)\right)\\
&= \sum_{r\in\mathcal{R}_n} 1_{\{T_n(X(r_1),\ldots,X(r_n))\leqslant x\}}\, P\left(\mathbf{R}=r\right) = \frac{1}{n!}\sum_{r\in\mathcal{R}_n} I_{\{T_n^r\leqslant x\}}.
\end{aligned}
$$

■

Moreover the next lemma states that the above test is exact if the observations are exchangeable under H_0, i.e. in particular if they are i.i.d. random variables. For dependent data, however, this is not true in general.

Lemma 1.2.2. *If $(X(1),\ldots,X(n))$ are exchangeable under H_0, the test $\varphi^{\mathbf{R}}$ is exact.*

Proof. First note that the exchangeability gives $T_n \overset{\mathcal{D}}{=} T_n^{\mathbf{R}}$. Hence it holds

$$
\begin{aligned}
&\mathrm{E}\left(\varphi^{\mathbf{R}}(X(1),\ldots,X(n))\right)\\
&= P\left(T_n > c(X(\cdot))\right) + \gamma P\left(T_n = c(X(\cdot))\right)\\
&= P\left(T_n^{\mathbf{R}} > c(X(\cdot))\right) + \gamma P\left(T_n^{\mathbf{R}} = c(X(\cdot))\right)\\
&= \mathrm{E}\left[P\left(T_n^{\mathbf{R}} > c(X(\cdot)) \mid X(1),\ldots,X(n)\right) + \gamma P\left(T_n^{\mathbf{R}} = c(X(\cdot)) \mid X(1),\ldots,X(n)\right)\right]\\
&= \alpha.
\end{aligned}
$$

■

Yet, the critical values depend on our observations, so we would like to verify that their limit behavior equals that of the unconditional critical values.
Also, the question remains how they behave (asymptotically) under alternatives. It turns out that the limit behavior under both – the null hypothesis as well as alternatives – matches the limit behavior of the unconditioned critical values under H_0.
Thus we get an approximation for the critical values corresponding to the null distribution, even if the observed data does follow an alternative.

To clarify matters: Usually there exists a random variable Y, such that under H_0 it holds $T_n \xrightarrow{\mathcal{D}} Y$. We then prove that under H_0 *as well as* H_1 it holds for $x \in C_Y$ (the points of continuity of the distribution function of Y)

$$
P\left(T_n^{\mathbf{R}} \leqslant x \mid X(1),\ldots,X(n)\right) \longrightarrow P\left(Y \leqslant x\right) \quad a.s.
$$

The main tool in proving such a limit theorem is the corresponding result for a linear rank statistic, where $X(i)$ is replaced by scores $a_n(i)$ with certain properties. Such results can be obtained by investigating functionals of simple linear rank statistics. There is a vast amount of literature on that topic, above all the book by Hájek et al. [40]. Once we have a limit theorem for the corresponding rank statistics it suffices to prove that $X(\cdot)$ fulfills the conditions on the scores in an $a.s.$-sense. This usually reduces to proving strong laws of large numbers. If these only hold true in a P-stochastic sense, it is usually still possible to obtain the above result but only in a P-stochastic sense using the subsequence principle.

This is the approach we use in Chapter 2. There we have possibly dependent observations, but they are – in a certain sense – still close to the independent case, so the above methods work as presented here. In Chapters 3 and 4 we consider the AMOC location model with dependent errors. There, we have to adapt the above method to allow for the dependency structure. In Chapter 3 this is done by using the so-called block permutation method, whereas in Chapter 4 the problem is solved by permuting the Fourier frequencies of the process rather than the observations themselves.

An Example: AMOC Location Model with i.i.d. Errors

Now we give an easy example to illustrate the above principles. Therefore we choose the AMOC location model with i.i.d. errors, i.e. $X(i) = \mu + d_n 1_{\{i>m\}} + e(i)$, $d_n \neq 0$. Here $\{e(\cdot)\}$ are i.i.d. random variables satisfying

$$\mathrm{E}\, e(1) = 0, \qquad 0 < \operatorname{var} e(1) < \infty, \qquad \mathrm{E}\, |e(1)|^{\nu} < \infty \quad \text{for some } \nu > 2.$$

The test problem we are interested in is

$$H_0 : m = n \qquad \text{against} \qquad H_1 : m < n.$$

We are illustrating the idea with the classical CUSUM statistic, i.e.

$$T_n := \max_{1 \leqslant m \leqslant n} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{m} (X(i) - \bar{X}_n) \right|,$$

where $\bar{X}_n := \frac{1}{n} \sum_{i=1}^{n} X(i)$.

Antoch and Hušková [2] introduced the permutation method to change-point analysis using the above problem but the weighted CUSUM statistic. This is somewhat more complicated than our example because it deals with an extreme-value statistic and the underlying rank theory is deeper than in our case. The proof for the permutation statistic, however, is the same in both cases.

It is a well-known fact (confer for example Csörgő and Horváth [19], Chapter 2) that it holds under H_0

$$\frac{T_n}{\widehat{\sigma}_n} \xrightarrow{\mathcal{D}} \sup_{0 \leqslant t \leqslant 1} |B(t)|,$$

where $\{B(\cdot)\}$ is a Brownian bridge and $\widehat{\sigma}_n^2 := \frac{1}{n} \sum_{j=1}^{n} (X(j) - \bar{X}_n)^2$.

Let $T_n^{\mathbf{a}} := \max_{1 \leqslant m \leqslant n} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{m} (a_n(R_i) - \bar{a}_n) \right|$, where $a_n(\cdot)$ are scores. The following lemma gives the limit behavior of the rank statistic.

Lemma 1.2.3. *If*

$$\sigma_n^2(\mathbf{a}) := \frac{1}{n} \sum_{i=1}^{n} (a_n(i) - \bar{a}_n)^2 \geqslant D_1 > 0 \tag{1.2.1}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}|a_n(i)-\bar{a}_n|^{\nu} < D_2 < \infty \qquad \textit{for some } \nu > 2, \tag{1.2.2}$$

it holds as $n \to \infty$

$$\frac{1}{\sigma_n(\mathbf{a})}T_n^{\mathbf{a}} = \max_{1\leqslant m\leqslant n}\frac{1}{\sqrt{n\,\sigma_n^2(\mathbf{a})}}\left|\sum_{i=1}^{m}(a_n(R_i)-\bar{a}_n)\right| \xrightarrow{\mathcal{D}} \sup_{0\leqslant t\leqslant 1}|B(t)|,$$

where $\{B(\cdot)\}$ *is a Brownian bridge.*

Proof. This follows for example from Billingsley [9], Theorem 24.2 (p. 212). For details confer Kirch [50], Corollary 5.3.1 b).
It is also possible to obtain the result using results from linear rank statistics (convergence in $C[0,1]$, tightness) in a similar way as in Section 4.5 or from Corollary D.2. ∎

We can now use the above rank statistic result and strong laws of large numbers to obtain the desired limit theorem for the permutation statistic under H_0 as well as H_1.

Let $T_n^{\mathbf{R}} = \max_{1\leqslant m\leqslant n}\frac{1}{\sqrt{n}}\left|\sum_{i=1}^{m}(X(R_i)-\bar{X}_n)\right|$ be the permutation statistic. The following theorem by Antoch and Hušková [2] gives its limit behavior.

Theorem 1.2.1. *Under the above assumptions and if additionally* $d_n \leqslant D_3 < \infty$*, it holds as* $n \to \infty$ *for all* $x \in \mathbb{R}$

$$P\left(\frac{1}{\widehat{\sigma}_n}T_n^{\mathbf{R}} \leqslant x \,\middle|\, X(1),\dots,X(n)\right) \longrightarrow P\left(\sup_{0\leqslant t\leqslant 1}|B(t)| \leqslant x\right) \qquad a.s.,$$

where $\{B(\cdot)\}$ *is a Brownian bridge and* $\widehat{\sigma}_n^2 := \frac{1}{n}\sum_{j=1}^{n}(X(j)-\bar{X}_n)^2$.

Proof. We can apply Lemma 1.2.3 with $a_n(i) := X(i)$, $i = 1,\dots,n$. To get the desired result it is sufficient to check that (1.2.1) and (1.2.2) are satisfied almost surely for these scores. Note

$$\frac{1}{n}\sum_{i=1}^{n}(X(i)-\bar{X}_n)^2 = \frac{1}{n}\sum_{i=1}^{n}(e(i)-\bar{e}_n)^2 - 2d_n\frac{1}{n}\sum_{i=1}^{m}(e(i)-\bar{e}_n) + d_n^2\frac{m(n-m)}{n^2}.$$

Hence the classical law of large numbers implies

$$\frac{1}{n}\sum_{i=1}^{n}(X(i)-\bar{X}_n)^2 \geqslant \operatorname{var}(e(1)) + o(1) \qquad a.s.$$

Similarly, we get

$$\frac{1}{n}\sum_{i=1}^{n}|X(i)-\bar{X}_n|^{\nu} \leqslant C\,(\mathrm{E}\,|e(1)|^{\nu} + D_3^{\nu}) + o(1) \qquad a.s.$$

for some constant $C < \infty$. For more details confer Kirch [50], Theorem 8.0.1. These relations ensure that the assumptions of Lemma 1.2.3 are fulfilled, thus the assertion follows. ∎

Bootstrap With Replacement

The permutation test from above can be interpreted as bootstrap without replacement. We are then interested in the conditional limit behavior of statistic $T_n^{\mathbf{U}} := T_n(X(U_1), \ldots, X(U_n))$, where $\mathbf{U} = (U_1, \ldots, U_n)$ forms a triangular array of row-wise i.i.d. random variables with $P(U_1 = j) = \frac{1}{n}$, $j = 1, \ldots, n$. Moreover we require that $\mathbf{U}$ is independent of $\{X(\cdot)\}$. Here, the corresponding test is usually not exact.

Usually this approach gives the same results as above. This is, however, hardly surprising considering that one often proves rank statistic results by deriving them from the corresponding results for the statistic with replacement. One example is the proof for the Lindeberg-Condition of rank statistics, cf. Theorems 3.1 and 4.1 of Hájek [39].

In this work we will concentrate on the permutation test. Yet, we always remark how to adapt the proofs to obtain the results for the bootstrap with replacement.

Small Sample Behavior and Simulations

After we have proven results of the above type for a given model (as e.g. Theorem 1.2.1 for the AMOC location model and the CUSUM statistic), we know that the permutation test is a valid test just like the asymptotic one. However, we still do not know whether it really is better than the asymptotic test regarding its small sample behavior if measured by α- respectively $\beta-$errors. To find out about that it is standard procedure to conduct a simulation study. This is done in Part II of this thesis.

There is little literature containing theoretical results showing that the small sample behavior of resampling methods is better than the one of the original asymptotic test. However, Berkes et al. [7] investigated convergence rates for different permutation statistics. For the weighted CUSUM statistic they obtained a better rate for the difference of the distribution functions of the permutation statistic and of the original statistic under H_0 than for the convergence under the null hypothesis.

1.3. Organization of the Material

In this thesis we investigate resampling methods for different change-point models with dependent errors.
In the main part we prove the asymptotic validity of the proposed methods. In Chapter 2 we investigate abrupt as well as gradual changes in a model of possibly dependent data satisfying a strong invariance principle. The statistics are based on the increments of the processes. Because the distance between observations increases over time the dependency structure is captured in the statistic. This is the reason why in this model the permutation method as described in the previous chapters works without adjustment. Many important examples satisfy the model assumptions including linear processes which we will investigate further in the following two chapters.

In Chapters 3 and 4 we turn our attention to the classical location model. We introduce common statistics that were originally developed for independent errors. They are used

frequently for practical purposes even when the assumption of independence of the observations is not realistic in many real-life situation. Antoch et al. [4] and Horváth [45] prove that the limit distributions of the statistics essentially remain the same if the errors are linear processes instead of independent random variables. Since convergence is rather slow and also the asymptotic quantiles are not always theoretically known, we propose resampling methods to derive critical values. Because just permuting the random variables does not sufficiently capture the dependency structure of the errors we have to adapt the methods.
In Chapter 3 we investigate a block permutation method. There we consider blocks of successive observations and permute the blocks but keep the order within the blocks. The idea is that the dependency structure is captured within the block.
In Chapter 4 we resample the Fourier frequencies of the estimated underlying linear process rather than the observations themselves. It is known that finitely many Fourier frequencies are asymptotically independent and normally distributed. Since resampling procedures work very well for independent data the idea is that permuting the Fourier frequencies will also work well.
In all three chapters we first prove the corresponding rank statistic result. We then derive the asymptotic of the permutation statistics from them. It turns out that the obtained critical values are indeed good approximations of the critical values for the original statistics under the null distribution. This is even true when the observations follow an alternative.
This confirms that the resampling methods are theoretically valid. We do, however, not know how well they perform in comparison to the asymptotic test.

This is why we give the results of a simulation study in a second part. The simulations belonging to the model from Chapter 2 implement partial sums as well as renewal processes and are given in Chapter 5.
The simulation study for the location model with an error sequence that forms a linear process is given in Chapter 6. We use the example of different causal AR(1)-sequences. It turns out that the resampling methods usually behave better than the asymptotic method if performance is measured by $\alpha-$ and $\beta-$errors, respectively.

Finally we summarize some frequently used results, inequalities and methods in an appendix. A first chapter gives a short introduction into stochastic Landau symbols and their properties. We use them frequently throughout this work without comment.
Moreover we develop some Hájek–Rényi–type inequalities for dependent random variables as well as moment inequalities for sums of dependent data which lead to (strong) laws of large numbers.
In Appendix C we give a short introduction into the Beveridge-Nelson decomposition which is a very useful tool in proving strong laws of large numbers for linear processes. We also give some results obtained by Phillips and Solo [71] by exploiting this decomposition.
The next two appendices deal with an embedding for permutation and exchangeable processes by Einmahl and Mason [28] and some properties of simple linear rank statistics. Both are useful to obtain the correct rank statistic results.
A final appendix states some results from change-point analysis that are used frequently in this work.

2. Change Analysis of Stochastic Processes under Strong Invariance

Developing testing procedures in models of dependent data is becoming more and more important considering nearly all real-life data is dependent.

In this chapter we develop a first permutation procedure for the change analysis of possibly dependent data. This model is, however, still close to the independent case due to the fact that the statistics consider observations that grow further and further apart. This is why we can use the permutation procedures as described in Section 1.2.

More precisely we deal with a change in the drift as well as variance of a stochastic process fulfilling a strong invariance principle. Abrupt as well as gradual changes are studied. The statistics involved are based on the increments of the underlying process which are asymptotically independent.

Simulation results in Chapter 5 confirm that the permutation tests usually behave better than the original tests if performance is measured by the $\alpha-$ and $\beta-$errors, respectively.

The rank statistic results as well as the results concerning the abrupt change have already been part of the author´s diploma thesis [50], although in a somewhat less refined version. The results of this chapter have been published as a joint article with Josef Steinebach, University of Cologne, [52].

This chapter is organized as follows:
The first section describes the models we are using and also specifies examples of processes included in this setting. Furthermore we introduce the statistics to detect abrupt or gradual changes as well as their null asymptotics.
Since the corresponding rank asymptotics are essential in the proof of the permutation tests we develop them in Section 2.2.
In a final section we then introduce the permutation statistics and give their limit distribution, which shows that the permutation test as described in Section 1.2 works.

2.1. Models, Statistics and their Null Asymptotics

First we give a mathematical description of the models we use and illustrate that they include a broad range of applications. Furthermore the statistics and their null asymptotics are given.

Abrupt Change in the Mean or Variance of a Stochastic Process Under Strong Invariance

A main tool in change-point analysis is to make use of invariance principles for the observed sequence and develop asymptotic tests based on the approximating process. This idea is also pursued in the following model by Horváth and Steinebach [46].

Suppose one observes a stochastic process $\{Z(t) : 0 \leqslant t < \infty\}$ having the following structure:

$$Z(t) = \begin{cases} at + bY(t), & 0 \leqslant t \leqslant T^*, \\ Z(T^*) + a^*(t - T^*) + b^*Y^*(t - T^*), & T^* < t \leqslant T, \end{cases} \tag{2.1.1}$$

where $a = a_T, a^* = a_T^*$ are unknown parameters, $b \neq 0, b^* \neq 0$ unknown constants, and $\{Y(t) : 0 \leqslant t < \infty\}$ resp. $\{Y^*(t) : 0 \leqslant t < \infty\}$ are (unobserved) stochastic processes satisfying the following strong invariance principles:

For every $T > 0$, there exist two independent Wiener processes $\{W_T(t) : 0 \leqslant t \leqslant T^*\}$ and $\{W_T^*(t) : 0 \leqslant t \leqslant T - T^*\}$, and some $\nu > 2$, such that, for $T \to \infty$,

$$\sup_{0 \leqslant t \leqslant T^*} |Y(t) - W_T(t)| = O\left(T^{1/\nu}\right) \qquad a.s. \tag{2.1.2}$$

and

$$\sup_{0 \leqslant t \leqslant T - T^*} |Y^*(t) - W_T^*(t)| = O\left(T^{1/\nu}\right) \qquad a.s. \tag{2.1.3}$$

Moreover, we assume $Y(0) = 0$ and $Y^*(0) = 0$. It should be noted that only weak invariance has been assumed in Horváth and Steinebach [46], instead of the strong rates of (2.1.2) and (2.1.3), which are required for later use here. Furthermore, the processes $\{Z(\cdot)\}, \{Y(\cdot)\}$, and $\{Y^*(\cdot)\}$ could be replaced by a family of processes $\{Z_T(\cdot)\}, \{Y_T(\cdot)\}$, and $\{Y_T^*(\cdot)\}$, $T > 0$, since the asymptotic analysis is merely based on the approximating family of Wiener processes $\{W_T(\cdot)\}$ and $\{W_T^*(\cdot)\}$, respectively.

One is interested in testing the hypothesis of "no change", i.e.

$$H_0 : T^* = T\,,$$

against the alternative of "a change in the mean at $T^* \in (0, T)$", i.e.

$$H_1^{(1)} : 0 < T^* < T \quad \text{and} \quad a \neq a^*\,,$$

respectively "a change in the variance at $T^* \in (0, T)$", i.e.

$$H_1^{(2)} : 0 < T^* < T \quad \text{and} \quad b \neq b^*, \;\text{ but } a = a^*.$$

We will now discuss some basic examples satisfying conditions (2.1.1) - (2.1.3) (for details we refer to Horváth and Steinebach [46]).

Example 2.1.1 (Partial sums). Let $\{X(i) : i \geqslant 1\}$ and $\{X^*(i) : i \geqslant 1\}$ be two independent sequences of i.i.d. random variables with $\mathrm{E}\, X(1) = \mu$, $\operatorname{var} X(1) = \sigma^2 > 0$

respectively $\mathrm{E}\, X^*(1) = \mu^*$ and $\operatorname{var} X^*(1) = \sigma^{*2} > 0$. Consider $Z(t) = S_{[t]}$, where $S_0 = 0$ and

$$S_k = \begin{cases} X(1) + X(2) + \ldots + X(k), & 1 \leqslant k \leqslant T^*, \\ S_{[T^*]} + X^*(1) + X^*(2) + \ldots + X^*(k - [T^*]), & T^* < k \leqslant T. \end{cases}$$

If $\mathrm{E}\,|X(1)|^\nu < \infty$ and $\mathrm{E}\,|X^*(1)|^\nu < \infty$ for some $\nu > 2$, then according to Komlós et al. and Major [54, 55, 62] there exist independent Wiener processes $\{W(t) : t \geqslant 0\}$, $\{W^*(t) : t \geqslant 0\}$, such that

$$\left| \frac{1}{\sigma} \sum_{i=1}^{k} [X(i) - \mu] - W(k) \right| = o\left(k^{1/\nu}\right), \quad \left| \frac{1}{\sigma^*} \sum_{i=1}^{k} [X^*(i) - \mu] - W^*(k) \right| = o\left(k^{1/\nu}\right) \quad a.s.$$

This yields (2.1.1) - (2.1.3) with $a = \mu$, $b = \sigma$, $Y(t) = (Z(t) - \mu t)/\sigma$, $a^* = \mu^*$, $b^* = \sigma^*$ and $Y^*(t - T^*) = (Z(t) - Z(T^*) - \mu^*(t - T^*))/\sigma^*$.

Example 2.1.2 (Renewal processes). Let $\{X(\cdot)\}$ and $\{X^*(\cdot)\}$ be as in Example 2.1.1. Furthermore $\mu, \mu^* > 0$. Consider

$$Z(t) = \begin{cases} N_1(t), & 0 \leqslant t \leqslant T^*, \\ N_1(T^*) + N_2(t - T^*), & T^* < t < \infty, \end{cases}$$

where for $0 \leqslant t < \infty$

$$N_1(t) = \min \left\{ k \geqslant 1; \sum_{i=1}^{k} X(i) > t \right\} - 1, \qquad N_2(t) = \min \left\{ k \geqslant 1; \sum_{i=1}^{k} X^*(i) > t \right\} - 1.$$

Then by Csörgő et al. [20] (confer also Csörgő and Horváth [18]; Steinebach [78]) the approximations in (2.1.1) - (2.1.3) hold with $a = 1/\mu$, $b = (\sigma^2/\mu^3)^{1/2}$, $Y(t) = (N_1(t) - at)/b$, $a^* = 1/\mu^*$, $b^* = (\sigma^{2*}/\mu^{*3})^{1/2}$ and $Y^*(t) = (N_2(t) - a^* t)/b^*$.

Example 2.1.3 (Dependent observations). Let $Z(t) = S_{[t]}$ be as in Example 2.1.1, but now we drop the assumption of independence of $X(1), X(2), \ldots$ and $X^*(1), X^*(2), \ldots$. Instead let $X(i) = \mu + \sigma e(i)$, $1 \leqslant i \leqslant T^*$, $X(i)^* = \mu^* + \sigma^* e([T^*] + i)$, $1 \leqslant i \leqslant T - [T^*]$, where $\{e(\cdot)\}$ fulfills a strong invariance principle. Precisely suppose there exists $\{W(t) : 0 \leqslant t < \infty\}$, such that for $k \to \infty$

$$\left| \sum_{i=1}^{k} e_i - \tau W(k) \right| = O(k^{1/\nu}) \quad a.s. \tag{2.1.4}$$

for some $\nu > 2$ and some $\tau > 0$. Such approximations have e.g. been obtained for weak Bernoulli processes (confer Eberlein [24]), α- and ϕ-mixing sequences, general Gaussian sequences and others (confer Philipp [70] for a comprehensive review). Aue et al. [5] show such approximations for squares of augmented GARCH sequences. Then (2.1.1) - (2.1.3) hold with $Y(t) = (Z(t) - \mu t)/b$ and $Y^*(t - T^*) = (Z(t) - Z(T^*) - \mu^*(t - T^*))/b^*$, where $b = \sigma\tau$ and $b^* = \sigma^*\tau$.

Example 2.1.4 (Linear processes). One sequence included in Example 2.1.3 deserves special attention, because we will also focus the research in Chapters 3 and 4 on it. Namely we assume that the sequence in Example 2.1.3 is a linear process, i.e.

$$e(i) = \sum_{s \geqslant 0} w_s\, \epsilon(i - s), \quad 1 \leqslant i < \infty,$$

where $\{\epsilon(\cdot)\}$ is a sequence of centered i.i.d. random variables with $\operatorname{var}\epsilon(0) = \sigma^2 > 0$ and $\mathrm{E}\,|\epsilon(0)|^\nu < \infty$ for some $\nu > 2$. If additionally (3.3.4) - (3.3.6) hold, then the invariance principle in (2.1.4) is fulfilled for $\tau := \sigma\left(\sum_{s\geqslant 0} w_s\right)$. For details confer Lemmas 2.1 and 2.2 in Horváth [45].

It is assumed that the process $\{Z(t) : t \geqslant 0\}$ has been observed at discrete time points $t_i = t_{i,N} = i\frac{T}{N}$, $1 \leqslant i \leqslant N = N(T)$. Let $\Delta Z_{i,T} = Z(t_i) - Z(t_{i-1})$ and $\widetilde{\Delta Z}_{i,T} = Z(t_i) - Z(t_{i-1}) - \overline{\Delta Z}_T$. Because $T/N \to \infty$ the intervals between observations become larger and larger. This is the reason why the asymptotic equals the one for independent observations and also why the permutation methods work just as described in Section 1.2. The block resampling methods of Chapter 3 are essentially based on the same idea (confer Section 3.1).

We use CUSUM statistics (for more details on the derivation of them confer Section 3.2), precisely:

$$M_T^{(1)} = \max_{1\leqslant k\leqslant N}\left\{\frac{1}{\sqrt{T}}\frac{1}{\widehat{b}_T}\left|\sum_{i=1}^{k}\left(\Delta Z_{i,T} - \overline{\Delta Z}_T\right)\right|\right\}, \tag{2.1.5}$$

where $\overline{\Delta Z}_T = \frac{1}{N}\sum_{i=1}^{N}\Delta Z_{i,T}$, and

$$\widehat{b}_T^2 = \frac{1}{T}\sum_{i=1}^{N}\left(\Delta Z_{i,T} - \overline{\Delta Z}_T\right)^2,$$

resp.

$$\widetilde{M}_T = \max_{1\leqslant k\leqslant N}\left\{\frac{1}{\sqrt{T}}\frac{1}{\widehat{c}_T}\left|\sum_{i=1}^{k}\left(\widetilde{\Delta Z}_{i,T}^2 - \overline{\widetilde{\Delta Z}_T^2}\right)\right|\right\}, \tag{2.1.6}$$

where $\overline{\widetilde{\Delta Z}_T^2} = \frac{1}{N}\sum_{i=1}^{N}\widetilde{\Delta Z}_{i,T}^2$, and

$$\widehat{c}_T^2 := \frac{1}{T}\sum_{i=1}^{N}\left((\Delta Z_{i,T} - \overline{\Delta Z}_T)^2 - \frac{1}{N}\sum_{l=1}^{N}\left(\Delta Z_{l,T} - \overline{\Delta Z}_T\right)^2\right)^2.$$

Remark 2.1.1. The statistic $\widetilde{M}_T$ uses a slightly different variance estimator $\widehat{c}_T^2$ than the one given in Horváth and Steinebach [46]. It possesses, however, the same asymptotic behavior, since the ratio of the two normalizations converges in probability to 1 under the null hypothesis, and to some positive constant under the alternative (cf. Theorem 4.5.2 and Remark 4.5.1 in Kirch [50], additionally to equations (2.3.5) and (2.3.6)).
The variance estimator we use for the permutation method is the variance of the corresponding rank statistic (alternatively we could only use an estimator that is asymptotically equivalent under both, the null hypothesis as well as alternatives). Using an estimator that is only equivalent under the null hypothesis would mean that the convergence of the permutation statistic under alternatives changes. The test remains consistent as long as under alternatives the permutation statistic does converge to some limit distribution conditionally on the observations. Still, it is nicer to have the same asymptotic behavior under both, the null hypothesis as well as alternatives.
This is why the modification is made, so we can use the same variance estimator for the original as well as the permutation statistic. Since it is invariant under permutations the

quality of the test does not depend anymore on the quality of the estimator. Simulations suggest that this is one of the main advantages of the permutation method.
It is worth mentioning that the ratio of the two estimators under alternatives converges to some positive constant. This means that the results concerning consistency (confer Theorem 3.3 of Horváth and Steinebach [46]) remain true.

The following null asymptotics hold under the above conditions:

Theorem 2.1.1. *a) If $N = N(T) \to \infty$ and $N = o\left(T^{1-2/\nu}\right)$ as $T \to \infty$, then, under H_0,*

$$M_T^{(1)} \xrightarrow{\mathcal{D}} \sup_{0 \leqslant t \leqslant 1} |B(t)|.$$

b) If $N = N(T) \to \infty$ and $N = o\left(T^{1/2-1/\nu}\right)$ as $T \to \infty$, then, under H_0,

$$\widetilde{M}_T \xrightarrow{\mathcal{D}} \sup_{0 \leqslant t \leqslant 1} |B(t)|.$$

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

Proof. Confer Theorems 2.1 and 2.2 of Horváth and Steinebach [46]. ■

Remark 2.1.2. Horváth and Steinebach [46] also show that under certain assumptions the tests are consistent (confer Theorems 2.3 and 3.3). This remains true for the new change estimator for statistic $\widetilde{M}_T$ (confer also Remark 2.1.1). The permutation tests are then automatically consistent under the same assumptions.

Gradual Change in the Mean of a Stochastic Process Under Strong Invariance

Steinebach [79] considers a similar model with a gradual change. Such changes are more realistic in many ways. Hušková and Steinebach [48] propose a testing procedures for a location model with such a change and independent observations. Permutation methods work also very well for that model (confer Kirch [50] or Kirch and Steinebach [52]).

As with the model with an abrupt change from the previous section, the model considered here allows dependency.

Suppose one observes a stochastic process $\{S(t) : 0 \leqslant t < \infty\}$ having the following structure:

$$S(t) := \begin{cases} at + bY(t), & 0 \leqslant t \leqslant T^*, \\ S(T^*) + a^*(t - T^*) + b^*Y^*(t - T^*), & T^* < t \leqslant T, \end{cases} \tag{2.1.7}$$

where a, b, b^* and $\{Y(\cdot)\}, \{Y^*(\cdot)\}$ are as in model 2.1.1 above, $a^*(t - T^*) = a(t - T^*) + d(t - T^*)^{1+\gamma}$, $\gamma > 0$ is known, $d = d_T$ is an unknown parameter. Note that – in contrast to abrupt changes – the biggest difference in the mean of the increments is not $a^* - a$ but depends on T, T^* and γ. Note that, instead of (2.1.2), Steinebach [79] assumes the following weak invariance principle for the process $\{Y(t) : 0 \leqslant t < \infty\}$, namely that, for every $T > 0$, there is a Wiener process $\{W_T(t) : 0 \leqslant t \leqslant T^*\}$ such that

$$\sup_{1 \leqslant t \leqslant T^*} |Y(T^*) - Y(T^* - t) - W_T(t)| / t^{1/\nu} = O_P(1) \qquad (T \to \infty). \tag{2.1.8}$$

The reason is that small approximation rates are required near the change-point T^*, but only in a weak sense, whereas we need strong approximations for our permutation principles below. Here, too, the processes $\{Z(\cdot)\}, \{Y(\cdot)\}$, and $\{Y^*(\cdot)\}$ could be replaced by a family of processes $\{Z_T(\cdot)\}, \{Y_T(\cdot)\}$, and $\{Y_T^*(\cdot)\},\ T > 0$.

We are now interested in testing the null hypothesis of "no change in the drift", i.e.

$$H_0 : T^* = T$$

against the alternative of "a smooth (gradual) change in the drift", i.e.

$$H_1 : 0 < T^* < T, \quad d \neq 0.$$

Basic examples fulfilling the conditions above are again partial sums of i.i.d. random variables, renewal processes based on i.i.d. waiting times and certain linear processes (confer Steinebach [79] for more details). As in model 2.1.1, we assume that we have observed $\{S(t) : t \geqslant 0\}$ at discrete time points $t_i = i\frac{T}{N}$, and set $\Delta S_{i,T} = S(t_i) - S(t_{i-1})$.

We will now work with the following statistic:

$$M_T^{(2)} = \sqrt{\frac{N}{T\widehat{b}_T^2}} \max_{1 \leqslant k < N} \frac{\left| \sum_{i=1}^{N} (i-k)_+^\gamma (\Delta S_{i,T} - \overline{\Delta S}_N) \right|}{\left(\sum_{i=1}^{N-k} i^{2\gamma} - \frac{1}{N} \left(\sum_{i=1}^{N-k} i^\gamma \right)^2 \right)^{1/2}}, \tag{2.1.9}$$

where $\overline{\Delta S}_T = \frac{1}{N} \sum_{i=1}^{N} \Delta S_{i,T}$, and $\widehat{b}_T^2 = \frac{1}{T} \sum_{i=1}^{N} \left(\Delta S_{i,T} - \overline{\Delta S}_T \right)^2$. The statistic was first proposed by Hušková and Steinebach in [48] in the context of a location model with gradual changes and independent errors. It is the maximum likelihood statistic for this model under normal errors.

Steinebach [79] assumes a slightly different weight, which is asymptotically equivalent to the one used above. Simulation studies, however, show that the above weight gives much better results for the permutation statistic. The results obtained in Steinebach [79] remain valid.

The following theorem gives the null asymptotics of the above statistic.

Theorem 2.1.2. *If (2.1.8) holds, $N = N(T) \to \infty$ and $N = O(T)$ as $T \to \infty$, then, under H_0, for all $x \in \mathbb{R}$:*

$$P\left(\alpha_N M_T^{(2)} - \beta_N \leqslant x \right) \to \exp\left(-2e^{-x} \right),$$

where $\alpha_N = \sqrt{2 \log \log N}$ and $\beta_N = \beta_N(\gamma)$ is as follows:

(i) for $\gamma > \frac{1}{2}$:

$$\beta_N = 2 \log \log N + \log \left(\frac{1}{4\pi} \left(\frac{2\gamma + 1}{2\gamma - 1} \right)^{1/2} \right);$$

(ii) for $\gamma = \frac{1}{2}$*:*

$$\beta_N = 2 \log \log N + \frac{1}{2} \log \log \log \log N - \log(4\pi);$$

(iii) for $0 < \gamma < \frac{1}{2}$*:*

$$\beta_N = 2 \log \log N + \frac{1 - 2\gamma}{2(2\gamma + 1)} \log \log \log N + \log \left(\frac{C_\gamma^{1/(2\gamma+1)} H_{2\gamma+1}}{\sqrt{\pi} 2^{2\gamma/(2\gamma+1)}} \right),$$

with H_γ *as in Remark* 12.2.10 *of Leadbetter et al. [59] (e.g.* $H_1 = 1$, $H_2 = 1/\sqrt{\pi}$*), and*

$$C_\gamma = -(2\gamma + 1) \int_0^\infty x^\gamma ((x+1)^\gamma - x^\gamma - \gamma x^{\gamma - 1})\, dx.$$

Proof. Confer Theorem 2.1 in Steinebach [79]. ■

Remark 2.1.3. Note that $H_{2\gamma+1}$ is explicitely known only for few values of γ. It is, however, necessary to know that value or have at least an approximation of it in order to use the asymptotic test. As a contrast the permutation test can be used nevertheless.

Remark 2.1.4. Steinebach [48] shows consistency of the asymptotic test under certain assumptions. Again this remains automatically true for the permutation test.

2.2. Asymptotics of the Corresponding Rank Statistics

In order to derive distributional asymptotics for the permutation statistics, we shall make use of the following theorems for the corresponding rank statistics. We begin with the corresponding rank asymptotic for the CUSUM statistic.

Theorem 2.2.1. *Let* $(R_1, \ldots, R_n)$ *be a random permutation of* $(1, \ldots, n)$*, and* $a_n(1), \ldots, a_n(n)$ *be scores satisfying the following conditions:*

$$\frac{1}{n} \sum_{i=1}^{n} (a_n(i) - \bar{a}_n)^2 \to 1, \tag{2.2.1}$$

where $\bar{a}_n := \frac{1}{n} \sum_{i=1}^{n} a_n(i)$*, and*

$$\frac{1}{n} \max_{1 \leqslant i \leqslant n} (a_n(i) - \bar{a}_n)^2 \to 0. \tag{2.2.2}$$

Then, as $n \to \infty$*,*

$$\max_{1 \leqslant k \leqslant n} \frac{1}{\sqrt{n}} \Big| \sum_{i=1}^{k} (a_n(R_i) - \bar{a}_n) \Big| \xrightarrow{\mathcal{D}} \sup_{0 \leqslant t \leqslant 1} |B(t)|,$$

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

Proof. It follows from Theorem 24.2 in Billingsley [9]. ■

Remark 2.2.1. We obtain an analogous result for score statistics, where we replace the permutations $(R_1, \ldots, R_n)$ by a triangular array $\{U_i : i = 1, \ldots, n\}$ of rowwise i.i.d. random variables that are uniformly distributed on $\{1, \ldots, n\}$. Then we obtain under (2.2.1) and (2.2.2)

$$\max_{1 \leqslant k \leqslant n} \frac{1}{\sqrt{n}} \Big| \sum_{i=1}^{k} (a_n(U_i) - \bar{a}_{U,n}) \Big| \xrightarrow{\mathcal{D}} \sup_{0 \leqslant t \leqslant 1} |B(t)| \,,$$

where $\bar{a}_{U,n} = \frac{1}{n} \sum_{i=1}^{n} a_n(U_i)$. This result can then be used to prove the validity of the bootstrap with replacement.

Proof of Remark 2.2.1. The result in Billingsley [9], Theorem 24.2, is already formulated for exchangeable random variables. It suffices to prove

$$\frac{1}{n} \sum_{i=1}^{n} (a_n(U_i) - \bar{a}_{U,n})^2 \xrightarrow{P} 1,$$
$$\frac{1}{n} \max_{i=1,\ldots,n} (a_n(U_i) - \bar{a}_{U,n})^2 \xrightarrow{P} 0.$$

First of all we obtain as in (D.5)

$$\bar{a}_{U,n} - \bar{a}_n = o_P(1),$$

thus it suffices to prove

$$\frac{1}{n} \sum_{i=1}^{n} (a_n(U_i) - \bar{a}_n)^2 = 1 + o_P(1). \tag{2.2.3}$$

We use Lemma B.2 with $Y_n(i) = (a_n(U_i) - \bar{a}_n)^2$ and $b_n = n$. Then (i) is obviously fulfilled by (2.2.2). Furthermore (ii) holds true because of (2.2.1) and (2.2.2), since

$$\frac{1}{n^2} \sum_{i=1}^{n} (a_n(i) - \bar{a}_n)^4 1_{\{(a_n(i) - \bar{a}_n)^2 \leqslant n\}} \leqslant \frac{1}{n^2} \sum_{i=1}^{n} (a_n(i) - \bar{a}_n)^4$$
$$\leqslant \frac{1}{n} \max_{i=1,\ldots,n} (a_n(i) - \bar{a}_n)^2 \, \frac{1}{n} \sum_{i=1}^{n} (a_n(i) - \bar{a}_n)^2 \to 0.$$

Finally it holds for large n because of (2.2.1) and (2.2.2)

$$\frac{1}{n} \sum_{i=1}^{n} (a_n(i) - \bar{a}_n)^2 1_{\{(a_n(i) - \bar{a}_n)^2 \leqslant n\}} = \frac{1}{n} \sum_{i=1}^{n} (a_n(i) - \bar{a}_n)^2.$$

Thus Lemma B.2 gives (2.2.3). ■

The next theorem deals with the rank asymptotics for the statistic for a gradual change. In the case $\gamma = 1$, it was already proven by Slabý [77].

Theorem 2.2.2. *Let* $\mathbf{R} = (R_1, \ldots, R_n)$ *be a random permutation of* $(1, \ldots, n)$, *and* $a_n(1), \ldots, a_n(n)$ *be scores satisfying*

$$\frac{1}{n} \sum_{i=1}^{n} (a_n(i) - \bar{a}_n)^2 \geqslant D_1, \tag{2.2.4}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}|a_n(i)-\bar{a}_n|^{\nu}\leqslant D_2, \tag{2.2.5}$$

where D_1, D_2 are some positive constants, $\nu > 2$, and $\bar{a}_n = \frac{1}{n}\sum_{i=1}^{n} a_n(i)$. Then, for fixed $\gamma > 0$ and all $x \in \mathbb{R}$, as $n \to \infty$

$$P\left(\alpha_n M_n^{(2)}(\mathbf{a}) - \beta_n \leqslant x\right) \to \exp\left(-2e^{-x}\right),$$

where

$$M_n^{(2)}(\mathbf{a}) = \frac{1}{\sigma_n(\mathbf{a})}\max_{1\leqslant k<n}\frac{\left|\sum_{i=1}^{n}(i-k)_+^{\gamma}(a_n(R_i)-\bar{a}_n)\right|}{\left(\sum_{i=1}^{n-k} i^{2\gamma}-\frac{1}{n}\left(\sum_{i=1}^{n-k} i^{\gamma}\right)^2\right)^{1/2}};$$

Here $\sigma_n^2(\mathbf{a}) = \frac{1}{n}\sum_{i=1}^{n}\left(a_n(i)-\bar{a}_n\right)^2$, the variance of $a_n(R_1)$, $\alpha_n = \sqrt{2\log\log n}$ and $\beta_n = \beta_n(\gamma)$ is as in Theorem 2.1.2.

Remark 2.2.2. We obtain an analogous result for score processes, where we replace the permutations $(R_1,\ldots,R_n)$ by a triangular array $\{U_i : i = 1,\ldots,n\}$ of rowwise i.i.d. random variables that are uniformly distributed on $\{1,\ldots,n\}$. Then we obtain under (2.2.4) and (2.2.5)

$$P\left(\alpha_n M_n^{(2,r)}(\mathbf{a}) - \beta_n \leqslant x\right) \to \exp\left(-2e^{-x}\right),$$

where

$$M_n^{(2,r)}(\mathbf{a}) = \frac{1}{\sigma_n(\mathbf{a})}\max_{1\leqslant k<n}\frac{\left|\sum_{i=1}^{n}(i-k)_+^{\gamma}(a_n(U_i)-\bar{a}_{U,n})\right|}{\left(\sum_{i=1}^{n-k} i^{2\gamma}-\frac{1}{n}\left(\sum_{i=1}^{n-k} i^{\gamma}\right)^2\right)^{1/2}},$$

where $\bar{a}_{U,n} = \frac{1}{n}\sum_{i=1}^{n} a_n(U_i)$. We just need to replace Corollary D.1 in the proof of the above theorem by Corollary D.2. This result can then be used to prove the validity of the bootstrap with replacement.

Proof of Theorem 2.2.2. Confer Kirch [50], Corollary 5.2.3. For the sake of completeness we will repeat the details here. First note that

$$\begin{aligned} n\sum_{i=1}^{n-k} i^{2\gamma} - \left(\sum_{i=1}^{n-k} i^{\gamma}\right)^2 &= (n-k)\sum_{i=1}^{n-k}\left(i^{\gamma}-\frac{1}{n-k}\sum_{j=1}^{n-k} j^{\gamma}\right)^2 + k\sum_{i=1}^{n-k} i^{2\gamma} \\ &\geqslant k\int_0^{n-k} x^{2\gamma}dx = k\frac{1}{2\gamma+1}(n-k)^{2\gamma+1}. \end{aligned} \tag{2.2.6}$$

It holds further as $n \to \infty$ uniformly in $k \leqslant n/2$

$$\begin{aligned} &\frac{1}{(n-k)^{2\gamma+1}}\sum_{i=1}^{n-k} i^{2\gamma} - \frac{1}{n(n-k)^{2\gamma+1}}\left(\sum_{i=1}^{n-k} i^{\gamma}\right)^2 \\ &\geqslant \frac{1}{(n-k)^{2\gamma+1}}\int_0^{n-k} x^{2\gamma}\,dx - \frac{1}{(n-k)^{2\gamma+2}}\left(\int_0^{n-k+1} x^{\gamma}\,dx\right)^2 \\ &= \frac{\gamma^2}{(2\gamma+1)(\gamma+1)^2} - \frac{1}{(\gamma+1)^2}\left[\left(1+\frac{1}{n-k}\right)^{2\gamma+2}-1\right] \\ &= \frac{\gamma^2}{(2\gamma+1)(\gamma+1)^2} + o(1). \end{aligned} \tag{2.2.7}$$

Putting together (2.2.6) (for $k > n/2$) and (2.2.7) (for $k \leqslant n/2$) we arrive at

$$\left(\sum_{i=1}^{n-k} i^{2\gamma} - \frac{1}{n}\left(\sum_{i=1}^{n-k} i^{\gamma}\right)^2\right)^{-1} = O\left((n-k)^{-(2\gamma+1)}\right) \tag{2.2.8}$$

uniformly in k. Now, from Corollary D.1 with $\mu = 0$, uniformly in $k \in [1, \frac{n}{2}]$:

$$\begin{aligned}\frac{1}{\sigma_n(\mathbf{a})}\sum_{i=1}^{k}\left(a_n\left(R_{n-i+1}\right) - \bar{a}_n\right) &\stackrel{\mathcal{D}}{=} \sqrt{n}B\left(\frac{k}{n}\right) + O_P\left(\sqrt{\frac{k(n-k)}{n}}\right)\\ &= \sqrt{n}B\left(\frac{k}{n}\right) + O_P\left(\sqrt{k}\right).\end{aligned}$$

It holds $\left\{\sqrt{n}B\left(\frac{k}{n}\right) : k = 0, \ldots, n\right\} \stackrel{\mathcal{D}}{=} \left\{W(k) - \frac{k}{n}W(n) : k = 0, \ldots, n\right\}$, where $\{W(t) : t \geqslant 0\}$ is a standard Wiener process. Since $\sum_{i=1}^{n}(i-k)_+^{\gamma}x_i = \sum_{i=1}^{n-k}(i^{\gamma}-(i-1)^{\gamma})\sum_{j=i+k}^{n}x_j$, we conclude

$$\begin{aligned}&\frac{1}{\sigma_n(\mathbf{a})}\max_{n-\log n<k<n}\frac{\left|\sum_{i=1}^{n}(i-k)_+^{\gamma}(a_n(R_i)-\bar{a}_n)\right|}{\left(\sum_{i=1}^{n-k} i^{2\gamma} - \frac{1}{n}\left(\sum_{i=1}^{n-k} i^{\gamma}\right)^2\right)^{1/2}}\\ &= \frac{1}{\sigma_n(\mathbf{a})}\max_{1<k<\log n}\frac{\left|\sum_{l=1}^{k}(l^{\gamma}-(l-1)^{\gamma})\sum_{i=1}^{k-l+1}(a_n(R_{n-i+1})-\bar{a}_n)\right|}{\left(\sum_{i=1}^{k} i^{2\gamma} - \frac{1}{n}\left(\sum_{i=1}^{k} i^{\gamma}\right)^2\right)^{1/2}}\\ &\stackrel{\mathcal{D}}{=} \max_{1<k<\log n}\frac{\left|\sum_{l=1}^{k}(l^{\gamma}-(l-1)^{\gamma})\left(W(k-l+1)-\frac{k-l+1}{n}W(n)\right)\right|}{\left(\sum_{i=1}^{k} i^{2\gamma} - \frac{1}{n}\left(\sum_{i=1}^{k} i^{\gamma}\right)^2\right)^{1/2}}\\ &\quad + O_P(1)\max_{1<k<\log n}\frac{\left|\sum_{l=1}^{k}(l^{\gamma}-(l-1)^{\gamma})\sqrt{k-l+1}\right|}{\left(\sum_{i=1}^{k} i^{2\gamma} - \frac{1}{n}\left(\sum_{i=1}^{k} i^{\gamma}\right)^2\right)^{1/2}}\\ &= o_P\left(\sqrt{\log\log n}\right).\end{aligned}$$

The last line follows because (2.2.8) and the law of iterated logarithm yield

$$\begin{aligned}&\max_{1<k<\log n}\frac{\left|\sum_{l=1}^{k}(l^{\gamma}-(l-1)^{\gamma})W(k-l+1)\right|}{\left(\sum_{i=1}^{k} i^{2\gamma} - \frac{1}{n}\left(\sum_{i=1}^{k} i^{\gamma}\right)^2\right)^{1/2}}\\ &= O_P(1)\max_{1<k<\log n}\frac{\left|\sum_{l=1}^{k}(l^{\gamma}-(l-1)^{\gamma})\sqrt{(k-l+1)\log\log(k-l+1)}\right|}{\left(\sum_{i=1}^{k} i^{2\gamma} - \frac{1}{n}\left(\sum_{i=1}^{k} i^{\gamma}\right)^2\right)^{1/2}}\\ &= O_P\left(\max_{1<k<\log n}\frac{k^{\gamma}\sqrt{k\log\log k}}{\sqrt{k^{2\gamma+1}}}\right) = O_P\left(\sqrt{\log\log\log n}\right),\end{aligned}$$

and similarly

$$\max_{1<k<\log n}\frac{\left|\sum_{l=1}^{k}(l^{\gamma}-(l-1)^{\gamma})\frac{k-l+1}{n}W(n)\right|}{\left(\sum_{i=1}^{k} i^{2\gamma} - \frac{1}{n}\left(\sum_{i=1}^{k} i^{\gamma}\right)^2\right)^{1/2}} = O_P\left(\sqrt{\frac{\log n\,\log\log n}{n}}\right),$$

$$\max_{1<k<\log n} \frac{\left|\sum_{l=1}^{k}(l^\gamma-(l-1)^\gamma)\sqrt{k-l+1}\right|}{\left(\sum_{i=1}^{k} i^{2\gamma}-\frac{1}{n}\left(\sum_{i=1}^{k} i^\gamma\right)^2\right)^{1/2}}=O(1).$$

Hence it suffices to investigate the maximum over $k\in[1,n-\log n]$ as Lemma F.1 shows.

We choose $X(i)$ such that $B\left(\frac{k}{n}\right)=\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{k}X(i)-k\bar{X}_n\right)$ with $\bar{X}_n=\frac{1}{n}\sum_{j=1}^{n}X(i)$, $\{B(\cdot)\}$ denoting the Brownian bridge of Corollary D.1. Let $Y_{in}:=\Pi_n(i)-\Pi_n(i-1)-\left(X(i)-\bar{X}_n\right)$, where $\{\Pi_n(\cdot)\}$ and $0<\mu<\min\left(\frac{\nu-2}{2\nu},\frac{1}{4}\right)$ are as in Corollary D.1 (with b_n replaced by a_n) and $S_n(l):=\sum_{i=1}^{l}Y_{in}$. Note that $\sum_{i=l}^{n}Y_{in}=-S_n(l-1)$. Now Corollary D.1 and (2.2.8) give

$$\begin{aligned}
&\max_{1\leqslant k\leqslant n-\log n}\sqrt{\frac{n}{n\sum_{i=1}^{n-k}i^{2\gamma}-\left(\sum_{i=1}^{n-k}i^\gamma\right)^2}}\left|\sum_{i=1}^{n}(i-k)_+^\gamma Y_{in}\right|\\
&\leqslant\max_{1\leqslant k\leqslant n-\log n}\sqrt{\frac{n}{n\sum_{i=1}^{n-k}i^{2\gamma}-\left(\sum_{i=1}^{n-k}i^\gamma\right)^2}}\sum_{l=1}^{n-k}|S_n(l+k-1)|(l^\gamma-(l-1)^\gamma)\\
&=O_P(1)\max_{1\leqslant k\leqslant n-\log n}n^\mu\sum_{l=1}^{n-k}\frac{((l+k-1)(n-l-k+1))^{1/2-\mu}}{\sqrt{n\sum_{i=1}^{n-k}i^{2\gamma}-\left(\sum_{i=1}^{n-k}i^\gamma\right)^2}}(l^\gamma-(l-1)^\gamma)\\
&=o_P((\log\log n)^{-1/2}),
\end{aligned}$$

where we used the fact that for $k\leqslant n/2$ it holds $(l+k-1)(n-l-k+1)\leqslant n^2/4$ and for $k>n/2$ it holds $(l+k-1)(n-l-k+1)\leqslant k(n-k)$.

Again an application of the law of the iterated logarithm gives similarly to above,

$$\max_{n-\log n\leqslant k\leqslant n}\frac{\left|\sum_{i=1}^{n}(i-k)_+^\gamma(X(i)-\frac{1}{n}\sum_{i=1}^{n}X(i))\right|}{\left(\sum_{i=1}^{n-k}i^{2\gamma}-\frac{1}{n}\left(\sum_{i=1}^{n-k}i^\gamma\right)^2\right)^{1/2}}=o_P\left(\sqrt{\log\log n}\right).$$

Thus Lemma F.1 yields that it is equivalent to consider the maximum over $1\leqslant k<n$. Theorem 3.3 of Hušková and Steinebach [48] shows

$$P\left(\alpha(n)\max_{1\leqslant k\leqslant n}\frac{\left|\sum_{i=1}^{n}(i-k)_+^\gamma(X(i)-\bar{X}_n)\right|}{\left(\sum_{i=1}^{n-k}i^{2\gamma}-\frac{1}{n}\left(\sum_{i=1}^{n-k}i^\gamma\right)^2\right)^{1/2}}-\beta(n)\leqslant x\right)\to\exp(-2e^{-x}),$$

which completes the proof. ■

2.3. Permutation Statistics and their Limit Distributions

In this chapter we show that the permutation statistics conditioned on the given data follow the same asymptotic as the original statistic under the null. This is true under the null hypothesis as well as under alternatives for almost all realizations. This shows that the permutation test as described in Section 1.2 works in this setting.
The main tool in the proofs are the rank statistic theorems developed in Section 2.2. We

will use them choosing the ranks essentially as the increments of our observed process. Then, we only need to verify the conditions on the scores which reduces to proving strong laws of large numbers for the increments.

To this end we will start with two lemmas. The first one is needed to get a law of logarithm for a sequence of Wiener processes. The second one contains some calculations for the increments of the underlying process in our model.

Lemma 2.3.1. *Let $\{W_n(t)\colon t \geqslant 0\}$, $n \in \mathbb{N}$, be Wiener processes and f be a function of n, then*

$$W_n(f(n)) = O\left(\sqrt{f(n)\log n}\right) \qquad a.s. \quad (n \to \infty).$$

Proof. It follows from a tail approximation for a normal distribution, namely it holds

$$P\left(\frac{W_n(f(n))}{\sqrt{f(n)\log n}} \geqslant 2\right) = \frac{1}{\sqrt{2\pi}}\int_{2\sqrt{\log n}}^{\infty} e^{-x^2/2}\,dx \ll \int_{2\sqrt{\log n}}^{\infty} xe^{-x^2/2}dx = n^{-2},$$

and $\sum n^{-2} < \infty$, so that the Borel-Cantelli lemma gives the assertion. ■

In the sequel we assume that there is a 1-1-correspondence between N and T, which is necessary to get a countable triangular array in N, and, in turn, allows us to use the preceding lemma.

Moreover, we assume $T^* = \theta T$, $0 < \theta \leqslant 1$, and $N = o(T^{1-2/\nu})$. Let $N^* = \lfloor \frac{NT^*}{T} \rfloor = \theta N(1 + o(1))$ and

$$\Delta Y_i = \begin{cases} b\left(Y\left(i\frac{T}{N}\right) - Y\left((i-1)\frac{T}{N}\right)\right) & , \quad i \leqslant N^*, \\ b(Y(T^*) - Y(\frac{N^*T}{N})) + b^*Y^*(\frac{(N^*+1)T}{N} - T^*) & , \quad i = N^*+1, \\ b^*\left(Y^*\left(i\frac{T}{N} - T^*\right) - Y^*\left((i-1)\frac{T}{N} - T^*\right)\right) & , \quad i \geqslant N^*+2. \end{cases} \tag{2.3.1}$$

The following lemma now gives some limit results concerning sums of functionals of the increments of the underlying processes $\{Y(\cdot)\}$ respectively $\{Y^*(\cdot)\}$.

Lemma 2.3.2. *a) It holds, as $N \to \infty$,*

$$\overline{\Delta Y} = \frac{1}{N}\sum_{i=1}^{N} \Delta Y_i = O\left(\frac{\sqrt{T\log N}}{N}\right) \qquad a.s.$$

b) (i) For $s = 2, 3, 4$, as $N \to \infty$,

$$\frac{N^{(s-2)/2}}{T^{s/2}}\sum_{i=1}^{N} (\Delta Y_i)^s \;\to\; \mathrm{E}\,W(1)^s\left(\theta b^s + (1-\theta)(b^*)^s\right) \qquad a.s.,$$

where $W(1)$ has a standard normal distribution.

(ii) For $\kappa > 0$, as $N \to \infty$,

$$\frac{N^{(\kappa-2)/2}}{T^{\kappa/2}}\sum_{i=1}^{N} \left|\Delta Y_i - \overline{\Delta Y}\right|^{\kappa} = O(1) \qquad a.s.$$

c) For $\kappa > 0$, as $N \to \infty$,

$$\frac{N^{(\kappa-2)/2}}{T^{\kappa/2}} \max_{1\leqslant i\leqslant N} \left|\Delta Y_i - \overline{\Delta Y}\right|^{\kappa} = o(1) \qquad a.s.$$

Proof. The proof makes use of (2.1.2) and (2.1.3) in combination with Lemma 2.3.1 and Corollary B.1.
First Lemma 2.3.1 and the invariance principles (2.1.2) for $\{Y(\cdot)\}$ respectively (2.1.3) for $\{Y^*(\cdot)\}$ yield

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N} \Delta Y_i = \frac{1}{N}\left(bY(T^*) + b^*Y^*(T-T^*)\right) \\
&= \frac{1}{N}\left((bW_T(T^*) + b^*W_T^*(T-T^*)\right) + O\left(\frac{T^{1/\nu}}{N}\right) = O\left(\frac{\sqrt{T\log N}}{N}\right) \qquad a.s.
\end{aligned}$$

Note that

$$\begin{aligned}
X_{iN} &:= \sqrt{\frac{N}{T}}\left(W_T\left(i\frac{T}{N}\right) - W_T\left((i-1)\frac{T}{N}\right)\right), \quad i = 1,\dots,N^*, \\
X^*_{iN} &:= \sqrt{\frac{N}{T}}\left(W^*_T\left(i\frac{T}{N} - T^*\right) - W^*_T\left((i-1)\frac{T}{N} - T^*\right)\right), \quad i = N^*+1,\dots,N
\end{aligned}$$

form a triangular array of i.i.d. standard normal random variables. Now Lemma 2.3.1 and Corollary B.1 give

$$\begin{aligned}
&\frac{1}{T}\sum_{i=1}^{N}(\Delta Y_i)^2 \\
&= b^2\frac{1}{N}\sum_{i=1}^{N^*} X_{iN}^2 + (b^*)^2\frac{1}{N}\sum_{i=N^*+2}^{N}(X^*_{iN})^2 + O\left(\frac{N}{T^{1-2/\nu}} + \left(\frac{N}{T^{1-2/\nu}}\right)^{1/2}\right) \\
&\quad + O\left(\frac{1}{T}\left[b\left(W_T(T^*) - W_T\left(N^*\frac{T}{N}\right)\right) + b^*W^*_T\left((N^*+1)\frac{T}{N} - T^*\right)\right]^2\right) \\
&= \theta b^2 + (1-\theta)(b^*)^2 + o(1) + O\left(\frac{\log N}{N}\right) \qquad a.s.,
\end{aligned}$$

because $|T^* - N^*\frac{T}{N}| \leqslant \frac{T}{N}$ and $|(N^*+1)\frac{T}{N} - T^*| \leqslant \frac{T}{N}$.

Since $\mathrm{E}\,W(1)^3 = 0$ we get analogously

$$\begin{aligned}
&\sqrt{\frac{N}{T^3}}\sum_{i=1}^{N}(\Delta Y_i)^3 \\
&= b^3\frac{1}{N}\sum_{i=1}^{N^*} X_{iN}^3 + (b^*)^3\frac{1}{N}\sum_{i=N^*+2}^{N}(X^*_{iN})^3 + O\left(\left(\frac{N}{T^{1-2/\nu}}\right)^{1/2} + \frac{(\log N)^{3/2}}{N}\right) \\
&= o(1) \qquad a.s.
\end{aligned}$$

Furthermore $\mathrm{E}\,W(1)^4 = 3$, hence

$$\begin{aligned}
&\frac{N}{T^2}\sum_{i=1}^{N}(\Delta Y_i)^4 \\
&= b^4\frac{1}{N}\sum_{i=1}^{N^*}X_{iN}^4 + (b^*)^4\frac{1}{N}\sum_{i=N^*+2}^{N}(X_{iN}^*)^4 + O\left(\left(\frac{N}{T^{1-2/\nu}}\right)^{1/2} + \frac{(\log N)^2}{N}\right) \\
&= 3\left(\theta b^4 + (1-\theta)(b^*)^4\right) + o(1) \qquad a.s.
\end{aligned}$$

Finally

$$\begin{aligned}
&\frac{N^{(\kappa-2)/2}}{T^{\kappa/2}}\sum_{i=1}^{N}\left|\Delta Y_i - \overline{\Delta Y}\right|^\kappa \ll \frac{N^{(\kappa-2)/2}}{T^{\kappa/2}}\sum_{i=1}^{N}|\Delta Y_i|^\kappa + \left(\frac{N}{T}\right)^{\kappa/2}|\overline{\Delta Y}|^\kappa \\
&\ll |b|^\kappa\frac{1}{N}\sum_{i=1}^{N^*}|X_{iN}|^\kappa + |b^*|^\kappa\frac{1}{N}\sum_{i=N^*+2}^{N}|X_{iN}^*|^\kappa \\
&\quad + O\left(\left(\frac{N}{T^{1-2/\nu}}\right)^{\kappa/2} + \frac{(\log N)^{\kappa/2}}{N}\right) + O\left(\left(\frac{\log N}{N}\right)^{\kappa/2}\right) \\
&\ll 1 \qquad a.s.
\end{aligned}$$

Concerning c) we obtain similarly

$$\begin{aligned}
&\frac{N^{(\kappa-2)/2}}{T^{\kappa/2}}\max_{1\leqslant i\leqslant N}\left|\Delta Y_i - \overline{\Delta Y}\right|^\kappa \ll \frac{N^{(\kappa-2)/2}}{T^{\kappa/2}}\max_{1\leqslant i\leqslant N}|\Delta Y_i|^\kappa + \frac{N^{(\kappa-2)/2}}{T^{\kappa/2}}|\overline{\Delta Y}|^\kappa \\
&\ll \frac{|b|^\kappa}{N}\max_{1\leqslant i\leqslant N^*}|X_{iN}|^\kappa + \frac{|b^*|^\kappa}{N}\max_{N^*+2\leqslant i\leqslant N}|X_{iN}^*|^\kappa + o(1) \\
&\leqslant \frac{|b|^\kappa}{N^{1/(\kappa+1)}}\left(\frac{1}{N}\sum_{i=1}^{N^*}|X_{iN}|^{\kappa+1}\right)^{\kappa/(\kappa+1)} + \frac{|b^*|^\kappa}{N^{1/(\kappa+1)}}\left(\frac{1}{N}\sum_{i=N^*+2}^{N}|X_{iN}^*|^{\kappa+1}\right)^{\kappa/(\kappa+1)} + o(1) \\
&= o(1) \qquad a.s.,
\end{aligned}$$

which completes the proof. ■

Abrupt Change in the Mean or Variance

We are now prepared to investigate the permutation statistics for an abrupt change in the mean or variance, i.e.

$$M_T^{(1)}(\mathbf{R}) = \max_{1\leqslant k\leqslant N}\left\{\frac{1}{\sqrt{T}}\frac{1}{\widehat{b}_T}\Big|\sum_{i=1}^{k}\left(\Delta Z_{R_i,T} - \overline{\Delta Z}_T\right)\Big|\right\},$$

and

$$\widetilde{M}_T(\mathbf{R}) = \max_{1\leqslant k\leqslant N}\left\{\frac{1}{\sqrt{T}}\frac{1}{\widehat{c}_T}\Big|\sum_{i=1}^{k}\left(\widetilde{\Delta Z}^2_{R_i,T} - \overline{\widetilde{\Delta Z}}^2_T\right)\Big|\right\}.$$

Here again, $\mathbf{R} = (R_1, \ldots, R_n)$ denotes a random permutation of $(1, \ldots, n)$ independent of $\{Z(\cdot)\}$.

Theorem 2.3.1. *Let $\{Z(t) : t \geq 0\}$ be a process according to model* (2.1.1). *Let $T^* = \theta T$, $0 < \theta \leqslant 1$, $N = o(T^{1-2/\nu})$, and in b) also $a = a^*$. In a) no restriction on a or a^* is necessary. Then, for all $x \in \mathbb{R}$, as $T \to \infty$,*

$$a) \qquad P\big(M_T^{(1)}(\boldsymbol{R}) \leqslant x \,\big|\, Z(t), 0 \leqslant t \leqslant T\big) \to P\big(\sup_{0 \leqslant t \leqslant 1} |B(t)| \leqslant x\big) \qquad a.s.$$

$$b) \qquad P\big(\widetilde{M}_T(\boldsymbol{R}) \leqslant x \,\big|\, Z(t), 0 \leqslant t \leqslant T\big) \to P\big(\sup_{0 \leqslant t \leqslant 1} |B(t)| \leqslant x\big) \qquad a.s.,$$

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

Remark 2.3.1. The corresponding bootstrap result with replacement also holds true. In that approach we sample with replacement from the increments of the observed process instead of permuting them. Because the corresponding score results holds true under the same assumptions on the scores (confer Remark 2.2.1), this follows immediately from the proof of Theorem 2.3.1.

Proof of Theorem 2.3.1. Kirch [50], Theorem 10.0.3, gives the proof for constant a and a^*. This is, however, not needed as we will see now. Let $d := a^* - a$. Then it holds for the increments of $\{Z(\cdot)\}$

$$\Delta Z_{i,T} = \Delta a_i + \Delta Y_i$$

with ΔY_i as in (2.3.1) and

$$\Delta a_i = \begin{cases} a\frac{T}{N}, & i \leqslant N^*, \\ a\frac{T}{N} + d\left((N^*+1)\frac{T}{N} - T^*\right), & i = N^*+1, \\ a\frac{T}{N} + d\frac{T}{N}, & i \geqslant N^*+2, \end{cases}$$

and $\overline{\Delta a} = \frac{1}{N}\sum_{i=1}^N \Delta a_i = \frac{1}{N}\left(aT + d(T - T^*)\right)$.

Now, for the proof of a), consider the scores $a_N(i) = \widehat{b}_T^{-1}\sqrt{\frac{N}{T}}\Delta Z_{i,T}$, $i = 1, \ldots, N$. Obviously, $\frac{1}{N}\sum_{i=1}^N (a_N(i) - \bar{a}_N)^2 = 1$, which means that it is sufficient to verify assumption (2.2.2) of Theorem 2.2.1.

Lemma 2.3.2 implies

$$\frac{1}{T}\sum_{i=1}^N \left(\Delta Y_i - \overline{\Delta Y}\right)^2 = \theta b^2 + (1-\theta)(b^*)^2 + o(1) \qquad a.s.,$$

and

$$\begin{aligned} &\frac{1}{T}\sum_{i=1}^N (\Delta Y_i - \overline{\Delta Y})(\Delta a_i - \overline{\Delta a}) = \frac{1}{T}\sum_{i=1}^N (\Delta Y_i - \overline{\Delta Y})\left[\Delta a_i - a\frac{T}{N}\right] \\ &\ll |d|\sqrt{\frac{T}{N}}\left(\frac{1}{\sqrt{NT}}|Y^*(T) - Y^*((N^*+1)T/N - T^*)| + \frac{1}{\sqrt{NT}}|\Delta Y_{N^*+1}| + \sqrt{\frac{N}{T}}\overline{\Delta Y}\right) \\ &= o\left(|d|\sqrt{T/N}\right) \qquad a.s. \end{aligned}$$

Moreover

$$\frac{1}{T}\sum_{i=1}^{N}(\Delta a_i - \overline{\Delta a})^2 = \frac{d^2 T}{N}\frac{N-N^*-1}{N} + d^2\frac{1}{T}\left(T\frac{N^*+1}{N} - T^*\right)^2 - d^2\frac{T}{N}\frac{(T-T^*)^2}{T^2}$$
$$= d^2\frac{T}{N}\,\theta(1-\theta) + o(d^2T/N).$$

Putting the above together we obtain

$$\widehat{b}_T^2 \geqslant \theta b^2 + (1-\theta)(b^*)^2 + o\left(|d|\sqrt{T/N}\right) + o(1) \qquad a.s. \tag{2.3.2}$$

as well as ($d \neq 0$)

$$\frac{N}{Td^2}\,\widehat{b}_T^2 \geqslant \theta(1-\theta) + o\left(\sqrt{\frac{N}{Td^2}}\right) + o(1) \qquad a.s. \tag{2.3.3}$$

Note that

$$\max_{i=1,\dots,n}(\Delta a_i - \overline{\Delta a})^2 \ll d^2\frac{T^2}{N^2}. \tag{2.3.4}$$

First consider the case $\frac{d^2T}{N} = O(1)$, which includes the null hypothesis. Then Lemma 2.3.2 and equations (2.3.2) respectively (2.3.4) give

$$\frac{1}{N}\max_{1\leqslant i\leqslant N}(a_N(i) - \bar{a}_N)^2 \ll \frac{1}{T\widehat{b}_T^2}\max_{1\leqslant i\leqslant N}(\Delta a_i - \overline{\Delta a})^2 + \frac{1}{T\widehat{b}_T^2}\max_{1\leqslant i\leqslant N}(\Delta Y_i - \overline{\Delta Y})^2$$
$$= o(1) \qquad a.s.$$

If, on the other hand, $\frac{N}{Td^2} = O(1)$, then Lemma 2.3.2 and equations (2.3.3) respectively (2.3.4) yield as well

$$\frac{1}{N}\max_{1\leqslant i\leqslant N}(a_N(i) - \bar{a}_N)^2 = o(1) \qquad a.s.$$

If neither $\frac{d^2T}{N} = O(1)$ nor $\frac{N}{Td^2} = O(1)$, we can divide the sequence into two subsequences each fulfilling one of the above conditions. This means we again get

$$\frac{1}{N}\max_{1\leqslant i\leqslant N}(a_N(i) - \bar{a}_N)^2 = o(1) \qquad a.s.,$$

which completes the proof of a).

For the proof of b), consider $a_N(i) = \widehat{c}_T^{-1}\sqrt{\frac{N}{T}}\left(\Delta Z_{i,T} - \overline{\Delta Z}_T\right)^2$. It suffices again to verify the assumptions of Theorem 2.2.1.

Again $\frac{1}{N}\sum_{i=1}^{N}(a_N(i) - \bar{a}_N)^2 = 1$. Note that because of $a = a^*$ it holds $\Delta Z_{i,T} - \overline{\Delta Z}_T = \Delta Y_i - \overline{\Delta Y}$. Like above, Lemma 2.3.2 gives

$$\frac{N}{T^2}\sum_{i=1}^{N}\left(\Delta Y_i - \overline{\Delta Y}\right)^4 \to 3\left(\theta b^4 + (1-\theta)(b^*)^4\right) \qquad a.s., \tag{2.3.5}$$

and

$$(\widehat{b}_T^2)^2 = \left(\frac{1}{T}\sum_{i=1}^{N}(\Delta Y_i)^2 - \frac{N}{T}\overline{\Delta Y}^2\right)^2 \to \left(\theta b^2 + (1-\theta)(b^*)^2\right)^2 \qquad a.s. \tag{2.3.6}$$

From Jensen's inequality we conclude

$$\begin{aligned}&\lim_{T\to\infty}\frac{N}{T}\widehat{c}_T^2 = \lim_{T\to\infty}\left(\frac{N}{T^2}\sum_{i=1}^{N}\left(\Delta Y_i - \overline{\Delta Y}\right)^4 - (\widehat{b}_T^2)^2\right)\\ &= 3(\theta b^4 + (1-\theta)(b^*)^4) - (\theta b^2 + (1-\theta)(b^*)^2)^2 \geqslant 2(\theta b^4 + (1-\theta)(b^*)^4) > 0 \qquad a.s.\end{aligned}$$

So, an application of Lemma 2.3.2 results in

$$\begin{aligned}&\frac{1}{N}\max_{1\leqslant k\leqslant N}(a_N(k) - \bar{a}_N)^2 = \frac{1}{T\widehat{c}_T^2}\max_{1\leqslant k\leqslant N}\left(\left(\Delta Y_k - \overline{\Delta Y}\right)^2 - \frac{1}{N}\sum_{i=1}^{N}\left(\Delta Y_i - \overline{\Delta Y}\right)^2\right)^2\\ &\ll \frac{N}{T^2}\max_{1\leqslant k\leqslant N}(\Delta Y_k - \overline{\Delta Y})^4 + \frac{1}{N}\left(\frac{1}{T}\sum_{i=1}^{N}(\Delta Y_i - \overline{\Delta Y})^2\right)^2 = o(1) \qquad a.s.,\end{aligned}$$

which completes the proof of b). ■

Gradual Change in the Mean

Finally we turn to model (2.1.7) and investigate the permutation analogue of (2.1.9), i.e. the statistic

$$M_T^{(2)}(\mathbf{R}) = \sqrt{\frac{N}{T\widehat{b}_T^2}}\max_{1\leqslant k<N}\left\{\frac{\left|\sum_{i=1}^{N}(i-k)_+^\gamma(\Delta S_{R_i,T} - \overline{\Delta S}_N)\right|}{\left(\sum_{i=1}^{N-k} i^{2\gamma} - \frac{1}{N}\left(\sum_{i=1}^{N-k} i^\gamma\right)^2\right)^{1/2}}\right\}.$$

The following asymptotic applies:

Theorem 2.3.2. *Let* $\{S(t) : t \geq 0\}$ *be a process according to model* (2.1.7). *Assume* $T^* = \theta T$, $0 < \theta \leqslant 1$, $N = o(T^{1-2/\nu})$, *no restrictions on* a *or* d *are necessary. Then, for all* $x \in \mathbb{R}$, *as* $T \to \infty$,

$$P\left(\alpha_N M_T^{(2)}(\mathbf{R}) - \beta_N \leqslant x \,\middle|\, S(t), 0 \leqslant t \leqslant T\right) \to \exp(-2e^{-x}) \qquad a.s.,$$

where α_N, $\beta_N = \beta_N(\gamma)$ *are as in Theorem 2.1.2.*

Remark 2.3.2. Again the corresponding bootstrap result with replacement also holds true. In that approach we sample with replacement from the increments of the observed process instead of permuting them. Because the corresponding score results holds true under the same assumptions on the scores (confer Remark 2.2.2), this follows immediately from the proof of Theorem 2.3.2.

Proof of Theorem 2.3.2. First note that, for the increments of $\{S(\cdot)\}$, we have

$$\Delta S_{i,T} = \Delta a_i + \Delta Y_i$$

with ΔY_i as in (2.3.1) and

$$\Delta a_i = \begin{cases} a\frac{T}{N}, & i \leqslant N^*, \\ a\frac{T}{N} + d\left(\frac{(N^*+1)T}{N} - T^*\right)^{1+\gamma}, & i = N^*+1, \\ a\frac{T}{N} + d\left(\left(\frac{iT}{N} - T^*\right)^{1+\gamma} - \left(\frac{(i-1)T}{N} - T^*\right)^{1+\gamma}\right), & i \geqslant N^*+2, \end{cases}$$

and $\overline{\Delta a} = \frac{1}{N}\sum_{i=1}^N \Delta a_i = \frac{1}{N}\left(aT + d(T-T^*)^{1+\gamma}\right)$.

Let $\widetilde{\Delta} a_i = \Delta a_i - aT/N$, then the mean value theorem gives uniformly in i

$$\widetilde{\Delta} a_i \ll |d| \frac{T^{1+\gamma}}{N}$$

and uniformly for $i \geqslant N^*+2$ depending on $\gamma \geqslant 1$ or $\gamma < 1$

$$\widetilde{\Delta} a_{i+1} - \widetilde{\Delta} a_i \ll |d| \left(\frac{T}{N}\right)^{1+\gamma} |i \pm 1 - \theta N|^{\gamma-1}.$$

Thus

$$\sum_{i=1}^N |\Delta a_i - \overline{\Delta a}|^\nu \ll |d|^\nu \frac{T^{(1+\gamma)\nu}}{N^{\nu-1}}. \tag{2.3.7}$$

Let $\Delta W_i = W_T^*(iT/N) - W_T^*((i-1)T/N)$, then Lemma 2.3.1 and partial summation show

$$\begin{aligned}
\sum_{i=N^*+2}^N \Delta Y_i \, \widetilde{\Delta} a_i &= \sum_{i=N^*+2}^N \Delta W_i \, \widetilde{\Delta} a_i + O(1)|d| \frac{T^{3/2+\gamma}}{N^{1/2}} \left(\frac{N}{T^{1-2/\nu}}\right)^{1/2} \\
&\ll \sqrt{T \log N} \widetilde{\Delta} a_N + \sum_{i=N^*+2}^{N-1} \sqrt{i \frac{T}{N} \log N} \, |\widetilde{\Delta} a_{i+1} - \widetilde{\Delta} a_i| + o\left(|d| \frac{T^{3/2+\gamma}}{N^{1/2}}\right) \\
&\ll |d| \frac{T^{3/2+\gamma}}{N^{1/2}} \sqrt{\frac{\log N}{N}} \frac{1}{N^{1/2+\gamma}} \sum_{i=N^*+2}^{N-1} i^{1/2} |i \pm 1 - \theta N|^{\gamma-1} + o\left(|d| \frac{T^{3/2+\gamma}}{N^{1/2}}\right) \\
&= o\left(|d| \frac{T^{3/2+\gamma}}{N^{1/2}}\right) \qquad a.s.
\end{aligned}$$

If one takes additionally Lemma 2.3.2 into account we arrive at

$$\sum_{i=1}^N (\Delta Y_i - \overline{\Delta Y})(\Delta a_i - \overline{\Delta a}) = o\left(|d| \frac{T^{3/2+\gamma}}{N^{1/2}}\right) \qquad a.s. \tag{2.3.8}$$

Next we have by the mean value theorem

$$\begin{aligned}
\sum_{i=1}^N (\Delta a_i - \overline{\Delta a})^2 &\geqslant \frac{|d|^2 T^{2+2\gamma}}{N^2} (1+\gamma)^2 \int_{N^*+1}^{N-1} \left(\frac{x}{N} - \theta\right)^{2\gamma} dx - |d|^2 \frac{(T-T^*)^{2+2\gamma}}{N} \\
&= \frac{|d|^2 T^{2+2\gamma}}{N} \left[\frac{(1+\gamma)^2}{2\gamma+1} (1-\theta)^{2\gamma+1} - (1-\theta)^{2\gamma+2} + o(1)\right] \\
&= \frac{|d|^2 T^{2+2\gamma}}{N} \left(\frac{(1-\theta)^{2\gamma+1}}{2\gamma+1} \left(\gamma^2 + \theta(2\gamma+1)\right) + o(1)\right).
\end{aligned} \tag{2.3.9}$$

Consider first the case where $|d|\,\frac{T^{1/2+\gamma}}{N^{1/2}} = O(1)$, which includes the null hypothesis. Then choose the scores $a_N(i) := \sqrt{N/T}\Delta S_{i,T}$. It suffices to prove the assumptions of Theorem 2.2.2. By Lemma 2.3.2 and (2.3.7) it holds

$$\frac{1}{N}\sum_{i=1}^{N}|a_N(i)-\bar{a}_N|^\nu \ll \left(\frac{N}{T}\right)^{\nu/2}\left[\frac{1}{N}\sum_{i=1}^{N}|\Delta a_i - \overline{\Delta a}|^\nu + \frac{1}{N}\sum_{i=1}^{N}|\Delta Y_i - \overline{\Delta Y}|^\nu\right] \ll 1 \qquad a.s.$$

Furthermore Lemma 2.3.2 and (2.3.8) show

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}(a_N(i)-\bar{a}_N)^2 \geqslant \frac{1}{T}\sum_{i=1}^{N}(\Delta Y_i - \overline{\Delta Y})^2 - 2\frac{1}{T}\sum_{i=1}^{N}(\Delta Y_i - \overline{\Delta Y})(\Delta a_i - \overline{\Delta a})\\
&= b^2\theta + (b^*)^2(1-\theta) + o(1) \qquad a.s.
\end{aligned}$$

If, on the other hand, $\frac{N^{1/2}}{|d|T^{1/2+\gamma}} = O(1)$, we choose the scores $a_N(i) := N/(|d|T^{1+\gamma})\Delta S_{i,T}$. Again it suffices to prove the assumptions of Theorem 2.2.2. By Lemma 2.3.2 and (2.3.7) it holds

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}|a_N(i)-\bar{a}_N|^\nu\\
&\ll \left(\frac{N}{|d|T^{1+\gamma}}\right)^{\nu}\left[\frac{1}{N}\sum_{i=1}^{N}|\Delta a_i - \overline{\Delta a}|^\nu + \frac{1}{N}\sum_{i=1}^{N}|\Delta Y_i - \overline{\Delta Y}|^\nu\right] \ll 1 \qquad a.s.
\end{aligned}$$

Furthermore (2.3.8) and (2.3.9) show

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}(a_N(i)-\bar{a}_N)^2\\
&\geqslant \frac{N}{|d|^2T^{2+2\gamma}}\sum_{i=1}^{N}(\Delta a_i - \overline{\Delta a})^2 - 2\frac{N}{|d|^2T^{2+2\gamma}}\sum_{i=1}^{N}(\Delta Y_i - \overline{\Delta Y})(\Delta a_i - \overline{\Delta a})\\
&\geqslant \frac{(1-\theta)^{2\gamma+1}}{2\gamma+1}\left(\gamma^2 + \theta(2\gamma+1)\right) + o(1) \qquad a.s.
\end{aligned}$$

If neither $|d|\,\frac{T^{1/2+\gamma}}{N^{1/2}} = O(1)$ nor $\frac{N^{1/2}}{|d|T^{1/2+\gamma}} = O(1)$ we can divide the sequence into two subsequences each fulfilling one of these conditions. This shows that the assumptions of Theorem 2.2.2 are also fulfilled in that case, completing the proof. ■

3. Block Resampling Methods for the Location Model of Linear Sequences

In the following two chapters we will use resampling methods to obtain better approximations of critical values in the case where the errors in the location model are dependent.

In the last chapter we have developed resampling procedures for possibly dependent data. The corresponding statistics are based on sums of increments of processes, where the increments are taken over larger and larger intervals. The idea behind that is similar to the idea of block resampling methods (confer the following section). But importantly the observations we use for the statistics in the last chapter change over time.

In many practical situation, however, the statistics originally developed for independent observations are used even when the observations are indeed dependent. Antoch et al. [4] and Horváth [45] showed that apart from some minor adjustment the asymptotics remain true for linear processes. The permutation test as described in Section 1.2 only works for independent errors. This means that, here, we have to somewhat change the technique to allow for dependency. In this chapter we will investigate block resampling methods, in the next one resampling methods in the frequency domain.

The first section in this chapter gives a general overview of the history of block bootstrapping techniques. We then give a short introduction into the most commonly used statistics for the location model in Section 3.2. This is followed by their null asymptotics under the given model.
Again the main tool in the proof of the validity of the permutation test is the limit behavior of the corresponding rank statistics. This is developed in Section 3.4. Then, in Section 3.5, we are ready to prove that the block permutation test works in this setting. Along the way we prove the correct asymptotic of a variance estimator based on blocks. Finally, we show that the block bootstrap with replacement also works by first proving the corresponding results for score processes and then concluding the validity of the bootstrap. We then finish with further examples where block resampling techniques may be useful.

3.1. Introduction

Since Efron [25] introduced the bootstrap in 1979, methods based on resampling have been applied to numerous statistical problems. For the bootstrap to work, however, one usually needs independent data. This was shown in 1981 by Singh [76] for m-dependent processes. At the same time he suggested intensive research for time series based on independent residuals. In that situation one can estimate the model parameters and bootstrap the estimated residuals. In the following years this idea was pursued for different models including linear regression (e.g. Freedman [31, 32]) and autoregressive time series (e.g. Bose [10]). Yet, this approach is restricted to situations where the model can be relied upon.

A more general approach that can also be applied if the model class or the model equations are unknown, resamples blocks of data rather than the data points themselves. The motivation is that the dependency structure is preserved within the blocks so that we get asymptotically correct estimates if the block size converges to infinity with the sample size. This was first suggested by Hall [41] in 1985 and Carlstein [15] in 1986. Both of them used the method to estimate variances and both of them used non-overlapping blocks. Later, Künsch [57] and Liu and Singh [60] introduced the so called "moving blocks" bootstrap, where overlapping blocks are used. Politis and Romano [72] 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, whereas otherwise the first and the last observations are underrepresented leading to some bias.

We will apply these techniques to our change-point problem with dependent innovations. In order to use permutations (confer Section 3.5) we have to work with non-overlapping blocks. Yet, for the bootstrap with replacement (confer Section 3.6) we focus our discussion on the circular approach of moving blocks by Politis and Romano [72]. The proofs can, however, be adapted to the regular moving blocks bootstrap (confer Remark 3.6.3).

Note that in the previous chapter we have already used a similar technique. There we deploy it for the null asymptotic not only the permutation test. To be more precise the statistics in Chapter 2 are based on sums of increments of processes between t_{i+1} and t_i. But their distance increases with time. The statistics in this chapter are also based on partial sums and the block technique essentially leads to partial sums of fewer elements, which themselves consist of sums of more and more observations. This is comparable to the increasing intervals from the previous chapter. The biggest difference being that there the original asymptotic was already based on "blocks" and the maximum taken over fewer elements (only at the points of complete "blocks"). The proof for the rank statistic in this chapter also depends crucially on the fact that it is asymptotically equivalent to take the maximum of partial sums including only complete blocks (confer Lemmas 3.4.1 and 3.4.2).

3.2. Statistics for the Location Model

In this section we give a short introduction into the most common statistics used for the location model. In particular we use the *pseudo maximum likelihood method* and the *pseudo Bayes method* to derive the weighted CUSUM statistic (as base type for *maximum-type statistics*) respectively a *sum-type statistic*, which is then generalized to a class of these.

These statistics were originally developed for the location model with i.i.d. errors. More precisely they are derived using the maximum likelihood or Bayes methods for independent normal errors. All the same it can be shown that the statistics work for all non-degenerate sequences of i.i.d. errors as long as the νth moment ($\nu > 2$) exists. For details confer Csörgő and Horváth [19].

Antoch et al. [4] and Horváth [45] showed that they also work for dependent errors that follow a linear process. For details confer Section 3.3.

Recall that the location model is given by

$$X(i) = \mu + d\,1_{\{i>m\}} + e(i), \qquad 1 \leqslant i \leqslant n,$$

and we are interested in testing the null hypothesis of no change

$$H_0 : m = n$$

against the alternative of an abrupt change in the mean, i.e.

$$H_1 : m < n, \qquad d \neq 0.$$

For the derivation of the test statistic we will further assume that the error sequence $\{e(\cdot)\}$ is standard normally distributed. Let $\phi(\cdot)$ be the density function of a standard normal distribution.

Weighted CUSUM Statistic

Suppose for the moment that the change-point m is known. The log-likelihood ratio for testing H_0 against H_1 is then given by

$$\begin{aligned}
&\Lambda_m(X(1), \ldots, X(n)) \\
&= \log \frac{\sup_{a,b} \prod_{i=1}^m \phi(X(i)-a) \prod_{j=m+1}^n \phi(X(j)-b)}{\sup_a \prod_{i=1}^n \phi(X(i)-a)} \\
&= \log \frac{\prod_{i=1}^m \phi(X(i)-\bar{X}_m) \prod_{j=m+1}^n \phi(X(j)-\bar{X}_m^*)}{\prod_{i=1}^n \phi(X(i)-\bar{X}_n)} \\
&= \frac{1}{2}\left(\sum_{i=1}^n (X(i)-\bar{X}_n)^2 - \sum_{i=1}^m (X(i)-\bar{X}_m)^2 - \sum_{i=m+1}^n (X(i)-\bar{X}_m^*)^2 \right) \\
&= \frac{1}{2} \frac{n}{m(n-m)} \left(\sum_{i=1}^m (X(i)-\bar{X}_n) \right)^2,
\end{aligned}$$

where $\bar{X}_m := \frac{1}{m}\sum_{i=1}^m X(i)$ and $\bar{X}_m^* := \frac{1}{n-m}\sum_{i=m+1}^n X(i)$.

Usually the change-point m will be unknown. Then we have to take the supremum of the log-likelihood ratio not only with respect to the means of the process before and after the change but also with respect to the change-point. Thus we get the maximum of the above expression, which is usually equivalently expressed as

$$T_n^{(1)} = \max_{1 \leqslant m < n} \left(\sqrt{\frac{n}{m(n-m)}}\, |S_m| \right), \tag{3.2.1}$$

where $S_m = \sum_{i=1}^m (X(i) - \bar{X}_n)$.

MOSUM Statistic

Here, we introduce a different type of statistic based on moving sums. It has proven useful in cases where there is more than one change. The MOSUM statistic is defined by

$$T_n^{(2)}(G) = \max_{G<m\leqslant n} \frac{1}{\sqrt{G}} \left| S_m - S_{m-G} \right|, \tag{3.2.2}$$

where again $S_m = \sum_{i=1}^m (X(i) - \bar{X}_n)$ and $G < n$. We assume that G/n is small, typically we choose $G/n \approx 0.05$ or 0.01. This is meant as a rule of thumb not in the correct asymptotic mathematical sense (confer assumptions on G in Theorem 3.3.2).

Remark 3.2.1. In recent years special attention has been paid to the following test statistic, which is related to $T_n^{(2)}(G)$ above:

$$\widetilde{T}_n^{(2)}(G) = \max_{G<m\leqslant n-G} \frac{1}{\sqrt{2G}} \left| S_{m+G} - 2S_m + S_{m-G} \right|.$$

It is especially suitable if we expect more than one change and it is useful as a diagnostic tool (cf. also Antoch et. al. [3], Chapter 4.1.3). Note that $T_n^{(2)}(G)$ is the first order difference of S_m´s, whereas $\widetilde{T}_n^{(2)}(G)$ corresponds to the second order difference. For reasons of simplicity we will here refrain from discussing it in detail, yet a similar asymptotic to that of $T_n^{(2)}(G)$ holds true and the proofs in this work can easily be modified for this statistic.

q-weighted CUSUM Statistics

The disadvantage of the above two statistics is that they converge almost surely to infinity as $n \to \infty$ respectively $G \to \infty$. This can be seen by the law of iterated logarithm. Thus, a limit distribution does not exist and the critical values also tend to infinity. Nevertheless, it is possible to use asymptotic results to obtain critical values that depend on n (confer Theorem 3.3.2), but the convergence is very slow.

The problem is caused by small and large values of m. This is why many authors prefer to trim the maximum, for example they use

$$\max_{\epsilon n\leqslant m<n-\epsilon n} \left(\sqrt{\frac{n}{m(n-m)}} \left| S_m \right| \right)$$

for some small $\epsilon > 0$ instead of the weighted CUSUM statistic. This trimmed version converges in distribution. In Chapter 4 we will also use trimmed versions, because it is not possible to prove the results we need for extreme-value asymptotics using the tools developed there (confer also Section 4.8.3).

Another possibility to alter the weighted CUSUM statistic such that it has a limiting distribution is the following:

$$T_n^{(3)}(q) = \max_{1\leqslant m<n} \left(\frac{1}{\sqrt{n}\, q(\frac{m}{n})} \left| S_m \right| \right), \tag{3.2.3}$$

where again $S_m = \sum_{i=1}^{m}(X(i) - \bar{X}_n)$ and $q(\cdot)$ is a weight function defined on $(0,1)$. We assume that the weight function q belongs to the class

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

For the case $q \equiv 1$ we will frequently refer to the statistic as classical CUSUM statistic.

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

$$I^*(q,c) = \int_0^1 \frac{1}{t(1-t)} \exp\left\{\frac{-cq^2(t)}{t(1-t)}\right\} dt. \tag{3.2.4}$$

We require the existence of a c such that $I^*(q,c) < \infty$ in order to show that the statistic converges in distribution (confer Theorem 3.3.2).

A typical class of weight functions fulfilling these assumptions is

$$q(t) = (t(1-t))^{\beta}, \qquad 0 \leqslant \beta < \frac{1}{2}.$$

This also shows the connection with the weighted CUSUM statistic.

For details and further references confer Csörgő and Horváth [18], Chapter 4.

Sum Statistics

We derive now a sum-type statistic using the pseudo Bayes method (for details confer e.g. Chernoff and Zacks [17]).

The method is based on the assumption that the unknown mean μ, the unknown mean change d and the unknown change-point m are independent random variables that are also independent of the error sequence. Let $\mathbf{m}$ be the random variable designating the change-point $m = 1, \ldots, n-1$ with

$$P(\mathbf{m} = m) = \frac{1}{n-1}, \qquad m = 1, \ldots, n-1.$$

The random variables designating the mean μ and mean change d are

$$\boldsymbol{\mu} \overset{\mathcal{D}}{\sim} N(0,\gamma^2) \qquad \text{respectively} \qquad \mathbf{d} \overset{\mathcal{D}}{\sim} N(0,\sigma^2).$$

The density of $(X(1), \ldots, X(n))$ under the null hypothesis given $\boldsymbol{\mu} = \mu$ is normal, more precisely

$$f_H(x_1, \ldots, x_n \,|\, \boldsymbol{\mu} = \mu) = \prod_{i=1}^{n} \phi(x_i - \mu).$$

Thus the unconditional density under the null hypothesis can be expressed by

$$
\begin{aligned}
f_H(x_1,\dots,x_n) &= \prod_{i=1}^{n} \int_{\mathbb{R}} \phi(x_i-\mu)\frac{1}{\gamma\sqrt{2\pi}} \exp\left(-\frac{\mu^2}{2\gamma^2}\right) d\mu \\
&= \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{1+n\gamma^2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2 + \frac{n^2}{2\left(\frac{1}{\gamma^2}+n\right)}\bar{x}_n^2\right),
\end{aligned}
$$

where $\bar{x}_n := \frac{1}{n}\sum_{i=1}^{n} x_i$. Under alternatives the conditional density is

$$
f_A\left(x_1,\dots,x_n \,|\, \mathbf{m}=m, \boldsymbol{\mu}=\mu, \mathbf{d}=d\right) = \prod_{i=1}^{m} \phi(x_i-\mu) \prod_{i=m+1}^{n} \phi(x_i-\mu-d).
$$

After some calculations we obtain for the corresponding unconditional density

$$
\begin{aligned}
&f_A(x_1,\dots,x_n) \\
&= \frac{1}{n-1}\sum_{m=1}^{n-1} \int_{\mathbb{R}}\int_{\mathbb{R}} \prod_{i=1}^{m}\phi(x_i-\mu) \prod_{j=m+1}^{n} \phi(x_j-\mu-d)\frac{1}{2\pi\gamma\sigma}\exp\left(-\frac{\mu^2}{2\gamma^2}-\frac{d^2}{2\sigma^2}\right) d\mu\, dd \\
&= \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right) \sqrt{\frac{1}{(1+n\gamma^2)(1+(n-m)\sigma^2)-(n-m)^2\gamma^2\sigma^2}} \\
&\quad \cdot \exp\left(\frac{1}{2\sigma^2}\left(1+(n-m)\sigma^2\right)c_1^2 + (n-m)c_1c_2 + \frac{1}{2}\left(\frac{1}{\gamma^2}+n\right)c_2^2\right),
\end{aligned}
$$

where

$$
\begin{aligned}
c_1 &\longrightarrow \sigma^2\frac{\sum_{i=m+1}^{n}(x_i-\bar{x}_n)}{1+\frac{m(n-m)}{n}\sigma^2}, \qquad \text{as } \gamma\to\infty, \\
c_2 &\longrightarrow \bar{x}_n - \frac{n-m}{n}c_1, \qquad \text{as } \gamma\to\infty.
\end{aligned}
$$

We look again at the likelihood ratio for fixed n. Because the mean before the change could be any value, we let $\gamma\to\infty$ and obtain:

$$
\begin{aligned}
\Lambda(X(1),\dots,X(n)) &= \frac{f_A(X(1),\dots,X(n))}{f_H(X(1),\dots,X(n))} \\
&\xrightarrow{\gamma\to\infty} \frac{1}{n-1}\sum_{m=1}^{n-1}\sqrt{\frac{1}{1+\frac{m(n-m)}{n}\sigma^2}}\exp\left[\frac{\sigma^2}{2}\frac{1+(n-m)\sigma^2}{\left(1+\frac{m(n-m)}{n}\sigma^2\right)^2}\left(\sum_{i=1}^{m}(X(i)-\bar{X}_n)\right)^2 + o(\sigma^2)\right],
\end{aligned}
$$

where $o(\sigma^2)$ is meant as $\sigma\to 0$.

Small values of σ correspond to weak alternatives, which are more difficult to detect. Applying a Taylor expansion to exp the above likelihood ratio can be written, for small values of σ, as

$$
1+\sigma^2\frac{1}{2(n-1)}\sum_{m=1}^{n-1}\left(\sum_{i=1}^{m}(X(i)-\bar{X}_n)\right)^2 + o(\sigma^2), \qquad \text{as } \sigma\to 0.
$$

Again we reject H_0 whenever the likelihood ratio is greater than an appropriate constant. Thus the above procedure yields the following test statistic:

$$\frac{1}{n}\sum_{m=1}^{n-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{m}(X(i)-\bar{X}_n)\right)^2.$$

Using weight functions as already for q-weighted CUSUM statistics we arrive at a generalized version – the class of *sum-type* statistics:

$$T_n^{(4)}(r)=\frac{1}{n}\sum_{m=1}^{n-1}\frac{1}{r(m/n)}\left(\frac{1}{\sqrt{n}}S_m\right)^2, \tag{3.2.5}$$

where again $S_m=\sum_{i=1}^{m}(X(i)-\bar{X}_n)$ and $r(\cdot)$ is a weight function defined on $(0,1)$.

Moreover we require that the weight function r fulfills for all $x\in(0,1)$

$$r(x)>0 \quad \text{and} \quad \int_0^1 \frac{t(1-t)}{r(t)}\,dt<\infty. \tag{3.2.6}$$

For these weight functions the above statistic has a distributional limit, confer Theorem 3.3.2.

Again typical weight functions fulfilling conditions 3.2.6 are

$$r(t)=(t(1-t))^{\beta}, \qquad 0\leqslant\beta<2.$$

For more details and further references confer Csörgő and Horváth [19], Chapter 2.

3.3. Model and Null Asymptotics

In this section we give a more detailed description of the model we use in this chapter and in Chapter 4. Moreover we state the null asymptotics of the statistics developed in the previous section.

As already mentioned we drop the assumption of independent errors in the location model. Instead we model the error sequence as linear processes.
Linear processes are used frequently in the modeling of time-series data. Consequently the above approach already captures a broad range of practical applications. One important example of linear processes is the so-called (causal) autoregressive moving average or ARMA process. For more details confer Brockwell and Davis [13], Chapter 3. Example 3.3.1 states some properties of the α-mixing coefficients of causal ARMA processes.

We use the following AMOC location model:

$$X(i)=\mu+d\,1_{\{i>m\}}+e(i), \quad 1\leqslant i\leqslant n, \tag{3.3.1}$$

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

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

with weights w_j and innovations $\{\epsilon(\cdot)\}$; $m = m(n)$ is the unknown change-point and $d = d_n$ the mean change.

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

$$\mathrm{E}\,\epsilon(i) = 0, \quad 0 < \sigma^2 = \mathrm{E}\,\epsilon(i)^2 < \infty, \quad \mathrm{E}\,|\epsilon(i)|^{\nu} < \infty \text{ for some } \nu > 2. \tag{3.3.2}$$

Mainly to ensure that the linear process is stationary and strong-mixing or fulfills the assumptions of the BN decomposition (confer Appendix C) we need to impose some more conditions on the weights as well as the innovations.

We suppose that the weights $\{w_s : s \geqslant 0\}$ satisfy

$$\sum_{s \geqslant 0} w_s \neq 0, \quad \sum_{s \geqslant 0} \sqrt{s}|w_s| < \infty. \tag{3.3.3}$$

The random variables $\epsilon(i)$ are smooth with density function f satisfying

$$\sup_{-\infty < s < \infty} \frac{1}{|s|} \int_{-\infty}^{\infty} |f(t+s) - f(t)|\, dt < \infty. \tag{3.3.4}$$

Let

$$g(z) = \sum_{s \geqslant 0} w_s z^s, \quad z \in \mathbb{C},$$

and assume

$$g(z) \neq 0 \qquad \text{for all } |z| \leqslant 1. \tag{3.3.5}$$

Also,

$$w_s = O(s^{-\beta}) \quad \text{as} \quad s \to \infty \quad \text{for some } \beta > 3/2. \tag{3.3.6}$$

In this chapter we investigate the statistics already derived in Section 3.2. In Chapter 4 we will mainly investigate the last two statistics and only use the trimmed versions of the first two. At a glance the 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 $S_m = \sum_{i=1}^{m} (X(i) - \bar{X}_n)$ and $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X(i)$.

Mixing Properties of Linear Processes

Gorodetskii [37] and Withers [82] investigate the strong-mixing properties of linear processes.

We now state a result showing that under the above assumptions the linear process is stationary and strong-mixing (for the definition of strong-mixing see Section B.2). We need it to verify the conditions of Theorem 3.5.1.

Theorem 3.3.1. *Under conditions (3.3.2), (3.3.4) - (3.3.6) we have for* $\beta + 1/2 > \nu > 2/(\beta - 1)$

$$\alpha_e(j) = O(j^{-\rho}), \qquad j \to \infty,$$

where $\rho = (\nu(\beta-1)-2)/(\nu+1)$*. Particularly we have* $\rho \geqslant 2$ *for* $\beta \geqslant 4$ *and* $\nu \geqslant 4$.

Moreover it holds for $w_s = O(e^{-\gamma s})$, $\gamma > 0$,

$$\alpha_e(j) = O\left(e^{-\gamma\lambda j}\right), \qquad j \to \infty,$$

where $\lambda = \nu(1+\nu)^{-1}$.

Proof. Confer Corollary 4 in Withers [82]. ■

Example 3.3.1 (ARMA processes). Causal ARMA processes are an important class of linear processes (confer Chapter 3 of Brockwell and Davis [13]). Equation (3.3.6) there shows that the weights fall exponentially, i.e. $w_s = O(e^{-\gamma s})$ for some $\gamma > 0$. Thus according to the above lemma the alpha-mixing coefficient also falls exponentially. Consequently conditions (3.5.2) respectively (3.5.3) are fulfilled for all δ and Δ, this is important in view of Theorem 3.5.1 .

Null asymptotics

We now state the null asymptotics of the statistics from Section 3.2 which go back to Antoch et al. [4] and Horváth [45]. The proof uses the Beveridge-Nelson decomposition (for details confer Appendix C) and can thus deduce the results from the corresponding ones for i.i.d. error sequences. We will give a short sketch of the proof for the classical CUSUM statistic with an error sequence that forms a linear process. A more detailed discussion of the problem can be found in Csörgő and Horváth [19], Section 4.1, or in the above mentioned articles.

Theorem 3.3.2. *Assume that (3.3.1) - (3.3.3) 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 $\widehat{\tau} - \tau = o_P((\log\log n)^{-1})$, $\tau^2 := \sigma^2\left(\sum_{s\geqslant 0} w_s\right)^2$ $(\tau > 0)$.

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

d) If r fulfills condition (3.2.6), 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 $\{B(t) : 0 \leqslant t \leqslant 1\}$ denotes a Brownian bridge and $\widehat{\tau} - \tau = o_P(1)$.

Remark 3.3.1. Horváth [45], Theorem 1.3, and Csörgő and Horváth [19], Theorem 4.1.2 (additionally to Example 4.1.1) and Theorem 4.1.3, prove the assertion for CUSUM statistics under (3.3.4) - (3.3.6) instead of (3.3.3). The latter also shows that the asymptotic for the q-weighted CUSUM statistics remains true for somewhat more general error sequences. The bootstrap methods then also hold if the necessary strong laws of large numbers are fulfilled, i.e. one only has to prove equations (3.5.4) - (3.5.8) respectively (3.6.7) - (3.6.8).

Proof of Theorem 3.3.2. Confer Theorem 2.1 in Antoch et al. [4]. The assumptions there are somewhat stronger, namely they require $\sum_{s\geqslant 0} s|w_s| < \infty$. This is, however, only needed to obtain the BN decomposition (confer C.1), which also holds under $\sum_{s\geqslant 0} \sqrt{s}|w_s| < \infty$.

To give an impression of the techniques we prove the result for the classical CUSUM statistic. This is the easiest example and one has to refine the methods to derive the results for different statistics. The proof essentially reduces to an application of the Beveridge-Nelson decomposition (confer Lemma C.1) additionally to an application of the corresponding results for i.i.d. sequences. The BN decomposition gives

$$e(i) = \epsilon(i) \sum_{s\geqslant 0} w_s - \widetilde{e}(i) + \widetilde{e}(i-1), \tag{3.3.7}$$

where $\{\widetilde{e}(\cdot)\}$ is a stationary process with $\mathrm{E}\,|\widetilde{e}(0)|^p < \infty$ for any $p < \nu$. Thus

$$\bar{e}_n - \bar{\epsilon}_n \sum_{s\geqslant 0} w_s = \frac{1}{n}(\widetilde{e}(0) - \widetilde{e}(n)) = o_P\left(n^{-1/2}\right).$$

Furthermore

$$\max_{k=1,\dots,n} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{k} \left[e(i) - \epsilon(i) \sum_{s\geqslant 0} w_s \right] \right| = \frac{1}{\sqrt{n}} \max_{k=1,\dots,n} |\widetilde{e}(0) - \widetilde{e}(k)| = o_P(1), \tag{3.3.8}$$

since the Markov inequality gives for $2 < p < \nu$

$$P\left(\frac{1}{\sqrt{n}} \max_{k=1,\ldots,n} |\widetilde{e}(k)| \geqslant \delta\right) \leqslant \frac{1}{\delta^{p/2}} \sum_{k=1}^{n} \frac{1}{n^{p/2}} \operatorname{E} |\widetilde{e}(k)|^p = o(1).$$

Putting together equations (3.3.7) and (3.3.8) we note that it suffices to investigate

$$\left|\sum_{s\geqslant 0} w_s\right| \max_{k=1,\ldots,n} \left|\frac{1}{\sqrt{n}} \sum_{i=1}^{k} (\epsilon(i) - \bar{\epsilon}_n)\right|.$$

Using the above theorem for i.i.d. errors (confer e.g. Csörgő and Horváth [19], Theorem 2.1.1) we arrive at the assertion. It also explains why the variance correction term is $\sigma^2 \left(\sum w_s\right)^2$ instead of $\sigma^2 \sum w_s^2$, which one might have suspected. ■

Remark 3.3.2. In very much the same way it is possible to obtain the consistency of the test from the consistency for the i.i.d. case. Theorem 2.4.13 in Csörgő and Horváth [19] ($h(x,y) = x - y$) shows the asymptotic consistency of the classical CUSUM test for certain alternatives. More precisely: If $d \neq 0$ is fixed and

$$\frac{m(n-m)}{n^{3/2}} \to \infty,$$

then it holds under H_A for all $x \in \mathbb{R}$ as $n \to \infty$

$$P\left(T_n^{(1)}(q_1) > x\right) \to 1,$$

where $q_1 \equiv 1$. The result holds if the error sequence $\{e(\cdot)\}$ is i.i.d. with existing νth moment ($\nu > 2$). Using the same arguments as in the proof of Theorem 3.3.2 we can derive the consistency also in the case where the error sequence forms a linear process fulfilling (3.3.2) and (3.3.3). This is true under the same alternatives as for the i.i.d. case. Note that for the alternatives where the above test is consistent and where the bootstrap is valid we automatically have consistency of the bootstrap test.

3.4. Asymptotics of the Corresponding Block Rank Statistics

It is well known that the rate of convergence of these statistics (especially the extreme value statistics) can be very slow. So we are interested in alternative methods to derive critical values, especially the permutation method by Hušková [47], confer also Section 1.2.

We assume that we split our 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.

Remark 3.4.1. 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. Also it is always possible to ignore the last ($n \mod L$) observations to use the below theory, so this is not much of a restriction.
For most results we need $L \to \infty$ as well as $K \to \infty$ as $n = KL \to \infty$. However, this cannot be simultaneously fulfilled due to prime numbers n.
We will assume that $L \to \infty$ and $K = K(L)$, $n = n(L) = KL$. Otherwise we have more than one K for each L, in which case we cannot use Corollary D.1 below anymore.
If there is only one K for each L the random sequence $\left\{ \frac{1}{\sqrt{K}} \sum_{k=1}^{K} e(Kl+k), 1 \leqslant l \leqslant L \right\}_L$ forms a triangular array in L instead of one in n. The same problem arises when proving equations (3.5.7) respectively (3.5.8) below.

We have already seen that the limit behavior of the corresponding rank statistics is crucial in proving the validity of permutation tests. On the other hand it is also of independent interest. In this section we prove the rank asymptotics from which we can deduce the limit behavior of the permutation statistics in the next section.

The corresponding rank statistics are based on partial sums

$$S^{\mathbf{a}}_{L,K}(l,k) := \sum_{i=1}^{l-1}\sum_{j=1}^{K}(a_n[K(R_i-1)+j] - \bar{a}_n) + \sum_{j=1}^{k}(a_n[K(R_l-1)+j] - \bar{a}_n),$$

where $\mathbf{R} = (R_1, \ldots, R_L)$ is a random permutation of $(1, \ldots, L)$. Precisely we are interested in:

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

$$T^{(2)}_{L,K}(G,\mathbf{a}) := \frac{1}{\sqrt{G}} \max_{\substack{1\leqslant l\leqslant L, 1\leqslant k\leqslant K\\ K(l-1)+k>G}} \left|S^{\mathbf{a}}_{L,K}(l,k) - S^{\mathbf{a}}_{L,K}(l^*,k^*)\right|,$$

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

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

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

Theorem 3.4.1. *Let $\mathbf{R} = (R_1, \ldots, R_L)$ be a random permutation of $(1, \ldots, L)$. Moreover let $a_n(1), \ldots, a_n(n)$ be scores satisfying*

$$\frac{1}{L}\sum_{l=0}^{L-1} \max_{k=0,\ldots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K}(a_n(Kl+j)-\bar{a}_n)\right|^{\kappa} \leqslant D_1 \tag{3.4.1}$$

for some $\kappa > 2$ and

$$\tau_n^2(\mathbf{a}) := \frac{1}{L}\sum_{l=0}^{L-1}\left[\frac{1}{\sqrt{K}}\sum_{k=1}^{K}(a_n(Kl+k)-\bar{a}_n)\right]^2 \geqslant D_2, \tag{3.4.2}$$

where $\bar{a}_n := \frac{1}{n}\sum_{i=1}^n a_n(i)$ *and* $D_1, D_2 > 0$ *are some constants. Let* $\alpha(x)$, $\beta(x)$ *be as in Theorem 3.3.2.*

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

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

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

$$G^{-1}L^{-2\mu} n \log(n/G) = o(1) \qquad \textit{for some} \quad 0 \leqslant \mu < \min\left(\frac{\kappa-2}{2\kappa}, \frac{1}{4}\right), \tag{3.4.3}$$

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

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

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

$$\frac{1}{L\,q^2\left(\frac{1}{KL}\right)} \to 0, \qquad \frac{1}{L\,q^2\left(1-\frac{1}{KL}\right)} \to 0, \tag{3.4.4}$$

then

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

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

d) If r *fulfills condition (3.2.6) and as* $L \to \infty$

$$\frac{1}{L^2K}\sum_{k=1}^{K} \frac{1}{r\left(\frac{k}{KL}\right)} \to 0, \qquad \frac{1}{L^2K}\sum_{k=1}^{K-1} \frac{1}{r\left(1-\frac{k}{KL}\right)} \to 0, \tag{3.4.5}$$

then

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

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

Remark 3.4.2. Concerning the weighted CUSUM-statistic $\widetilde{T}^{(1)}_{L,K}(\mathbf{a})$ 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 by Lemma F.1, if

$$\max_{1\leqslant k\leqslant K}\left|\frac{1}{\sqrt{k}}\sum_{j=1}^{k}(a_n(K(R_1-1)+j)-\bar{a}_n)\right| = o_P\left(\sqrt{\log\log n}\right)$$

$$\text{and} \quad \max_{1\leqslant k\leqslant K}\left|\frac{1}{\sqrt{k}}\sum_{j=K-k+1}^{K}(a_n(K(R_1-1)+j)-\bar{a}_n)\right| = o_P\left(\sqrt{\log\log n}\right).$$

The Markov inequality implies that this is fulfilled if

$$\begin{aligned} &\frac{1}{(\log\log n)^{\mu/2}}\frac{1}{L}\sum_{l=1}^{L}\max_{k=1,\dots,K}\left|\frac{1}{\sqrt{k}}\sum_{j=1}^{k}(a_n(K(l-1)+j)-\bar{a}_n)\right|^{\mu}\to 0\\ \text{and}\quad &\frac{1}{(\log\log n)^{\mu/2}}\frac{1}{L}\sum_{l=1}^{L}\max_{k=1,\dots,K}\left|\frac{1}{\sqrt{k}}\sum_{j=K-k+1}^{K}(a_n(K(l-1)+j)-\bar{a}_n)\right|^{\mu}\to 0 \end{aligned} \tag{3.4.6}$$

for some $\mu > 0$.

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

Lemma 3.4.1. *Let $\mathbf{R} = (R_1, \dots, R_L)$ be a random permutation of $(1, \dots, L)$. Moreover let $a_n(1), \dots, a_n(n)$ be scores satisfying*

$$\frac{1}{L}\sum_{l=0}^{L-1}\max_{k=1,\dots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K}(a_n(Kl+j)-\bar{a}_n)\right|^{\kappa}\leqslant D \quad \textit{for some constant } D, \tag{3.4.7}$$

where $\kappa > 2$ and $\bar{a}_n := \frac{1}{n}\sum_{i=1}^{n} a_n(i)$. Then we have for all $\mu < \min\left(\frac{\kappa-2}{2\kappa}, \frac{1}{4}\right)$

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

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

$$\begin{aligned} &P\left(\max_{1\leqslant l\leqslant L-1}\max_{1\leqslant k\leqslant K}\left(\frac{L}{l(L-l)}\right)^{1/2-\mu}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K}(a_n[K(R_l-1)+j]-\bar{a}_n)\right|\geqslant C\right)\\ &\leqslant\sum_{l=1}^{L-1}P\left(\max_{1\leqslant k\leqslant K}\left(\frac{L}{l(L-l)}\right)^{1/2-\mu}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K}(a_n[K(R_l-1)+j]-\bar{a}_n)\right|\geqslant C\right)\\ &\leqslant\sum_{l=1}^{L-1}\frac{1}{C^{\kappa}}\left(\frac{L}{l(L-l)}\right)^{(1/2-\mu)\kappa}\mathrm{E}\left(\max_{1\leqslant k\leqslant K}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K}(a_n[K(R_l-1)+j]-\bar{a}_n)\right|^{\kappa}\right)\\ &\leqslant\epsilon. \end{aligned}$$

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

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

■

The next lemma corresponds to the one above, since it deals with the increments of Brownian bridges. It is based on results of Csörgő and Révész [21].

Lemma 3.4.2. *Let $\{B(t) : 0 \leqslant t \leqslant 1\}$ be a Brownian bridge. Then it holds:*

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

b) $$\max_{\substack{(\log L)^s\leqslant l\leqslant L-(\log L)^s\\ 1\leqslant k\leqslant K}} \frac{L}{\sqrt{l(L-l)}}\left|B\left(\frac{l}{L}\right) - B\left(\frac{K(l-1)+k}{KL}\right)\right|$$
$$= O((\log L)^{\frac{1-s}{2}}) \quad a.s., \qquad \text{for any } s \geqslant 0.$$

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 [21], and implies b).
Note that b) implies c) for $\log L \leqslant l \leqslant L - \log L$.

Further 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 now gives the assertion for $(K-k)/(LK)W(L)$ so that it suffices to investigate the maximum of $|W(l) - W(l - (K-k)/K)|$. The assertion for $l = 1$ 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 and Révész [21] gives now

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

because

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

Since $\{B(t)\} \stackrel{\mathcal{D}}{=} \{B(1-t)\}$, we have

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

Now we get analogously to above

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

Putting everything together we arrive at the assertion. ■

Now, we are ready to prove the main theorem of this section. In addition to the above lemmas we use Corollary D.1. The techniques to obtain the correct limit results from this kind of weighted embeddings are well established in change-point analysis.

Proof of Theorem 3.4.1. The idea of the proof is the following: We use Corollary D.1 for

$$b_n(i) := \frac{1}{\sqrt{K}}\sum_{k=1}^{K} a_n(K(i-1)+k).$$

To be able to do that, we need Lemma 3.4.1 to deal with the difference between the statistic we get this way and $T_{L,K}(\mathbf{a})$. Very much in the same way we need Lemma 3.4.2 to deal with the difference between $B(l/L)$ and $B([K(l-1)+k]/[LK])$.

For notational convenience let $\widetilde{a}_n(i,j) := a_n(K(i-1)+j)-\bar{a}_n$. Corollary D.1 states for $\mu < \min\left(\frac{\kappa-2}{2\kappa}, \frac{1}{4}\right)$

$$\max_{1\leqslant l\leqslant L-1}\left(\frac{l(L-l)}{L}\right)^{\mu}\frac{L}{\sqrt{l(L-l)}}\left|\frac{1}{\sqrt{L}}\Pi(l)-B\left(\frac{l}{L}\right)\right|=O_P(1), \tag{3.4.8}$$

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

$$\{\Pi(l) : 1\leqslant l\leqslant L\} \stackrel{\mathcal{D}}{=} \left\{\frac{1}{\sqrt{K\tau_n^2(\mathbf{a})}}\sum_{i=1}^{l}\sum_{j=1}^{K}\widetilde{a}_n(R_i,j) : 1\leqslant l\leqslant L\right\}. \tag{3.4.9}$$

First we prove assertion a). Since

$$\max_{2\leqslant l\leqslant L-1}\ \max_{1\leqslant k\leqslant K}\frac{l(L-l)}{(l-1+\frac{k}{K})(L-(l-1)-\frac{k}{K})}=O(1), \tag{3.4.10}$$

equation (3.4.8) yields

$$\begin{aligned}\max_{2\leqslant l\leqslant L-1}\ \max_{1\leqslant k\leqslant K}&\left(\frac{(l-1+\frac{k}{K})(L-(l-1)-\frac{k}{K})}{L}\right)^{\mu}\\ &\frac{L}{\sqrt{(l-1+k/K)(L-(l-1)-k/K)}}\left|\frac{1}{\sqrt{L}}\Pi(l)-B(l/L)\right|=O_P(1),\end{aligned} \tag{3.4.11}$$

Moreover assumption (3.4.2) and Lemma 3.4.1 give

$$\begin{aligned}
&\max_{2\leqslant l\leqslant L-1}\max_{1\leqslant k\leqslant K}\left(\frac{(l-1+\frac{k}{K})(L-l+1-\frac{k}{K})}{L}\right)^{\mu-1/2}\frac{1}{\tau_n(\mathbf{a})}\\
&\qquad\cdot\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K}\widetilde{a}_n(R_l,j)\right|\\
&\ll\max_{2\leqslant l\leqslant L-1}\max_{1\leqslant k\leqslant K}\left(\frac{l(L-l)}{(l-1+\frac{k}{K})(L-(l-1)-\frac{k}{K})}\right)^{1/2-\mu}\\
&\qquad\cdot\max_{2\leqslant l\leqslant L-1}\max_{1\leqslant k\leqslant K}\left(\frac{L}{l(L-l)}\right)^{1/2-\mu}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K}\widetilde{a}_n(R_l,j)\right|=O_P(1).
\end{aligned}\tag{3.4.12}$$

Using (3.4.11) and (3.4.12) with $\mu=0$ we get by the law of iterated logarithm

$$\begin{aligned}
&\max_{2\leqslant l\leqslant(\log L)^2}\max_{1\leqslant k\leqslant K}\sqrt{\frac{KL}{(K(l-1)+k)(KL-K(l-1)-k)}}\frac{1}{\tau_n(\mathbf{a})}\\
&\qquad\cdot\left|\sum_{i=1}^{l-1}\sum_{j=1}^{K}\widetilde{a}_n(R_i,j)+\sum_{j=1}^{k}\widetilde{a}_n(R_l,j)\right|\\
&=\max_{2\leqslant l\leqslant(\log L)^2}\max_{1\leqslant k\leqslant K}\sqrt{\frac{KL}{(K(l-1)+k)(KL-K(l-1)-k)}}\frac{1}{\tau_n(\mathbf{a})}\left|\sum_{i=1}^{l}\sum_{j=1}^{K}\widetilde{a}_n(R_i,j)\right|\\
&\qquad+O_P(1)\\
&=\max_{2\leqslant l\leqslant(\log L)^2}\max_{1\leqslant k\leqslant K}\sqrt{\frac{L}{(l-1+\frac{k}{K})(L-l+1-\frac{k}{K})}}\sqrt{L}B\left(\frac{l}{L}\right)+O_P(1)\\
&=O_P(1)+O_P\left(\max_{2\leqslant l\leqslant(\log L)^2}\sqrt{\frac{L}{l(L-l)}}\left(\sqrt{l\log\log l}+\frac{l}{L}\sqrt{L\log\log L}\right)\right)\\
&=o_P(\sqrt{\log\log n}).
\end{aligned}$$

Analogously we get

$$\begin{aligned}
&\max_{L-(\log L)^2\leqslant l\leqslant L-1}\max_{1\leqslant k\leqslant K}\sqrt{\frac{KL}{(K(l-1)+k)(KL-K(l-1)-k)}}\frac{1}{\tau_n(\mathbf{a})}\\
&\qquad\cdot\left|\sum_{i=1}^{l-1}\sum_{j=1}^{K}\widetilde{a}_n(R_i,j)+\sum_{j=1}^{k}\widetilde{a}_n(R_l,j)\right|=o_P(\sqrt{\log\log n}).
\end{aligned}$$

Lemma F.1 shows that it suffices to consider the maximum over $(\log L)^2<l<L-(\log L)^2$.

Using (3.4.11) and (3.4.12) again we arrive now at

$$\begin{aligned}
&\max_{(\log L)^2<l\leqslant L-(\log L)^2}\ \max_{1\leqslant k\leqslant K}\sqrt{\frac{KL}{(K(l-1)+k)(KL-K(l-1)-k)}}\frac{1}{\tau_n(\mathbf{a})}\\
&\qquad\cdot\left|\sum_{i=1}^{l-1}\sum_{j=1}^{K}\widetilde{a}_n(R_i,j)+\sum_{j=1}^{k}\widetilde{a}_n(R_l,j)\right|\\
&=\max_{(\log L)^2<l\leqslant L-(\log L)^2}\ \max_{1\leqslant k\leqslant K}\sqrt{\frac{KL}{(K(l-1)+k)(KL-K(l-1)-k)}}\frac{1}{\tau_n(\mathbf{a})}\\
&\qquad\cdot\left|\sum_{i=1}^{l}\sum_{j=1}^{K}\widetilde{a}_n(R_i,j)\right|+O_P\left((\log L)^{-2\mu}\right)\\
&=\max_{(\log L)^2<l\leqslant L-(\log L)^2}\ \max_{1\leqslant k\leqslant K}\sqrt{\frac{L}{(l-1+\frac{k}{K})(L-l+1-\frac{k}{K})}}\sqrt{L}\left|B\left(\frac{l}{L}\right)\right|\\
&\qquad+o_P\left((\log\log n)^{-1/2}\right).
\end{aligned}$$

This means that $\alpha(n)\, T_{L,K}^{(1)}(\mathbf{a})/\tau_n(\mathbf{a})-\beta(n)$ has the same limit distribution as

$$\alpha(n)\max_{(\log L)^2<l<L-(\log L)^2}\ \max_{1\leqslant k\leqslant K}\sqrt{\frac{L}{(l-1+\frac{k}{K})(L-l+1-\frac{k}{K})}}\sqrt{L}\left|B\left(\frac{l}{L}\right)\right|-\beta(n).$$

Lemma 3.4.2 b) and (3.4.10) now give

$$\begin{aligned}
&\max_{\substack{(\log L)^2<l\leqslant L-(\log L)^2\\1\leqslant k\leqslant K}}\sqrt{\frac{L}{(l-1+\frac{k}{K})(L-l+1-\frac{k}{K})}}\sqrt{L}\left|B\left(\frac{l}{L}\right)\right|\\
&=\max_{\substack{(\log L)^2<l\leqslant L-(\log L)^2\\1\leqslant k\leqslant K}}\sqrt{\frac{KL}{(Kl-K+k)(KL-Kl+K-k)}}\sqrt{KL}\left|B\left(\frac{K(l-1)+k}{KL}\right)\right|\\
&\quad+o_P((\log\log n)^{-1/2})\\
&\stackrel{\mathcal{D}}{=}\max_{(\log L)^2K<m\leqslant n-(\log L)^2K}\sqrt{\frac{n}{m(n-m)}}\left|W(m)-\frac{m}{n}W(n)\right|+o_P((\log\log n)^{-1/2}).
\end{aligned}$$

The law of iterated logarithm, $K=(\log n)^\gamma$, and Lemma F.1 yield as above that it is equivalent to investigate the maximum over $[1,\ldots,n-1]$. Assertion a) then follows from Theorem 3.3.2 for i.i.d. standard normal random variables.

Now we prove assertion b). First we have

$$\frac{T_{L,K}^{(2)}(G,\mathbf{a})}{\tau_n(\mathbf{a})}=\frac{1}{\sqrt{K\tau_n^2(\mathbf{a})}}\max_{\substack{1\leqslant l\leqslant L,1\leqslant k\leqslant K\\K(l-1)+k>G}}\sqrt{\frac{K}{G}}\left|\sum_{i=l^*+1}^{l}\sum_{j=1}^{K}\widetilde{a}_n(R_i,j)\right|+o_P((\log(n/G))^{-1/2}),$$

since (3.4.1), (3.4.2) and the Markov inequality give analogously to the proof of Lemma 3.4.1

$$\begin{aligned}
&\frac{1}{\sqrt{\tau_n^2(\mathbf{a})}}\max_{1\leqslant l\leqslant L}\sqrt{\frac{K}{G}}\max_{1\leqslant k\leqslant K}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K}\widetilde{a}_n(R_l,j)\right|=L\left(\frac{K}{G}\right)^{\kappa/2}O_P(1)\\
&=o_P((\log(n/G))^{-1/2}).
\end{aligned}$$

Note that

$$L\left(\frac{K}{G}\right)^{\kappa/2} = \left(\frac{n\log(n/G)}{GL^{2\mu}}\right)^{\kappa/2}\frac{1}{\log(n/G)^{\kappa/2}L^{\kappa/2-1-\mu\kappa}} = o((\log(n/G))^{-1/2}),$$

because of (3.4.3) and $\kappa/2-1-\mu\kappa>0$. Equation (3.4.8) now gives

$$\max_{1\leqslant l\leqslant L}\sqrt{\frac{K}{G}}\left|\Pi(l)-\sqrt{L}B(l/L)\right| = O_P(1)\max_{1\leqslant l\leqslant L}\sqrt{\frac{K}{G}}\left(\frac{l(L-l)}{L}\right)^{1/2-\mu} = O_P\left(\sqrt{\frac{n}{GL^{2\mu}}}\right)$$
$$= o_P((\log(n/G))^{-1/2}).$$

Thus we deduce from the triangle inequality

$$\frac{T^{(2)}_{L,K}(G,\mathbf{a})}{\tau_n(\mathbf{a})} \overset{\mathcal{D}}{=} \sqrt{\frac{K}{G}}\max_{l,k}\sqrt{L}\,|B(l/L)-B(l^*/L)| + o_P((\log(n/G))^{-1/2})$$
$$= \sqrt{\frac{K}{G}}\max_{l,k}\sqrt{L}\left|B\left(\frac{K(l-1)+k}{KL}\right)-B\left(\frac{K(l^*-1)+k^*}{KL}\right)\right| + o_P((\log(n/G))^{-1/2}),$$

since Lemma 3.4.2 a) yields

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

Note that (3.4.3) implies $\frac{K}{G}\log L\,\log(n/G)=o(1)$. The assertion now follows as above from the case of independent standard normally distributed random variables (confer e.g. Theorem 3.3.2), since (3.4.3) implies $G^{-1}n^{2/\kappa}\log n\to 0$.

Next we prove assertion c). Since $\max_{\eta<t<1-\eta}q(t)\geqslant C(\eta)>0$ for all $0<\eta<1/2$, (3.4.8) gives as $n\to\infty$

$$\max_{\eta L\leqslant l\leqslant(1-\eta)L}\max_{1\leqslant k\leqslant K}\frac{1}{q\left(\frac{K(l-1)+k}{KL}\right)}\left|\frac{1}{\sqrt{L}}\Pi(l)-B\left(\frac{l}{L}\right)\right| = O_P(L^{-\mu}) = o_P(1),$$

where the constants depend on η.

Furthermore we get uniformly in L using again (3.4.10) and Lemma F.3 a)

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

The term is equal to 0 for $l = L$ and for $l = 1$ assumption (3.4.4) and (3.4.8) give as $L \to \infty$

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

since $\frac{1}{Lq^2(1/KL)} \to 0$ and q is non-decreasing in a neighborhood of 0.

On first choosing η small enough and then in dependence of η an L_0 big enough, we now get

$$\max_{\substack{1 \leqslant l \leqslant L, 1 \leqslant k \leqslant K \\ (l,k) \neq (L,K)}} \frac{1}{q\left(\frac{K(l-1)+k}{KL}\right)} \left| \frac{1}{\sqrt{L}} \Pi(l) - B\left(\frac{l}{L}\right) \right| = o_P(1) \quad \text{as } L \to \infty. \tag{3.4.13}$$

Lemma 3.4.1 and assumptions (3.4.1) respectively (3.4.2) show as $L \to \infty$

$$\frac{1}{\sqrt{\tau_n^2(\mathbf{a})}} \max_{\substack{1 \leqslant l \leqslant L-1 \\ 1 \leqslant k \leqslant K}} \left(\frac{l(L-l)}{L} \right)^{\mu} \frac{L}{\sqrt{l(L-l)}} \left| \frac{1}{\sqrt{LK}} \sum_{j=k+1}^{K} \widetilde{a}_n(R_l, j) \right| = O_P(1).$$

Analogously to above this yields for $L \to \infty$

$$\max_{\substack{1 \leqslant l \leqslant L, 1 \leqslant k \leqslant K \\ (l,k) \neq (L,K)}} \frac{1}{\sqrt{LK\tau_n^2(\mathbf{a})}} \frac{1}{q\left(\frac{K(l-1)+k}{KL}\right)} \left| \sum_{j=k+1}^{K} \widetilde{a}_n(R_l, j) \right| = o_P(1), \tag{3.4.14}$$

where an application of the Markov inequality gives the assertion for $l = L$, because of assumptions (3.4.1) and (3.4.4) and the fact that q is non-increasing in a neighborhood of 1.

Now we deduce from (3.4.13) and (3.4.14) as in the proof of a)

$$\frac{T_{L,K}^{(3)}(q, \mathbf{a})}{\tau_n(\mathbf{a})} \stackrel{\mathcal{D}}{=} \max_{\substack{1 \leqslant l \leqslant L, 1 \leqslant k \leqslant K \\ (l,k) \neq (L,K)}} \frac{\left| B\left(\frac{l}{L}\right) \right|}{q\left(\frac{K(l-1)+k}{KL}\right)} + o_P(1). \tag{3.4.15}$$

Lemma F.3 a) together with the fact that $\inf_{\eta \leqslant t \leqslant 1-\eta} q(t) > 0$, for $0 < \eta < 1/2$ yields

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

Thus Lemma 3.4.2 b) and equations (3.4.10) imply

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

We realize that assumption (3.4.4) respectively Lemma F.3 a) yield in view of (3.4.10)

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

Using Lemma 3.4.2 c) we thus obtain

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

Finally we have for $l=L$ with assumption (3.4.4)

$$\begin{aligned}
&\max_{1\leqslant k<K}\frac{\left|B\left(1-\frac{K-k}{KL}\right)\right|}{q\left(1-\frac{K-k}{KL}\right)}\\
&\leqslant \frac{1}{\sqrt{L}\,q\left(1-\frac{1}{KL}\right)}\max_{1\leqslant k<K}\sqrt{L}\left|B\left(1-\frac{K-k}{KL}\right)\right| = o_P(1),
\end{aligned}$$

since $\sqrt{L}\,B[1-(K-k)/(KL)] = \sqrt{L}\,\widetilde{B}[(K-k)/(KL)] = W_L[(K-k)/K]-\frac{K-k}{K\sqrt{L}}\widetilde{W}(1)$ for a suitable Brownian bridge $\{\widetilde{B}(t):0\leqslant t\leqslant 1\}$ and suitable Wiener processes $\{W_L(t):t\geqslant 0\}$ respectively $\{\widetilde{W}(t):t\geqslant 0\}$.

This yields

$$\max_{\substack{1\leqslant l\leqslant L,1\leqslant k\leqslant K\\ (l,k)\neq(L,K)}}\frac{\left|B\left(\frac{l}{L}\right)\right|}{q\left(\frac{K(l-1)+k}{KL}\right)} = \max_{\substack{1\leqslant l\leqslant L,1\leqslant k\leqslant K\\ (l,k)\neq(L,K)}}\frac{\left|B\left(\frac{K(l-1)+k}{KL}\right)\right|}{q\left(\frac{K(l-1)+k}{KL}\right)}+o_P(1). \tag{3.4.16}$$

Equations (3.4.15) and (3.4.16) show that it suffices to consider the asymptotic behavior of $\max_{1\leqslant m<n}\frac{1}{\sqrt{n}\,q(m/n)}\left|\sum_{i=1}^m(X_i-\bar{X}_n)\right|$, where $\{X_i\colon i\geqslant 1\}$ are i.i.d. standard normal random variables. Theorem 3.3.2 now gives assertion c).

To prove assertion d) we first realize

$$\frac{1}{KL}\sum_{\substack{1\leqslant l\leqslant L,1\leqslant k\leqslant K\\ (l,k)\neq(L,K)}}\frac{1}{r\left(\frac{K(l-1)+k}{KL}\right)}\left(\frac{1}{\sqrt{L}}\Pi(l)-B(l/L)\right)^2 = o_P(1), \tag{3.4.17}$$

where again $\{\Pi(l):1\leqslant l\leqslant L\}$ is as in (3.4.9).
We deduce this assertion from (3.4.8) in a way similar to the proof of c). Assumptions (3.2.6) and (3.4.10) give for $0<\eta<1/2$

$$\begin{aligned}
&\frac{1}{n}\sum_{l=\eta L}^{(1-\eta)L}\sum_{k=1}^{K}\frac{1}{r\left(\frac{K(l-1)+k}{KL}\right)}\left(\frac{1}{\sqrt{L}}\Pi(l)-B(l/L)\right)^2\\
&= O_P(1)\cdot\frac{1}{KL}\sum_{l=\eta L}^{(1-\eta)L}\sum_{k=1}^{K}\frac{\frac{l}{L}\,\frac{L-l}{L}}{r\left(\frac{K(l-1)+k}{KL}\right)}\left(\frac{L}{l(L-l)}\right)^{2\mu} = O_P(L^{-2\mu}) = o_P(1),
\end{aligned}$$

where the constants depend on η. Moreover (3.4.10) yields

$$\frac{1}{KL}\sum_{l=2}^{\eta L}\sum_{k=1}^{K}\frac{1}{r\left(\frac{K(l-1)+k}{KL}\right)}\left(\frac{1}{\sqrt{L}}\Pi(l)-B(l/L)\right)^2$$
$$\ll \max_{1\leqslant l<L}\frac{L^2}{l(L-l)}\left(\frac{1}{\sqrt{L}}\Pi(l)-B(l/L)\right)^2\frac{1}{KL}\sum_{j=2}^{\eta L}\sum_{k=1}^{K}\frac{\frac{K(j-1)+k}{KL}\,\frac{KL-K(j-1)-k}{KL}}{r\left(\frac{K(j-1)+k}{KL}\right)}.$$

Now (3.4.8) with $\mu=0$ implies

$$\max_{1\leqslant l<L}\frac{L^2}{l(L-l)}\left(\frac{1}{\sqrt{L}}\Pi(l)-B(l/L)\right)^2$$
$$\cdot\left|\frac{1}{KL}\sum_{j=2}^{\eta L}\sum_{k=1}^{K}\frac{\frac{K(j-1)+k}{KL}\,\frac{KL-K(j-1)-k}{KL}}{r\left(\frac{K(j-1)+k}{KL}\right)}-\eta\int_0^{\eta}\frac{t(1-t)}{r\left(t\right)}\,dt\right|=o_P(1)$$

as $L\to\infty$ for any $\eta>0$ (constants may depend on η). Finally assumption (3.2.6) together with (3.4.8) ($\mu=0$) give

$$\max_{1\leqslant l<L}\frac{L^2}{l(L-l)}\left(\frac{1}{\sqrt{L}}\Pi(l)-B(l/L)\right)^2\eta\int_0^{\eta}\frac{t(1-t)}{r\left(t\right)}\,dt=o_P(1)$$

as $\eta\to 0$ uniformly in L.

We get an analogous expressions for the sum of l over $[(1-\eta)L,L-1]$.

Choosing first η small enough and then L big enough we arrive at (3.4.17), since the term is equal to 0 for $l=L$ and for $l=1$ condition (3.4.5) and equation (3.4.8) yield

$$\frac{1}{KL}\sum_{k=1}^{K}\frac{1}{r\left(\frac{k}{KL}\right)}\left(\frac{1}{\sqrt{L}}\Pi(1)-B(1/L)\right)^2=O_P(1)\cdot\frac{1}{L^2K}\sum_{k=1}^{K}\frac{1}{r\left(\frac{k}{KL}\right)}=o_P(1).$$

Analogously we deduce from Lemma 3.4.1 and assumptions (3.4.1) respectively (3.4.2)

$$\frac{1}{\tau_n^2(\mathbf{a})}\frac{1}{KL}\sum_{\substack{1\leqslant l\leqslant L,1\leqslant k\leqslant K\\(l,k)\neq(L,K)}}\frac{1}{r\left(\frac{K(l-1)+k}{KL}\right)}\left(\frac{1}{\sqrt{KL}}\sum_{j=k+1}^{K}\widetilde{a}_n(R_l,j)\right)^2=o_P(1),$$

where the Markov inequality gives the assertion for $l=L$ because of conditions (3.4.1) (which remains true with κ replaced by 2 and D_1 replaced by $1+D_1$) and (3.4.5).

It holds by the Minkovski inequality

$$\sqrt{\frac{T_{L,K}^{(4)}(r,\mathbf{a})}{\tau_n^2(\mathbf{a})}}\overset{\mathcal{D}}{\equiv}\sqrt{\frac{1}{KL}\sum_{\substack{1\leqslant l\leqslant L,1\leqslant k\leqslant K\\(l,k)\neq(L,K)}}\frac{1}{r\left(\frac{K(l-1)+k}{KL}\right)}B^2\left(\frac{l}{L}\right)}+o_P(1).$$

Lemma 3.4.2 b) and (3.2.6) yield

$$\frac{1}{KL}\sum_{l=\lceil\log^2 L\rceil}^{\lfloor L-\log^2 L\rfloor}\sum_{k=1}^{K}\frac{1}{r\left(\frac{K(l-1)+k}{KL}\right)}\left(B\left(\frac{l}{L}\right)-B\left(\frac{K(l-1)+k}{KL}\right)\right)^2$$
$$=O_P((\log L)^{-1})\cdot\frac{1}{KL}\sum_{l=\lceil\log^2 L\rceil}^{\lfloor L-\log^2 L\rfloor}\sum_{k=1}^{K}\frac{\frac{l}{L}\,\frac{L-l}{L}}{r\left(\frac{K(l-1)+k}{KL}\right)}=o_P(1).$$

Condition (3.2.6) implies for all $\vartheta \geqslant \frac{\log^2 L}{L}$

$$\frac{1}{KL}\sum_{l=2}^{\lceil \log^2 L\rceil -1}\sum_{k=1}^{K}\frac{\frac{K(l-1)+k}{KL}}{r\left(\frac{K(l-1)+k}{KL}\right)} \leqslant \vartheta\frac{1}{\vartheta KL}\sum_{l=2}^{\vartheta L}\sum_{k=1}^{K}\frac{\frac{K(l-1)+k}{KL}}{r\left(\frac{K(l-1)+k}{KL}\right)}.$$

Again

$$\vartheta\frac{1}{\vartheta KL}\sum_{l=2}^{\vartheta L}\sum_{k=1}^{K}\frac{\frac{K(l-1)+k}{KL}}{r\left(\frac{K(l-1)+k}{KL}\right)} - \vartheta\int_0^{\vartheta}\frac{t(1-t)}{r(t)}\,dt = o(1)$$

as $L \to \infty$ for all $\vartheta > 0$ and

$$\vartheta\int_0^{\vartheta}\frac{t(1-t)}{r(t)}\,dt = o(1) \quad \text{as } \vartheta \to 0.$$

This means together with (3.4.5) and (3.4.10)

$$\frac{1}{KL}\sum_{l=1}^{\lceil \log^2 L\rceil -1}\sum_{k=1}^{K}\frac{\frac{l(L-l)}{L^2}}{r\left(\frac{K(l-1)+k}{KL}\right)} = o(1) \quad \text{as } L \to \infty.$$

Thus Lemma 3.4.2 c) gives

$$\begin{aligned}
&\frac{1}{KL}\sum_{l=1}^{\lceil \log^2 L\rceil -1}\sum_{k=1}^{K}\frac{1}{r\left(\frac{K(l-1)+k}{KL}\right)}\left(B\left(\frac{l}{L}\right)-B\left(\frac{K(l-1)+k}{KL}\right)\right)^2\\
&= O_P(1)\cdot\frac{1}{KL}\sum_{l=1}^{\lceil \log^2 L\rceil -1}\sum_{k=1}^{K}\frac{\frac{l(L-l)}{L^2}}{r\left(\frac{K(l-1)+k}{KL}\right)} = o_P(1)
\end{aligned}$$

as $L \to \infty$. We get an analogous result for the sum over $[\lfloor L-\log^2 L\rfloor+1, L-1]$ and for $l = L$ the fact that $\max_k \sqrt{L}B(1-k/(KL)) \stackrel{\mathcal{D}}{=} \max_k(W(k/K)-k/(KL)W(L)) = O_P(1)$ gives

$$\frac{1}{L^2K}\sum_{k=1}^{K-1}\frac{1}{r\left(1-\frac{k}{KL}\right)}\left(\sqrt{L}\,B\left(1-\frac{k}{KL}\right)\right)^2 = o_P(1).$$

Putting everything together we see by the Minkovski inequality

$$\sqrt{\frac{T_{L,K}^{(4)}(r,\mathbf{a})}{\tau_n^2(\mathbf{a})}} \stackrel{\mathcal{D}}{=} \sqrt{\frac{1}{n}\sum_{m=1}^{n-1}\frac{1}{r\left(\frac{m}{n}\right)}B^2\left(\frac{m}{n}\right)} + o_P(1).$$

Theorem 3.3.2 now yields the assertion. ∎

3.5. Block Permutation Statistics and their Limit Distributions

In this section we prove that the block permutation test is indeed valid, i.e. it holds the chosen level asymptotically. Precisely we show that the quantiles from the permutation statistics given our observed data approximate the critical values corresponding to the null distribution 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. for certain alpha-mixing sequences, for details confer Appendix B.2. We have already seen that many linear sequences are alpha-mixing, causal ARMA sequences even with an exponentially decaying mixing coefficient (confer Section 3.3). This last property is nice in view of the assumptions of Theorem 3.5.1, because then the conditions concerning the alpha-mixing coefficients are fulfilled (for any $\delta, \Delta > 0$).

For the permutation result to hold true we standardize using the variance of the block rank statistic. Then we verify that this variance converges to

$$\tau^2 := \sigma^2 \left(\sum_{j \geqslant 0} w_j \right)^2$$

under H_0, if we replace the ranks by our observations. The convergence is even sufficiently fast for Theorem 3.3.2 under suitable conditions. The estimator we obtain that way is ($\widehat{\tau}_{LK} > 0$)

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

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. Indeed, that is one of the major advantages of the permutation test. The simulation study in Chapter 6 suggests that the main problem with the asymptotic test is the performance of the estimator for τ^2.

We begin with a lemma verifying that $\widehat{\tau}^2_{LK}$ converges to τ^2.

Lemma 3.5.1. *Under and (3.3.1) - (3.3.3) and* H_0

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

where $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X(i)$ *and* $\mu < \min(\nu, 4)$.

Remark 3.5.1. It is easy to check that for independent errors the corresponding unbiased estimate is

$$\frac{1}{K(L-1)}\sum_{l=0}^{L-1}\left[\sum_{k=1}^{K}(X(Kl+k)-\bar{X}_n)\right]^2, \tag{3.5.1}$$

which has asymptotically the same behavior – no matter which model $\{X(i)\}$ follows. Under H_0 it also does not change the convergence rates. These two features make it a good candidate as variance estimator as well.

Remark 3.5.2. We only have the correct asymptotic behavior of the above estimator with $K = 1$ for independent observations (Marcinkiewicz–Zygmund). In general, the following asymptotic is valid under H_0 (confer Theorems C.1 and C.3), as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n}(X(i)-\bar{X}_n)^2 \to \sigma^2 \sum_{s\geqslant 0} w_s^2 \quad \text{a.s.}$$

Proof of Lemma 3.5.1. The proof makes use of the Beveridge-Nelson decomposition outlined in Appendix C.

First of all we have (it holds $X(i) = e(i)$ under H_0):

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

For the second term Theorem C.2 yields

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

The BN decomposition (confer Lemma C.1) gives

$$e(j) = \epsilon(j)\left(\sum_{s\geqslant 0} w_s\right) + \widetilde{e}(j-1) - \widetilde{e}(j)$$

and thus for the first term

$$\begin{aligned}
&\frac{1}{KL}\sum_{l=0}^{L-1}\left(\sum_{k=1}^{K}e(Kl+k)\right)^2 \\
&= \left(\sum_{s\geqslant 0} w_s\right)^2 \frac{1}{KL}\sum_{l=0}^{L-1}\left(\sum_{k=1}^{K}\epsilon(Kl+k)\right)^2 + \frac{1}{KL}\sum_{l=0}^{L-1}\left(\widetilde{e}(Kl)-\widetilde{e}(K(l+1))\right)^2 \\
&\quad + \sum_{s\geqslant 0} w_s\,\frac{2}{KL}\sum_{l=0}^{L-1}\Big(\sum_{k=1}^{K}\epsilon(Kl+k)\Big)\left(\widetilde{e}(Kl)-\widetilde{e}(K(l+1))\right) \\
&=: D_1(L,K) + D_2(L,K) + D_3(L,K),
\end{aligned}$$

where $\widetilde{e}(\cdot)$ is another stationary linear process with existing second moment (for details confer Lemma C.1).

Concerning the first term we deduce from the law of Marcinkiewicz (cf. e.g. Loève [61], p. 254, Moments Lemma 4°)

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

where the last line follows from the Markov inequality, since

$$\begin{aligned}
&\mathrm{var}\Big(\frac{1}{KL}\sum_{l=0}^{L-1}\sum_{\substack{k_1\neq k_2\\1}}^{K}\epsilon(Kl+k_1)\,\epsilon(Kl+k_2)\Big)\\
&= \frac{1}{(KL)^2}\sum_{l=0}^{L-1}\sum_{\substack{k_{11}\neq k_{12}\\1}}^{K}\sum_{\substack{k_{21}\neq k_{22}\\1}}^{K}\mathrm{E}(\epsilon(Kl+k_{11})\,\epsilon(Kl+k_{12})\,\epsilon(Kl+k_{21})\,\epsilon(Kl+k_{22}))\\
&\ll \sigma^4\frac{1}{L},
\end{aligned}$$

because

$$\begin{aligned}
&\mathrm{E}(\epsilon(Kl+k_{11})\,\epsilon(Kl+k_{12})\,\epsilon(Kl+k_{21})\,\epsilon(Kl+k_{22}))\\
&= \begin{cases}\sigma^4, & (k_{11}=k_{21}\wedge k_{12}=k_{22})\vee(k_{11}=k_{22}\wedge k_{12}=k_{21}).\\ 0, & \text{else.}\end{cases}
\end{aligned}$$

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 yields

$$D_2(L,K)=O_P\left(\frac{1}{K}\right).$$

Concerning $D_3(L,K)$ we get by the Cauchy-Schwartz inequality

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

Again the Markov inequality gives

$$D_3(L,K)=O_P\left(\sqrt{\frac{1}{K}}\right).$$

Putting everything together we arrive now at the assertion. ■

Now we discuss the asymptotic behavior of the block permutation statistic conditionally on the given observations.

We assume that $L \to \infty$, $K = K(L) \to \infty, n = n(L) = KL$ and $K/L = O(1)$ (confer Remark 3.4.1).

Let $\mathbf{R} = (R_1, \ldots, R_L)$ be a random permutation of $(1, \ldots, L)$ independent of $\{X(\cdot)\}$. We are interested in the permutation statistics

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

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

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

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

where $K(l^*-1)+k^* = K(l-1)+k-G$, i.e. $l^*-1 = \left\lfloor \frac{K(l-1)+k-G}{K} \right\rfloor$, $k^* = (K(l-1)+k-G) \mod K$ and

$$S_{L,K}^{\mathbf{R}}(l,k) := \sum_{i=1}^{l-1} \sum_{j=1}^{K} (X_{K(R_i-1)+j} - \bar{X}_n) + \sum_{j=1}^{k} (X_{K(R_l-1)+j} - \bar{X}_n).$$

Now we prove that these statistics conditioned on the observations have the same asymptotic behavior as the statistics under the null hypothesis (cf. Theorem 3.3.2). It does not matter whether our observations follow the null hypothesis or an alternative. This is true under either of the following assumptions on the error sequence $\{e(i) : i \geqslant 1\}$ for certain δ, Δ. Theorem 3.3.1 gives conditions under which linear sequences are strong mixing and even provides the mixing coefficients. The alpha-mixing coefficients of causal ARMA sequences decay exponentially (confer Remark 3.3.1), thus the below conditions on the mixing coefficients are fulfilled for any δ, Δ.

$\mathcal{A}.1$ Let $\{Z_i : i \in \mathbb{Z}\}$ be a random sequence with $\mathrm{E}\, Z_i = 0$, $i \in \mathbb{Z}$. Assume there is a $\delta, \Delta > 0$, $2l \leqslant \delta \leqslant 2(l+1)$, $l = 0, 1, 2 \ldots$, with

$$\mathrm{E}\,|Z_i|^{2+\delta+\Delta} \leqslant D_1 \quad \text{for all } i \in \mathbb{Z}$$

and

$$\sum_{k=0}^{\infty} (k+1)^{2l+2} \alpha_Z(k)^{\Delta/(2l+4+\Delta)} \leqslant D_2(\delta, \Delta), \tag{3.5.2}$$

where α_Z is the corresponding α-mixing coefficient.

$\mathcal{A}.2$ Let $\{Z_i : i \in \mathbb{Z}\}$ be a strictly stationary sequence with $\mathrm{E}\, Z_i = 0$, $i \in \mathbb{Z}$. Assume there are $\delta, \Delta > 0$ with

$$\mathrm{E}\,|Z_i|^{2+\delta+\Delta} \leqslant D_1 \quad \text{for all } i \in \mathbb{Z}$$

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

$$\sum_{k=0}^{\infty} (k+1)^{\delta/2} \alpha(k)^{\Delta/(2+\delta+\Delta)} \leqslant D_2, \tag{3.5.3}$$

where α_Z is the corresponding α-mixing coefficient.

Now we state the main theorem showing that the block permutation method is valid.

Theorem 3.5.1. *Assume that $\{X(i) : 1 \leqslant i \leqslant n\}$ fulfills (3.3.1) - (3.3.3) with $\nu > 4$. Let $0 < \tilde{\delta} < (\nu - 4)/2$ and let the sequence $\{e(i) : i \geqslant 1\}$ fulfill assumptions $\mathcal{A}.1$ or $\mathcal{A}.2$ 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* $\frac{d^2 K}{L} = o(1)$ *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 3.3.2 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 $L \to \infty$.

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

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

b) If $G = G(n) \to \infty$, $G/n \to 0$, and (3.4.3), then we have for all $x \in \mathbb{R}$ as $L \to \infty$

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

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

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

then we have for all $x \in \mathbb{R}$ as $L \to \infty$

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

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

d) If r fulfills condition (3.2.6) and

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

then we have for all $x\in\mathbb{R}$ as $L\to\infty$

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

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

Remark 3.5.3. It is worth mentioning that the proof does not exploit the fact that the underlying error sequence forms a linear process. In fact we only need that the variance of the sums is asymptotically positive (confer Lemma 3.5.2), the existence of a νth moment ($\nu>4$) and conditions $\mathcal{A}.1$ respectively $\mathcal{A}.2$ on the mixing coefficients for appropriate $\delta,\Delta>0$. We can then use the results of Appendix B.2.

Remark 3.5.4. 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 (3.5.2) respectively (3.5.3) for $\tilde{\delta}<\delta^{(1)}<\nu-2$, $\Delta^{(1)}=\nu-2-\delta^{(1)}$ and $\delta^{(2)}>2+2\tilde{\delta}$, $\Delta^{(2)}>0$. The proofs are analogous; we just replace the random variables by suitably truncated versions to get equations (3.5.7) respectively (3.5.8) in a P-stochastic sense. As an example we will now show the analogue of (3.5.8). Let

$$\tilde{Y}_L(l):=\max_{k=0,\dots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K}e(Kl+k)\right|^{2+\tilde{\delta}}.$$

Note that by Theorem B.8 a) it holds $\mathrm{E}\,|\tilde{Y}_L(l)|\leqslant D$ for all l,L and also $\mathrm{E}\,|\tilde{Y}_L(l)|^{1+\delta}<\tilde{D}$ for $\delta=\frac{\delta^{(1)}-\tilde{\delta}}{2+\tilde{\delta}}$ and all l,L. Let $Y_L(l):=\tilde{Y}_L(l)-\mathrm{E}\,\tilde{Y}_L(l)$. Note that $\{Y_L(l)\}$ is uniformly integrable. We show that $\frac{1}{L}\sum_{l=1}^{L}|Y_L(l)|=o_P(1)$. Indeed the Markov inequality and the uniform integrability give ($\delta=\frac{2+\delta^{(2)}}{2+\tilde{\delta}}-2$)

$$\begin{aligned}
&P\left(\frac{1}{L}\sum_{l=0}^{L-1}|Y_L(l)|\geqslant 2\epsilon\right)\\
&\leqslant P\left(\frac{1}{L}\sum_{l=0}^{L-1}|Y_L(l)1_{\{|Y_L(l)|>\log L\}}|\geqslant\epsilon\right)+P\left(\frac{1}{L}\sum_{l=0}^{L-1}|Y_L(l)1_{\{|Y_L(l)|\leqslant\log L\}}|\geqslant\epsilon\right)\\
&\ll_\epsilon \max_{l=0,\dots,L-1}\mathrm{E}\,|Y_L(l)1_{\{|Y_L(l)|>\log L\}}|+\mathrm{E}\left|\frac{1}{L}\sum_{l=0}^{L-1}|Y_L(l)1_{\{|Y_L(l)|\leqslant\log L\}}|\right|^{2+\delta}\\
&\ll_\epsilon o(1)+\frac{(\log L)^{2+\delta}}{L^{(2+\delta)/2}}=o(1)
\end{aligned}$$

as $L\to\infty$. The last line follows by Theorems B.5 respectively B.6, since for all $r>0$ (in particular for $r=2+\delta^{(2)}+\Delta^{(2)}$)

$$\mathrm{E}\left(\frac{|Y_L(l)1_{\{|Y_L(l)|\leqslant\log L\}}|}{\log L}\right)^r\leqslant 1.$$

Remark 3.5.5. a) Analogously to (4.5) respectively (4.6) in Lemma 4.3 in Antoch et al. [4] one can show that under the null hypothesis it is possible to just consider the maximum over $C(\log n)^\gamma \leqslant m \leqslant n - C(\log n)^\gamma$, $C, \gamma > 0$ in the statistic $T_n^{(1)}$. This means that the corresponding permutation statistic has, conditionally on the observations, exactly the same limit distribution as in the null asymptotic of the original statistic. This shows that the critical values of the permutation distribution are good approximations of the critical values of the distribution under the null hypothesis no matter whether the observed sequence follows the null hypothesis or some alternative.

b) Here, however, it is also possible to use $\widetilde{T}_n^{(1)}$, where we take the maximum over the whole range $1 \leqslant K(l-1)+k < n$. The reason is that equation (3.4.6) in Remark 3.4.2 is fulfilled *a.s.* for logarithmic K.

Proof of Remark 3.5.5b). Analogously to equations (3.5.9) respectively (3.5.10) it suffices to show that as $L \to \infty$

$$\frac{1}{(\log\log n)^{(2+\tilde{\delta})/2}} \frac{1}{L} \sum_{l=0}^{L-1} \max_{k=1,\dots,K} \left| \frac{1}{\sqrt{k}} \sum_{j=1}^{k} e(Kl+j) \right|^{2+\tilde{\delta}} \to 0 \qquad a.s.$$

and
$$\frac{1}{(\log\log n)^{(2+\tilde{\delta})/2}} \frac{1}{L} \sum_{l=0}^{L-1} \max_{k=1,\dots,K} \left| \frac{1}{\sqrt{k}} \sum_{j=K-k+1}^{K} e(Kl+j) \right|^{2+\tilde{\delta}} \to 0 \qquad a.s.$$

Similarly to equation (3.5.8), the proof of Corollary B.1 shows that the first equation is fulfilled, if uniformly in l

$$\frac{1}{(\log\log n)^{(2+\tilde{\delta})(2+\delta^{(2)})/2}} \operatorname{E} \max_{k=1,\dots,K} \left| \frac{1}{\sqrt{k}} \sum_{j=1}^{k} e(Kl+j) \right|^{(2+\tilde{\delta})(2+\delta^{(2)})} = O(1)$$

and
$$\frac{1}{(\log\log n)^{(2+\tilde{\delta})/2}} \operatorname{E} \max_{k=1,\dots,K} \left| \frac{1}{\sqrt{k}} \sum_{j=1}^{k} e(Kl+j) \right|^{2+\tilde{\delta}} = o(1).$$

However Theorems B.3 and B.5 respectively B.6 give under the assumptions of Theorem 3.5.1 for $2 < \eta \leqslant 2 + \delta^{(1)}$

$$\operatorname{E} \max_{k=1,\dots,K} \left| \frac{1}{\sqrt{k}} \sum_{j=1}^{k} e(Kl+j) \right|^{\eta} \ll \sum_{k=1}^{K} \frac{1}{(\sqrt{k})^\eta} k^{\eta/2-1} = \sum_{k=1}^{K} \frac{1}{k} \ll \log K.$$

This gives the assertion if it holds $\frac{\log K}{(\log\log n)^{(2+\tilde{\delta})/2}} = o(1)$. Since $(2+\tilde{\delta})/2 > 1$ this is always fulfilled for logarithmic K. The proof for the second expression is analogous. ∎

Remark 3.5.6. Note that we do not need $K \to \infty$ for this theorem, however, in the case of dependent errors $\widehat{\tau}_{LK} \xrightarrow{P} \tau$ only if $K \to \infty$. This means, we need $K \to \infty$ to be able to use $\widehat{\tau}_{LK}$ as an estimator in the original statistics (confer Lemma 3.5.1).
For independent errors, however, the permutation statistic for $K = 1$ has the correct asymptotic behavior and, as the simulation study in Section 6.2 suggests, $K = 1$ is the best choice of the block length.

In order to prove the theorem we first need the following lemma.

Lemma 3.5.2. *Under conditions (3.3.2) (the existence of the second moment suffices) and (3.3.3) we have as $n \to \infty$*

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

Proof. It holds

$$\operatorname{var}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e(i)\right) = \frac{1}{n}\sum_{l=1}^{n}\sum_{k=1}^{n}\sum_{i\geqslant 0} w_i \sum_{j\geqslant 0} w_j \operatorname{cov}(\epsilon(l-j), \epsilon(k-i))$$
$$= \frac{1}{n}\sum_{l=1}^{n}\sum_{k=1}^{n}\sum_{i\geqslant 0} w_i \sum_{j\geqslant 0} w_j\, \sigma^2\, 1_{\{l-j=k-i\}} \to \sigma^2 \sum_{i\geqslant 0} w_i \sum_{j\geqslant 0} w_j,$$

since

$$\left|\frac{1}{n}\sum_{i\geqslant 0}\sum_{j\geqslant 0} w_i\, w_j \sum_{l=1}^{n}\left(1 - \sum_{k=1}^{n} 1_{\{l-j=k-i\}}\right)\right|$$
$$= \left|\frac{1}{n}\sum_{i\geqslant 0}\sum_{j\geqslant 0} w_i\, w_j \left(\sum_{l=1}^{\min(j-i,n)} 1 + \sum_{l=\max(n-i+j+1,1)}^{n} 1\right)\right|$$
$$= \frac{1}{n}\left|\sum_{i\geqslant 0}\sum_{j\geqslant i+1} w_i\, w_j\, \min(j-i,n) + \sum_{i\geqslant 0}\sum_{j=0}^{i-1} w_i\, w_j\, \min(n, i-j)\right|$$
$$\leqslant \frac{2}{\sqrt{n}}\sum_{i\geqslant 0} |w_i| \sum_{j\geqslant 0} \sqrt{j}|w_j| \to 0,$$

as $n \to \infty$. ■

We are now ready to prove the main theorem.

Proof of Theorem 3.5.1. In the following we will repeatedly use D as a constant. It may be different in every inequality.

The Minkovski inequality and the monotone convergence theorem give for all $i \geqslant 0$

$$\mathrm{E}\,|e(i)|^\nu \leqslant D\left(\sum_{j\geqslant 0} |w_j|\right)^\nu,$$

for some constant D.

First of all Theorem B.8 b) gives as $L \to \infty$

$$\frac{1}{n}\sum_{j=1}^{n} e(j) = O\left(\sqrt{\frac{\log n}{n}}\right) \quad a.s. \tag{3.5.4}$$

Since $\mathrm{E}\left(\max_{k=1,\dots,K}\left|\sum_{j=1}^{k} e(K\hat{l}_n+j)\right|^{2+\delta^{(1)}}\right) = \mathrm{E}\left(\max_{k=1,\dots,K}\left|\sum_{j=1}^{k} e(j)\right|^{2+\delta^{(1)}}\right)$ for all $\hat{l}_n$, Theorem B.8 and Remark B.2 give

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

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

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

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

$$\frac{1}{\sqrt{L}}\left|\sum_{k=1}^{K} e(K\hat{l}_n+k)\right| = O\left(\sqrt{\log K}\right) \quad a.s. \tag{3.5.5}$$

for $K \to \infty$ as well as K bounded.

Similarly we deduce for $l^* := \lceil m/K \rceil$ as $L \to \infty$ $(m \neq n)$

$$\frac{1}{\sqrt{L(n-m)}}\left|\sum_{j=Kl^*+1}^{n} e(j)\right| = o(1) \quad a.s. \tag{3.5.6}$$

Since the α-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, assumptions (3.5.2) respectively (3.5.3) are uniformly fulfilled in K. Also this sequence remains stationary for stationary $\{e(\cdot)\}$.

Consequently Theorem B.8 a) shows that there is a $D > 0$ such that

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

uniformly in $l \geqslant 0$ and K.

Note that $2(2+\delta^{(2)}+\Delta^{(2)}) < (2+\tilde{\delta})(2+\delta^{(2)}+\Delta^{(2)}) = 2+\delta^{(1)}$. The conditions of Corollary B.1 are fulfilled and we conclude as $L \to \infty$

$$\begin{aligned}
&\frac{1}{L}\sum_{l=0}^{L-1}\Big(\frac{1}{\sqrt{K}}\sum_{k=1}^{K} e(Kl+k)\Big)^2 \\
&= \frac{1}{L}\sum_{l=0}^{L-1}\left[\Big(\frac{1}{\sqrt{K}}\sum_{k=1}^{K} e(Kl+k)\Big)^2 - \operatorname{var}\Big(\frac{1}{\sqrt{K}}\sum_{k=1}^{K} e(k)\Big)\right] \\
&\qquad + \operatorname{var}\Big(\frac{1}{\sqrt{K}}\sum_{k=1}^{K} e(k)\Big) \;\to\; C > 0 \quad a.s.,
\end{aligned} \tag{3.5.7}$$

where we either use Lemma 3.5.2 (if $K \to \infty$) or the fact that $\operatorname{var}\left(\frac{1}{\sqrt{K}}\sum_{k=1}^{K} e(k)\right) \geqslant c > 0$, if K is bounded.

Moreover we get

$$\frac{1}{L}\sum_{l=0}^{L-1}\left[\max_{k=0,\ldots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K} e(Kl+j)\right|^{2+\tilde{\delta}} - \mu_K(2+\tilde{\delta})\right] + \mu_K(2+\tilde{\delta}) \tag{3.5.8}$$
$$\leqslant D \quad a.s.$$

Now we are ready to use Theorem 3.4.1 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 $\mu = 0$.

Note that

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

Moreover we have $\bar{X}_n = d\,\frac{n-m}{n} + \bar{e}_n$ so that equation (3.5.4) yields

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

Furthermore equation (3.5.4) gives as $L \to \infty$

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

since $\frac{d^2K}{L} = o(1)$ 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 (3.5.5) and (3.5.6) now imply for $l^* := \lceil m/K \rceil$

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

Putting everything together equation (3.5.7) gives as $L \to \infty$

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

Now note that $X(i) = (\mu+d) - d1_{\{i\leqslant m\}} + e(i)$. The same calculation shows as $L \to \infty$

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

which yields

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

Second of all it holds using equations (3.5.4) and (3.5.8)

$$\begin{aligned}
&\frac{1}{L}\sum_{l=0}^{L-1}\max_{k=0,\dots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K}\left(X(Kl+j)-\bar{X}_n\right)\right|^{2+\tilde{\delta}} \\
&\ll \frac{1}{L}\sum_{l=0}^{L-1}\max_{k=0,\dots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K}e(Kl+j)\right|^{2+\tilde{\delta}}+|\sqrt{K}\bar{e}_n|^{2+\tilde{\delta}} \qquad (3.5.9)\\
&\quad +|\sqrt{K}d|^{2+\tilde{\delta}}((n-m)/n)^{2+\tilde{\delta}}+|\sqrt{K}d|^{2+\tilde{\delta}}(n-m)/n \\
&\ll 1+|\sqrt{K}d|^{2+\tilde{\delta}}(n-m)/n \quad a.s.
\end{aligned}$$

As above we get using the fact that $X(i)=(\mu+d)-d1_{\{i\leqslant m\}}+e(i)$

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

which gives

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

For the proof of (ii) we distinguish the two main cases $Kd_n^2=O(1)$ and $\frac{1}{Kd_n^2}=O(1)$. The first one is included in (i), so let us assume now that $\frac{1}{d^2K}=O(1)$. Then choose the scores $a_n(i):=X(i)/\sqrt{d^2K}$. Without loss of generality assume $\mu=0$.

Similarly as above we have

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

Moreover $\bar{X}_n=d\,\frac{n-m}{n}+\bar{e}_n$ so that equation (3.5.4) gives

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

Furthermore equation (3.5.4) yields as $L\to\infty$

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

since

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

and with $l^* := \lceil \frac{m}{K} \rceil$ we obtain by equations (3.5.5) respectively (3.5.6)

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

Altogether we have as $L \to \infty$

$$\begin{aligned}
&\frac{1}{L}\sum_{l=0}^{L-1}\left(\frac{1}{|d|K}\sum_{k=1}^{K}\left(X(Kl+k) - \bar{X}_n\right)\right)^2 \\
&\geqslant (n-m)/n - ((n-m)/n)^2 + o(1) \geqslant \frac{\epsilon}{2} + o(1) \quad a.s.
\end{aligned}$$

Finally it holds using equations (3.5.4) and (3.5.8)

$$\begin{aligned}
&\frac{1}{L}\sum_{l=0}^{L-1}\max_{k=0,\dots,K-1}\left|\frac{1}{|d|K}\sum_{j=k+1}^{K}\left(X(Kl+j) - \bar{X}_n\right)\right|^{2+\tilde{\delta}} \\
&\ll \left(\frac{1}{|d|\sqrt{K}}\right)^{2+\tilde{\delta}}\frac{1}{L}\sum_{l=0}^{L-1}\max_{k=0,\dots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K} e(Kl+j)\right|^{2+\tilde{\delta}} \\
&\quad + \left(\frac{1}{|d|\sqrt{K}}\right)^{2+\tilde{\delta}}|\sqrt{K}\,\bar{e}_n|^{2+\tilde{\delta}} + ((n-m)/n)^{2+\tilde{\delta}} + \frac{n-m}{n} \\
&\ll 1 \quad a.s.
\end{aligned} \tag{3.5.10}$$

In the case, where $d = d_n$ is such that it follows neither of the above possibilities, we have infinitely many n with $K_n d_n^2 \leqslant 1$ and also infinitely many with $K_n d_n^2 > 1$. Then just choose the scores

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

As above the assumptions of Theorem 3.4.1 are fulfilled for both subsequences, hence also for the complete sequence.

The assertion now follows from Theorem 3.4.1. ■

3.6. Block Bootstrap with Replacement

In the previous sections we have proven that the block permutation test works in this setting. In this section we study the classical block-bootstrap (with replacement) and show that it also yields asymptotically correct critical values. We only sketch the proofs, because they are very close to the ones for the permutation tests.

A block length of K 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 [72] 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. So we glue the last and the first observation together and use all n blocks for the bootstrap, i.e. we have the following blocks $\{(X_{l+1},\ldots,X_{l+K}),\ l=0,\ldots,n-1\}$, $X_i = X_{i-n}$, $i>n$. We concentrate on this method, yet it is also possible to prove the validity of the first approach using the same methods.

Again we give first the asymptotic results for the corresponding score-processes, which is the replacement analogue to rank statistics. Then we use these results to obtain the asymptotics for the block bootstrap statistics.

3.6.1. Asymptotics of the Corresponding Score-Processes

Analogously to the rank asymptotics we derive now a corresponding result for the bootstrap with replacement. Here, it should also be possible to derive the asymptotics directly using results of Einmahl [26], since we have statistics based on sums of independent random variables, which, however, form a triangular array.

Again $L\to\infty$, $K=K(L)$, $n=n(L)=KL$, confer also Remark 3.4.1.

Now we are ready to study the asymptotics for the following corresponding score-processes, which are also based on partial sums

$$\widetilde{S}^{\mathbf{a}}_{L,K}(l,k) := \sum_{i=1}^{l-1}\sum_{j=1}^{K}(a_n[U(i)+j]-\bar{a}_{\mathbf{U},n}) + \sum_{j=1}^{k}(a_n[U(l)+j]-\bar{a}_{\mathbf{U},n}),$$

where $\mathbf{U}=(U(1),\ldots,U(L))$ is a vector of i.i.d. uniformly distributed r.v.´s on $(0,\ldots,n-1)$. Precisely we are interested in

$$\widetilde{T}^{(1)}_{L,K}(\mathbf{a}) := \max_{2\leqslant l\leqslant L-1}\ \max_{1\leqslant k\leqslant K}\sqrt{\frac{LK}{(K(l-1)+k)(LK-K(l-1)-k)}}\left|\widetilde{S}^{\mathbf{a}}_{L,K}(l,k)\right|,$$

$$\widetilde{T}^{(2)}_{L,K}(G,\mathbf{a}) := \frac{1}{\sqrt{G}}\max_{\substack{1\leqslant l\leqslant L,1\leqslant k\leqslant K\\ K(l-1)+k>G}}\left|\widetilde{S}^{\mathbf{a}}_{L,K}(l,k)-\widetilde{S}^{\mathbf{a}}_{L,K}(l^*,k^*)\right|,$$

$$\widetilde{T}^{(3)}_{L,K}(q,\mathbf{a}) := \max_{\substack{1\leqslant l\leqslant L,1\leqslant k\leqslant K\\ (l,k)\neq(L,K)}}\frac{1}{\sqrt{KL}\,q\left(\frac{K(l-1)+k}{KL}\right)}\left|\widetilde{S}^{\mathbf{a}}_{L,K}(l,k)\right|,$$

$$\widetilde{T}^{(4)}_{L,K}(r,\mathbf{a}) := \frac{1}{(KL)^2}\sum_{\substack{1\leqslant l\leqslant L,1\leqslant k\leqslant K\\ (l,k)\neq(L,K)}}\frac{1}{r\left(\frac{K(l-1)+k}{KL}\right)}\left(\widetilde{S}^{\mathbf{a}}_{L,K}(l,k)\right)^2,$$

where $K(l^*-1)+k^* = K(l-1)+k-G$, i.e. $l^*-1=\left\lfloor\frac{K(l-1)+k-G}{K}\right\rfloor$, $k^*=(K(l-1)+k-G)\mod K$.

Theorem 3.6.1. *Let* $\mathbf{U} = (U(1), \ldots, U(L))$ *be a vector of i.i.d. uniformly distributed random variables on* $\{0, \ldots, n-1\}$*. Moreover let* $a_n(1), \ldots, a_n(n)$ *be scores satisfying*

$$\frac{1}{n}\sum_{l=0}^{n-1} \max_{k=0,\ldots,K-1} \left| \frac{1}{\sqrt{K}} \sum_{j=k+1}^{K} (a_n(l+j) - \bar{a}_n) \right|^{\kappa} \leqslant D_1 \tag{3.6.1}$$

for some $2 < \kappa \leqslant 4$ *and*

$$\tau_n^2(\mathbf{a}) := \frac{1}{n}\sum_{l=0}^{n-1} \left[\frac{1}{\sqrt{K}} \sum_{k=1}^{K} (a_n(l+k) - \bar{a}_n) \right]^2 \geqslant D_2, \tag{3.6.2}$$

where $\bar{a}_n := \frac{1}{n}\sum_{i=1}^{n} a_n(i)$ *and* $D_1, D_2 > 0$ *are some constants. Let* $\alpha(x)$*,* $\beta(x)$ *be as in Theorem 3.3.2.*

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

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

b) If, as $L \to \infty$*,* $G = G(n) \to \infty$*,* $G/n \to 0$*, and (3.4.3), then we have for all* $x \in \mathbb{R}$

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

c) If $q \in Q_{0,1}$ *and* $I^*(q, c) < \infty$ *for some* $c > 0$ *and (3.4.4), then*

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

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

d) If r fulfills condition (3.2.6) and (3.4.5), then

$$\frac{\widetilde{T}_{L,K}^{(4)}(r, \mathbf{a})}{\tau_n^2(\mathbf{a})} \xrightarrow{\mathcal{D}} \int_0^1 \frac{B^2(t)}{r(t)}\, dt \quad \text{as } L \to \infty,$$

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

Proof. Instead of Corollary D.1 we use Theorem D.1 with

$$Z_L(l) = \frac{1}{\sqrt{\tau_n^2(\mathbf{a})K}} \sum_{k=1}^{K} (a_n(U(l)+k) - \bar{a}_{\mathbf{U},n}).$$

Analogous arguments as in the proof of Corollary D.2 show that under (3.6.1) and (3.6.2) the assumptions of Theorem D.1 are fulfilled. The proof is then completely analogous to the one of Theorem 3.4.1, if one replaces Lemma 3.4.1 with Lemma 3.6.1 below, and therefore omitted. ■

Remark 3.6.1. Concerning the weighted CUSUM-statistic $\tilde{\tilde{T}}^{(1)}_{L,K}(\mathbf{a})$ 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

$$\begin{aligned} &\frac{1}{(\log\log n)^{\mu/2}}\frac{1}{n}\sum_{l=0}^{n-1}\max_{k=1,\dots,K}\left|\frac{1}{\sqrt{k}}\sum_{j=1}^{k}(a_n(l+j)-\bar{a}_n)\right|^{\mu} \to 0 \\ \text{and}\quad &\frac{1}{(\log\log n)^{\mu/2}}\frac{1}{n}\sum_{l=0}^{n-1}\max_{k=1,\dots,K}\left|\frac{1}{\sqrt{k}}\sum_{j=K-k+1}^{K}(a_n(l+j)-\bar{a}_n)\right|^{\mu} \to 0 \end{aligned} \tag{3.6.3}$$

for some $\mu > 0$. This is analogous to Remark 3.4.2, however one has to take into account, that as in equation (3.6.5) we have $\max_{1\leqslant k\leqslant K}\sqrt{k}(\bar{a}_{\mathbf{U},n}-\bar{a}_n) = \sqrt{K}(\bar{a}_{\mathbf{U},n}-\bar{a}_n) = o_P(\sqrt{\log\log n})$.

We prove now that Lemma 3.4.1 remains true in our situation:

Lemma 3.6.1. *Let $\mathbf{U} = (U(1),\dots,U(L))$ be a vector of i.i.d. uniformly distributed r.v.´s on $(0,\dots,n-1)$. Moreover let $a_n(1),\dots,a_n(n)$ be scores satisfying*

$$\frac{1}{n}\sum_{l=0}^{n-1}\max_{k=0,\dots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K}(a_n(l+j)-\bar{a}_n)\right|^{\kappa} \leqslant D \quad \textit{for some constant } D, \tag{3.6.4}$$

where $\kappa > 2$, $a_n(j) = a_n(j-n), j > n$, and $\bar{a}_n := \frac{1}{n}\sum_{i=1}^{n}a_n(i)$. Then we have for all $\mu < \min\left(\frac{\kappa-2}{2\kappa},\frac{1}{4}\right)$

$$\max_{\substack{1\leqslant l\leqslant L-1\\1\leqslant k\leqslant K}}\left(\frac{l(L-l)}{L}\right)^{\mu}\frac{L}{\sqrt{l(L-l)}}\left|\frac{1}{\sqrt{LK}}\sum_{j=k+1}^{K}(a_n[U(l)+j]-\bar{a}_{\mathbf{U},n})\right| = O_P(1),$$

where $\bar{a}_{\mathbf{U},n} = \frac{1}{n}\sum_{l=1}^{L}\sum_{k=1}^{K}a_n(U(l)+k)$.

Proof. Note that

$$\begin{aligned} &\max_{\substack{1\leqslant l\leqslant L-1\\1\leqslant k\leqslant K}}\left(\frac{l(L-l)}{L}\right)^{\mu}\frac{L}{\sqrt{l(L-l)}}\left|\frac{1}{\sqrt{LK}}\sum_{j=k+1}^{K}(a_n[U(l)+j]-\bar{a}_{\mathbf{U},n})\right| \\ &\leqslant \max_{\substack{1\leqslant l\leqslant L-1\\1\leqslant k\leqslant K}}\left(\frac{l(L-l)}{L}\right)^{\mu}\frac{L}{\sqrt{l(L-l)}}\left|\frac{1}{\sqrt{LK}}\sum_{j=k+1}^{K}(a_n[U(l)+j]-\bar{a}_n)\right| \\ &\quad + \max_{1\leqslant l\leqslant L-1}\left(\frac{L}{l(L-l)}\right)^{\frac{1}{2}-\mu}\sqrt{K}|\bar{a}_n-\bar{a}_{\mathbf{U},n}| \end{aligned}$$

The proof for the first part is analogous to the proof of Lemma 3.4.1 and is therefore omitted. Concerning the second part, note that $\max_{1\leqslant l\leqslant L-1}\left(\frac{L}{l(L-l)}\right)^{\frac{1}{2}-\mu} \leqslant 2$ and

$$\begin{aligned} \mathrm{E}\left|\sqrt{K}(\bar{a}_n-\bar{a}_{\mathbf{U},n})\right| &\leqslant \frac{1}{L}\sum_{l=1}^{L}\mathrm{E}\left|\frac{1}{\sqrt{K}}\sum_{k=1}^{K}(a_n(U(l)+k)-\bar{a}_n)\right| \\ &= \frac{1}{n}\sum_{j=0}^{n-1}\left|\frac{1}{\sqrt{K}}\sum_{k=1}^{K}(a_n(j+k)-\bar{a}_n)\right| \leqslant 1+D, \end{aligned} \tag{3.6.5}$$

because of assumption (3.6.4). The Markov inequality now gives the assertion. ■

3.6.2. Limit Distributions of the Block Bootstrap Statistics

We prove now that the block bootstrap also gives asymptotically correct critical values. Main tools are again the results developed in the previous subsection for score processes together with the laws of large numbers of Appendix B.2. We need some assumptions on the alpha-mixing coefficients of the linear process, which can be checked by Theorem 3.3.1. Example 3.3.1 states that causal ARMA sequences with appropriate innovations always fulfill the assumptions on the mixing coefficients.
As with the permutation method we obtain an estimator for τ^2 along the way. Since all proofs of this section are very close to the corresponding ones for permutation tests we only sketch them.

First we study the following estimator for τ^2 $(\tau, \widehat{\tau}_n(\mathbf{X}) > 0)$:

$$\widehat{\tau}_n^2(\mathbf{X}) := \frac{1}{n}\sum_{l=0}^{n-1}\left[\frac{1}{\sqrt{K}}\sum_{k=1}^{K}(X(l+k)-\bar{X}_n)\right]^2.$$

Again the block bootstrap test is independent of the actual value of that estimator for the given observations. Thus the performance of the test does not depend on the performance of an estimator, which is in fact the biggest problem of the asymptotic test (confer Chapter 6).

Lemma 3.6.2. *Under (3.3.1) - (3.3.3) and H_0*

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

where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X(i)$ and $\mu <\in (\nu, 4)$.

Proof. The proof is analogous to that of Lemma 3.5.1. Since the blocks are not mutually independent anymore, we need a different argument for the mixed term of the analogue to $D_1(L,K)$.

$$\begin{aligned}
&\operatorname{var}\Big(\frac{1}{Kn}\sum_{l=0}^{n-1}\sum_{\substack{k_1\neq k_2\\1}}^{K}\epsilon(l+k_1)\epsilon(l+k_2)\Big)\\
&= \frac{1}{(Kn)^2}\sum_{l_1=0}^{n-1}\sum_{l_2=0}^{n-1}\sum_{\substack{k_{11}\neq k_{12}\\1}}^{K}\sum_{\substack{k_{21}\neq k_{22}\\1}}^{K}\mathrm{E}(\epsilon(l_1+k_{11})\,\epsilon(l_1+k_{12})\,\epsilon(l_2+k_{21})\,\epsilon(l_2+k_{22}))\\
&\ll \sigma^4\frac{1}{L},
\end{aligned}$$

because

$$\begin{aligned}
&\mathrm{E}(\epsilon(l_1+k_{11})\,\epsilon(l_1+k_{12})\,\epsilon(l_2+k_{21})\,\epsilon(l_2+k_{22}))\\
&= \begin{cases}\sigma^4, & (l_1+k_{11}=l_2+k_{21}\wedge l_1+k_{12}=l_2+k_{22})\\ & \vee(l_1+k_{11}=l_2+k_{22}\wedge l_1+k_{12}=l_2+k_{21}).\\ 0, & \text{else.}\end{cases}
\end{aligned}$$

The rest of the proof is analogous and therefore omitted. ■

We discuss now the following bootstrap statistics

$$T^{(1)}_{L,K}(\mathbf{U}) := \max_{2\leqslant l\leqslant L-1}\max_{1\leqslant k\leqslant K}\sqrt{\frac{LK}{(K(l-1)+k)(LK-K(l-1)-k)}}\left|\widetilde{S}^{\mathbf{U}}_{L,K}(l,k)\right|,$$

$$T^{(2)}_{L,K}(G,\mathbf{U}) := \frac{1}{\sqrt{G}}\max_{\substack{1\leqslant l\leqslant L, 1\leqslant k\leqslant K\\ K(l-1)+k>G}}\left|\widetilde{S}^{\mathbf{U}}_{L,K}(l,k)-\widetilde{S}^{\mathbf{U}}_{L,K}(l^*,k^*)\right|,$$

$$T^{(3)}_{L,K}(q,\mathbf{U}) := \max_{\substack{1\leqslant l\leqslant L, 1\leqslant k\leqslant K\\ (l,k)\neq(L,K)}}\frac{1}{\sqrt{LK}\,q\left(\frac{K(l-1)+k}{KL}\right)}\left|\widetilde{S}^{\mathbf{U}}_{L,K}(l,k)\right|,$$

$$T^{(4)}_{L,K}(r,\mathbf{U}) := \frac{1}{(KL)^2}\sum_{\substack{1\leqslant l\leqslant L, 1\leqslant k\leqslant K\\ (l,k)\neq(L,K)}}\frac{1}{r\left(\frac{K(l-1)+k}{KL}\right)}\left(\widetilde{S}^{\mathbf{U}}_{L,K}(l,k)\right)^2,$$

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

$$\widetilde{S}^{\mathbf{U}}_{L,K}(l,k) := \sum_{i=1}^{l-1}\sum_{j=1}^{K}(X_{U(i)+j}-\bar{X}_{\mathbf{U},n}) + \sum_{j=1}^{k}(X_{U(l)+j}-\bar{X}_{\mathbf{U},n}).$$

$\mathbf{U} = (U_1,\ldots,U_L)$ is a vector of i.i.d. random variables uniformly distributed on $\{0,\ldots,n-1\}$ independent of $X_1,\ldots,X_n$.

We prove now, that – under certain assumptions on $\{X(i) : i \in \mathbb{N}\}$, – these statistics have, conditioned on the observations, exactly the same limit behavior as the original statistic under the null. It does not matter whether our observations do follow the null hypothesis or an alternative.

Theorem 3.6.2. *Assume that the assumptions of Theorem 3.5.1 hold and*

$$K \leqslant L^{\frac{\delta(2)}{2}-\epsilon} \qquad \textit{for some } \epsilon > 0. \tag{3.6.6}$$

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

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

b) If $G = G(n) \to \infty$, $G/n \to 0$, and (3.4.3), then we have for all $x \in \mathbb{R}$ as $L \to \infty$

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

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

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

then we have for all $x \in \mathbb{R}$ as $L \to \infty$

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

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

d) If r fulfills condition (3.2.6) and

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

then we have for all $x \in \mathbb{R}$ as $L \to \infty$

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

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

If assumption (3.6.6) is not fulfilled, the above asymptotics remain true, but only in the sense of P-stochastic convergence (instead of a.s.-convergence).

Remark 3.6.2. If $2 < \nu \leqslant 4$ we get as in Remark 3.5.4 the above assertion in a P-stochastic sense. Note that

$$\tilde{Y}_L(l) := \max_{k=0,\ldots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K} e(l+k)\right|^{2+\tilde{\delta}}$$

still has a uniformly existing moment larger than the first one, hence is also uniformly integrable. As in Remark 3.5.4 the uniform integrability and an argument as in (3.6.9) gives

$$P\left(\frac{1}{n}\sum_{l=0}^{n-1}|Y_L(l)| \geqslant 2\epsilon\right) \ll_\epsilon o(1) + \frac{(\log L)^{2+\delta}}{L^{\delta/2}}\frac{K}{L} = o(1)$$

as $L \to \infty$.

Remark 3.6.3. It is also possible to use the bootstrap, where one does not use the circular approach of Politis and Romano [72], which effectively gives $(n-K)$ blocks to choose from (instead of n blocks). The proof uses the same methods, but one has to take the new situation into account and therefore go through the proof carefully.

Remark 3.6.4. Here, however, it is also possible to use $\widetilde{T}^{(1)}_{L,K}(\mathbf{U})$, where we take the maximum over the whole range $1 \leqslant K(l-1)+k < n$. The reason is that equation (3.6.3) in Remark 3.6.1 is fulfilled $a.s.$ for logarithmic K. The proof is analogous to the one of Remark 3.5.5 b), if one takes into account equation (3.6.9).

Proof of Theorem 3.6.2. The proof is analogous to that of Theorem 3.5.1, yet one has to be very careful, which is why we will sketch it for this situation.

We are using repeatedly D as a constant, which may be different in every inequality.

In the present situation one has to be careful with equations (3.5.7) and (3.5.8). We still want to use Theorems B.5 respectively B.6, however the α-mixing condition is not fulfilled, because we have overlapping blocks. Using the following reasoning it is possible to avoid that problem.

First of all note that for the glued parts it still holds uniformly in K and $l^* = n - K + 1, \ldots, n-1$ using Theorems B.5 respectively B.6

$$\mathrm{E}\left|\frac{1}{\sqrt{K}}\left(\sum_{j=1}^{n-l^*} e(l^*+j) + \sum_{j=1}^{K-(n-l^*)} e(j)\right)\right|^{2+\delta^{(1)}} \leqslant D.$$

Now use the following argument to arrive at the equivalent of equations (3.5.7) respectively (3.5.8), i.e.

$$\frac{1}{n}\sum_{l=0}^{n-1}\left(\frac{1}{\sqrt{K}}\sum_{k=1}^{K} e(l+k)\right)^2 \to C > 0 \quad a.s. \tag{3.6.7}$$

respectively

$$\frac{1}{n}\sum_{l=0}^{n-1}\max_{k=0,\ldots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K} e(l+j)\right|^{2+\tilde{\delta}} = \mu_K(2+\tilde{\delta}) + o(1) \leqslant D \quad a.s. \tag{3.6.8}$$

where again $e(j) = e(j-n)$ for $j > n$.

The Markov inequality and Theorems B.5 respectively B.6 give for any $\epsilon > 0$

$$\begin{aligned} P\left(\frac{1}{n}\sum_{j=1}^{n}|Y(j)| \geqslant \epsilon\right) &= P\left(\sum_{k=1}^{K}\frac{1}{L}\sum_{l=0}^{L-1}|Y(Kl+k)| \geqslant K\epsilon\right) \\ \leqslant P\left(\max_{k=1,\ldots,K}\frac{1}{L}\sum_{l=0}^{L-1}|Y(Kl+k)| \geqslant \epsilon\right) &\leqslant \sum_{k=1}^{K} P\left(\frac{1}{L}\sum_{l=0}^{L-1}|Y(Kl+k)| \geqslant \epsilon\right) \\ \ll \frac{1}{\epsilon^{2+\delta^{(2)}}}\frac{K}{L^{1+\delta^{(2)}/2}}, \end{aligned} \tag{3.6.9}$$

where

$$Y(s) := \left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K} e(s-1+j)\right)^2 - \operatorname{var}\left(\frac{1}{\sqrt{K}}\sum_{k=1}^{K} e(k)\right)$$

respectively

$$Y(s) := \max_{i=0,\ldots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=i+1}^{K} e(s-1+j)\right|^{2+\tilde{\delta}} - \mu_K(2+\tilde{\delta}).$$

This converges to 0 sufficiently fast under (3.6.6) to imply $a.s.$-convergence, since

$$\sum_{L\geqslant 1}\frac{K}{L^{1+\delta^{(2)}/2}} \leqslant \sum_{L\geqslant 1}\frac{1}{L^{1+\epsilon}} < \infty.$$

If assumption (3.6.6) is not fulfilled, the above convergence still holds in a stochastic sense. The following reasoning in addition to the subsequence principle then also gives the desired result.

Theorems B.5 respectively B.6 yield for $l_1(n) < l_2(n)$ and $0 \leqslant c_l \leqslant C(n)$

$$\sum_{l=l_1(n)+1}^{l_2(n)} c_l\, e(l) \ll C(n) \sum_{l=l_1(n)+1}^{l_2(n)} |e(l)| = O(C(n)(l_2(n) - l_1(n))) \quad a.s., \tag{3.6.10}$$

if $l_2(n) - l_1(n) \to \infty$. Otherwise if $c_l = O(1)$ and $l_2(n) - l_1(n) = O(1)$, the Markov inequality yields

$$\frac{1}{\sqrt{L}} \left| \sum_{l=l_1(n)+1}^{l_2(n)} c_l\, e(l) \right| = o(1) \quad a.s. \tag{3.6.11}$$

Now we are ready to use Theorem 3.6.1 to arrive at the assertion. We only sketch the proof for (i), because it should be clear then how to adapt the proof of Theorem 3.4.1 (ii).

Again choose the scores $a_n(i) := X(i)$. Without loss of generality assume $\mu = 0$. Let $X(j) = X(j-n)$ for $j > n$. Note that it holds

$$\frac{1}{n} \sum_{l=0}^{n-1} \left(\frac{1}{\sqrt{K}} \sum_{k=1}^{K} \left(X(l+k) - \bar{X}_n \right) \right)^2 = \frac{1}{Kn} \sum_{l=0}^{n-1} \left(\sum_{k=1}^{K} X(l+k) \right)^2 - K\bar{X}_n^2.$$

Equation (3.5.4) gives

$$K\bar{X}_n^2 = Kd^2 \left(\frac{n-m}{n} \right)^2 + o\left(\sqrt{K} d \frac{n-m}{n} \right) + o(1) \quad a.s.$$

Again we get as $L \to \infty$

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

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

$$\begin{aligned} &\frac{1}{Kn} \sum_{l=0}^{n-1} \left(\sum_{k=1}^{K} d1_{\{n \geqslant l+k > m\}} \right)^2 = \frac{d^2\, K^2}{Kn} (n-K-m+1)_+ + 2 \frac{d^2}{Kn} \sum_{j=1}^{K-1} j^2 \\ &= d^2\, K \frac{n-m}{n} + o(1). \end{aligned}$$

Similarly to equation (3.5.6)

$$\frac{1}{\sqrt{L(n-m)}} \sum_{j=m+K}^{n-K+1} e(j) = o(1) \quad a.s.$$

This and equations (3.6.10) and (3.6.11) now imply since $K/L = O(1)$

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

The rest is analogous to the proof of Theorem 3.5.1. ■

3.7. Future Research

The question arises whether this approach could be used for different problems in change-point analysis involving dependent data.

In fact the same statistics are used in diverse contexts. One possibility is to use the above model with an error sequence other than linear processes. Theorem 4.1.2 of Csörgő and Horváth [19] gives such a result for the q-CUSUM statistic, confer also Remark 3.3.1. A different example can be found in Aue et al. [5]. There they derive an invariance principle for augmented GARCH sequences. One of the main concerns in econometrics is to decide whether the volatility remains stable over time or whether it changes in the observation period. They propose to use statistics as in Section 3.2 for this problem and derive their null asymptotics, which turn out to be the same as in Theorem 3.3.2. But they do not clarify alternatives nor show consistency of the test.

Since the asymptotic distribution is the same as in Theorem 3.4.1, one can immediately use the rank statistic result derived there. It then suffices to prove that assumptions (3.4.1) and (3.4.2) are fulfilled almost surely under the null hypothesis as well as alternatives for scores of the type $a_n(i) := c(K,n)X(i)$, where $c(K,n)$ is some function of K and n but not on i.

For the example in Remark 3.3.1 it would suffice to prove equations (3.5.4) - (3.5.8) respectively (3.6.7) - (3.6.8) in order to obtain the validity of the bootstrap. In many cases the proof of those equations also remains true, because it only uses certain conditions on the mixing coefficients; confer Remark 3.5.3. The rest of the proof is then analogous to the one of Theorem 3.5.1.

The problem is more difficult for the model proposed by Aue et al. [5]. First one has to clarify alternatives, then prove that assumptions (3.4.1) and (3.4.2) are fulfilled almost surely for scores of the type $a_n(i) := c(K, n)X(i)$ under the null hypothesis as well as suitable alternatives.

4. Resampling Methods in the Frequency Domain for Linear Sequences

In this chapter we continue to investigate resampling methods for the AMOC location model with error sequences that form a linear process. Instead of using blocking techniques we resample now in the frequency domain rather than the time domain. This approach depends crucially on the fact that the error sequence is a linear process, whereas the blocking techniques also hold for a larger class of error sequences (confer Remark 3.5.3).
Nevertheless, the frequency bootstrap is a very interesting approach with the advantage that it does not depend on a free parameter such as the block bootstrap does on the block length.

The Chapter is organized as follows:
First we give a short introduction into the history of resampling in the frequency domain. Then we thoroughly describe the algorithm for the frequency bootstrap test before giving a mathematical formulation of the problem. We turn our attention to proving that the test described in the first section is indeed asymptotically correct. Because the bootstrap statistics are based on trigonometric polynomials we first have to state some properties of trigonometric functions.
The proofs for the validity of the tests are again based on the corresponding rank statistic asymptotics. Thus we will develop them in Section 4.5, yet this time the proofs are not based on the embedding by Einmahl and Mason, Theorem D.1, but follow from simple linear rank statistic results (confer Appendix E).

This enables us to prove in Section 4.6 that the critical values obtained by the frequency bootstrap are in fact approximations of the quantiles of the distribution under the null hypothesis. Again this is true even when the observations follow an alternative. This time the approach makes use of change-point estimators. This is why we first prove some of their properties. Finally we give a short outline of the proof of the validity for the bootstrap with replacement.

We conclude the chapter with a discussion of some problems arising in the proofs, some possible variations of the method and last but not least some future fields of applications for the frequency bootstrap in change-point analysis.

4.1. Introduction

We have seen in the previous chapter that block bootstrapping techniques are established methods if the data at hand is dependent. There is, however, one major drawback namely the choice of block length. This is motivated by the observation that the performance of block bootstrapping procedures is sensitive to it. The simulation study in Section 6.2 confirms this observation. Recently some literature about the optimal choice of block length has appeared. Yet, this is only available for some examples and the optimal choice depends on the problem. By minimizing the mean squared error Hall et al. [41] show for example that the optimal block length for a series of length n is equal to multiples of $n^{\frac{1}{3}}$, $n^{\frac{1}{4}}$, and $n^{\frac{1}{5}}$ in the cases of variance or bias estimation, estimation of a one-sided distribution function, and estimation of a two-sided distribution function, respectively. The constant depends not only on the statistic of interest or the context, but also on the generally unknown auto-covariance function of the underlying process.

For those reasons in 1992 Franke and Härdle [30] proposed a different approach of bootstrapping kernel spectral density estimates based on resampling from the periodogram of the original data. The idea behind that approach is that a random vector of the periodograms of finitely many frequencies is approximately independent and exponentially distributed (cf. e.g. Brockwell and Davis [13], Theorem 10.3.1). Later this approach was also pursued for different models, e.g. for ratio statistics such as autocorrelations by Dahlhaus and Janas [22] or in regression models by Hidalgo [44].

In the above papers the estimation problem as a whole was transformed into the frequency domain. As a contrast we backtransform the bootstrapped coefficients and look at the new sequences – back in the time domain – as new pseudo-time series. Using empirical distribution functions we then construct the estimator. A similar approach was used by Braun and Kulperger [11]. Nevertheless their bootstrapping in the frequency domain is quite different from ours. Above all their bootstrapped coefficients do not capture the properties of Fourier coefficients, such as the fact that the last ones are the conjugated complex of the first ones. As a result Braun and Kulperger [11] do not necessarily get real numbers after backtransforming and are thus forced to just work with the real part of the backtransformed sequence.

4.2. Idea and Algorithm

In this section we explain the idea behind the bootstrap in the frequency domain and give a thorough description of the algorithm.

First Step: Stationarization

First we need to get closer to the actual error sequence which we know forms a linear process. So we use estimators for the change-point and the mean before and after the change and can thus estimate the error sequence:

$$\widetilde{X}(i) := (X(i) - \bar{X}_{\widehat{m}})1_{[1,\widehat{m}]}(i) + (X(i) - \bar{X}^*_{\widehat{m}})1_{[\widehat{m}+1,n]}(i), \tag{4.2.1}$$

where e.g. $\widehat{m} = \min(\arg\max\{|S_k|; k = 1, \dots, n\})$, $S_k = \sum_{i=1}^{k}(X_i - \bar{X})$, and $\bar{X}^*_{\widehat{m}} = \frac{1}{n-\widehat{m}} \sum_{i=\widehat{m}+1}^{n} X_i$.

Second Step: Fourier Transform and Resampling of Fourier Coefficients

Then we compute the Fourier coefficients of $\{\widetilde{X}(i) : 1 \leqslant i \leqslant n\}$:

$$\omega(j) := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \widetilde{X}(k) \exp(-2\pi ijk/n).$$

It is known that these coefficients are asymptotically (complex) normally distributed and independent with mean 0 and variance $2\pi f(\lambda_j)$; f is the spectral density of $\widetilde{X}$; $\lambda_j = 2\pi ij/n$. This is true in the following sense (for details on the assumptions on the underlying process and the proof confer Brillinger [12], Theorem 4.4.1, p. 94):

Theorem 4.2.1. *Let $s_j(n)$ be an integer with $\lambda_j(n) = 2\pi s_j(n)/n \to \lambda_j$ as $n \to \infty$ for $j = 1, \dots, J$, where $2\lambda_j(n), \lambda_j(n) \pm \lambda_k(n) \not\equiv 0 \mod 2\pi$ for $1 \leqslant j < k \leqslant J$. Under some conditions on the moments and cumulants of the stationary sequence $\{e(i) : i \geqslant 0\}$, which does not need to be a linear process, the real and imaginary part of the Fourier coefficients $(Re[\omega(\lambda_j(n))], Im[\omega(\lambda_j(n))])$, $j = 1, \dots, J$, are asymptotically independent two-dimensionally normally distributed with mean zero and covariance matrix Σ, where*

$$\Sigma = \begin{pmatrix} \pi f(\lambda) & 0 \\ 0 & \pi f(\lambda) \end{pmatrix}$$

and $\omega(\lambda) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \widetilde{X}(k) \exp(-ik\lambda)$.

Brockwell and Davis [13], Theorem 10.3.1, give a similar result for the periodograms. This is less general because it specifically deals with linear processes but the conditions on the weights and moments are less stringent. Namely they assume (3.3.2) with $\nu = 4$ and (3.3.3). Our proofs are closely related to theirs and so we need the same assumptions although it is possible to relax the first one somewhat, namely it suffices $\nu > 2$.

Note that $\omega(n-k) = \overline{\omega(k)}$.

Let $g(1) := \mathrm{Re}(\omega(1)), g(2) := \mathrm{Im}(\omega(1)), \dots, g(\tilde{n}-1) := \mathrm{Re}\left(\omega\left(\frac{\tilde{n}}{2}\right)\right), g(\tilde{n}) := \mathrm{Im}\left(\omega\left(\frac{\tilde{n}}{2}\right)\right)$, where $\tilde{n} := n-1$ for n odd and $\tilde{n} := n-2$ for n even. We bootstrap the centered coefficients $g(i) - \frac{1}{\tilde{n}} \sum_{j=1}^{\tilde{n}} g(j)$ either with or without replacement.
We then choose $\omega(n-1), \dots, \omega(n-\tilde{n}/2)$ in the corresponding way.
For n even set $\omega(n/2) = 0$ (or keep it). Both methods work in simulations. Asymptotically there is usually no difference. It holds that $T_n(X) = T_n^{(1)}(X) + T_n^{(2)}(X)$, where $T_n^{(1)}(X)$ is the statistic for the choice $\omega(n/2) = 0$ and $T_n^{(2)}(X) \to 0$, as $n \to \infty$, conditionally on X in a P-stochastic sense for most statistics T.
Note that $\omega(n)$ is the mean of the sequence (both before and after bootstrapping). Since all the statistics we use center the input sequence, the mean is irrelevant, so we might just set $\omega(n) = 0$.

Remark 4.2.1. a) The covariance structure of the original sequence is coded in the variances of the coefficients. Bootstrapping in the above way will destroy that, but correspond to a similar sequence with independent errors and variance $\sigma^2 \sum w_s^2$, where w_j are the weights of the given linear process. This is why we still need an estimator for $\sigma^2 \left(\sum w_s\right)^2$ in order to use the critical values of the bootstrap statistic for the null hypothesis. Confer also Section 4.8.1 below.

b) For independent errors: $f(\lambda_j)$ (hence the variances) does not depend on j.

c) The permuted coefficients will be similar to normally distributed r.v. with mean 0 and variance $\pi n\sigma^2 \sum w_j^2$, since $\int_0^{2\pi} f(\lambda)d\lambda = \sigma^2 \sum w_j^2$ (see e.g. Neuhaus and Kreiss [66], p. 26, Theorem 3.4.). Hence they will correspond to a sequence of i.i.d. errors with variance $\sigma^2 \sum w_j^2$. Remark 4.5.2 gives the mathematically correct explanation for this.

Third Step: Backtransformation

Use inverse Fourier transform to obtain a similar sample as the original one:

$$X_{\mathbf{R}}(l) := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \omega_{\mathbf{R}}(k) \exp(2\pi i l k/n), \tag{4.2.2}$$

where

$$\omega_{\mathbf{R}}(l) = g(R_l) - \bar{g} + i\left(g(R_{\tilde{n}+1-l}) - \bar{g}\right), \quad l = 1, \ldots, \frac{\tilde{n}}{2},$$

$\bar{g} = \frac{1}{\tilde{n}} \sum_{j=1}^{\tilde{n}} g(j)$. Moreover $\omega_{\mathbf{R}}(n) = 0$, $\omega_{\mathbf{R}}(n/2) = 0$ (for n even);
$\omega_{\mathbf{R}}(n-l) = \overline{\omega_{\mathbf{R}}(l)}$, $l = 1, \ldots, \frac{\tilde{n}}{2}$, the conjugated complex of $\omega_{\mathbf{R}}(l)$.
Here, $\mathbf{R} = (R_1, \ldots, R_{\tilde{n}})$ is a random permutation of $(1, \ldots, \tilde{n})$.

Now calculate the value of the chosen statistic for sample $X_{\mathbf{R}}$. This seems to work in simulations. Our proof, however, only holds true if we just use $\frac{n}{\alpha(n)}$ of the n components of $X_{\mathbf{R}}$ where $\alpha(n) \to \infty$, no matter how slowly (for $\nu \geqslant 4$). It is also not important which ones we use, although it is reasonable to use successive ones (confer also Section 4.8.1).

The permutation statistic is then standardized with the exact variance of the corresponding rank statistic, confer also Remark 4.5.2, i.e.

$$\frac{2}{\tilde{n}} \sum_{l=1}^{\tilde{n}} \left(g(l) - \frac{1}{\tilde{n}} \sum_{k=1}^{\tilde{n}} g(l) \right)^2 .$$

This is essentially the factor needed to standardize the Fourier coefficients $g(\cdot)$. The factor 2 is needed because we use each of these coefficients twice (in $\omega_{\mathbf{R}}(\cdot)$ as well as $\omega_{\mathbf{R}}(n - \cdot)$)

Remark 4.2.2. It is also possible to use the uncentered Fourier coefficients. Since they converge uniformly to 0, the proofs remain the same. Just for some weights q respectively r we need a good enough rate for the uniform convergence. To get that we need somewhat stronger moment conditions on the linear process. It is noted in a remark where necessary.

Forth Step: Calculation of Critical Values

Then we repeat the second and third step t times and calculate the α-quantile of the statistic based on these t "realizations".

We reject the null hypothesis if the value of the statistic for the original sample (here we have to divide by the asymptotic variance $\sigma^2 \left(\sum w_s\right)^2$ or an appropriate estimate) is larger than the above α-quantile.

Some Thoughts and Remarks

The advantage of this method over the block permutation is that it does not depend on the choice of a free parameter such as the block length. The block permutation, however, does dependent crucially on the "correct" choice of it (confer the simulation study in Section 6.2).

The disadvantage is that we have to have good estimates for $\sigma^2 \left(\sum w_s\right)^2$. If those estimates are not good the quality of the test also degrades. This is the same as with the asymptotic test. Unfortunately the estimates so far do not work too well. The block permutation method, however, does not depend on it, because the null statistic as well as each of the realizations of the permutation statistic is divided by the same value of it. Further it can be expected that the block bootstrap is more robust if the error sequence is in fact not a linear process but some other stationary sequence (confer Remark 3.5.3).

The simulation study in Section 6.2 shows that the frequency permutation method gives better results than the asymptotic one. There we use the correct value of the variance instead of an estimator, since we need to estimate the same value in both tests and we are interested in evaluating the performance of the method developed in this chapter not the performance of the estimator.

4.3. Mathematical Formulation of the Problem

In this section we give an exact mathematical formulation of the bootstrapped sequence. This is important for the following sections which prove the validity of the resampling procedure. For practical purposes, however, the representations in (4.3.1) respectively (4.3.3) are not very well suited because the calculation is too slow. It is much more efficient to use a Fast-Fourier-Transform (FFT) Algorithm.

The algorithm in Section 4.2 gives for n odd the following random sequence as bootstrap approximation of the original observations:

$$\begin{aligned} X^o_{\mathbf{R}}(s) &:= \frac{2}{n}\sum_{j=1}^{n} \widetilde{X}(j)\left\langle c^o_j(\mathbf{R}) - \frac{1}{n-1}\sum_{k=1}^{n-1} c^o_j(k), c^o_s\right\rangle \\ &= \frac{2}{n}\sum_{l=1}^{n-1} c^o_s(l)\left(\sum_{j=1}^{n}\widetilde{X}(j)c^o_j(R_l) - \frac{1}{n-1}\sum_{k=1}^{n-1}\sum_{i=1}^{n}\widetilde{X}(i)c^o_i(k)\right), \end{aligned} \tag{4.3.1}$$

where $\langle\, ,\,\rangle$ is the standard scalar product, $\mathbf{R} = (R_1, \ldots, R_{n-1})$ is a random permutation of $(1, \ldots, n-1)$, independent of $X_1, \ldots, X_n$, and for $j = 1, \ldots, n$

$$\begin{aligned} c^o_j &= (0, c^o_j(1), \ldots, c^o_j(n-1))^T \\ &= (0, \cos(2\pi j/n), \sin(-2\pi j/n), \cos(2\cdot 2\pi j/n), \ldots, \sin(-(n-1)/2\cdot 2\pi j/n))^T, \end{aligned} \tag{4.3.2}$$

$$c^o_j(R) = (0, c^o_j(R_1), \ldots, c^o_j(R_{n-1}))^T.$$

Proof of representation (4.3.1). The defining equation (4.2.2) gives, for n odd,

$$\begin{aligned}
nX^o_{\mathbf{R}}(s) &= \sqrt{n}\sum_{l=1}^{\tilde{n}/2}\omega_{\mathbf{R}}(l)\exp(s\lambda_l) + \overline{\sqrt{n}\sum_{l=1}^{\tilde{n}/2}\omega_{\mathbf{R}}(l)\exp(s\lambda_l)} \\
&= 2\sqrt{n}\,\mathrm{Re}\sum_{l=1}^{\tilde{n}/2}\omega_{\mathbf{R}}(l)\exp(s\lambda_l) \\
&\overset{\mathcal{D}}{=} 2\sum_{l=1}^{\tilde{n}/2}\cos(2\pi sl/n)\left(\sum_{j=1}^{n}\widetilde{X}(j)c^o_j(R_l) - \frac{1}{n-1}\sum_{k=1}^{n-1}\sum_{i=1}^{n}\widetilde{X}(i)c^o_i(k)\right) \\
&\quad - 2\sum_{l=1}^{\tilde{n}/2}\sin(2\pi sl/n)\left(\sum_{j=1}^{n}\widetilde{X}(j)c^o_j(R_{n-l}) - \frac{1}{n-1}\sum_{k=1}^{n-1}\sum_{i=1}^{n}\widetilde{X}(i)c^o_i(k)\right) \\
&\overset{\mathcal{D}}{=} 2\sum_{l=1}^{n-1}c^o_s(l)\left[\sum_{j=1}^{n}\widetilde{X}(j)c^o_j(R_l) - \frac{1}{n-1}\sum_{t=1}^{n-1}\sum_{i=1}^{n}\widetilde{X}(i)c^o_i(k)\right],
\end{aligned}$$

where $\lambda_j = 2\pi ij/n$ and $\overline{x+iy} = x - iy$ is the conjugated complex. $\mathbf{R} = (R_1,\ldots,R_{n-1})$ is a random permutation of $(1,\ldots,n-1)$ independent of $X(1),\ldots,X(n)$. However due to a different order of the $c^o_j(l)$, we have only distributional equality, where noted, since a different permutation might lead to that representation. ■

For n even and $\omega(\frac{n}{2}) = 0$, we get

$$\begin{aligned}
X^e_{\mathbf{R}}(s) &:= \frac{2}{n}\sum_{j=1}^{n}\widetilde{X}(j)\left\langle c^e_j(\mathbf{R}) - \frac{1}{n-2}\sum_{k=1}^{n-2}c^e_j(k), c^e_s\right\rangle \\
&\overset{\mathcal{D}}{=} \frac{2}{n}\sum_{l=1}^{n-2}c^e_s(l)\left(\sum_{j=1}^{n}\widetilde{X}(j)c^e_j(R_l) - \frac{1}{n-2}\sum_{k=1}^{n-2}\sum_{j=1}^{n}\widetilde{X}(j)c^e_j(k)\right),
\end{aligned} \tag{4.3.3}$$

where $\mathbf{R} = (R_1,\ldots,R_{n-2})$ is a random permutation of $(1,\ldots,n-2)$ and

$$\begin{aligned}
c^e_j &= (0,0,c^e_j(1),\ldots,c^e_j(n-2))^T \\
&= (0,0,\cos(2\pi j/n),\sin(-2\pi j/n),\cos(2\cdot 2\pi j/n),\ldots,\sin(-(n/2-1)\cdot 2\pi j/n))^T,
\end{aligned} \tag{4.3.4}$$

$j = 1,\ldots,n$, $c^e_j(R) = (0,0,c^e_j(R_1),\ldots,c^e_j(R_{n-2}))^T$.

The proof is analogous to the one for n odd.

If one does not center the Fourier coefficients, we get the above expressions without the mean term.

If instead we choose to keep our middle term, i.e. choose $\omega(n/2) = \sum_{j=1}^{n}\widetilde{X}(j)(-1)^j$, we get

$$\begin{aligned}
X^e_{\mathbf{R}}(s) =& \frac{2}{n}\sum_{l=1}^{n-2}c^e_s(l)\left(\sum_{j=1}^{n}\widetilde{X}(j)c^e_j(R_l) - \frac{1}{n-2}\sum_{k=1}^{n-2}\sum_{i=1}^{n}\widetilde{X}(i)c^e_i(k)\right) \\
&+ \frac{1}{n}(-1)^s\sum_{j=1}^{n}\widetilde{X}(j)(-1)^j.
\end{aligned}$$

The limit (in the $C[0,1]$–sense), however, remains usually the same. Going through the proofs the assertion of Lemma 4.3.1 is normally sufficient. For the statistics with weight functions we sometimes need somewhat stronger conditions (often involving the weight functions). This will be noted in a remark where necessary.

In the following we suppress the upper index of c_j for the sake of simplicity, i.e. $c_j = c_j^o$ for n odd and $c_j = c_j^e$ for n even.

Remark 4.3.1. Define $b_j := c_j^o + \frac{1}{\sqrt{2}} e_1$ for n odd and $b_j := c_j^e + \frac{1}{\sqrt{2}} e_1 + \frac{(-1)^j}{\sqrt{2}} e_2$ for n even, where $e_1 = (1, 0, \ldots, 0)^T$, $e_2 = (0, 1, 0 \ldots, 0)^T$. Note that $\left\{ \sqrt{2/n}\, b_j \right\}_{j=1,\ldots,n}$ is an ON-Basis (confer Theorem 4.4.1) for n even and odd with $\left\langle \sqrt{2/n}\, b_s, e_1 \right\rangle = \sqrt{1/n}$. This gives immediately that the mean of the bootstrap sample is 0:

$$\begin{aligned} \sum_{s=1}^{n} X_{\mathbf{R}}(s) &= \sqrt{2} \sum_{s=1}^{n} \left\langle \sum_{j=1}^{n} \widetilde{X}(j)(c_j(\mathbf{R}) - \bar{c}_j), \sqrt{\frac{2}{n}} b_s \right\rangle \left\langle \sqrt{\frac{2}{n}} b_s, e_1 \right\rangle \\ &= \sqrt{2} \sum_{j=1}^{n} \widetilde{X}(j) \left\langle (c_j(\mathbf{R}) - \bar{c}_j), e_1 \right\rangle = 0, \end{aligned} \tag{4.3.5}$$

since for every ON-Basis $\tilde{b}$ and every x it holds $x = \sum_{s=1}^{n} \langle x, \tilde{b}_s \rangle \tilde{b}_s$.
Here $\bar{c}_j := \frac{1}{\tilde{n}} \sum_{k=1}^{\tilde{n}} c_j(k)(0, 1, \ldots, 1)^T$, $\tilde{n} := n - 1$ for n odd, respectively $\bar{c}_j := \frac{1}{\tilde{n}} \sum_{k=1}^{\tilde{n}} c_j(k)(0, 0, 1, \ldots, 1)^T$, $\tilde{n} := n - 2$ for n even.

We use this bootstrap sequence to determine critical values. We choose $n/\alpha(n)$ of the n bootstrap variables, and calculate the value of the statistic for these. Repeating this we can then calculate critical values.
We need $\alpha(n) \to \infty$, $\frac{\alpha(n) \log^2(n)}{n} = o(1)$, which means that $\alpha(n)$ converges to infinity, but not too fast. It does, however, not matter how slowly. If we only have an existing moment smaller than the fourth, we need that $\alpha(n)$ is also not too small (e.g. it could be of logarithmic order with appropriate exponent) in order to get tightness (cf. Theorem 4.5.3 and Remark 4.5.8).

We are interested in choosing $\alpha(n)$ as small as possible, preferably $\alpha(n) = 1$. Unfortunately Section 4.8.2 shows that it is not possible to prove the result for $\alpha(n) = 1$ using the same methods. However the simulations in Section 6.3 illustrate that the performance of the test does not depend on the choice of $\alpha(n)$.

For notational reasons choose $\alpha(n)$ such that $\frac{n}{\alpha(n)}$ is an integer.

Define

$$\widetilde{Z}_n(u) = \sqrt{\frac{\alpha(n)}{n}} \sum_{s \leqslant \frac{n}{\alpha(n)} u} X_{\mathbf{R}}(\beta(s)), \quad \text{for } u = \frac{\alpha(n)}{n}, \frac{2\alpha(n)}{n}, \ldots, 1,$$

$\widetilde{Z}_n(0) = 0$ and let $\widetilde{Z}_n(t)$ be linearly interpolated between $(i-1)/n$ and i/n for $i = 1, \ldots, n$. The function β defines which $\frac{n}{\alpha(n)}$ of the n bootstrap variables we choose, e.g. $\beta(s) = s$.

Now we are ready to state the above mentioned lemma:

Lemma 4.3.1. *a) For any triangular array* $\{x_{i,n} : i = 1, \ldots, n\}_{n \in \mathbb{N}}$ *with either* *(i)* $\frac{1}{\sqrt{n}} \max_{j=1,\dots,n} |x_{j,n}| \to 0 \qquad$ *(as $n \to \infty$) or*

(ii) $\frac{1}{n^{\frac{3}{2}}} \sum_{j=1}^{n} x_{j,n}(-1)^j = o(1)$

it holds as $n \to \infty$

$$\frac{1}{n^{\frac{3}{2}}} \sup_{0 \leqslant u \leqslant 1} \left| \sum_{s=1}^{\lfloor un \rfloor} (-1)^s \sum_{j=1}^{n} x_{j,n}(-1)^j \right| \to 0.$$

b) For any triangular array $\{x_{i,n} : i = 1, \dots, n\}_{n \in \mathbb{N}}$ with either

(i) $\sup_{0 \leqslant u \leqslant 1} \left| \sum_{s=1}^{\lfloor \frac{n}{\alpha(n)} u \rfloor} (-1)^{\beta(s)} \right| = O(1)$ *and* $\sqrt{\frac{\alpha(n)}{n}} \max_{j=1,\dots,n} |x_{j,n}| \to 0 \qquad$ *or*

(ii) $\frac{1}{n} \sum_{j=1}^{n} x_{j,n}(-1)^j = o\left(\sqrt{\frac{\alpha(n)}{n}}\right)$

it holds as $n \to \infty$

$$\frac{\sqrt{\alpha(n)}}{n^{\frac{3}{2}}} \sup_{0 \leqslant u \leqslant 1} \left| \sum_{s=1}^{\lfloor u \frac{n}{\alpha(n)} \rfloor} (-1)^{\beta(s)} \sum_{j=1}^{n} x_{j,n}(-1)^j \right| \to 0.$$

In many situations the maximum-type condition on the scores is fulfilled in a P-stochastic sense if we replace $x_{j,n}$ by $\widetilde{X}(j)$ as in (4.6.24), where the estimators are given in (4.6.11) for $\gamma < \frac{2}{3}$. This is e.g. the case if there exists a moment larger than the fourth one and if under the alternative $\min\left(\frac{m}{n}, \frac{n-m}{n}\right) \geqslant \epsilon > 0$ and $|d| \leqslant D < \infty$. Furthermore $\alpha(\cdot)$ has to be of sufficiently small order, e.g. logarithmic.
The main argument is

$$\frac{\alpha(n)}{n} \max_{i=1,\dots,n} x_{i,n}^2 \leqslant \frac{1}{n^{\frac{\kappa-2}{\kappa}}} \left(\frac{1}{n} \sum_{i=1}^{n} |x_{i,n}|^\kappa \right)^{2/\kappa}.$$

The proof of Theorem 3.5.1 gives that the sum is bounded in an *a.s.*-sense for linear processes. We then have to adapt the argument for $\widetilde{X}$. This is, however, analogous to arguments given in the proof of Theorem 4.6.2.

Proof of Lemma 4.3.1. For the proof of a), note that $\sum_{s=1}^{\lfloor un \rfloor} (-1)^s = 0$, if $\lfloor un \rfloor$ is even, and otherwise $\sum_{s=1}^{\lfloor un \rfloor} (-1)^s = -1$. This gives

$$\frac{1}{n^{\frac{3}{2}}} \sup_{0 \leqslant u \leqslant 1} \left| \sum_{s=1}^{\lfloor un \rfloor} (-1)^s \sum_{j=1}^{n} x_{j,n}(-1)^j \right| \ll \frac{1}{\sqrt{n}} \max_{j=1,\dots,n} |x_{j,n}| \to 0.$$

The proof of b)(i) is analogous. For the proof of b)(ii) note that $\sum_{s=1}^{\lfloor u \frac{n}{\alpha(n)} \rfloor} (-1)^{\beta(s)} = O\left(\frac{n}{\alpha(n)}\right)$. ■

In the following sections we will prove

$$P\left(h((\widetilde{Z}_n(\mathrm{id}) - \mathrm{id}\, \widetilde{Z}_n(1))/\widetilde{\sigma}) \leq x \middle| X_1, \dots, X_n\right) \xrightarrow{P} P\left(h(B(\cdot)) \leqslant x\right)$$

for all continuous $h : C[0,1] \to \mathbb{R}$, for all $x \in \mathbb{R}$ and

$$\widetilde{\sigma}^2 := \frac{2}{n\tilde{n}} \sum_{l=1}^{\tilde{n}} \left(\sum_{i=1}^{n} \widetilde{X}(i) c_i(l) - \frac{1}{\tilde{n}} \sum_{k=1}^{\tilde{n}} \sum_{j=1}^{n} \widetilde{X}(j) c_j(k) \right)^2 .$$

From this we can derive the asymptotics for the statistics proposed in Section 3.3 by choosing h appropriately (confer the proof of Corollary 4.5.1). Instead of using the extreme value statistics we will work with the trimmed version.

This shows that we get an approximation for the critical values corresponding to the null distribution, even if the observed data does follow an alternative.

4.4. Some Properties of Trigonometric Functions

The previous section has shown that the frequency permutation statistics are based on trigonometric polynomials. In the proofs for the rank statistic asymptotics we use the special structure of trigonometric polynomials. In this section we, therefore, state and prove the properties of trigonometric functions that we need in the following.

The first theorem shows that $\{\sqrt{2/n}\, b_j\}$ forms an orthonormal basis.

Theorem 4.4.1. *a) For $j = 1, \dots, n$ it holds*

$$\sum_{l=1}^{n-1} c_j^o(l)^2 = \frac{n-1}{2}, \qquad \sum_{l=1}^{n-2} c_j^e(l)^2 = \frac{n-2}{2}.$$

b) For $i, j = 1, \dots, n, i \neq j$ it holds

$$\sum_{l=1}^{n-1} c_j^o(l)\, c_i^o(l) = -\frac{1}{2}, \qquad \sum_{l=1}^{n-2} c_j^e(l)\, c_i^e(l) = -\frac{1}{2}(1 + (-1)^{i-j}) = \begin{cases} 0 & i-j \text{ odd}, \\ -1 & i-j \text{ even}. \end{cases}$$

This shows that $\{\sqrt{2/n}\, b_j\}_{j=1,\dots,n}$ forms an ON-Basis.

Remark 4.4.1. Note that $\{\sqrt{\frac{2}{n}} c_\diamondsuit(l), l = 1, \dots, \tilde{n}\}$ forms an ON-System as well (confer e.g. Brockwell and Davis [13], page 333).

Proof. Assertion a) follows, since $\sin^2 x + cos^2 x = 1$ and $\sin(-x) = -\sin(x)$.

To prove assertion b) let $\lambda_s := 2\pi s/n$ and

$$\tilde{x}_s := \left(\cos(\lambda_s), \dots, \cos\Big(\frac{\tilde{n}}{2}\lambda_s\Big), \sin(-\lambda_s), \dots, \sin\Big(-\frac{\tilde{n}}{2}\lambda_s\Big) \right)^T \in \mathbb{R}^{\tilde{n}},$$

$$\tilde{y}_s := \left(\sin(\lambda_s), \dots, \sin\Big(\frac{\tilde{n}}{2}\lambda_s\Big), -\cos(-\lambda_s), \dots, -\cos\Big(-\frac{\tilde{n}}{2}\lambda_s\Big) \right)^T ,$$

where $\tilde{n} = n-1$ for n odd and $\tilde{n} = n-2$ for n even. First of all we realize

$$\langle \tilde{x}_s, \tilde{x}_t \rangle - \langle \tilde{y}_s, \tilde{y}_t \rangle = 0. \tag{4.4.1}$$

Let $s \neq t$. Since $\sin(-x) - i\cos(-x) = -i\exp(-ix)$ and $\overline{-i\exp(-ix)} = i\exp(ix)$, we get by the geometric sum for n odd

$$\begin{aligned}
\langle \tilde{x}_s + i\,\tilde{y}_s, \tilde{x}_t + i\,\tilde{y}_t\rangle &= \sum_{l=1}^{\tilde{n}/2} \exp(il(\lambda_s - \lambda_t)) + \sum_{l=1}^{\tilde{n}/2} \exp(-il(\lambda_s - \lambda_t)) \\
&= \sum_{l=1}^{n-1} \exp(il(\lambda_s - \lambda_t)) = -1,
\end{aligned}$$

which gives

$$\langle \tilde{x}_s, \tilde{x}_t\rangle + \langle \tilde{y}_s, \tilde{y}_t\rangle = \mathrm{Re}(\langle \tilde{x}_s + i\tilde{y}_s, \tilde{x}_t + i\tilde{y}_t\rangle) = -1. \tag{4.4.2}$$

Putting together equation (4.4.1) and (4.4.2) we now arrive at the assertion for n odd.

The proof for n even is analogous with

$$\langle \tilde{x}_s + i\,\tilde{y}_s, \tilde{x}_t + i\,\tilde{y}_t\rangle = \sum_{\substack{l=1 \\ l\neq\frac{n}{2}}}^{n-1} \exp(il(\lambda_s - \lambda_t)) = -1 - (-1)^{s-t}.$$

To obtain that $\{\sqrt{2/n}\, b_j\}_{j=1,\dots,n}$ forms an ON–Basis, note that

$$i - j \text{ even} \quad \Leftrightarrow \quad i + j \text{ even.}$$

■

The following lemma gives a similar result for the sum of the product of four such trigonometric functions.

Lemma 4.4.1. *It holds for n odd*

$$\sum_{l=1}^{n-1} c^o_{s_1}(l)c^o_{s_2}(l)c^o_{s_3}(l)c^o_{s_4}(l) = \begin{cases} O(n), & \sum_{j=1}^4 \delta_\pm^{(j)} s_j = in,\ i \in \mathbb{Z},\ \delta_\pm^{(j)} = \pm 1 \\ & \quad \textit{with } \sum_{j=1}^n \delta_\pm^{(j)} = 0 \textit{ or } \sum_{j=1}^n \delta_\pm^{(j)} = 4, \\ -\frac{1}{2}, & \textit{otherwise,} \end{cases}$$

and for n even

$$\sum_{l=1}^{n-2} c^e_{s_1}(l)c^e_{s_2}(l)c^e_{s_3}(l)c^e_{s_4}(l) = \begin{cases} O(n), & \sum_{j=1}^4 \delta_\pm^{(j)} s_j = in,\ i \in \mathbb{Z},\ \delta_\pm^{(j)} = \pm 1 \\ & \quad \textit{with } \sum_{j=1}^n \delta_\pm^{(j)} = 0 \textit{ or } \sum_{j=1}^n \delta_\pm^{(j)} = 4, \\ 0, & s_1 + s_2 + s_3 + s_4 \textit{ odd,} \\ -1, & \textit{otherwise.} \end{cases}$$

Note that the condition above means that the sum is only of linear order if s_4 is determined by a finite number of combinations of s_1, s_2, s_3.

Proof. Let again $\tilde{n} = n - 1$ for n odd and $\tilde{n} = n - 2$ for n even and $\lambda_s := 2\pi s/n$. Define

$$\begin{aligned}
(x_s(1), \dots, x_s(\tilde{n}))^T &:= \left(\cos(\lambda_s), \dots, \cos\left(\frac{\tilde{n}}{2}\lambda_s\right), \sin(-\lambda_s), \dots, \sin\left(-\frac{\tilde{n}}{2}\lambda_s\right)\right)^T \\
(y_s(1), \dots, y_s(\tilde{n}))^T &:= \left(\sin(\lambda_s), \dots, \sin\left(\frac{\tilde{n}}{2}\lambda_s\right), -\cos(-\lambda_s), \dots, -\cos\left(-\frac{\tilde{n}}{2}\lambda_s\right)\right)^T.
\end{aligned}$$

Moreover $e_s := x_s + iy_s$ and $\overline{e}_s = x_s - iy_s$. Note that

$$
\begin{aligned}
S_1 &:= \sum_{l=1}^{\tilde{n}} x_{s_1}(l)\, x_{s_2}(l)\, x_{s_3}(l)\, x_{s_4}(l) = \sum_{l=1}^{\tilde{n}} y_{s_1}(l)\, y_{s_2}(l)\, y_{s_3}(l)\, y_{s_4}(l),\\
S_2 &:= \sum_{l=1}^{\tilde{n}} x_{s_1}(l)\, x_{s_2}(l)\, y_{s_3}(l)\, y_{s_4}(l) = \sum_{l=1}^{\tilde{n}} y_{s_1}(l)\, y_{s_2}(l)\, x_{s_3}(l)\, x_{s_4}(l),\\
S_3 &:= \sum_{l=1}^{\tilde{n}} x_{s_1}(l)\, y_{s_2}(l)\, x_{s_3}(l)\, y_{s_4}(l) = \sum_{l=1}^{\tilde{n}} y_{s_1}(l)\, x_{s_2}(l)\, y_{s_3}(l)\, x_{s_4}(l),\\
S_4 &:= \sum_{l=1}^{\tilde{n}} x_{s_1}(l)\, y_{s_2}(l)\, y_{s_3}(l)\, x_{s_4}(l) = \sum_{l=1}^{\tilde{n}} y_{s_1}(l)\, x_{s_2}(l)\, x_{s_3}(l)\, y_{s_4}(l).
\end{aligned}
$$

For n odd we get

$$
\begin{aligned}
&2(S_1 - S_2 - S_3 - S_4) = \sum_{l=1}^{\tilde{n}} \mathrm{Re}(e_{s_1}(l)\, e_{s_2}(l)\, e_{s_3}(l)\, e_{s_4}(l))\\
&= \sum_{l=1}^{\tilde{n}/2} (\exp(il(\lambda_{s_1} + \lambda_{s_2} + \lambda_{s_3} + \lambda_{s_4})) + \exp(-il(\lambda_{s_1} + \lambda_{s_2} + \lambda_{s_3} + \lambda_{s_4})))\\
&= \sum_{l=1}^{\tilde{n}} \exp(il(\lambda_{s_1} + \lambda_{s_2} + \lambda_{s_3} + \lambda_{s_4})) = \begin{cases} -1, & s_1 + s_2 + s_3 + s_4 \neq jn,\\ \tilde{n}, & \text{otherwise.} \end{cases}
\end{aligned}
$$

and analogously

$$
\begin{aligned}
2(S_1 - S_2 + S_3 + S_4) &= \sum_{l=1}^{\tilde{n}} \mathrm{Re}(\overline{e}_{s_1}(l)\, \overline{e}_{s_2}(l)\, e_{s_3}(l)\, e_{s_4}(l))\\
&= \begin{cases} -1, & -s_1 - s_2 + s_3 + s_4 \neq jn,\\ \tilde{n}, & \text{otherwise,} \end{cases}\\
2(S_1 + S_2 - S_3 + S_4) &= \sum_{l=1}^{\tilde{n}} \mathrm{Re}(\overline{e}_{s_1}(l)\, e_{s_2}(l)\, \overline{e}_{s_3}(l)\, e_{s_4}(l))\\
&= \begin{cases} -1, & -s_1 + s_2 - s_3 + s_4 \neq jn,\\ \tilde{n}, & \text{otherwise,} \end{cases}\\
2(S_1 + S_2 + S_3 - S_4) &= \sum_{l=1}^{\tilde{n}} \mathrm{Re}(\overline{e}_{s_1}(l)\, e_{s_2}(l)\, e_{s_3}(l)\, \overline{e}_{s_4}(l))\\
&= \begin{cases} -1, & -s_1 + s_2 + s_3 - s_4 \neq jn,\\ \tilde{n}, & \text{otherwise.} \end{cases}
\end{aligned}
$$

Solving the above system of equations for the different cases we get $S_1 = -\frac{1}{2}$, if all of the equations above are equal to -1, otherwise $S_1 = O(n)$, hence the assertion. Note that the sum of the above expressions gives $8S_1$.

More precisely $S_1 = \frac{n-4}{8}$ if exactly one of the equations equals $\tilde{n}$, $S_1 = \frac{n-2}{4}$ if exactly two of the equations equal $\tilde{n}$, $S_1 = \frac{3n-4}{8}$ if exactly three of the equations equal $\tilde{n}$ and $S_1 = \frac{\tilde{n}}{2}$ if all four of the equations equal $\tilde{n}$.

For n even it holds

$$\begin{aligned}
&2(S_1+S_2+S_3+S_4)\\
&=\sum_{\substack{l=1\\ l\neq \frac{n}{2}}}^{n-1}\exp(il(\lambda_{s_1}+\lambda_{s_2}+\lambda_{s_3}+\lambda_{s_4}))\\
&=\begin{cases}-1-(-1)^{s_1+s_2+s_3+s_4}, & s_1+s_2+s_3+s_4\neq jn,\\ n-2, & \text{otherwise,}\end{cases}
\end{aligned}$$

the other equations analogous.

Again note that

$$\begin{aligned}
& s_1+s_2+s_3+s_4 \text{ odd} &\Leftrightarrow\quad& -s_1-s_2+s_3+s_4 \text{ odd}\\
\Leftrightarrow\quad & -s_1+s_2-s_3+s_4 \text{ odd} &\Leftrightarrow\quad& -s_1+s_2+s_3-s_4 \text{ odd.}
\end{aligned}$$

Solving the system of equations (again the sum gives $8S_1$) we get the assertion. More precisely for $s_1+s_2+s_3+s_4$ even we get $S_1=\frac{n-8}{8}$ if exactly one of the equations equals $n-2$, $S_1=\frac{n-4}{4}$ if exactly two of the equations equal $n-2$, $S_1=\frac{3n-8}{8}$ if exactly three of the equations equal $n-2$ and $S_1=\frac{n-2}{2}$ if all four of the equations equal $n-2$. ■

Lemma 4.4.2. *For $s\neq jn$, $j\in\mathbb{Z}$, it holds*

$$\begin{aligned}
&\sum_{l=1}^{m}\cos(2\pi sl/n)=O\left(\max\left(\frac{n}{s},\frac{n}{n-s}\right)\right)\\
\textit{and}\qquad &\sum_{l=1}^{m}\sin(2\pi sl/n)=O\left(\max\left(\frac{n}{s},\frac{n}{n-s}\right)\right)
\end{aligned}$$

uniformly in m. Note that for all $k\leqslant n$ it holds

$$\sum_{s=1}^{k}\max\left(\frac{n}{s},\frac{n}{n-s}\right)\leqslant 2n\sum_{s=1}^{n}\frac{1}{s}=O(n\log n).$$

Remark 4.4.2. Most of the time we use Lemma 4.4.2 for $\sum_{l=1}^{\tilde{n}/2}(\cos(2\pi sl/n)+\sin(2\pi sl/n))$ with $\tilde{n}=n-1$ for n odd and $\tilde{n}=n-2$ for n even. Even though it is then possible to give the exact value for the sum, the rate will not be better (with the exception of n and s simultaneously even).
More precisely one can show (using equation 1.342 of Gradshteyn and Ryzhik [38] and some well-known facts about trigonometric functions) for $s\neq jn, j\in\mathbb{Z}$, n odd,

$$\begin{aligned}
&\sum_{l=1}^{\frac{n-1}{2}}\cos(2\pi sl/n)=-\frac{1}{2},\\
&\sum_{l=1}^{\frac{n-1}{2}}\sin(2\pi sl/n)=\begin{cases}-\frac{1}{2}\tan\left(\frac{\pi s}{2n}\right), & s \text{ even,}\\ -\frac{1}{2}\tan\left(\frac{\pi s}{2n}+\frac{\pi}{2}\right), & s \text{ odd.}\end{cases}
\end{aligned}$$

For n even we get analogously

$$\sum_{l=1}^{\frac{n}{2}-1} \cos(2\pi sl/n) = \begin{cases} -1, & s \text{ even}, \\ 0, & s \text{ odd}. \end{cases}$$

$$\sum_{l=1}^{\frac{n}{2}-1} \sin(2\pi sl/n) = \begin{cases} 0, & s \text{ even}, \\ \operatorname{cotan}\left(\frac{\pi s}{n}\right), & s \text{ odd}. \end{cases}$$

Proof Lemma 4.4.2. First equation 1.342 of Gradshteyn and Ryzhik [38] gives

$$\sum_{k=1}^{m} \sin(2\pi sk/n) = \frac{1}{\sin(\pi s/n)} \sin((m+1)\pi s/n)\sin(m\pi s/n) = O\left(\frac{1}{\sin(\pi s/n)}\right),$$
$$\sum_{k=1}^{m} \cos(2\pi sk/n) = \frac{1}{\sin(\pi s/n)} \cos((m+1)\pi s/n)\sin(m\pi s/n) = O\left(\frac{1}{\sin(\pi s/n)}\right)$$

uniformly in m. Moreover $\max_{x\in(0,\pi)} \frac{\min(x,\pi-x)}{\sin(x)} = O(1)$ (since $\frac{\sin(x)}{x} \to 1$ as $x \to 0$ and $\frac{\sin(x)}{\pi-x} \to 1$ as $x \to \pi$). This gives now $\frac{1}{\sin(\pi s/n)} = O\left(\max\left(\frac{n}{s}, \frac{n}{n-s}\right)\right)$, hence the assertion.
■

4.5. Corresponding Rank Statistics

As in the previous chapters the derivation of the permutation asymptotics is based on the corresponding results of rank asymptotics. In this section we analyze the corresponding frequency rank statistics. The structure of the frequency rank statistics is, however, much more complicated than before. Thus it is not possible to use the relatively strong embedding of Einmahl and Mason, Theorem D.1. Instead we use simple linear rank statistic results as outlined in Appendix E. That works fine for the q-weighted CUSUM, the Sum, and the trimmed versions of the weighted CUSUM and MOSUM statistics, but it is too weak for the extreme value statistics (confer also Section 4.8.3).

We prove convergence in $C[0,1]$. Therefore we show the convergence of the finite-dimensional distribution in a first subsection and tightness of the process in a second subsection. In Subsection 4.5.3 we finally state the main result, i.e. convergence in $C[0,1]$, and deduce the rank asymptotics for our statistics of choice from it.

The frequency rank statistics are essentially determined by $\widetilde{Z}_n(u) := Z_n(u) - \operatorname{E} Z_n(u)$, where

$$\begin{aligned} Z_n(u) &= \frac{2\sqrt{\alpha(n)}}{n^{3/2}} \sum_{l=1}^{\tilde{n}} \sum_{s\leqslant \frac{n}{\alpha(n)}u} c_{\beta(s)}(l) \sum_{j=1}^{n} x_{j,n} c_j(R_l) \\ &= \sqrt{\frac{\alpha(n)}{n}} \sum_{s\leqslant \frac{n}{\alpha(n)}u} x_{\mathbf{R}}(\beta(s)) \end{aligned} \tag{4.5.1}$$

for $u = j\frac{\alpha(n)}{n}$ and linearly interpolated in between. Again $\omega(\frac{n}{2}) = 0$, $c_j = c_j^o$ and $\tilde{n} = n-1$ for n odd respectively $c_j = c_j^e$ and $\tilde{n} = n-2$ for n even.

We assume that the scores fulfill the following conditions:

$$
\begin{aligned}
&\frac{2}{n\tilde{n}}\sum_{l=1}^{\tilde{n}}\left(\sum_{i=1}^{n}x_{i,n}c_i(l)-\frac{1}{\tilde{n}}\sum_{k=1}^{\tilde{n}}\sum_{j=1}^{n}x_{j,n}c_j(k)\right)^2=1\\
&\frac{1}{n}\sum_{l=1}^{\tilde{n}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_{i,n}c_i(l)-\frac{1}{\tilde{n}}\sum_{k=1}^{\tilde{n}}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}x_{j,n}c_j(k)\right|^{\kappa}=O(1),\qquad 2<\kappa\leqslant 4.
\end{aligned}
\tag{4.5.2}
$$

The last condition can be weakened to $= o(g(n))$ for some $g(n)\to\infty$, confer Remarks 4.5.3 and 4.5.8.

Remark 4.5.1. If we do not subtract the mean of the Fourier coefficients we additionally need (cf. Remarks 4.5.4, 4.5.6, 4.5.7)

$$
\frac{1}{\tilde{n}}\sum_{l=1}^{\tilde{n}}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}x_jc_j(l)=o\left(\frac{1}{\log n\sqrt{\alpha(n)}}\right). \tag{4.5.3}
$$

This is e.g. fulfilled for $\frac{\alpha(n)\log^4(n)}{n}\max_{1\leqslant i\leqslant n}x_{i,n}^2\to 0$, which follows for appropriate (e.g. logarithmic $\alpha(n)$) from condition

$$
\frac{1}{n}\sum_{i=1}^{n}|x_{i,n}|^{\kappa}=O(1),
$$

since

$$
\frac{1}{n}\max_{1\leqslant i\leqslant n}x_{i,n}^2\leqslant\frac{1}{n^{\frac{\kappa-2}{\kappa}}}\left(\frac{1}{n}\sum_{i=1}^{n}|x_{i,n}|^{\kappa}\right)^{\frac{2}{\kappa}}
$$

In case of weighted statistics we need an even stronger condition, which can be derived as above for appropriate weights (cf. Remark 4.5.9).

In the following we write $x_j := x_{j,n}$ for the sake of simplicity.

Remark 4.5.2. Note that by Theorem 4.4.1 it holds for n odd

$$
\frac{2}{n\tilde{n}}\sum_{l=1}^{\tilde{n}}\left(\sum_{j=1}^{n}x_jc_j(l)\right)^2=\frac{1}{n}\sum_{j=1}^{n}x_j^2-\frac{1}{n\tilde{n}}\sum_{\substack{i\neq j\\1}}^{n}x_ix_j=\frac{1}{\tilde{n}}\sum_{j=1}^{n}x_j^2-\frac{1}{n\tilde{n}}\left(\sum_{j=1}^{n}x_j\right)^2,
$$

respectively for n even in the same way

$$
\frac{2}{n\tilde{n}}\sum_{l=1}^{\tilde{n}}\left(\sum_{j=1}^{n}x_jc_j(l)\right)^2=\frac{1}{\tilde{n}}\sum_{j=1}^{n}x_j^2-\frac{2}{n\tilde{n}}\left(\sum_{j=1}^{n/2}x_{2j}\right)^2-\frac{2}{n\tilde{n}}\left(\sum_{j=0}^{n/2-1}x_{2j+1}\right)^2.
$$

If we replace x_j by $\widetilde{X}(j)$ the three squared means above will converge to 0 in a P-stochastic sense under appropriate conditions as the proof of Theorems 4.6.1 and 4.6.2 state.
Theorem C.3 states that as $n\to\infty$

$$
\frac{1}{\tilde{n}}\sum_{j=1}^{n}e^2(j)\to\sigma^2\sum_{s\geqslant 0}w_s^2\quad a.s.
$$

for a linear process with weights w_j and innovations having variance σ^2. The proof of Theorem 4.6.2 also shows $\frac{1}{n^{3/2}}\sum_l\sum_j(\widetilde{X}(j)-e(j))c_j(l)=o_P(1)$. This gives together with equation (4.6.3) that

$$\frac{2}{n\tilde{n}}\sum_{l=1}^{\tilde{n}}\Big(\sum_{i=1}^{n}\widetilde{X}(i)c_i(l)-\frac{1}{\tilde{n}}\sum_{k=1}^{\tilde{n}}\sum_{j=1}^{n}\widetilde{X}(j)c_j(k)\Big)^2\xrightarrow{P}\sigma^2\sum_{s\geqslant 0}w_s^2$$

under the same assumptions as in Theorem 4.6.2.
Thus we standardize the bootstrap statistic asymptotically with $\sigma\sqrt{\sum w_s^2}$. On the other hand the original statistic is asymptotically standardized with $\tau=\sigma|\sum w_s|$. This shows that our bootstrap sample rather corresponds to an independent sequence with variance $\sigma^2\sum w_s^2$. That confirms the statement in Remark 4.2.1 a) respectively c) .

4.5.1. Convergence of the Finite-Dimensional Distributions

The next theorem shows that the finite dimensional distributions of Z_n converge to those of a Wiener process (for $\alpha(n)\to\infty$). For $\alpha(n)=1$ the covariance structure converges to that of a Brownian bridge, but it is not clear whether the limit distribution is normal (confer also Section 4.8.2).

Theorem 4.5.1. *Under* (4.5.2) *it holds for any* $0\leqslant u_1<\ldots<u_k\leqslant 1,\ k\geqslant 1,$

$$(Z_n(u_1)-\operatorname{E}Z_n(u_1),\ldots,Z_n(u_k)-\operatorname{E}Z_n(u_k))\xrightarrow{\mathcal{D}}(W(u_1),\ldots,W(u_k)),$$

where $\{W(t):0\leqslant t\leqslant 1\}$ *is a Wiener process, if* $\alpha(n)\to\infty$, *but* $\frac{\alpha(n)\log^2(n)}{n}\to 0$.

Remark 4.5.3. Note that the above theorem remains true, if one replaces the condition

$$\frac{1}{n}\sum_{l=1}^{\tilde{n}}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}x_jc_j(l)-\frac{1}{\tilde{n}}\sum_{k=1}^{\tilde{n}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_ic_i(k)\right|^{\kappa}=O(1)$$

by the following weaker condition

$$\frac{1}{n}\sum_{l=1}^{\tilde{n}}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}x_jc_j(l)-\frac{1}{\tilde{n}}\sum_{k=1}^{\tilde{n}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_ic_i(k)\right|^{\kappa}=o\left(\alpha(n)^{\frac{\kappa-2}{2}}\right).$$

This will be useful later because of the characterization of P-stochastic convergence via *a.s.*-convergence (subsequence principle). The problem with the first condition is that there is no such characterization for O_P.

Remark 4.5.4. Under (4.5.3) the assertion of Theorem 4.5.1 remains true if we do not center the Fourier coefficients since Lemma 4.4.2 gives uniformly in u

$$\begin{aligned}&\operatorname{E}Z_n\left(\lfloor Nu\rfloor/N\right)=2\frac{\sqrt{\alpha(n)}}{n^{3/2}}\sum_{s=1}^{\lfloor Nu\rfloor}\sum_{l=1}^{\tilde{n}}c_{\beta(s)}(l)\sum_{j=1}^{n}x_j\frac{1}{\tilde{n}}\sum_{k=1}^{\tilde{n}}c_j(k)\\&\ll\sqrt{\alpha(n)}\log n\frac{1}{n^{\frac{3}{2}}}\sum_{k=1}^{\tilde{n}}\sum_{j=1}^{n}x_jc_j(k)=o(1).\end{aligned}\tag{4.5.4}$$

The following theorem might be useful in some situations. It states that any finite sequence of bootstrapped samples belonging to any frequency is asymptotically independent, standard normally distributed.

Theorem 4.5.2. *For any $M > 0$ and any $0 < \lambda_1 < \ldots < \lambda_M < 1$ let*

$$x^*_{s,n} := \frac{2}{n} \sum_{l=1}^{\tilde{n}} c_{\lfloor n\lambda_s \rfloor}(l) \left(\sum_{j=1}^{n} x_j c_j(R_l) - \frac{1}{\tilde{n}} \sum_{k=1}^{\tilde{n}} \sum_{j=1}^{n} x_j c_j(k) \right),$$

then it holds under conditions (4.5.2) as $n \to \infty$

$$(x^*_{1,n}, \ldots, x^*_{M,n}) \xrightarrow{\mathcal{D}} (Y_1, \ldots, Y_M),$$

where $(Y_1, \ldots, Y_M) \stackrel{\mathcal{D}}{=} N(0, I_M)$.

Remark 4.5.5. As in Remark 4.5.3 we can also weaken the condition by the following one:

$$\frac{1}{n^{\frac{\kappa}{2}}} \sum_{l=1}^{\tilde{n}} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} x_j c_j(l) - \frac{1}{\tilde{n}} \sum_{k=1}^{\tilde{n}} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i c_i(k) \right|^{\kappa} = o(1).$$

Proof of Theorem 4.5.2. This goes along the lines of the proof of Theorem 4.5.1 below with $d_n(l) = \sqrt{\frac{2}{n}} \sum_{i=1}^{M} \gamma_i c_{\lfloor n\lambda_i \rfloor}(l)$. ■

For the proof of Theorem 4.5.1 we need the next lemma. It states that the linearly interpolated part of $Z_n(\cdot)$ can be neglected to derive the asymptotics.

Lemma 4.5.1. *Under conditions (4.5.2) it holds as $n \to \infty$ uniformly in $0 \leqslant u \leqslant 1$*

$$Z_n(u) - \mathrm{E}\, Z_n(u) - Z_n \left(\left\lfloor \frac{n}{\alpha(n)} u \right\rfloor \frac{\alpha(n)}{n} \right) + \mathrm{E}\, Z_n \left(\left\lfloor \frac{n}{\alpha(n)} u \right\rfloor \frac{\alpha(n)}{n} \right) = O_P \left(\sqrt{\frac{\alpha(n)}{n}} \right).$$

This remains true for $\alpha(n) = 1$.

Proof. Let $N := \frac{n}{\alpha(n)} \to \infty$ as $n \to \infty$. Then it holds

$$Z_n(u) - Z_n \left(\frac{\lfloor Nu \rfloor}{N} \right) = (Nu - \lfloor Nu \rfloor) \frac{1}{\sqrt{N}} x_{\mathbf{R}}(\beta(\lceil Nu \rceil)).$$

We note that $x_{\mathbf{R}}(\beta(\lceil Nu \rceil))$ as defined in equation (4.5.1) is a linear rank statistic with the following variance (confer Lemma E.1)

$$\begin{aligned}
&\operatorname{var} \left(\frac{2}{n} \sum_{l=1}^{\tilde{n}} c_{\beta(\lceil Nu \rceil)}(l) \sum_{j=1}^{n} x_j c_j(R_l) \right) \\
&= \frac{4}{n^2(\tilde{n}-1)} \sum_{k=1}^{\tilde{n}} \left(c_{\beta(\lceil Nu \rceil)}(k) - \frac{1}{\tilde{n}} \sum_{i=1}^{\tilde{n}} c_{\beta(\lceil Nu \rceil)}(i) \right)^2 \sum_{l=1}^{\tilde{n}} \left(\sum_{j=1}^{n} x_j c_j(l) - \frac{1}{\tilde{n}} \sum_{t=1}^{\tilde{n}} \sum_{i=1}^{n} x_i c_i(t) \right)^2 \\
&\ll \frac{1}{n} \sum_{i=1}^{\tilde{n}} (c_{\beta(\lceil Nu \rceil)}(i))^2 + \left(\frac{1}{\tilde{n}} \sum_{i=1}^{\tilde{n}} c_{\beta(\lceil Nu \rceil)}(i) \right)^2 = O(1),
\end{aligned}$$

where we used the fact that $\frac{2}{n\tilde{n}} \sum_{l=1}^{\tilde{n}} \left(\sum_{i=1}^{n} x_i c_i(l) - \frac{1}{\tilde{n}} \sum_{t=1}^{\tilde{n}} \sum_{j=1}^{n} x_j c_j(t) \right)^2 = 1$. The Markov inequality now gives the assertion. ■

Remark 4.5.6. If we do not center the Fourier coefficients, it still holds that the difference between $Z_n(u)$ and $Z_n(\lfloor Nu \rfloor/N)$ converges to 0, but possibly with a slower rate because of equation (4.5.4). By using the mean of the difference we obtain analogously $\mathrm{E}\left(\frac{1}{\sqrt{N}} x_{\mathbf{R}}\left(\beta(\lceil Nu \rceil)\right)\right) \ll \sqrt{\alpha(n)} \frac{1}{n^{3/2}} \sum_k \sum_j x_j c_j(k)$.

Now we are ready to prove Theorem 4.5.1.

Proof of Theorem 4.5.1. By the Cramer-Wold device it suffices to prove that for $\gamma_1, \dots, \gamma_k \neq 0$ we have $\sum_{i=1}^k \gamma_i (Z_n(u_i) - \mathrm{E}\, Z_n(u_i)) \xrightarrow{\mathcal{D}} \sum_{i=1}^k \gamma_i W(u_i)$ (w.l.o.g. $u_1 \neq 0$).

Again let $N = \frac{n}{\alpha(n)}$. First of all Lemma 4.5.1 shows that it suffices to prove $\sum_{i=1}^k \gamma_i \left(Z_n\left(\lfloor Nu_i \rfloor/N\right) - \mathrm{E}\, Z_n\left(\lfloor Nu_i \rfloor/N\right)\right) \xrightarrow{\mathcal{D}} \sum_{i=1}^k \gamma_i W(u_i)$.

Regarding the variance we first realize using Theorem 4.4.1

$$
\begin{aligned}
&\frac{2}{nN} \sum_{l=1}^{\tilde{n}} \left(\sum_{i=1}^k \gamma_i \sum_{s=1}^{\lfloor Nu_i \rfloor} c_{\beta(s)}(l) \right)^2 \\
&= 2 \sum_{i=1}^k \sum_{j=1}^k \gamma_i \gamma_j \frac{1}{nN} \sum_{s_1=1}^{\lfloor Nu_i \rfloor} \sum_{s_2=1}^{\lfloor Nu_j \rfloor} \sum_{l=1}^{\tilde{n}} c_{\beta(s_1)}(l)\, c_{\beta(s_2)}(l) \\
&= \sum_{i=1}^k \sum_{j=1}^k \gamma_i \gamma_j \frac{\tilde{n}}{n} \min\left(\frac{\lfloor Nu_i \rfloor}{N}, \frac{\lfloor Nu_j \rfloor}{N} \right) + O\left(\sum_{i=1}^k \sum_{j=1}^k |\gamma_i \gamma_j| \frac{\lfloor Nu_i \rfloor \lfloor Nu_j \rfloor}{Nn} \right) \\
&\to \sum_{i=1}^k \sum_{j=1}^k \gamma_i \gamma_j \min(u_i, u_j).
\end{aligned}
\tag{4.5.5}
$$

Moreover with Lemma 4.4.2 we get

$$
\frac{1}{\sqrt{N}n} \sum_{l=1}^{\tilde{n}} \sum_{i=1}^k \gamma_i \sum_{s=1}^{\lfloor Nu_i \rfloor} c_{\beta(s)}(l) \ll \frac{\log(n)}{\sqrt{N}} \to 0 \quad (n \to \infty). \tag{4.5.6}
$$

Now Lemma E.1 together with (4.5.5) and (4.5.6) yield as $n \to \infty$

$$
\operatorname{var}\left(\sum_{i=1}^k \gamma_i Z_n\left(\lfloor Nu_i \rfloor/N\right) \right) \to \sum_{i=1}^k \sum_{j=1}^k \gamma_i \gamma_j \min(u_i, u_j), \tag{4.5.7}
$$

where we used (4.5.2) again.

Since the Lindeberg condition is fulfilled as we will see below, this gives asymptotic normality with mean 0 and variance $\sum_{i=1}^k \sum_{j=1}^k \gamma_i \gamma_j \min(u_i, u_j) = \operatorname{var}\left(\sum_{i=1}^k \gamma_i W(u_i)\right)$.

Now we verify the Lindeberg condition for rank statistics, confer Theorem E.1.
We consider the linear rank statistic $S_n = \sum_{i=1}^n d_n(i) a_n(R_n(i))$ with $d_n(l) := \sqrt{\frac{2}{nN}} \sum_{i=1}^k \gamma_i \sum_{s=1}^{\lfloor Nu_i \rfloor} c_{\beta(s)}(l)$ and $a_n(l) := \sqrt{\frac{2}{n}} \sum_{j=1}^n x_j c_j(l)$. (4.5.5) and (4.5.6) show that

$$
\sum_{l=1}^{\tilde{n}} \left(d_n(l) - \bar{d}_n\right)^2 \to \sum_{i=1}^k \sum_{j=1}^k \gamma_i \gamma_j \min(u_i, u_j) > 0. \tag{4.5.8}
$$

Moreover

$$\frac{1}{nN}\max_{1\leqslant l\leqslant \tilde{n}}\left(\sum_{i=1}^{k}\gamma_i\sum_{s=1}^{\lfloor Nu_i\rfloor}c_{\beta(s)}(l)\right)^2 \ll \frac{N^2}{nN}=\frac{1}{\alpha(n)}\to 0,$$

hence because of (4.5.6)

$$\max_{1\leqslant l\leqslant \tilde{n}}\left(d_n(l)-\bar{d}_n\right)^2 = O\left(\frac{1}{\alpha(n)}\right)\to 0. \tag{4.5.9}$$

By (4.5.8) and conditions (4.5.2) it holds

$$\frac{1}{\tilde{n}}\sum_{i=1}^{\tilde{n}}(d_n(i)-\bar{d}_n)^2\sum_{j=1}^{\tilde{n}}(a_n(j)-\bar{a}_n)^2 \to \sum_{i=1}^{k}\sum_{j=1}^{k}\gamma_i\gamma_j\min(u_i,u_j)>0, \tag{4.5.10}$$

and (4.5.9) gives

$$\{|d_n(j)-\bar{d}_n||a_n(i)-\bar{a}_n|>\tau\}\subset\{|a_n(i)-\bar{a}_n|>\epsilon\sqrt{\alpha(n)}\},$$

where $\epsilon = C^{-1}\tau$ for an appropriate constant $C>0$. This together with (4.5.8) - (4.5.10) means that the Lindeberg condition

$$\lim_{n\to\infty}\frac{1}{\tilde{n}}\sum_{|\delta_{nij}|>\tau}\delta_{nij}^2 = 0 \qquad \text{for any } \tau>0,$$

where

$$\delta_{nij}=(d_n(j)-\bar{d}_n)(a_n(i)-\bar{a}_n)\left[\frac{1}{\tilde{n}}\sum_{j=1}^{\tilde{n}}(d_n(j)-\bar{d}_n)^2\sum_{i=1}^{\tilde{n}}(a_n(i)-\bar{a}_n)^2\right]^{-\frac{1}{2}},$$

can be reduced to

$$\frac{1}{\tilde{n}}\sum_{|a_n(i)-\bar{a}_n|>\epsilon\sqrt{\alpha(n)}}(a_n(i)-\bar{a}_n)^2\to 0 \qquad \text{for any } \epsilon>0.$$

This yields a Lyapunov-type condition as follows

$$\frac{1}{\alpha(n)^{\frac{\kappa-2}{2}}}\frac{1}{\tilde{n}}\sum_{i=1}^{\tilde{n}}|a_n(i)-\bar{a}_n|^{\kappa}\to 0 \qquad \text{for some } \kappa>2. \tag{4.5.11}$$

Since this condition is fulfilled by conditions (4.5.2), this completes the proof. ■

4.5.2. Tightness

We have seen that the finite-dimensional distributions are asymptotically normal, so we only need tightness to get convergence in $C[0,1]$. This is the assertion of the following theorem.

Theorem 4.5.3. *Under conditions (4.5.2) the sequence of processes* $\{Z_n(u)-\mathrm{E}\,Z_n(u):0\leqslant u\leqslant 1\}$ *is tight for* $\kappa<4$ *if* $\frac{\log^2(n)\alpha(n)}{n}=O(1)$ *and* $\frac{\log^2 n}{\alpha(n)^{\frac{\kappa-2}{4-\kappa}}}=o(1)$. *For* $\kappa=4$ *the above sequence is tight for any* $\alpha(n)$ *including* $\alpha(n)=1$.

Remark 4.5.7. If one does not center the Fourier coefficients the above theorem remains true. Equation (4.5.4) together with the linearity of the expectation gives (Z_n is linearly interpolated) that $\sup_{0\leqslant s\leqslant t\leqslant 1} |\mathrm{E}\, Z_n(t) - \mathrm{E}\, Z_n(s)| \to 0$, which shows that $\{Z_n(u) : 0 \leqslant u \leqslant 1\}$ is also tight. Note that Theorem 8.2 in Billingsley [9] states that tightness for a process $\{Y_n(u) : 0 \leqslant u \leqslant 1\}$ means that $\{Y_n(0)\}$ is tight and for all $\epsilon, \eta > 0$, there exists $0 < \delta < 1$ and n_0 such that

$$P\left(\sup_{|s-t|<\delta} |Y_n(s) - Y_n(t)| \geqslant \epsilon\right) \leqslant \eta \quad \text{for all } n \geqslant n_0.$$

Remark 4.5.8. As in Remark 4.5.3 for $\kappa = 4$ we can here substitute the condition

$$\frac{1}{n}\sum_{l=1}^{\tilde{n}}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n} x_j c_j(l) - \frac{1}{\tilde{n}}\sum_{k=1}^{\tilde{n}}\frac{1}{\sqrt{n}}\sum_{j=1}^{n} x_j c_j(k)\right)^4 = O(1)$$

by

$$\frac{1}{\alpha(n)n}\sum_{l=1}^{\tilde{n}}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n} x_j c_j(l) - \frac{1}{\tilde{n}}\sum_{k=1}^{\tilde{n}}\frac{1}{\sqrt{n}}\sum_{j=1}^{n} x_j c_j(k)\right)^4 = O(1),$$

if $\frac{\log^2(n)\alpha(n)}{n} = O(1)$, since then it holds in the proof for Theorem 4.5.3 below $z_{4d} \ll nN^3(t-u)^2$.
For $\kappa < 4$ we can also substitute the condition by

$$\frac{1}{n}\sum_{l=1}^{\tilde{n}}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n} x_j c_j(l) - \frac{1}{\tilde{n}}\sum_{k=1}^{\tilde{n}}\frac{1}{\sqrt{n}}\sum_{j=1}^{n} x_j c_j(k)\right|^{\kappa} = O\left(\alpha(n)^{\delta}\right)$$

for some $0 < \delta < 1$, if $(\log n)^2 = o\Big(\alpha(n)^{\frac{(\kappa-2)(1-\delta)}{4-\kappa} - \delta}\Big)$. The proof is analogous to the one for Theorem 4.5.3 below with $A_n := \alpha(n)^{\frac{1-\delta}{4-\kappa}}$.

Proof of Theorem 4.5.3. Let $N = \frac{n}{\alpha(n)}$ again. First note that $Z_n(t) - \mathrm{E}\, Z_n(t) - Z_n(u) + \mathrm{E}\, Z_n(u) = \frac{2}{n\sqrt{N}}\sum_{l=1}^{\tilde{n}} d_n(l)\ (a_n(R_l) - \bar{a}_n)$ is a linear rank statistic with

$$d_n(l) = \sum_{s=\lceil Nu\rceil+1}^{\lfloor Nt\rfloor} c_{\beta(s)}(l) + (Nt - \lfloor Nt\rfloor)\, c_{\beta(\lceil Nt\rceil)}(l) + (\lceil Nu\rceil - Nu)\, c_{\beta(\lceil Nu\rceil)}(l),$$

$$a_n(l) = \sum_{i=1}^{n} x_i\, c_i(l).$$

Here $\bar{a}_n := \frac{1}{\tilde{n}}\sum_{l=1}^{\tilde{n}} a_n(l)$ and an equivalent expression for $\bar{d}_n$.

Define

$$z_{2d} := \sum_{l=1}^{\tilde{n}}\left(d_n(l) - \bar{d}_n\right)^2 = \sum_{l=1}^{\tilde{n}} d_n^2(l) - \tilde{n}(\bar{d}_n)^2 \leqslant \sum_{l=1}^{\tilde{n}} d_n^2(l),$$

$$z_{4d} := \sum_{l=1}^{\tilde{n}}\left(d_n(l) - \bar{d}_n\right)^4,$$

z_{2a} analogously. For $\kappa < 4$ we need the following decomposition

$$
\begin{aligned}
&Z_n(t) - \mathrm{E}\, Z_n(t) - Z_n(u) + \mathrm{E}\, Z_n(u) \\
&= \frac{2}{n\sqrt{N}} \sum_{l=1}^{\tilde{n}} d_n(l) \Bigg[(a_n(R_l) - \bar{a}_n)\, 1_{\{\frac{1}{\sqrt{n}}|a_n(R_l)-\bar{a}_n|>A_n\}} \\
&\qquad\qquad\qquad\qquad - \frac{1}{n} \sum_{k=1}^{\tilde{n}} (a_n(k) - \bar{a}_n)\, 1_{\{\frac{1}{\sqrt{n}}|a_n(k)-\bar{a}_n|>A_n\}} \Bigg] \\
&\quad + \frac{2}{n\sqrt{N}} \sum_{l=1}^{\tilde{n}} d_n(l) \Bigg[(a_n(R_l) - \bar{a}_n)\, 1_{\{\frac{1}{\sqrt{n}}|a_n(R_l)-\bar{a}_n|\leqslant A_n\}} \\
&\qquad\qquad\qquad\qquad - \frac{1}{n} \sum_{k=1}^{n} (a_n(k) - \bar{a}_n)\, 1_{\{\frac{1}{\sqrt{n}}|a_n(k)-\bar{a}_n|\leqslant A_n\}} \Bigg] \\
&=: S^c(t,u) + S(t,u).
\end{aligned}
$$

Choose $A_n := \alpha(n)^{1/(4-\kappa)}$. We will first prove the tightness of $S^c(t,0)$. According to Billingsley [9], Theorem 8.3, it suffices to prove that for each $\epsilon, \eta > 0$, there exists $0 < \delta < 1$, n_0 with

$$\frac{1}{\delta} P\left(\sup_{u \leqslant t \leqslant u+\delta} |S^c(t,u)| \geqslant \epsilon \right) \leqslant \eta, \qquad n \geqslant n_0,$$

for all u. Because of the linear interpolation it suffices to show

$$\frac{1}{\delta} P\left(\max_{j \leqslant i \leqslant \max(N, j+1+\delta N)} |S^c(i/N, j/N)| \geqslant \epsilon \right) \leqslant \eta, \qquad n \geqslant n_0,$$

for all $1 \leqslant j \leqslant N$.

Theorem 4.4.1 gives

$$z_{2d} \leqslant \sum_{l=1}^{\tilde{n}} d_n^2(l) \ll nN(t-u) + N^2(t-u)^2 \ll nN(t-u). \tag{4.5.12}$$

Hence Lemma E.1, (4.5.2) and the fact that $\sum (b_i - \bar{b})^2 \leqslant \sum b_i^2$ show

$$
\begin{aligned}
\mathrm{E}\,|S^c(t,u)|^2 &\leqslant \frac{1}{nN} z_{2d} \frac{1}{\tilde{n}n} \sum_{k=1}^{\tilde{n}} (a_n(k) - \bar{a}_n)^2\, 1_{\{\frac{1}{\sqrt{n}}|a_n(k)-\bar{a}_n|>A_n\}} \\
&\ll (t-u) A_n^{-(\kappa-2)} \frac{1}{\tilde{n}} \sum_{k=1}^{\tilde{n}} \left| \frac{1}{\sqrt{n}} (a_n(k) - \bar{a}_n) \right|^{\kappa} \ll (t-u) A_n^{-(\kappa-2)}.
\end{aligned}
\tag{4.5.13}
$$

Now Theorem B.4 gives

$$
\begin{aligned}
&\frac{1}{\delta} P\left(\max_{j \leqslant i \leqslant \max(N, j+1+\delta N)} |S^c(i/N, j/N)| \geqslant \epsilon \right) \\
&\leqslant \frac{1}{\epsilon^2 \delta} \mathrm{E} \max_{j \leqslant i \leqslant \max(N, j+1+\delta N)} |S^c(i/N, j/N)|^2 \ll \frac{1}{\epsilon^2} \frac{(\log n)^2}{A_n^{\kappa-2}} \ll \eta, \quad n \geqslant n_0.
\end{aligned}
$$

To get the tightness of $\{S(t,0) : 0 \leqslant t \leqslant 1\}$ it suffices to show according to Billingsley [9], Theorem 12.3, that for some $C > 0$ and $\gamma > 1$ and all $0 \leqslant u \leqslant t \leqslant 1$

$$\mathrm{E}[S(t,u)]^4 \leqslant C(t-u)^\gamma.$$

This together with the tightness of $S^c(t,0)$ gives then the tightness of $Z_n(t) - \mathrm{E}\, Z_n(t)$.

Now Lemma E.1 yields

$$\mathrm{E}[S(t,u)]^4 \ll \frac{1}{N^2 n^6} z_{2\tilde{a}}^2 z_{2d}^2 + \frac{1}{N^2 n^5} z_{4\tilde{a}} z_{4d} + \frac{1}{N^2 n^6} z_{4\tilde{a}} z_{2d}^2 + \frac{1}{N^2 n^6} z_{2\tilde{a}}^2 z_{4d}, \tag{4.5.14}$$

where $\tilde{a}_n(l) := (a_n(l) - \bar{a}_n)\, 1_{\{\frac{1}{\sqrt{n}}|a_n(l)-\bar{a}_n| \leqslant A_n\}}$, $z_{2\tilde{a}}, z_{4\tilde{a}}$ analogous to above.

First note that $z_{2\tilde{a}} \leqslant z_{2a}$. By assumption (4.5.2) we have

$$z_{2a} = O(n^2). \tag{4.5.15}$$

Let $s_{4\tilde{a}} = \sum_{l=1}^{\tilde{n}} (a_n(l) - \bar{a}_n)^4\, 1_{\{\frac{1}{\sqrt{n}}|a_n(l)-\bar{a}_n| \leqslant A_n\}}$, then

$$\begin{aligned} z_{4\tilde{a}} &\ll s_{4\tilde{a}} + n^3 \left(\frac{1}{n} \sum_{l=1}^{\tilde{n}} \frac{1}{\sqrt{n}} \tilde{a}_n(l) \right)^4 \ll s_{4\tilde{a}} + n^3 \left(1 + \frac{1}{n^2} z_{2a} \right)^4 \\ &\ll s_{4\tilde{a}} + n^3 \ll \alpha(n) n^3, \end{aligned} \tag{4.5.16}$$

since

$$s_{4\tilde{a}} \ll A_n^{4-\kappa} n^2 \sum_{l=1}^{\tilde{n}} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} x_j c_j(l) - \frac{1}{\tilde{n}} \sum_{k=1}^{\tilde{n}} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i c_i(k) \right|^\kappa \ll \alpha(n) n^3.$$

Next Lemma 4.4.2 gives

$$\bar{d}_n \ll \sum_{s=1}^{n} \max\left(\frac{1}{s}, \frac{1}{n-s} \right) \ll \log n,$$

and anyway

$$\bar{d}_n \ll N(t-u),$$

which together means

$$n(\bar{d}_n)^4 \ll nN^2 (t-u)^2 \log^2(n) \ll nN^3 (t-u)^2.$$

Moreover Lemma 4.4.1 shows

$$\sum_{l=1}^{n} d_n^4(l) \ll nN^3 (t-u)^3 + N^4 (t-u)^4 \ll nN^3 (t-u)^2.$$

This finally gives

$$z_{4d} \ll nN^3 (t-u)^2. \tag{4.5.17}$$

Putting together equations (4.5.12) and (4.5.14) to (4.5.17), we realize

$$\mathrm{E}[S(t,u)]^4 \leqslant C(t-u)^2, \tag{4.5.18}$$

which gives the assertion.

For $\kappa = 4$ the argument is as for $S(t,u)$. Note that then $z_{4a} = O(n^3)$, also we replace equation (4.5.17) by $z_{4d} \ll n^2 N^2 (t-u)^2$. This way we do not need the assumption that $\frac{\log^2(n)\alpha(n)}{n} = O(1)$. ■

4.5.3. Rank Asymptotics

Now we can use the results of the previous two subsections to derive the main theorem of this section. It shows that the process $\{\widetilde{Z}_n(\cdot)\}$ converges in $C[0,1]$ to a Wiener process. In a corollary we deduce then the asymptotics for the frequency rank statistics belonging to the statistics we investigate. It is the main tool in the proof of the validity of the frequency bootstrap test.

Theorem 4.5.4. *Under the conditions of Theorems 4.5.1 and 4.5.3, it holds*

$$\{\widetilde{Z}_n(u) : 0 \leqslant u \leqslant 1\} \overset{C[0,1]}{\longrightarrow} \{W(u) : 0 \leqslant u \leqslant 1\},$$

where $\widetilde{Z}_n(u) := Z_n(u) - \operatorname{E} Z_n(u)$ *and* $\{W(u) : 0 \leqslant u \leqslant 1\}$ *is a standard Wiener process.*

Proof. This follows immediately from Billingsley [9], Theorem 8.1, in regard of Theorems 4.5.1 and 4.5.3. ∎

From the above theorem we can now derive the asymptotics for the frequency rank statistics we are interested in.

Corollary 4.5.1. *Let the conditions of Theorems 4.5.1 and 4.5.3 be fulfilled and* $\widetilde{Z}_n(t) := Z_n(t) - \operatorname{E} Z_n(t)$.

a) For all $\epsilon > 0$ *we get*

$$T_n^{(1f)}(\mathbf{x}) := \sup_{\epsilon \leqslant t \leqslant 1-\epsilon} \sqrt{\frac{1}{t(1-t)}} |\widetilde{Z}_n(t) - t\widetilde{Z}_n(1)| \overset{\mathcal{D}}{\longrightarrow} \sup_{\epsilon \leqslant t \leqslant 1-\epsilon} \sqrt{\frac{1}{t(1-t)}} |B(t)|.$$

b) It holds for any $\epsilon > 0$

$$T_n^{(2f)}(\mathbf{x}) := \sup_{\epsilon \leqslant t \leqslant 1} |\widetilde{Z}_n(t) - \widetilde{Z}_n(t-\epsilon) - \epsilon\widetilde{Z}_n(1)| \overset{\mathcal{D}}{\longrightarrow} \sup_{\epsilon \leqslant t \leqslant 1} |B(t) - B(t-\epsilon)|.$$

c) If $q \in FC_0^1$ *and*

(i)

$$\int_0^1 \frac{1}{q^2(t)}\, dt < \infty \qquad \textit{or}$$

(ii) $\kappa = 4$,

$$\int_0^1 \frac{t(1-t)}{q^4(t)}\, dt < \infty$$

and $\frac{t}{q^4(t)}$ *is non-increasing in a neighborhood of* 0 *and* $\frac{1-t}{q^4(t)}$ *is non-decreasing in a neighborhood of* 1,

then it holds

$$T_n^{(3f)}(\mathbf{x}, q) := \max_{1 \leqslant k < N} \frac{1}{q\left(\frac{k}{N}\right)} \left| \widetilde{Z}_n\left(\frac{k}{N}\right) - \frac{k}{N}\widetilde{Z}_n(1) \right| \overset{\mathcal{D}}{\longrightarrow} \sup_{0 \leqslant t \leqslant 1} \frac{|B(t)|}{q(t)},$$

where $N = \frac{n}{\alpha(n)}$.

d) For $\int_0^1 \frac{(t(1-t))^\kappa}{r(t)}\,dt < \infty$ for some $0 \leqslant \kappa < 1$, it holds

$$T_n^{(4f)}(\mathbf{x}, r) := \int_0^1 \frac{1}{r(t)} |\widetilde{Z}_n(t) - t\widetilde{Z}_n(1)|^2\,dt \xrightarrow{\mathcal{D}} \int_0^1 \frac{B^2(t)}{r(t)}\,dt.$$

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

Remark 4.5.9. If we want to keep the middle term of the Fourier coefficients, i.e. choose $\omega(n/2) = \sum_{j=1}^n \widetilde{X}(j)(-1)^j$, the bootstrap sample $X_{\mathbf{R}}(s)$ has an additional term, namely $\frac{1}{n}(-1)^s \sum_{j=1}^n \widetilde{X}(j)(-1)^j$. We need to make sure that it does not change the convergence of the statistic.
Lemma 4.3.1 gives conditions that are sufficient for a) respectively b). Yet for general weight functions $q(\cdot)$ and $r(\cdot)$ we need in the case of c) respectively d) a somewhat stronger condition, namely

$$\frac{1}{n} \sup_{0 \leqslant u \leqslant 1} \left| \sum_{s=1}^{\lfloor uN \rfloor} (-1)^{\beta(s)} \sum_{j=1}^n x_j (-1)^j \right| = O(1). \tag{4.5.19}$$

This is, however, fulfilled in many interesting cases. For c) Lemma F.3 a) gives – similarly to the proof of the corollary below – that the part of the statistic involving the middle term converges to 0 under (4.5.19), more precisely

$$\max_{1 \leqslant k < N} \frac{1}{n\sqrt{N} q\left(\frac{k}{N}\right)} \left| \sum_{s=1}^{k} (-1)^{\beta(s)} \sum_{j=1}^n x_j (-1)^j - \frac{k}{N} \sum_{t=1}^{N} (-1)^{\beta(t)} \sum_{j=1}^n x_j (-1)^j \right| = o(1).$$

Under d) analogous arguments as in the proof of the corollary below give

$$\begin{aligned}
&\int_0^1 \frac{1}{r(t)} (\widetilde{Y}(t) - t\widetilde{Y}(1))^2\,dt = \int_{1/N}^{1-1/N} \frac{1}{r(t)} (\widetilde{Y}(t) - t\widetilde{Y}(1))^2\,dt + o(1) \\
&= o(1) \int_0^1 \frac{(t(1-t))^\kappa}{r(t)}\,dt + o(1) = o(1),
\end{aligned}$$

where $\widetilde{Y}\left(\frac{k}{N}\right) = \frac{1}{n\sqrt{N}} \sum_{s=1}^k (-1)^{\beta(s)} \sum_{j=1}^n x_j (-1)^j$, $k = 1, \ldots, N$, and linearly interpolated in between.

It is also possible to use the uncentered Fourier coefficients. We then need under c) additionally as $N \to \infty$

$$\begin{aligned}
&\frac{\log^2(n)}{\max(q(1/N), q(1-1/N))\sqrt{N}} \max |x_j| \to 0 \qquad \text{or} \\
&\frac{\log n \sqrt{\alpha(n)}}{\max(q(1/N), q(1-1/N))} \frac{1}{n^{3/2}} \sum_{l=1}^{\tilde{n}} \sum_{j=1}^n x_j c_j(l) \to 0.
\end{aligned}$$

The above assumptions are somewhat stronger than the ones in Remark 4.5.1. Analogous to equation (4.5.4) it then holds

$$\max_{1 \leqslant k < N} \frac{1}{q\,(k/N)} |\mathrm{E}(Z_n(k/N))| \to 0.$$

Note that there exists an $\eta > 0$, such that $\sup_{\eta \leqslant \frac{k}{N} \leqslant 1-\eta} \frac{1}{q\left(\frac{k}{N}\right)} \leqslant C < \infty$ and for $\frac{k}{N} \leqslant \eta$ respectively $\frac{k}{N} \geqslant 1-\eta$ it holds $\frac{1}{q\left(\frac{k}{N}\right)} \leqslant \frac{1}{q\left(\frac{1}{N}\right)}$.
Under d) we need the above condition with $q(t)$ replaced by $(t(1-t))^{\kappa/2}$. Similarly to above we then get

$$\int_0^1 \frac{1}{r(t)} \operatorname{E}\left(\widetilde{Z}_n(t) - t\widetilde{Z}_n(1)\right)^2 = o(1).$$

Proof of Corollary 4.5.1. The results for a) and b) follow immediately from Theorem 4.5.4, since we can deduce from the Portmanteau theorem (cf. e.g. Billingsley [9], Theorem 2.1)

$$f(Z_n(\cdot)) \xrightarrow{\mathcal{D}} f(W(\cdot)),$$

where $\{W(t) : 0 \leqslant t \leqslant 1\}$ is a Wiener process and $f : C[0,1] \to \mathbb{R}$ any continuous function. Note that $C[0,1]$ is provided with the sup-norm. Thus both of the above transformations are continuous, which gives the assertion.

For the proof of c) note that in the same way we get for any $\eta > 0$

$$\max_{N\eta \leqslant k \leqslant N - N\eta} \frac{1}{q\left(\frac{k}{N}\right)} \left| \widetilde{Z}_n\left(\frac{k}{N}\right) - \frac{k}{N}\widetilde{Z}_n(1)\right| \xrightarrow{\mathcal{D}} \sup_{\eta \leqslant t \leqslant 1-\eta} \frac{|B(t)|}{q(t)}, \tag{4.5.20}$$

because $\inf_{\eta \leqslant t \leqslant 1-\eta} q(t) > 0$.

Moreover Lemma F.2 shows that $I^*(q,c) < \infty$ for all $c > 0$ so that Lemma F.3 b) gives

$$\lim_{\eta \to 0} \sup_{0<t<\eta, 1-\eta<t<1} \frac{|B(t)|}{q(t)} = 0 \qquad a.s.,$$

which implies for all $x > 0$

$$P\left(\sup_{0<t<\eta, 1-\eta<t<1} \frac{|B(t)|}{q(t)} > x\right) \to 0 \qquad \text{as } \eta \to 0. \tag{4.5.21}$$

The proof for tightness (Theorem 4.5.3, equation (4.5.18)) has shown

$$N^2 \operatorname{E}|S(j/N, i/N)|^4 \leqslant C(j-i+1)^2.$$

In a neighborhood of 0, i.e. for η small enough, $q(\cdot)$ is non-decreasing. Now the Markov inequality together with Theorem B.3 gives for all $x > 0$ and all $\epsilon > 0$

$$P\left(\max_{1 \leqslant k < \eta N} \frac{1}{q\left(\frac{k}{N}\right)} |S(k/N, 0)| > x\right) \ll \frac{1}{x^4} \frac{1}{N^2} \sum_{k=1}^{\eta N} \frac{k}{q^4\left(\frac{k}{N}\right)} \leqslant \epsilon \tag{4.5.22}$$

for $\eta \leqslant \eta_0$ and all n. Under (i) Lemma F.2 gives that $I^*(q,c) < \infty$ for all $c > 0$ so that Lemma F.3 a) yields

$$\frac{1}{N^2} \sum_{k=1}^{\eta N} \frac{k}{q^4\left(\frac{k}{N}\right)} \leqslant \sup_{0 \leqslant t \leqslant \eta} \frac{t}{q^2(t)} \int_0^\eta \frac{1}{q^2(t)}\, dt \to 0 \quad \text{as } \eta \to 0. \tag{4.5.23}$$

The proof of tightness (Theorem 4.5.3, equation (4.5.13)) and Theorem B.4 show that

$$P\left(\max_{1\leqslant k<\eta N}\frac{1}{q\left(\frac{k}{N}\right)}|S^c(k/N,0)|>x\right)\ll\frac{1}{x^2}\frac{\log^2(n)}{\alpha(n)^{\frac{\kappa-2}{4-\kappa}}}\int_0^\eta\frac{1}{q^2(t)}\,dt\to 0, \tag{4.5.24}$$

uniformly in n as $\eta\to 0$.

For $\kappa=4$ under c)(i) the same arguments as in (4.5.22) and (4.5.23) hold, if we replace $S(t,u)$ by $Z_n(t)-\mathrm{E}\,Z_n(t)-Z_n(u)+\mathrm{E}\,Z_n(u)$. Under c)(ii) it holds

$$\frac{1}{N^2}\sum_{k=1}^{\eta N}\frac{k}{q^4\left(\frac{k}{N}\right)}\leqslant\int_0^\eta\frac{t}{q^4(t)}\,dt\to 0\quad\text{as }\eta\to 0.$$

Equation (4.5.7) gives $\mathrm{E}(Z_n(1)-\mathrm{E}\,Z_n(1))^2=O(1)$ and by Lemma F.3 a)

$$\max_{1\leqslant k\leqslant\eta N}\frac{\sqrt{k/N}}{q(k/N)}\to 0$$

as $\eta\to 0$ uniformly in n, thus the Chebyshev inequality gives for all $x>0,\delta>0$

$$P\left(\max_{1\leqslant k\leqslant\eta N}\frac{k/N}{q(k/N)}|Z_n(1)-\mathrm{E}\,Z_n(1)|>x\right)\leqslant\delta \tag{4.5.25}$$

for all n, for all $\eta\leqslant\eta_0(x,\delta)$.

Putting together equations (4.5.22), (4.5.24) and (4.5.25) we finally get for all $x>0$

$$P\left(\max_{1\leqslant k<\eta N}\frac{1}{q\left(\frac{k}{N}\right)}\left|Z_n\left(\frac{k}{N}\right)-\mathrm{E}\,Z_n\left(\frac{k}{N}\right)-\frac{k}{N}(Z_n(1)-\mathrm{E}\,Z_n(1))\right|>x\right)\leqslant\epsilon \tag{4.5.26}$$

for $\eta\leqslant\eta_0$ and all n.

Using the same arguments we deduce

$$P\left(\max_{1-\eta\leqslant k<1}\frac{1}{q\left(\frac{k}{N}\right)}\left|\widetilde{Z}_n\left(\frac{k}{N}\right)-\frac{k}{N}\widetilde{Z}_n(1)\right|>x\right)\leqslant\epsilon \tag{4.5.27}$$

for $\eta\leqslant\eta_0$ and all n.

Moreover $|P(\max(X,Y)\leqslant x)-P(X\leqslant x)|\leqslant P(Y>x)$ so that equations (4.5.20), (4.5.21), (4.5.26) and (4.5.27) give for all $x>0$ as $n\to\infty$

$$P\left(\max_{1\leqslant k<N}\frac{1}{q\left(\frac{k}{N}\right)}\left|\widetilde{Z}_n\left(\frac{k}{N}\right)-\frac{k}{N}\widetilde{Z}_n(1)\right|\leqslant x\right)\to P\left(\sup_{0\leqslant t\leqslant 1}\frac{|B(t)|}{q(t)}\,dt\leqslant x\right)$$

by first choosing η small enough and then n big enough in dependence of η.

For the proof of d) first note that $\int_0^1\frac{1}{(t(1-t))^\kappa}\,dt<\infty$, hence by c)

$$\max_{1\leqslant k\leqslant N}\frac{|\widetilde{Z}_n(k/N)-k/N\widetilde{Z}_n(1)|}{(k/N(1-k/N))^{\kappa/2}}=O_P(1). \tag{4.5.28}$$

Since $\widetilde{Z}_n(t) - t\widetilde{Z}_n(1) = Nt\left(\widetilde{Z}_n\left(\frac{1}{N}\right) - \frac{1}{N}\widetilde{Z}_n(1)\right)$ for $0 \leqslant t \leqslant \frac{1}{N}$, we get as $N \to \infty$

$$\int_0^{\frac{1}{N}} \frac{\left(\widetilde{Z}_n(t) - t\widetilde{Z}_n(1)\right)^2}{r(t)}\, dt \leqslant \frac{|\widetilde{Z}_n(1/N) - 1/N\widetilde{Z}_n(1)|^2}{(1/N)^\kappa} \int_0^{\frac{1}{N}} \frac{t^\kappa}{r(t)}\, dt = o_P(1).$$

Since an analogous expression for $\int_{1-\frac{1}{N}}^1 \frac{\left(\widetilde{Z}_n(t)-t\widetilde{Z}_n(1)\right)^2}{r(t)}\, dt$ holds, it suffices to consider $\int_{\frac{1}{N}}^{1-\frac{1}{N}} \frac{1}{r(t)}|\widetilde{Z}_n(t) - t\widetilde{Z}_n(1)|^2\, dt$.

Due to Theorem 4.5.4 it holds for any $\eta > 0$

$$C_n(\eta) := \int_\eta^{1-\eta} \frac{1}{r(t)}|\widetilde{Z}_n(t) - t\widetilde{Z}_n(1)|^2\, dt \xrightarrow{\mathcal{D}} \int_\eta^{1-\eta} \frac{B^2(t)}{r(t)}\, dt =: C(\eta), \tag{4.5.29}$$

because $\int_\eta^{1-\eta} \frac{1}{r(t)}\, dt < \infty$.

Moreover

$$\begin{aligned} A_n(\eta) &:= \int_{\{(1/N,\eta)\cup(1-\eta,1-1/N)\}} \frac{1}{r(t)}|\widetilde{Z}_n(t) - t\widetilde{Z}_n(1)|^2\, dt \\ &\leqslant \left(\sup_{1/N\leqslant t\leqslant 1-1/N} \frac{|\widetilde{Z}_n(t) - t\widetilde{Z}_n(1)|}{(t(1-t))^{\kappa/2}}\right)^2 \int_{\{(0,\eta)\cup(1-\eta,1)\}} \frac{(t(1-t))^\kappa}{r(t)}\, dt = o_P(1) \end{aligned} \tag{4.5.30}$$

as $\eta \to 0$ uniformly in n, since

$$\sup_{\frac{1}{N}\leqslant t\leqslant 1-\frac{1}{N}} \frac{|\widetilde{Z}_n(t) - t\widetilde{Z}_n(1)|}{(t(1-t))^{\kappa/2}} = O_P(1)$$

because of (4.5.28) and the linear interpolation of $\widetilde{Z}_n(t)$ between $\frac{k}{N}$ and $\frac{k+1}{N}$.

Furthermore Lemma F.3 c) gives

$$\lim_{\eta\to 0} C(\eta) = \lim_{\eta\to 0} \int_\eta^{1-\eta} \frac{B^2(t)}{r(t)}\, dt = \int_0^1 \frac{B^2(t)}{r(t)}\, dt =: C \qquad a.s., \tag{4.5.31}$$

since $\int_0^1 \frac{t(1-t)}{r(t)}\, dt < \infty$.

Putting together the above results we derive now the assertion. Let $\epsilon > 0$, $x \geqslant 0$. Choose $\delta > 0$ such that $|P(C \leqslant x + 2\delta) - P(C \leqslant x)| \leqslant \epsilon$, η such that $P(|A_n(\eta)| \geqslant \delta) \leqslant \epsilon$, for all n, as well as $P(|C(\eta) - C| \geqslant \delta) \leqslant \epsilon$. Finally choose n_1, such that $|P(C_n(\eta) \leqslant x + \delta) - P(C(\eta) \leqslant x + \delta)| \leqslant \epsilon$, for all $n \geqslant n_1$. Then it holds for all $n \geqslant n_1$

$$\begin{aligned} &P\left(\int_{\frac{1}{N}}^{1-\frac{1}{N}} \frac{1}{r(t)}|\widetilde{Z}_n(t) - t\widetilde{Z}_n(1)|^2\, dt \leqslant x\right) \\ &= P\left(A_n(\eta) + C_n(\eta) \leqslant x, |A_n(\eta)| < \delta\right) + P\left(A_n(\eta) + C_n(\eta) \leqslant x, |A_n(\eta)| \geqslant \delta\right) \\ &\leqslant P(C_n(\eta) \leqslant x + \delta) + \epsilon \\ &\leqslant P(C(\eta) \leqslant x + \delta) + 2\epsilon \\ &= P(C(\eta) \leqslant x + \delta, |C(\eta) - C| < \delta) + P(C(\eta) \leqslant x + \delta, |C(\eta) - C| \geqslant \delta) + 2\epsilon \\ &\leqslant P(C \leqslant x + 2\delta) + 3\epsilon \end{aligned}$$

$$\leqslant P\left(\int_0^1 \frac{B^2(t)}{r(t)}\,dt \leqslant x\right) + 4\epsilon.$$

This shows that $\limsup_{n\to\infty} P(\int_{\frac{1}{N}}^{1-\frac{1}{N}} \frac{1}{r(t)}|\widetilde{Z}_n(t) - t\widetilde{Z}_n(1)|^2\,dt \leqslant x) \leqslant P(\int_0^1 \frac{B^2(t)}{r(t)}\,dt \leqslant x)$.

An analogous argument (now choosing δ such that $|P(C \leqslant x - 2\delta) - P(C \leqslant x)| \leqslant \epsilon$ and n_1 such that $|P(C_n(\eta) \leqslant x - \delta) - P(C(\eta) \leqslant x - \delta)| \leqslant \epsilon$) gives for all $n \geqslant n_1$

$$P\left(\int_{\frac{1}{N}}^{1-\frac{1}{N}} \frac{1}{r(t)}|\widetilde{Z}_n(t) - t\widetilde{Z}_n(1)|^2\,dt \leqslant x\right) \geqslant P\left(\int_0^1 \frac{B^2(t)}{r(t)}\,dt \leqslant x\right) - 2\epsilon,$$

hence $\liminf_{n\to\infty} P(\int_{\frac{1}{N}}^{1-\frac{1}{N}} \frac{1}{r(t)}|\widetilde{Z}_n(t) - t\widetilde{Z}_n(1)|^2\,dt \leqslant x) \geqslant P(\int_0^1 \frac{B^2(t)}{r(t)}\,dt \leqslant x)$. This completes the proof. ■

4.6. Limit Distributions of the Frequency Permutation Statistics

In this section we finally prove that the critical values obtained by the frequency permutation method as proposed in Section 4.2 are asymptotically correct. Precisely we prove that the quantiles we obtain from the bootstrap are asymptotically the same as the ones corresponding to the distribution of the original statistic under the null hypothesis. Thus, even if our observations follow an alternative we get a good approximation of the critical values corresponding to the null distribution.

In the previous chapters it was always possible to obtain such approximations by bootstrapping the raw observation data. As a contrast for the frequency method we first need to estimate the underlying linear sequence and then bootstrap the Fourier coefficients of the estimated linear process. Note that the other approach will in general fail in this context (confer Remark 4.6.2).

To estimate the underlying linear process we need estimators for the change-point, the mean of the observations before the change and the mean after the change. In a first subsection we prove that the frequency permutation method works if we have observed a linear process. This is essentially what we had if the estimators were always correct.
We then show in a second step that the estimators are good enough in the sense that using them will not change the limit behavior. This shows that the frequency permutation method as described in Section 4.2 works in this setting. The proof can be found in Subsection 4.6.3.
In order to be able to prove this important theorem we first need to investigate some properties of the change estimators we use, which can be found in Subsection 4.6.2.

4.6.1. Permutation Statistics of a Linear Sequence

Now we investigate the sequence corresponding to $\widetilde{X}(\cdot)$ if we used the correct values for the change-point and the mean before and after the change instead of estimators. This is then just the underlying linear process

$$e(i) := (X(i) - \mu)1_{[1,m]}(i) + (X(i) - d - \mu)1_{[m+1,n]}(i),$$

where m is the real change point ($m = n$ under the null hypothesis), μ the mean of the sequence before the change and $d + \mu$ the mean after the change.

Theorem 4.6.1. *Let (3.3.2) and (3.3.3) be fulfilled,* $\alpha(n) \to \infty$ *but* $\frac{\alpha(n)\log^2(n)}{n} = O(1)$. *For* $\nu < 4$ *additionally assume* $\log^2(n) = o\left(\alpha(n)^{(\nu-2)(1-\delta)/(4-\nu)-\delta}\right)$ *for some* $0 < \delta < 1$. *Then it holds*

$$P\left(f[(\widetilde{Z}_n^{\mathbf{e}}(id, \mathbf{R}) - id\, \widetilde{Z}_n^{\mathbf{e}}(1, \mathbf{R}))/\widehat{\sigma}_n] \le x \,\middle|\, X(1), \ldots, X(n)\right) \xrightarrow{P} P\left(f(B(\cdot)) \leqslant x\right).$$

for all continuous $f : C[0,1] \to \mathbb{R}$ *and for all* $x \in \mathbb{R}$. *Here*

$$\widehat{\sigma}_n^2 := \frac{2}{n\tilde{n}} \sum_{l=1}^{\tilde{n}} \left(\sum_{i=1}^{n} e(i) c_i(l) - \frac{1}{\tilde{n}} \sum_{k=1}^{\tilde{n}} \sum_{j=1}^{n} e(j) c_j(k) \right)^2.$$

and

$$\widetilde{Z}_n^{\mathbf{e}}(u, \mathbf{R}) = \sqrt{\frac{\alpha(n)}{n}} \sum_{s \leqslant \frac{n}{\alpha(n)} u} e_{\mathbf{R}}(\beta(s))$$

for $u = \frac{\alpha(n)}{n}, \frac{2\alpha(n)}{n}, \ldots, 1$; $\widetilde{Z}_n^{\mathbf{e}}(0, R) = 0$. $\widetilde{Z}_n^{\mathbf{e}}(t, R)$ *is linearly interpolated between* $(i-1)/n$ *and* i/n *for* $i = 1, \ldots, n$, *where* $e_{\mathbf{R}}(\cdot)$ *is as in equation (4.3.1) respectively (4.3.3) with* $\widetilde{X}(\cdot)$ *replaced by* $e(\cdot)$.

If additionally the error-sequence $\{\epsilon(i) : -\infty < i < \infty\}$ *of the linear process is i.i.d. normally distributed, the above assertion holds in an almost sure sense.*

Remark 4.6.1. It is possible to keep the middle term of the Fourier coefficients instead of setting it equal to 0 if e.g. $\sup_{0 \leqslant u \leqslant 1} \left| \sum_{s=1}^{\lfloor Nu \rfloor} (-1)^{\beta(s)} \right| = O(1)$ even for the corresponding result to Corollary 4.5.1. Then (4.6.5) shows that condition (4.5.19) in Remark 4.5.9 is fulfilled.
If we do not center the Fourier coefficients, Remark 4.5.9 gives a sufficient additional condition. For appropriate $\alpha(\cdot)$ and appropriate weight functions this is fulfilled because of equation (4.6.8). Alternatively one can confer Remark B.1 which states $\frac{1}{n} \sum_{j=1}^{n} |e(j)|^{\mu} = O(1)$ *a.s.* for all $2 < \mu < \nu$.

Remark 4.6.2. In contrast to the block bootstrap the below proof will not work in this setting if we work with the observed random variables as they are under alternatives. Essentially, this means instead of proving the result for $e(i)$ we had to prove it for $e(i) + d\,1_{\{i > m\}}$. This is not possible in the below way, because e.g. for $l = 1$ and a change at $n/4$

$$d^4 \frac{1}{n^3} \left(\sum_{j > n/4} c_j(1) \right)^4 \approx c n d^4$$

for some constant $c > 0$, which is not bounded. The mixed terms also will in general not be able to compensate for that.

Proof of Theorem 4.6.1. In view of Theorem 4.5.4 and the Portmanteau theorem (cf. e.g. Billingsley [9], Theorem 2.1) it suffices to prove conditions (4.5.2) in the $O_P(1)$–sense for

$$x_i := e(i)\left(\frac{2}{n\tilde{n}}\sum_{l=1}^{\tilde{n}}\left(\sum_{i=1}^{n} e(i)c_i(l) - \frac{1}{\tilde{n}}\sum_{k=1}^{\tilde{n}}\sum_{j=1}^{n} e(j)c_j(k)\right)^2\right)^{-\frac{1}{2}}.$$

This implies $= o_P(g(n))$ for some $g(n) \to \infty$ as in Remarks 4.5.3 respectively 4.5.8. Thus we can use the subsequence principle to obtain the assertions by an application of Theorem 4.5.4.
For normal errors we prove the above assertion in an almost sure sense.

First Theorem C.1 yields

$$\frac{1}{n}\sum_{j=1}^{n} e(j) = o(1) \quad a.s. \tag{4.6.1}$$

In view of Remark 4.5.2 it therefore suffices to prove

$$\frac{1}{n}\sum_{j=1}^{n} e(j)^2 \xrightarrow{P} c > 0, \tag{4.6.2}$$

$$\frac{1}{n^{3/2}}\sum_{l=1}^{\tilde{n}}\sum_{j=1}^{n} e(j)c_j(l) = o_P(1), \tag{4.6.3}$$

$$\frac{1}{n}\sum_{l=1}^{\tilde{n}}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n} e(j)c_j(l)\right|^{\kappa} = O_P(1) \tag{4.6.4}$$

for $\kappa = \min(4, \nu)$. Additionally, we need for n even

$$\frac{2}{n}\sum_{j=1}^{n/2} e(2j) = o_P(1), \qquad \frac{2}{n}\sum_{j=1}^{n/2} e(2j-1) = o_P(1). \tag{4.6.5}$$

For normally distributed errors we need the above asymptotics in an $a.s.$-sense. For (4.6.3) this holds because of Corollary B.1 and the argument given in Remark 4.5.1.

Concerning conditions (4.6.5) note that

$$e(2l) = \sum_{s\geqslant 0} w_s\epsilon(2l-s) = \sum_{s\geqslant 0}[w_{2s}\epsilon(2l-2s) + w_{2s+1}\epsilon(2l-2s-1)] =: e_1(l) + e_2(l).$$

Both $e_1(\cdot)$ as well as $e_2(\cdot)$ fulfill now the conditions of Theorem C.1, which gives the assertion in an $a.s.$-sense. The second condition follows in the same way.

Condition (4.6.2) follows immediately from Theorem C.3 in an $a.s.$-sense for both normal as well as general innovations.

The proof of conditions (4.6.3) respectively (4.6.4) for both cases goes along the lines of the proof of Theorem 10.3.1 in Brockwell and Davis [13]. It is based on the following decomposition

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{j=1}^{n} e(j)\exp(-2\pi ijl/n)\\
&= \frac{1}{\sqrt{n}}\sum_{k\geqslant 0} w_k \exp(-2\pi ilk/n)\left(\sum_{j=1-k}^{n-k}\epsilon(j)\exp(-2\pi ilj/n)\right)\\
&= \Psi(l)\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\epsilon(j)\exp(-2\pi ilj/n)+Y_n(l),
\end{aligned}
$$

where

$$
\begin{aligned}
\Psi(l) &:= \sum_{k\geqslant 0} w_k \exp(-2\pi ilk/n),\\
Y_n(l) &:= \frac{1}{\sqrt{n}}\sum_{k\geqslant 0} w_k \exp(-2\pi ikl/n)U_{n,j}(k),\\
U_{n,l}(k) &:= \sum_{j=1-k}^{n-k}\epsilon(j)\exp(-2\pi ilj/n)-\sum_{j=1}^{n}\epsilon(j)\exp(-2\pi ilj/n).
\end{aligned}
$$

This now gives

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{j=1}^{n} e(j)\cos(2\pi jl/n)\\
&\quad= \mathrm{Re}(\Psi(l))\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\epsilon(j)\cos(2\pi jl/n)-\mathrm{Im}(\Psi(l))\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\epsilon(j)\sin(-2\pi jl/n)\\
&\qquad+\mathrm{Re}(Y_n(l)),\\
&\frac{1}{\sqrt{n}}\sum_{j=1}^{n} e(j)\sin(-2\pi jl/n)\\
&\quad= \mathrm{Re}(\Psi(l))\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\epsilon(j)\sin(-2\pi jl/n)+\mathrm{Im}(\Psi(l))\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\epsilon(j)\cos(2\pi jl/n)\\
&\qquad+\mathrm{Im}(Y_n(l)).
\end{aligned}
$$

First we look at condition (4.6.3). It holds

$$
\begin{aligned}
&\frac{1}{n}\sum_{l=1}^{\tilde{n}}\frac{1}{\sqrt{n}}\sum_{j=1}^{n} e(j)c_j(l)\\
&= \frac{1}{n}\sum_{j=1}^{n}\epsilon(j)\left[\frac{1}{\sqrt{n}}\sum_{l=1}^{\tilde{n}}c_j(l)\,\mathrm{Re}\left(\Psi\left(\left\lfloor\frac{l+1}{2}\right\rfloor\right)\right)\right]\\
&\quad+\frac{1}{n}\sum_{j=1}^{n}\epsilon(j)\left[\frac{1}{\sqrt{n}}\sum_{l=1}^{\tilde{n}}c_j(l)(-1)^{l+1}\,\mathrm{Im}\left(\Psi\left(\left\lfloor\frac{l+1}{2}\right\rfloor\right)\right)\right]\\
&\quad+\frac{1}{n}\sum_{l=1}^{\tilde{n}/2}(\mathrm{Re}(Y_n(l))+\mathrm{Im}(Y_n(l))).
\end{aligned}
$$

For the first term the Chebyshev inequality gives for any $\tau > 0$

$$\begin{aligned}
&P\left(\left|\frac{1}{n}\sum_{j=1}^{n}\epsilon(j)\left[\frac{1}{\sqrt{n}}\sum_{l=1}^{\tilde{n}}c_j(l)\mathrm{Re}\left(\Psi\left(\left\lfloor\frac{l+1}{2}\right\rfloor\right)\right)\right]\right|\geqslant\tau\right)\\
&\leqslant\frac{\sigma^2}{\tau^2n^2}\sum_{j=1}^{n}\frac{1}{n}\left[\sum_{l=1}^{\tilde{n}}c_j(l)\mathrm{Re}\left(\Psi\left(\left\lfloor\frac{l+1}{2}\right\rfloor\right)\right)\right]^2\\
&\leqslant\frac{\sigma^2}{\tau^2n^3}\left[\sum_{l=1}^{\tilde{n}}\mathrm{Re}^2\left(\Psi\left(\left\lfloor\frac{l+1}{2}\right\rfloor\right)\right)\sum_{j=1}^{n}c_j(l)^2\right.\\
&\quad\left.+\sum_{l_1\neq l_2}\mathrm{Re}\left(\Psi\left(\left\lfloor\frac{l_1+1}{2}\right\rfloor\right)\right)\mathrm{Re}\left(\Psi\left(\left\lfloor\frac{l_2+1}{2}\right\rfloor\right)\right)\sum_{j=1}^{n}c_j(l_1)c_j(l_2)\right]\\
&\ll\frac{1}{\tau^2n^2}\sum_{l=1}^{\tilde{n}/2}\mathrm{Re}^2(\Psi(l)),
\end{aligned}$$

where the last line follows because $\{\sqrt{\frac{2}{n}}c_\Diamond(l), l = 1,\ldots,\tilde{n}\}$ is an ON-System (confer e.g. Remark 4.4.1), where $c_\Diamond(l) = (c_1(l),\ldots,c_n(l))^T$.

Because $\sum_{j\geqslant 0}\sqrt{j}|w_j| < \infty$ it holds

$$\max_{l=1,\ldots,\tilde{n}}|\Psi(l)| = \max_{l=1,\ldots,\tilde{n}}\left|\sum_{j\geqslant 0}w_j\exp(-2\pi ijl/n)\right| \leqslant \sum_{j\geqslant 0}|w_j| < \infty, \tag{4.6.6}$$

hence $\frac{1}{n^2}\sum_{l=1}^{\tilde{n}/2}\mathrm{Re}^2(\Psi(l)) \ll \frac{1}{n}$. This gives

$$\frac{1}{n}\sum_{j=1}^{n}\epsilon(j)\left[\frac{1}{\sqrt{n}}\sum_{l=1}^{\tilde{n}}c_j(l)\mathrm{Re}\left(\Psi\left(\left\lfloor\frac{l+1}{2}\right\rfloor\right)\right)\right] = O_P\left(\frac{1}{\sqrt{n}}\right)$$

and an analogous argument yields

$$\frac{1}{n}\sum_{j=1}^{n}\epsilon(j)\left[\frac{1}{\sqrt{n}}\sum_{l=1}^{\tilde{n}}c_j(l)(-1)^{l+1}\mathrm{Im}\left(\Psi\left(\left\lfloor\frac{l+1}{2}\right\rfloor\right)\right)\right] = O_P\left(\frac{1}{\sqrt{n}}\right).$$

Note that $\mathrm{Re}(U_{n,l}(k))$ respectively $\mathrm{Im}(U_{n,l}(k))$ is a sum of $2k$ independent random variables for $k < n$ and for $k \geqslant n$ a sum of $2n$ independent r.v.´s with mean 0 and a uniformly bounded κth moment. Thus it holds

$$\mathrm{E}\,\mathrm{Re}^2(U_{n,l}(k)) \ll \min(k,n), \qquad \mathrm{E}\,\mathrm{Im}^2(U_{n,l}(k)) \ll \min(k,n)$$

and hence the Minkowski inequality gives uniformly in l

$$\mathrm{E}\left(\mathrm{Re}^2(Y_n(l))\right) \ll \frac{1}{n}\left(\sum_{k\geqslant 0}|w_k|\min(k,n)^{\frac{1}{2}}\right)^2 \ll \frac{1}{n}\left(\sum_{k\geqslant 0}k^{\frac{1}{2}}|w_k|\right)^2 \ll \frac{1}{n}, \tag{4.6.7}$$

analogously $\mathrm{E}(\mathrm{Im}^2(Y_n(l))) \ll \frac{1}{n}$. The Chebyshev and Cauchy-Schwarz inequalities now imply

$$P\left(\frac{1}{n}\left|\sum_{l=1}^{\tilde{n}/2}\mathrm{Re}(Y_n(l))\right| \geqslant \tau\right) \leqslant \frac{1}{\tau^2 n^2}\sum_{l,k}\mathrm{cov}(\mathrm{Re}(Y_n(k)),\mathrm{Re}(Y_n(l)))$$
$$\leqslant \frac{1}{\tau^2 n^2}\sum_{l,k}\sqrt{\mathrm{var}\,\mathrm{Re}(Y_n(k))\,\mathrm{var}\,\mathrm{Re}(Y_n(l))} \ll \frac{1}{\tau^2 n}.$$

Analogously we get $P\left(\frac{1}{n}\left|\sum_{l=1}^{\tilde{n}/2}\mathrm{Im}(Y_n(l))\right| \geqslant \tau\right) \ll \frac{1}{\tau^2 n}$.

Putting everything together we have

$$\frac{1}{n^{3/2}}\sum_{l=1}^{\tilde{n}}\sum_{j=1}^{n} e(j)c_j(l) = O_P(n^{-1/2}), \tag{4.6.8}$$

hence condition (4.6.3).

For the proof of (4.6.4) we need the following, which we get using a similar argument as above. Theorem B.7 shows

$$\mathrm{E}\,|U_{n,l}(k)|^{\kappa} \ll \min(k,n)^{\frac{\kappa}{2}}$$

and hence the Minkowski inequality gives uniformly in l

$$\mathrm{E}\,|Y_n(l)|^{\kappa} \ll \frac{1}{n^{\frac{\kappa}{2}}}\left(\sum_{k\geqslant 0}|w_k|\min(k,n)^{\frac{1}{2}}\right)^{\kappa} \ll \frac{1}{n^{\frac{\kappa}{2}}}\left(\sum_{k\geqslant 0}|w_k|k^{\frac{1}{2}}\right)^{\kappa} \ll \frac{1}{n^{\frac{\kappa}{2}}}.$$

The Markov inequality implies

$$P\left(\frac{1}{n}\sum_{l=1}^{\tilde{n}/2}|Y_n(l)|^{\kappa} \geqslant \tau\right) \leqslant \frac{1}{\tau}\max_l \mathrm{E}\,|Y_n(l)|^{\kappa} \ll \frac{1}{\tau}\frac{1}{n^{\frac{\kappa}{2}}}, \tag{4.6.9}$$

which converges to 0 sufficiently fast so that we have almost sure convergence.

Furthermore it holds

$$\frac{1}{n}\sum_{l=1}^{\tilde{n}}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n} e(j)c_j(l)\right|^{\kappa}$$
$$\ll \max_{l=1,\ldots,\tilde{n}/2}|\Psi(l)|^{\kappa}\;\frac{1}{n}\sum_{l=1}^{\tilde{n}}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\epsilon(j)c_j(l)\right|^{\kappa} + \frac{1}{n}\sum_{l=1}^{\tilde{n}/2}|Y_n(l)|^{\kappa}.$$

Putting together (4.6.6) and (4.6.9) shows that it suffices to prove the assertion for a sequence of i.i.d. r.v.´s.

For $\{\epsilon(i) : 1 \leqslant i \leqslant n\}$ i.i.d. $N(0,\sigma^2)$ r.v.´s, $\left\{\sqrt{2/n}\,\sum_{j=1}^{n}\epsilon(j)c_j(l) : 1 \leqslant l \leqslant \tilde{n}\right\}$ is also i.i.d. $N(0,\sigma^2)$–distributed (confer e.g. Brockwell and Davis [13], p. 344). Then the law of large numbers (for the triangular case confer e.g. Corollary B.1) gives

$$\frac{1}{n^3}\sum_{l=1}^{\tilde{n}}\left(\sum_{j=1}^{n}\epsilon(j)c_j(l)\right)^4 = O(1) \quad a.s. \tag{4.6.10}$$

For a general sequence of innovations the Markov inequality gives the assertion, since

$$P\left(\frac{1}{n^{1+\kappa/2}}\sum_{l=1}^{\tilde{n}}\left|\sum_{j=1}^{n}\epsilon(j)c_j(l)\right|^{\kappa}\geqslant C\right)\leqslant\frac{1}{C}\max_{1\leqslant l\leqslant\tilde{n}}\mathrm{E}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\epsilon(j)c_j(l)\right|^{\kappa}\ll\frac{1}{C},$$

where the last line can be seen via Theorem B.7. ■

4.6.2. Some Properties of the Change Estimators

In this subsection we will discuss estimators $\widehat{m}$, $\widehat{\mu}_1$ respectively $\widehat{\mu}_2$ for the unknown variables m, $\mu_1 := \mu$, $\mu_2 := \mu + d$, where $d \equiv 0$ under H_0. Specifically we are interested in the following estimators:

$$\begin{aligned}
&\widehat{m}=\widehat{m}(\gamma)=\min(\arg\max(|S_k(\gamma)|,k=1,\dots,n-1)),\\
&\qquad\text{where } S_k(\gamma)=\left(\frac{n}{k(n-k)}\right)^{\gamma}\sum_{i=1}^{k}(X(i)-\bar{X}),\qquad 0\leqslant\gamma<1.\\
&\widehat{\mu}_1=\widehat{\mu}_1(\gamma)=\frac{1}{\widehat{m}}\sum_{i=1}^{\widehat{m}}X(i),\\
&\widehat{\mu}_2=\widehat{\mu}_2(\gamma)=\frac{1}{n-\widehat{m}}\sum_{i=\widehat{m}+1}^{n}X(i).
\end{aligned}\tag{4.6.11}$$

The most interesting estimators being the ones with $0 \leqslant \gamma \leqslant \frac{1}{2}$, and there more specifically $\gamma = 0$ and $\gamma = \frac{1}{2}$.

We establish some properties of these estimators in order to prove that the frequency permutation methods still gives good approximations of the critical values, if we do not know the underlying linear process but need to estimate it. The estimators suitable for this purpose have to fulfill the following conditions as the proof of Theorem 4.6.2 will show.
Lemma 4.6.1 proves that the above estimators fulfill these conditions under certain assumptions on the alternatives.

Under H_0 $\widehat{m}$ still has to behave reasonably, more precisely

$$\begin{aligned}
&|\mu_j-\widehat{\mu}_j|=o_P\left(\frac{\sqrt{n}}{\log n}\right),\qquad j=1,2,\\
&\frac{\widehat{m}}{n}|\mu_1-\widehat{\mu}_1|^r+\frac{n-\widehat{m}}{n}|\mu_2-\widehat{\mu}_2|^r 1_{\{\widehat{m}<n\}}=o_P(1),\qquad r=1,2,\\
&\frac{\widehat{m}^3}{n^2}(\mu_1-\widehat{\mu}_1)^4+\frac{(n-\widehat{m})^3}{n^2}(\mu_2-\widehat{\mu}_2)^4 1_{\{\widehat{m}<n\}}=O_P(1).
\end{aligned}\tag{4.6.12}$$

Under H_1 the following conditions are sufficient

$$\begin{aligned}
&|\mu_i-\widehat{\mu}_j|=o_P\left(\frac{\sqrt{n}}{\log n}\right),\qquad i,j=1,2,\\
&n^{\frac{1}{4}}|\mu_j-\widehat{\mu}_j|=O_P(1),\qquad j=1,2,\\
&(\mu_j-\widehat{\mu}_j)^4\frac{|m-\widehat{m}|^3}{n^2}+|d|^4\frac{|m-\widehat{m}|^3}{n^2}=O_P(1),\qquad j=1,2,\\
&\frac{|m-\widehat{m}|}{n}|d|^r=o_P(1),\qquad r=1,2.
\end{aligned}\tag{4.6.13}$$

For local alternatives we need a somewhat weaker set of conditions:

$$\begin{aligned}
&|\mu_i-\widehat{\mu}_j| = o_P\left(\frac{\sqrt{n}}{\log n}\right), \qquad i,j=1,2,\\
&\frac{m\wedge\widehat{m}}{n}|\mu_1-\widehat{\mu}_1|^r = o_P(1), \quad \frac{n-(m\vee\widehat{m})}{n}|\mu_2-\widehat{\mu}_2|^r 1_{\{\widehat{m}<n\}} = o_P(1),\\
&\frac{(m\wedge\widehat{m})^3}{n^2}(\mu_1-\widehat{\mu}_1)^4 = O_P(1), \quad \frac{(n-(m\vee\widehat{m}))^3}{n^2}(\mu_2-\widehat{\mu}_2)^4 1_{\{\widehat{m}<n\}} = O_P(1),\\
&(\mu_2-\widehat{\mu}_1)^4\frac{(\widehat{m}-m)_+^3}{n^2} = O_P(1), \quad (\mu_1-\widehat{\mu}_2)^4\frac{(m-\widehat{m})_+^3}{n^2}1_{\{\widehat{m}<n\}} = O_P(1),\\
&\frac{(\widehat{m}-m)_+}{n}|\mu_2-\widehat{\mu}_1|^r = o_P(1), \quad \frac{(m-\widehat{m})_+}{n}|\mu_1-\widehat{\mu}_2|^r 1_{\{\widehat{m}<n\}} = o_P(1),
\end{aligned} \tag{4.6.14}$$

for $r=1,2$, where $a_+=\max(a,0)$, $a\wedge b=\min(a,b)$, $a\vee b=\max(a,b)$.

Remark 4.6.3. The conditions given in the first lines must be stronger if we do not center the Fourier coefficients. Confer also Remark 4.6.7, this is why we give the exact convergence rates in Lemma 4.6.1.

Now we prove that the estimators given in equation (4.6.11) fulfill the above conditions (and thus work as estimators for the frequency bootstrap). We need to impose certain conditions on the change-point m and the mean change d first.

Let $\widetilde{\vartheta} := \min\left(\frac{m}{n},\frac{n-m}{n}\right)$. The first possibility is that we are under the null or a local alternative, fulfilling

$$|d|\frac{\log n}{\sqrt{n}} = o(1), \quad |d|^2\widetilde{\vartheta} = o(1), \quad \text{and} \quad |d|^4 n\widetilde{\vartheta}^3 = O(1). \tag{4.6.15}$$

Under the following alternatives the conditions are also fulfilled:

a) For $0\leqslant\gamma<\frac{1}{2}$:

$$\widetilde{\vartheta}\geqslant\delta>0, \qquad |d|\sqrt{n}\to\infty, \qquad |d|\frac{\log n}{\sqrt{n}} = o(1). \tag{4.6.16}$$

b) For $\gamma=\frac{1}{2}$:

$$\widetilde{\vartheta}\geqslant\delta>0, \qquad |d|\sqrt{\frac{n}{\log n}}\to\infty, \qquad |d|\sqrt{\frac{\log^3 n}{n}} = O(1). \tag{4.6.17}$$

c) For $\frac{1}{2}<\gamma\leqslant\frac{3}{4}$:

$$\begin{aligned}
&\widetilde{\vartheta}\geqslant\delta>0, \qquad |d|n^{1-\gamma}\to\infty, \qquad |d|n^{\gamma-1} = o(1),\\
&\frac{|d|}{n^{2-3\gamma}} = O(1).
\end{aligned} \tag{4.6.18}$$

Remark 4.6.4. If $\min\left(\frac{m}{n}, \frac{n-m}{n}\right) \geqslant \delta > 0$, then it suffices that the mean change d does not converge to infinity too fast. Precisely:

$$0 \leqslant \gamma < \tfrac{1}{2}: \qquad \frac{|d| \log n}{\sqrt{n}} = o(1),$$

$$\gamma = \tfrac{1}{2} \qquad |d| \sqrt{\frac{\log^3(n)}{n}} = O(1).$$

This shows that under the usual assumptions $m = \vartheta n$, $0 < \vartheta < 1$, and $|d(n)| \leqslant D < \infty$ the estimators given in (4.6.11), $0 \leqslant \gamma \leqslant \frac{1}{2}$, fulfill the above conditions. This remains true for $\gamma < \frac{2}{3}$.
The reason is that each sequence can be divided into two sub-sequences (possibly one of which is empty), one fulfilling $|d|^4 n = O(1)$, the other one $|d|\sqrt{n} \to \infty$, $|d|\sqrt{\frac{n}{\log n}} \to \infty$, $|d| n^{1-\gamma} \to \infty$, respectively.

The following lemma states that for the above alternatives and under the null hypothesis the estimators (4.6.11) fulfill the condition we need to prove the validity of the bootstrap.

Lemma 4.6.1. *i) Under H_0 the estimators (4.6.11) fulfill (4.6.12). Moreover it holds $|\mu_j - \widehat{\mu}_j| = O_P(1)$, $j = 1, 2$.*

ii) Under local alternatives that fulfill (4.6.15), the set of assumptions given in (4.6.14) are valid. Besides, $|\mu_i - \widehat{\mu}_j| = O_P(1 + |d|) = o_P(\sqrt{n}/\log(n))$, $i, j = 1, 2$.

iii) Under alternatives fulfilling (4.6.16)- (4.6.18), respectively, equations (4.6.13) hold true. Additionally $|\mu_i - \widehat{\mu}_j| = O_P(1 + |d|) = o_P(\sqrt{n}/\log(n))$, $i, j = 1, 2$.

Before we can prove the above lemma we first need a result from Kokoszka and Leipus [53].

Lemma 4.6.2. *For the estimators $\widehat{m}$ in equation (4.6.11) it holds*

$$\begin{aligned} &|d| \min\left(\frac{m}{n}, \frac{n-m}{n}\right)^{1-\gamma} \left|\frac{\widehat{m}}{n} - \frac{m}{n}\right| \\ &\leqslant C n^{\gamma-1} \max_{1 \leqslant k < n} \left(\frac{1}{k^\gamma} \left| \sum_{j=1}^{k} (X(j) - \mathrm{E}\, X(j)) \right| + \frac{1}{(n-k)^\gamma} \left| \sum_{j=k+1}^{n} (X(j) - \mathrm{E}\, X(j)) \right| \right), \end{aligned}$$

where $C := 2 \max\left(\frac{m}{n}, \frac{n-m}{n}\right)^\gamma / (1-\gamma)$.

Proof. Confer equation (3.11) and (3.12) in the proof of Theorem 1.1 in Kokoszka and Leipus [53]. ■

Now we are ready to prove that the estimators (4.6.11) are suitable for the frequency bootstrap.

Proof of Lemma 4.6.1. It holds $e(j) = X(j) - \mathrm{E}\, X(j)$. First we prove the assertion for the null hypothesis. We have to verify conditions (4.6.12). Concerning the second equation

$$\begin{aligned}
\left(\frac{\widehat{m}}{n}\right)^{\frac{1}{2}} |\mu_1 - \widehat{\mu}_1| &= \frac{1}{n^{\frac{1}{2}}} \left| \frac{1}{\widehat{m}^{\frac{1}{2}}} \sum_{j=1}^{\widehat{m}} (X(j) - \mathrm{E}\, X(j)) \right| \leqslant \frac{1}{n^{\frac{1}{2}}} \max_{1 \leqslant k \leqslant n} \left| \frac{1}{k^{\frac{1}{2}}} \sum_{j=1}^{k} e(j) \right| \\
&= O_P\left(\sqrt{\frac{\log n}{n}}\right) = o_P(1),
\end{aligned}$$

because the Hájek-Renyi inequality in Lemma B.1 gives:

$$P\left(\max_{1 \leqslant k \leqslant n} \left| \frac{1}{k^{\frac{1}{2}}} \sum_{j=1}^{k} e(j) \right| \geqslant C \right) \ll \frac{1}{C^2} \sum_{k=1}^{n} k^{-1} \ll \frac{1}{C^2} \log n$$

Analogously

$$\frac{n - \widehat{m}}{n} |\mu_2 - \widehat{\mu}_2|^2 1_{\{\widehat{m} < n\}} = o_P(1).$$

In the exact same way we get the other assertions, precisely

$$\begin{aligned}
&\frac{\widehat{m}^3}{n^2} (\mu_1 - \widehat{\mu}_1)^4 + \frac{(n - \widehat{m})^3}{n^2} (\mu_2 - \widehat{\mu}_2)^4 1_{\{\widehat{m} < n\}} = O_P\left(\frac{1}{n}\right) \\
&|\mu_j - \widehat{\mu}_j| = O_P(1), \\
&\frac{\widehat{m}}{n} |\mu_1 - \widehat{\mu}_1| = O_P\left(n^{-\frac{1}{2}}\right) = o_P(1), \quad \frac{n - \widehat{m}}{n} |\mu_2 - \widehat{\mu}_2| 1_{\{\widehat{m} < n\}} = o_P(1).
\end{aligned}$$

This gives the assertion for the null hypothesis.

Next we prove that under (4.6.15), the estimators fulfill assumptions (4.6.14). Note that

$$\widehat{\mu}_1 - \mu_1 = \frac{1}{\widehat{m}} \sum_{j=1}^{\widehat{m}} (X(j) - \mathrm{E}\, X(j)) + d\, \frac{(\widehat{m} - m)_+}{\widehat{m}}.$$

The proofs for $\frac{1}{\widehat{m}} \sum_{j=1}^{\widehat{m}} (X(j) - \mathrm{E}\, X(j))$ are analogous to the ones for the null hypothesis. For the second term note that

$$\begin{aligned}
&\frac{\min(m, \widehat{m})}{n} \left| d\, \frac{(\widehat{m} - m)_+}{\widehat{m}} \right|^r \ll |d|^r \min\left(\frac{m}{n}, \frac{(\widehat{m} - m)_+}{n}\right) \ll |d|^r \min\left(\frac{m}{n}, \frac{n - m}{n}\right) \\
&= o(1), \qquad r = 1, 2,
\end{aligned}$$

since

$$|d| \min\left(\frac{m}{n}, \frac{n - m}{n}\right) \leqslant \sqrt{|d|^2 \min\left(\frac{m}{n}, \frac{n - m}{n}\right)} \sqrt{\min\left(\frac{m}{n}, \frac{n - m}{n}\right)} = o(1).$$

In the same way we get for the third line of (4.6.14)

$$\frac{\min(m, \widehat{m})^3}{n^2} \left| d\, \frac{(\widehat{m} - m)_+}{\widehat{m}} \right|^4 \ll |d|^4 \min\left(\frac{m^3}{n^2}, \frac{(\widehat{m} - m)_+^3}{n^2}\right) \ll |d|^4 n \widetilde{\vartheta}^3 = O(1).$$

Concerning the last two conditions it holds for $\widehat{m} \geqslant m$

$$\widehat{\mu}_1 - \mu_2 = \frac{1}{\widehat{m}} \sum_{j=1}^{\widehat{m}} (X(i) - \mathrm{E}\, X(i)) - \frac{m}{\widehat{m}} d.$$

Again the proofs for $\frac{1}{\widehat{m}}\sum_{j=1}^{\widehat{m}}(X(i)-\operatorname{E}X(i))$ are as above, if one takes into account that $(\widehat{m}-m)_+ \leqslant \widehat{m}$. So we are interested in the second term. For the fourth condition we get

$$\frac{(\widehat{m}-m)_+^3}{n^2}\left|d\,\frac{m}{\widehat{m}}\right|^4 \ll |d|^4 \min\left(\frac{m^3}{n^2},\frac{(\widehat{m}-m)_+^3}{n^2}\right) = O(1).$$

And similarly for the last condition

$$\frac{(\widehat{m}-m)_+}{n}\left|d\,\frac{m}{\widehat{m}}\right|^r \ll |d|^r \min\left(\frac{m}{n},\frac{(\widehat{m}-m)_+}{n}\right) = o(1), \qquad r=1,2.$$

The proofs of the results involving $\widehat{\mu}_2$ are analogous and therefore omitted. Moreover we have shown $|\mu_i - \widehat{\mu}_j| = O_P(1+|d|)$, $i,j=1,2$.

Finally we prove that conditions (4.6.13) are fulfilled, if the alternative fulfills assumption (4.6.16) - (4.6.18), respectively. The first line of (4.6.13) follows analogous to local alternatives.

Lemma B.1 states

$$P\left(n^{\gamma-1}\max_{1\leqslant k\leqslant n}\frac{1}{k^\gamma}\left|\sum_{j=1}^{k}(X(j)-\operatorname{E}X(j))\right|\geqslant C\right) \ll \frac{n^{2(\gamma-1)}}{C^2}\sum_{k=1}^{n}\frac{1}{k^{2\gamma}}$$
$$\ll \frac{1}{C^2}\cdot\begin{cases} n^{-1}, & \gamma<\frac{1}{2},\\ n^{-1}\log n, & \gamma=\frac{1}{2},\\ n^{2\gamma-2}, & \gamma>\frac{1}{2},\end{cases}$$

and an analogous expression for $\max_{1\leqslant k<n}\frac{1}{(n-k)^\gamma}\sum_{j=k+1}^{n}(X(j)-\operatorname{E}X(j))$. This together with Lemma 4.6.2 gives now in case of

$$\begin{aligned}
\gamma<\frac{1}{2}: &\qquad |d|\frac{|\widehat{m}-m|}{n} = O_P\left(\sqrt{\frac{1}{n}}\right),\\
\gamma=\frac{1}{2}: &\qquad |d|\frac{|\widehat{m}-m|}{n} = O_P\left(\sqrt{\frac{\log n}{n}}\right),\\
\gamma>\frac{1}{2}: &\qquad |d|\frac{|\widehat{m}-m|}{n} = O_P\left(n^{\gamma-1}\right),
\end{aligned}\tag{4.6.19}$$

since $\widetilde{\vartheta}\geqslant\delta>0$. Hence under the following condition on d

$$\begin{aligned}
\gamma<\frac{1}{2}: &\qquad |d|\sqrt{n}\to\infty,\\
\gamma=\frac{1}{2}: &\qquad |d|\sqrt{\frac{n}{\log n}}\to\infty,\\
\gamma>\frac{1}{2}: &\qquad |d|n^{1-\gamma}\to\infty,
\end{aligned}\tag{4.6.20}$$

it holds $\frac{\widehat{m}-m}{n} = o_P(1)$. This gives

$$\frac{n}{\min(\widehat{m},n-\widehat{m})} = O_P(1), \tag{4.6.21}$$

since

$$\frac{n}{\widehat{m}} = \left(\frac{\widehat{m}-m}{n} + \frac{m}{n} \right)^{-1} \leqslant \left(\delta + \frac{\widehat{m}-m}{n} \right)^{-1} \xrightarrow{P} \delta^{-1} < \infty$$

and an analogous argument for $\frac{n}{n-\widehat{m}} \leqslant \delta^{-1} + o_P(1)$. As before $\widehat{\mu}_1 - \mu_1 = \frac{1}{\widehat{m}} \sum_{j=1}^{\widehat{m}} (X(j) - \mathrm{E}\, X(j)) + d\, \frac{(\widehat{m}-m)_+}{\widehat{m}}$. For the second term it holds under (4.6.20) because of (4.6.19) respectively (4.6.21)

$$\frac{d(\widehat{m}-m)_+}{\widehat{m}} = \begin{cases} O_P\left(\sqrt{\frac{1}{n}}\right), & \gamma < \frac{1}{2}, \\ O_P\left(\sqrt{\frac{\log n}{n}}\right), & \gamma = \frac{1}{2}, \\ O_P\left(n^{\gamma-1}\right), & \gamma > \frac{1}{2}, \end{cases} \tag{4.6.22}$$

Moreover by Lemma B.1

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{\widehat{m}} (X(i) - \mathrm{E}\, X(i)) \leqslant \frac{1}{\sqrt{n}} \max_{1 \leqslant k < n} \sum_{j=1}^{k} (X(i) - \mathrm{E}\, X(i)) = O_P(1).$$

This together with (4.6.21) and (4.6.22) now gives under (4.6.20) in case of

$$\begin{aligned} \gamma < \frac{1}{2}: &\quad \sqrt{n} |\widehat{\mu}_j - \mu_j| = O_P(1), \\ \gamma = \frac{1}{2}: &\quad \sqrt{\frac{n}{\log n}} |\widehat{\mu}_j - \mu_j| = O_P(1), \\ \gamma > \frac{1}{2}: &\quad n^{1-\gamma} |\widehat{\mu}_j - \mu_j| = O_P(1), \qquad j = 1, 2. \end{aligned} \tag{4.6.23}$$

Here the assertion for $j = 2$ follows in the same way. Putting together (4.6.19) and (4.6.23) we realize that under (4.6.16) - (4.6.18), respectively, assumptions (4.6.13) hold. ■

4.6.3. Convergence of the Permutation Statistics Using Estimators

We are now ready to state the main theorem. It shows that the procedure described in Section 4.2 gives asymptotically the correct critical values. The theorem states convergence of the permutation processes conditioned on the observations in $C[0,1]$ whereas the corollary specifies the results for the statistics we are interested in.

Consider now

$$\widetilde{X}(i) := (X(i) - \widehat{\mu}_1) 1_{[1,\widehat{m}]}(i) + (X(i) - \widehat{\mu}_2) 1_{[\widehat{m}+1,n]}(i), \tag{4.6.24}$$

for suitable estimators $\widehat{m}$, $\widehat{\mu}_1$ and $\widehat{\mu}_2$, confer also equation (4.2.1). Usually we will choose the estimators defined in equation (4.6.11) with $0 \leqslant \gamma \leqslant 1/2$. Lemma 4.6.1 proves that they fulfill the assumptions of the following theorem for a wide range of alternatives.

Theorem 4.6.2. *Let the conditions of Theorem 4.6.1 be fulfilled. Let under H_0 assumptions (4.6.12) and under H_1 assumptions (4.6.13) respectively (4.6.14) be fulfilled. Then*

$$P\left(f[(\widetilde{Z}_n^{\mathbf{X}}(id, \mathbf{R}) - id\, \widetilde{Z}_n^{\mathbf{X}}(1, \mathbf{R}))/\widehat{\sigma}_n] \leq x \,\middle|\, X(1), \ldots, X(n) \right) \xrightarrow{P} P\left(f(B(\cdot)) \leqslant x \right)$$

for all continuous $f : C[0,1] \to \mathbb{R}$ *and for all* $x \in \mathbb{R}$. *Here*

$$\widehat{\sigma}_n^2 := \frac{2}{n\tilde{n}} \sum_{l=1}^{\tilde{n}} \left(\sum_{i=1}^{n} \widetilde{X}(i) c_i(l) - \frac{1}{\tilde{n}} \sum_{k=1}^{\tilde{n}} \sum_{j=1}^{n} \widetilde{X}(j) c_j(k) \right)^2 ,$$

$$\widetilde{Z}_n^{\mathbf{X}}(u, \mathbf{R}) = \sqrt{\frac{\alpha(n)}{n}} \sum_{s \leqslant \frac{n}{\alpha(n)} u} X_{\mathbf{R}}(\beta(s)), \quad \text{for } t = \frac{\alpha(n)}{n}, \frac{2\alpha(n)}{n}, \dots, 1,$$

$\widetilde{Z}_n^{\mathbf{X}}(0, R) = 0$ *and* $\widetilde{Z}_n^{\mathbf{X}}(t, R)$ *is linearly interpolated between* $(i-1)/n$ *and* i/n *for* $i = 1, \dots, n$, $X_{\mathbf{R}}(\cdot)$ *is as in equation (4.3.1) respectively (4.3.3).*

Remark 4.6.5. If we have estimators with appropriate convergence rates in an $a.s.$-sense, we get for normal innovations as in Theorem 4.6.1 the convergence in an $a.s.$-sense.

Remark 4.6.6. It is again possible to keep the middle term of the Fourier coefficients instead of setting it equal to 0. It is also possible to work with uncentered Fourier coefficients under appropriate conditions. For more details confer Remark 4.6.7.

The following corollary states the limit behavior of our statistics of interest.

Corollary 4.6.1. *Under the conditions of Theorem 4.6.2 the following holds:*

a) For all $\epsilon > 0$ *we get for all* $x \in \mathbb{R}$

$$P\left(T_n^{(1f)}(\mathbf{R}) \leqslant x \,|\, X(1), \dots, X(n)\right) \xrightarrow{P} P\left(\sup_{\epsilon \leqslant t \leqslant 1-\epsilon} \sqrt{\frac{1}{t(1-t)}} |B(t)| \leqslant x\right),$$

$$\text{where } \; T_n^{(1f)}(\mathbf{R}) := \sup_{\epsilon \leqslant t \leqslant 1-\epsilon} \sqrt{\frac{1}{t(1-t)}} |\widetilde{Z}_n^{\mathbf{X}}(t, \mathbf{R}) - t\widetilde{Z}_n^{\mathbf{X}}(1, \mathbf{R})|.$$

b) It holds for any $0 < \epsilon < 1$

$$P\left(T_n^{(2f)}(\mathbf{R}) \leqslant x \,|\, X(1), \dots, X(n)\right) \xrightarrow{P} P\left(\sup_{\epsilon \leqslant t \leqslant 1} |B(t) - B(t-\epsilon)| \leqslant x\right),$$

$$\text{where } \; T_n^{(2f)}(\mathbf{R}) := \sup_{\epsilon \leqslant t \leqslant 1} |\widetilde{Z}_n^{\mathbf{X}}(t, \mathbf{R}) - \widetilde{Z}_n^{\mathbf{X}}(t-\epsilon, \mathbf{R}) - \epsilon \widetilde{Z}_n^{\mathbf{X}}(1, \mathbf{R})|.$$

c) If $q \in FC_0^1$ *and*

i)

$$\int_0^1 \frac{1}{q^2(t)}\, dt < \infty \quad \text{or}$$

ii) $\kappa = 4$,

$$\int_0^1 \frac{t(1-t)}{q^4(t)}\, dt < \infty$$

and $\frac{t}{q^4(t)}$ *is non-increasing in a neighborhood of* 0 *and* $\frac{1-t}{q^4(t)}$ *is non-decreasing in a neighborhood of* 1, *then it holds*

$$P\left(T_n^{(3f)}(\mathbf{R},q)\leqslant x\,|\,X(1),\ldots,X(n)\right)\xrightarrow{P}P\left(\sup_{0\leqslant t\leqslant 1}\frac{|B(t)|}{q(t)}\leqslant x\right),$$

$$\text{where } T_n^{(3f)}(\mathbf{R},q):=\max_{1\leqslant k<N}\frac{1}{q\left(\frac{k}{N}\right)}\left|\widetilde{Z}_n^{\mathbf{X}}\left(\frac{k}{N},\mathbf{R}\right)-\frac{k}{N}\widetilde{Z}_n^{\mathbf{X}}(1,\mathbf{R})\right|,\quad N=\frac{n}{\alpha(n)}.$$

d) For $\int_0^1\frac{(t(1-t))^s}{r(t)}\,dt<\infty$ *for some* $0\leqslant s<1$*, it holds*

$$P\left(T_n^{(4f)}(\mathbf{R})\leqslant x\,|\,X(1),\ldots,X(n)\right)\xrightarrow{P}P\left(\int_0^1\frac{B^2(t)}{r(t)}\,dt\leqslant x\right),$$

$$\text{where } T_n^{(4f)}(\mathbf{R},r):=\int_0^1\frac{1}{r(t)}|\widetilde{Z}_n^{\mathbf{X}}(t,\mathbf{R})-t\widetilde{Z}_n^{\mathbf{X}}(1,\mathbf{R})|^2\,dt.$$

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

Proof. The proof is analogous to that of Theorem 4.6.2, below, using Corollary 4.5.1 instead of Theorem 4.5.4. ■

Remark 4.6.7. As in Remark 4.6.1 in view of Remark 4.5.9 it is possible to keep the middle term if $\sup_{0\leqslant u\leqslant 1}\left|\sum_{s=1}^{\lfloor Nu\rfloor}(-1)^{\beta(s)}\right|=O(1)$.
It is also possible to work with uncentered Fourier coefficients if

$$\frac{\sqrt{\alpha(n)}\,\log n}{\max(q(1/N),q(1-1/N))}\frac{1}{n^{3/2}}\sum_{l=1}^{\tilde{n}}\sum_{j=1}^{n}\widetilde{X}(j)c_j(l)\xrightarrow{P}0.$$

Equations (4.6.8), (4.6.25) and Lemma 4.6.1 show that

$$\frac{1}{n^{3/2}}\sum_{l=1}^{\tilde{n}}\sum_{j=1}^{n}\widetilde{X}(j)c_j(l)=O_P\left(\frac{\log n}{\sqrt{n}}(1+|d|)\right),$$

where $d\equiv 0$ under H_0.

Also the following result can be proven in exactly the same way using Theorem 4.5.2. It gives the validity of bootstrapping procedures where one is only interested in functionals of a finite sequence. It might be useful in some situations.

Theorem 4.6.3. *For any* $M>0$ *and any* $0<\lambda_1<\ldots<\lambda_M<1$ *let*

$$X_{s,n}^*:=\frac{2}{n}\sum_{l=1}^{\tilde{n}}c_{\lfloor\lambda_s\cdot n\rfloor}(l)\sum_{j=1}^{n}\widetilde{X}(j)c_j(R_l).$$

Under the conditions of Theorem 4.6.2 it holds as $n\to\infty$ *for all* $x\in\mathbb{R}$ *and all* $f:\mathbb{R}^M\to\mathbb{R}$ *continuous:*

$$P\left(f(X_{s,n}^*,\ldots,X_{M,n}^*)\leqslant x\,|\,X(1),\ldots,X(n)\right)\xrightarrow{P}P(f(Y_1,\ldots,Y_M)\leqslant x),$$

where $(Y_1,\ldots,Y_M)\overset{\mathcal{D}}{=}N(0,I_M)$.

Now we prove the main theorem of this section.

Proof of Theorem 4.6.2. Again we have to verify conditions (4.5.2) in a P-stochastic sense keeping in mind the conditions in Remarks 4.5.3 respectively 4.5.8 for

$$x_i := \widetilde{X}(i)\left(\frac{2}{n\tilde{n}}\sum_{l=1}^{\tilde{n}}\left(\sum_{i=1}^{n}\widetilde{X}(i)c_i(l)-\frac{1}{\tilde{n}}\sum_{k=1}^{\tilde{n}}\sum_{j=1}^{n}\widetilde{X}(j)c_j(k)\right)^2\right)^{-\frac{1}{2}}.$$

Using (4.6.1) - (4.6.5) and the triangle inequality, it suffices to show that

$$\frac{1}{n}\sum_{j=1}^{n}(\widetilde{X}(j)-e(j)) = o_P(1) \quad \text{for } n \text{ odd;}$$

$$\frac{1}{n}\sum_{i=1}^{n}|\widetilde{X}(i)-e(i)|^2 = o_P(1);$$

$$\frac{1}{n^{3/2}}\sum_{l=1}^{\tilde{n}}\sum_{j=1}^{n}(\widetilde{X}(j)-e(j))c_j(l) = o_P(1);$$

$$\frac{1}{n}\sum_{l=1}^{\tilde{n}}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}c_j(l)(\widetilde{X}(j)-e(j))\right|^{\kappa} = O_P(1) \quad \text{for } \kappa=\min(4,\nu);$$

$$\frac{1}{n}\sum_{j=1}^{n/2}(\widetilde{X}(2j)-e(2j)) = o_P(1), \quad \frac{1}{n}\sum_{j=1}^{n/2}(\widetilde{X}(2j-1)-e(2j-1)) = o_P(1) \quad \text{for } n \text{ even.}$$

Concerning the second equality note that $(a+b)^2 \leqslant 2a^2+2b^2$, hence $a^2 \leqslant 2(a-b)^2 + 2\,b^2$, which gives $(a-b)^2 \geqslant \frac{1}{2}a^2 - b^2$. This shows in our situation $\frac{1}{n}\sum_{i=1}^{n}|\widetilde{X}(i)|^2 \geqslant \frac{1}{2n}\sum_{i=1}^{n}|e(i)|^2 - \frac{1}{n}\sum_{i=1}^{n}|\widetilde{X}(i)-e(i)|^2$.

Because of Lemma 4.4.2

$$\frac{1}{n^{3/2}}\sum_{l=1}^{\tilde{n}}\sum_{j=1}^{n}(\widetilde{X}(j)-e(j))c_j(l) \ll \frac{\log n}{\sqrt{n}}\max_{j=1,\ldots,n}|\widetilde{X}(j)-e(j)|, \tag{4.6.25}$$

and

$$\frac{1}{n}\sum_{l=1}^{\tilde{n}}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}c_j(l)(\widetilde{X}(j)-e(j))\right|^{\kappa} \ll 1+\frac{1}{n}\sum_{l=1}^{\tilde{n}}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}c_j(l)(\widetilde{X}(j)-e(j))\right|^{4}.$$

This means it suffices to prove the boundedness for $\kappa = 4$. Moreover we will verify

$$\frac{\log n}{\sqrt{n}}\max_{j=1,\ldots,n}|\widetilde{X}(j)-e(j)| = o_P(1).$$

Under H_0 it holds ($\mu_1 = \mu_2 = \mu$)

$$\begin{aligned}\widetilde{X}(i)-e(i) &= (X(i)-\widehat{\mu}_1)1_{[1,\widehat{m}]}(i) + (X(i)-\widehat{\mu}_2)1_{[\widehat{m}+1,n]}(i) - (X(i)-\mu)\\ &= (\mu_1-\widehat{\mu}_1)1_{[1,\widehat{m}]}(i) + (\mu_2-\widehat{\mu}_2)1_{[\widehat{m}+1,n]}(i).\end{aligned}$$

By Lemma 4.4.1 it holds

$$\frac{1}{n^3}\sum_{l=1}^{\tilde{n}}(\mu_1-\widehat{\mu}_1)^4\left(\sum_{j=1}^{\widehat{m}}c_j(l)\right)^4+\frac{1}{n^3}\sum_{l=1}^{\tilde{n}}(\mu_2-\widehat{\mu}_2)^4\left(\sum_{j=\widehat{m}+1}^{n}c_j(l)\right)^4$$
$$\ll\frac{\widehat{m}^3n}{n^3}(\mu_1-\widehat{\mu}_1)^4+\frac{(n-\widehat{m})^3n}{n^3}(\mu_2-\widehat{\mu}_2)^4 1_{\{\widehat{m}<n\}}=O_P(1).$$

This shows that the conditions are fulfilled under H_0, since assumptions (4.6.12) are fulfilled.

Under H_1 we have

$$\begin{aligned}&\widetilde{X}(i)-e(i)\\&=(X(i)-\widehat{\mu}_1)1_{[1,\widehat{m}]}(i)+(X(i)-\widehat{\mu}_2)1_{[\widehat{m}+1,n]}(i)-(X(i)-\mu_1)1_{[1,m]}(i)\\&\quad-(X(i)-\mu_2)1_{[m+1,n]}(i)\\&=(\mu_1-\widehat{\mu}_1)1_{[1,m\wedge\widehat{m}]}(i)+(\mu_2-\widehat{\mu}_2)1_{(m\vee\widehat{m},n]}(i)+(\mu_1-\widehat{\mu}_2)1_{(\widehat{m},m]}(i)+(\mu_2-\widehat{\mu}_1)1_{(m,\widehat{m}]}(i).\end{aligned}$$

The triangle inequality together with assumptions (4.6.13) now give for $r=1,2$

$$\frac{m\wedge\widehat{m}}{n}|\mu_1-\widehat{\mu}_1|^r+\frac{n-(m\vee\widehat{m})}{n}|\mu_2-\widehat{\mu}_2|^r+\frac{(m-\widehat{m})_+}{n}|\mu_1-\widehat{\mu}_2|^r$$
$$+\frac{(\widehat{m}-m)_+}{n}|\mu_2-\widehat{\mu}_1|^r\ll o_P(1)+\frac{|m-\widehat{m}|}{n}|d|^r=o_P(1),$$

where $a_+=\max(a,0)$. By Lemma 4.4.1

$$\frac{1}{n^3}\sum_{l=1}^{\tilde{n}}(\mu_1-\widehat{\mu}_1)^4\left(\sum_{j=1}^{m\wedge\widehat{m}}c_j(l)\right)^4\ll\frac{\min(m,\widehat{m})^3}{n^2}(\mu_1-\widehat{\mu}_1)^4\ll n(\mu_1-\widehat{\mu}_1)^4=O_P(1).$$

Analogously we get

$$\frac{1}{n^3}\sum_{l=1}^{\tilde{n}}(\mu_2-\widehat{\mu}_2)^4\left(\sum_{j=m\vee\widehat{m}}^{n}c_j(l)\right)^4=O_P(1).$$

Moreover for $\widehat{m}<m\leqslant n$

$$(\mu_1-\widehat{\mu}_2)^4\frac{1}{n^3}\sum_{l=1}^{\tilde{n}}\left(\sum_{j=\widehat{m}+1}^{m}c_j(l)\right)^4\ll(\mu_1-\widehat{\mu}_2)^4\frac{(m-\widehat{m})_+^3n}{n^3}$$
$$\ll(\mu_2-\widehat{\mu}_2)^4\frac{|m-\widehat{m}|^3}{n^2}+|d|^4\frac{|m-\widehat{m}|^3}{n^2}=O_P(1).$$

Analogously for $m<\widehat{m}$

$$(\mu_2-\widehat{\mu}_1)^4\frac{1}{n^3}\sum_{l=1}^{\tilde{n}}\left(\sum_{j=m+1}^{\widehat{m}}c_j(l)\right)^4=O_P(1).$$

Under assumptions (4.6.14) analogous arguments hold. This completes the proof. ■

4.7. Frequency Bootstrap with Replacement

The question remains whether the bootstrap with replacement is also possible in this setup. So instead of permuting the Fourier coefficients we sample from them with replacement. It turns out that this approach gives the same results as above. This is, however, hardly surprising considering that one often proves rank statistic results by deriving them from the corresponding results for the statistic with replacement. One example is the proof of the Lindeberg condition for rank statistics, confer Theorems 3.1 and 4.1 of Hájek [39].

The proofs for the frequency bootstrap with replacement are analogous although one has to use the following two lemmas instead of the rank statistic results of Appendix E.

Lemma 4.7.1. *Let $\{U_i\}_{i=1,\ldots,n}$ be a triangular array of row-wise i.i.d. r.v.´s fulfilling $P(U_i = k) = \frac{1}{n}$ for $k = 1, \ldots, n$.*

a) $\mathrm{E}\left(\sum_{i=1}^n d_i a(U_i)\right) = n\bar{d}_n\bar{a}_n$, where $\bar{d}_n = \frac{1}{n}\sum_{i=1}^n d_i$ and an analogous expression for $\bar{a}_n$.

b) $\mathrm{var}\left(\sum_{i=1}^n d_i a(U_i)\right) = \sum_{i=1}^n d_i^2 \frac{1}{n}\sum_{j=1}^n \left(a(j) - \bar{a}_n\right)^2$.

c) $\mathrm{E}\left(\sum_{i=1}^n d_i(a(U_i) - \bar{a}_n)\right)^4 \ll \frac{1}{n^2}s_{2d}^2 z_{2a}^2 + \frac{1}{n}s_{4d}z_{4a}$, where $s_{2d} := \sum_{j=1}^n d_i^2$, $z_{2a} := \sum_{j=1}^n (a(i) - \bar{a}_n)^2$, $s_{4d} := \sum_{j=1}^n d_i^4$ and $z_{4a} := \sum_{j=1}^n (a(i) - \bar{a}_n)^4$.

Proof. The first two assertions are obvious. For c) note that

$$\begin{aligned}
&\mathrm{E}\left((a(U_{i_1}) - \bar{a}_n)(a(U_{i_2}) - \bar{a}_n)(a(U_{i_3}) - \bar{a}_n)(a(U_{i_4}) - \bar{a}_n)\right) \\
&= \begin{cases} \frac{1}{n}\sum_{j=1}^n (a(j) - \bar{a}_n)^4, & i_1 = i_2 = i_3 = i_4, \\ \frac{1}{n^2}\left(\sum_{j=1}^n (a(j) - \bar{a}_n)^2\right)^2, & (i_{\pi(1)} = i_{\pi(2)}) \wedge (i_{\pi(3)} = i_{\pi(4)}) \wedge (i_{\pi(1)} \neq i_{\pi(3)}), \\ 0, & \textit{else}, \end{cases}
\end{aligned}$$

π is a permutation of $\{1, 2, 3, 4\}$. This yields

$$\begin{aligned}
&\mathrm{E}\left(\sum_{i=1}^n d_i(a(U_i) - \bar{a}_n)\right)^4 \\
&= \sum_{i_1,i_2,i_3,i_4} d_{i_1}d_{i_2}d_{i_3}d_{i_4}\,\mathrm{E}\left((a(U_{i_1}) - \bar{a}_n)(a(U_{i_2}) - \bar{a}_n)(a(U_{i_3}) - \bar{a}_n)(a(U_{i_4}) - \bar{a}_n)\right) \\
&\ll \frac{1}{n^2}s_{2d}^2 z_{2a}^2 + \frac{1}{n}s_{4d}z_{4a}.
\end{aligned}$$

■

The central limit theorem for a triangular array of row-wise independent r.v.´s states:

Lemma 4.7.2. *If the Lindeberg condition*

$$\frac{1}{s_{2d}z_{2a}} \sum_{\substack{i,j \\ |d_i(a(j) - \bar{a}_n)| \geqslant \epsilon\sqrt{s_{2d}z_{2a}/n}}} \left(d_i(a(j) - \bar{a}_n)\right)^2 \to 0$$

for all $\epsilon > 0$ is fulfilled, then $\sum_{i=1}^n d_i a(U_i)$ is asymptotically normal.

Note that the Lindeberg condition is essentially the same as before, the only difference being that we are working with the uncentered sums s_{2d}, s_{4d} instead of the corresponding centered sums z_{2d}, z_{4d}. Lemma 4.7.1 shows that the moments also have the same upper bounds if one replaces the centered sums z_{2d}, z_{4d} by their uncentered counterpart s_{2d}, s_{4d}. Thus going carefully through the proofs of Lemma 4.5.1 as well as Theorems 4.5.1, 4.5.2, and 4.5.3 one notes that the proofs remain valid for the above case, i.e. for the score processes corresponding to the bootstrap with replacement, under the same assumptions on the scores as before. In fact the proofs are even somewhat easier.

Consequently, Theorem 4.5.4 and Corollary 4.5.1 remain true when we replace the permutations $(R_1, \ldots, R_n)$ by random vectors $(U_1, \ldots, U_n)$ as above. This finally shows that the bootstrap with replacement also has the wanted asymptotic behavior. More precisely Theorem 4.6.2, 4.6.3 as well as Corollary 4.6.1 remain true if again we replace the permutations $\mathbf{R}$ by random vectors $(U_1, \ldots, U_n)$ as above and independent of $X_1, \ldots, X_n$.

4.8. Future Research

In this section we discuss some possible variations of the method. For example the above algorithm gives a bootstrap sample that is close to the independent case. It may be possible to first estimate the spectral density and then get a bootstrap sample that preserves the covariance structure much better. Also the question remains whether we can or cannot use $\alpha(n) = 1$. Finally we discuss some other areas from change-point analysis where the method could also be useful.

4.8.1. Frequency Bootstrap Under Knowledge of Spectral Density?

We have seen theoretically as well as in the simulation study that the above approach essentially creates another sample of the time series close to an independent one. The reason supposedly is that permuting the Fourier coefficients destroys the covariance structure of the original sequence, which is coded in the variances of the Fourier coefficients – multiples of the spectral density at that point.

So it stands to reason that we might get an even better bootstrap if we divide the coefficients by the square root of the spectral density, permute and then multiply with that value again. Dahlhaus and Janas [22] follow that approach in the context of ratio statistics, so do Franke and Härdle [30] for kernel spectral density estimates. Of course it is immediately obvious that it is an even more complicated task to get good estimators for the spectral density than for the single value $\sigma^2 \left(\sum_{s \geqslant 0} w_s\right)^2$. And of course the correct density will usually be unknown in applications.

Let $\tilde{g}$ be the spectral density and $g(2j-1) = g(2j) = \tilde{g}(\frac{2\pi j}{n})$ the spectral density at the point $\frac{2\pi j}{n}$. Then the above bootstrap gives the following bootstrap sequence

$$\check{X}_{\mathbf{R}}(s) = \frac{2}{n} \sum_{l=1}^{\tilde{n}} c_s(l) \sqrt{g(l)} \left(\sum_{j=1}^{n} \widetilde{X}(j) \frac{c_j(R_l)}{\sqrt{g(R_l)}} - \frac{1}{\tilde{n}} \sum_{k=1}^{\tilde{n}} \sum_{i=1}^{n} \widetilde{X}(i) \frac{c_i(k)}{\sqrt{g(k)}} \right), \quad (4.8.1)$$

which can be seen analogously to equation (4.3.1).

What we would like now is that the above sequence has asymptotically the correct covariance structure conditionally on the given data and not only the covariance structure of a corresponding independent sequence. This would then give the desired improvement. The algorithm then compares the value of the null statistic without use of a variance estimator with the value of the corresponding bootstrap statistic involving

$$\check{Z}_n^{\mathbf{X}}(u) = \frac{1}{N} \sum_{s \leqslant Nu} \check{X}_{\mathbf{R}}(\beta(s))$$

again without standardizing.

To get the asymptotic behavior we want we need to choose successive random variables from the bootstrap sample, i.e. $\beta(s) = c + s$ for some constant $0 \leqslant c < n - N$. This is not surprising if we keep in mind that heuristically the bootstrap sample is close to an AR(1) series. Taking only every second element of an AR(1) sequence in general has a different covariance structure than the original AR(1) sequence. For the sake of simplicity we assume $c = 0$.

The conditional covariance structure is then given by $\mathrm{E}(\check{Z}_n^{\mathbf{X}}(u)\check{Z}_n^{\mathbf{X}}(v)|X_1, \ldots, X_n)$ and we would like this value to converge to $\min(u, v)\sigma^2 \left(\sum w_s\right)^2$. It requires quite complicated calculations (e.g. multiple sums of products of the spectral density (or its inverse) with trigonometric functions) to obtain the correct asymptotic for the above covariance. Here, we will only discuss the case of an AR(1)-time series $X(i) = \rho X(i-1) + \epsilon(i)$, where the spectral density is known (confer e.g. Brockwell and Davis [13]), i.e.

$$\tilde{g}(x) = \frac{\sigma^2}{2\pi}(1 - 2\rho\cos(x) + \rho^2)^{-1}.$$

This gives already a very good impression of the difficulties that arise in the general case.

To prove the validity of the bootstrap we need similar results as in the proof of the following lemma. The important tools are results for sums of trigonometric functions multiplied by different powers (positive and negative) of the spectral density, similar to the results obtained in Section 4.4. It seems difficult to get such results for the spectral density of a general linear process, even more difficult to get them for an estimator thereof. The following lemma only needs a small part of these results – and only for the known density of an AR(1)-process. It is, however, already complicated to get these results, so that one can expect difficulties in the general case.

Remark 4.8.1. There is one thing noteworthy about the proof concerning the question about $\alpha(n) = 1$ discussed in more detail in Section 4.8.2. For the bootstrap without the frequency density we get the covariance structure of a Brownian bridge instead of a Wiener Process if $\alpha(n) = 1$. The reason is that the part where $s_1 \neq s_2$ converges to uv if $\alpha(n) = 1$ and to 0 otherwise. Now using the above approach we note that even for $\alpha(n) \to \infty$, which is needed for this proof to hold true, the part where $s_1 \neq s_2$ does not converge to 0 but is essential to obtain the correct asymptotic. Without that term the asymptotic covariance would correspond again to the independent case. We can see this by equation (4.8.4). This might be a hint that we need $\alpha(n) \to \infty$ also in the other bootstrap.

Lemma 4.8.1. *Let $X(\cdot)$ be an AR(1) time series with parameter ρ ($0 < |\rho| < 1$, $w_s = \rho^s$) and $\check{X}_{\mathbf{R}}(s)$ as in (4.8.1), where we replace $\widetilde{X}(\cdot)$ by $X(\cdot)$. Let the innovations be i.i.d. fulfilling (3.3.2) with $\nu > 4$ and (3.3.4), furthermore $\frac{\log n}{\sqrt{N}} \to 0$ and $\frac{\log n}{\alpha(n)} \to 0$. Then as $n \to \infty$*

$$\mathrm{E}(\check{Z}_n^{\mathbf{X}}(u)\check{Z}_n^{\mathbf{X}}(v)|X_1,\ldots,X_n) \to \min(u,v)\frac{\sigma^2}{(1-\rho)^2} \qquad a.s.$$

Proof. First of all Lemma E.1 shows

$$\begin{aligned}
&\mathrm{E}(\check{Z}_n^{\mathbf{X}}(u)\check{Z}_n^{\mathbf{X}}(v)|X_1,\ldots,X_n)\\
&= \frac{2}{n(\tilde{n}-1)}\sum_{l=1}^{\tilde{n}}\left(\sum_{j=1}^{n}X(j)\frac{c_j(l)}{\sqrt{g(l)}} - \frac{1}{\tilde{n}}\sum_{t=1}^{\tilde{n}}\sum_{i=1}^{n}X(i)\frac{c_i(t)}{\sqrt{g(t)}}\right)^2\\
&\quad\cdot\frac{2}{nN}\sum_{k=1}^{\tilde{n}}\left(\sum_{s_1\leqslant Nu}c_{s_1}(k)\sqrt{g(k)} - \frac{1}{\tilde{n}}\sum_{t=1}^{\tilde{n}}\sum_{s_{11}\leqslant Nu}c_{s_{11}}(t)\sqrt{g(t)}\right)\\
&\qquad\cdot\left(\sum_{s_2\leqslant Nv}c_{s_2}(k)\sqrt{g(k)} - \frac{1}{\tilde{n}}\sum_{t=1}^{\tilde{n}}\sum_{s_{21}\leqslant Nv}c_{s_{21}}(t)\sqrt{g(t)}\right) \qquad a.s.
\end{aligned} \tag{4.8.2}$$

Concerning the second sum we note that uniformly in u

$$\frac{1}{\sqrt{N}n}\sum_{l=1}^{\tilde{n}}\sum_{s\leqslant Nu}c_s(l)\sqrt{g(l)} = o(1). \tag{4.8.3}$$

For the proof note that g is monotone in $[0,\pi]$ and for $|\rho| < 1$ there exist m, M with $0 < m \leqslant \tilde{g}(x) \leqslant M < \infty$ for all x. Lemma 4.4.2 and partial summation give

$$\begin{aligned}
&\frac{1}{\sqrt{N}n}\sum_{l=1}^{\tilde{n}}\sum_{s\leqslant Nu}c_s(l)\sqrt{g(l)}\\
&= \frac{1}{\sqrt{N}n}\sum_{s\leqslant Nu}\sum_{l=1}^{\tilde{n}/2}\sqrt{\tilde{g}(l)}\left[\cos\left(\frac{2\pi ls}{n}\right) - \sin\left(\frac{2\pi ls}{n}\right)\right]\\
&= \frac{1}{\sqrt{N}n}\sum_{s\leqslant Nu}\left[\sum_{l=1}^{\tilde{n}/2}\left[\cos\left(\frac{2\pi ls}{n}\right) - \sin\left(\frac{2\pi ls}{n}\right)\right]\sqrt{\tilde{g}(\tilde{n}/2)}\right.\\
&\qquad\left. - \sum_{k\leqslant\frac{\tilde{n}}{2}-1}\left(\sum_{t=1}^{k}\left[\cos\left(\frac{2\pi ts}{n}\right) - \sin\left(\frac{2\pi ts}{n}\right)\right]\right)\left[\sqrt{\tilde{g}(k+1)} - \sqrt{\tilde{g}(k)}\right]\right]\\
&\ll \frac{\log n}{\sqrt{N}}\left(1 + \sum_{k\leqslant\frac{\tilde{n}}{2}-1}\left|\sqrt{\tilde{g}(k+1)} - \sqrt{\tilde{g}(k)}\right|\right) \ll \frac{\log n}{\sqrt{N}} = o(1)
\end{aligned}$$

uniformly in u.

Concerning the main part of the second sum we can replace the sum (over k) by an integral, which we can then calculate. For this we need the following equality, which e.g. can be found in Gradshteyn et al. [38] 1.352

$$\sum_{k=1}^{m-1} k\sin(kx) = \frac{\sin(mx)}{4\sin^2(x/2)} - \frac{m\cos((2m-1)x/2)}{2\sin(x/2)} = \begin{cases} O\left(\frac{m}{x}\right), & 0 < x < \frac{\pi}{2}, \\ O(m), & \frac{\pi}{2} < x < \pi. \end{cases}$$

Note that $\sup_{0<x<\pi/4} \frac{x}{\sin(x)} = O(1)$, $\sup_{x>0} \left|\frac{\sin(x)}{x}\right| = O(1)$, $\sup_{\pi/2<x<\pi} \frac{1}{\sin^2(x/2)} \leqslant \frac{1}{\sin^2(\pi/4)}$, and $|\sin(m(\pi-x))| = |\sin(mx)|$.

Noting that $\cos(x)\cos(y) + \sin(x)\sin(y) = \cos(x-y)$ the mean value theorem gives

$$\begin{aligned}
&\frac{1}{N}\sum_{s_1,s_2}\left(\frac{2\pi}{n}\sum_{k=1}^{\tilde{n}/2}\frac{\cos(2\pi k(s_1-s_2)/n)}{1-2\rho\cos(2\pi k/n)+\rho^2} - \int_0^\pi \frac{\cos(x(s_2-s_1))}{1-2\rho\cos x+\rho^2}\,dx\right) \\
&= \frac{1}{N}\sum_{s_1,s_2}\left(\sum_{k=1}^{\tilde{n}/2}\int_{2\pi(k-1)/n}^{2\pi k/n}\left(\frac{\cos(2\pi k(s_1-s_2)/n)}{1-2\rho\cos(2\pi k/n)+\rho^2} - \frac{\cos(x(s_2-s_1))}{1-2\rho\cos x+\rho^2}\right)dx\right. \\
&\qquad \left. - \int_{\pi\tilde{n}/n}^{\pi}\frac{\cos(x(s_2-s_1))}{1-2\rho\cos x+\rho^2}\,dx\right) \\
&\ll \frac{1}{\alpha(n)} + \frac{1}{Nn}\sum_{k=2}^{\tilde{n}/2-1}\int_{2\pi(k-1)/n}^{2\pi k/n}\left|\sum_{s_1,s_2}(s_2-s_1)\sin(\xi(s_2-s_1))\right|dx \\
&\ll \frac{1}{\alpha(n)} + \frac{N}{n}\sum_{k=1}^{\tilde{n}/4}\frac{1}{k} \ll \frac{\log(n)}{\alpha(n)} = o(1).
\end{aligned}$$

It holds (cf. e.g. Bronstein et al. [14], p. 1053) for $|\rho| < 1$

$$\int_0^\pi \frac{\cos(x(s_2-s_1))}{1-2\rho\cos x+\rho^2}\,dx = \frac{\pi\rho^{|s_2-s_1|}}{1-\rho^2}.$$

Thus we get (w.l.o.g. $u \leqslant v$)

$$\begin{aligned}
&\frac{\sigma^2}{2\pi^2 N}\sum_{s_1,s_2}\int_0^\pi \frac{\cos(x(s_2-s_1))}{1-2\rho\cos(x)+\rho^2}\,dx \\
&= \frac{\sigma^2}{2\pi(1-\rho^2)N}\sum_{s_1=1}^{\lfloor Nu\rfloor}\left(\sum_{s_2=1}^{s_1}\rho^{s_1-s_2} + \sum_{s_2=s_1+1}^{\lfloor Nv\rfloor}\rho^{s_2-s_1}\right) \\
&= \frac{\sigma^2}{2\pi(1-\rho^2)N}\sum_{s_1=1}^{\lfloor Nu\rfloor}\left(\frac{1-\rho^{s_1}}{1-\rho} + \rho\,\frac{1-\rho^{\lfloor Nv\rfloor - s_1}}{1-\rho}\right) \\
&= \frac{\sigma^2}{2\pi(1-\rho^2)}\left(u\frac{1}{1-\rho} + u\frac{\rho}{1-\rho}\right) + o(1) = \frac{\sigma^2\min(u,v)}{2\pi(1-\rho)^2} + o(1).
\end{aligned} \tag{4.8.4}$$

Noting that $\sum_{i=1}^n (X(i)-\bar{X})(Y(i)-\bar{Y}) = \sum_{i=1}^n X(i)Y(i) - n\bar{X}\bar{Y}$ equations (4.8.3) and

(4.8.4) give

$$\frac{2}{nN}\sum_{k=1}^{\tilde n}\left(\sum_{s_1\leqslant Nu} c_{s_1}(k)\sqrt{g(k)}-\frac{1}{\tilde n}\sum_{t=1}^{n}\sum_{s_{11}\leqslant Nu} c_{s_{11}}(t)\sqrt{g(t)}\right)$$
$$\cdot\left(\sum_{s_2\leqslant Nv} c_{s_2}(k)\sqrt{g(k)}-\frac{1}{\tilde n}\sum_{t=1}^{n}\sum_{s_{21}\leqslant Nv} c_{s_{21}}(t)\sqrt{g(t)}\right) \tag{4.8.5}$$
$$=\frac{\sigma^2\min(u,v)}{2\pi(1-\rho)^2}+o(1).$$

Concerning the first sum of the covariance decomposition we get analogous to the proof of equation (4.8.3)

$$\frac{1}{n^{3/2}}\sum_{t=1}^{\tilde n}\sum_{i=1}^{n}X(i)\frac{c_i(t)}{\sqrt{g(t)}}\ll\frac{1}{\sqrt n}\sum_{i=1}^{n}|X(i)|\left(\frac{1}{i}+\frac{1}{n-i}\right)$$
$$\ll\frac{\log n}{\sqrt n}\max_{1\leqslant i\leqslant n}|X(i)|\ll\frac{\log n}{n^{\tilde\delta/(2(2+\tilde\delta))}}\left(\frac{1}{n}\sum_{i=1}^{n}|X(i)|^{2+\tilde\delta}\right)^{1/(2+\tilde\delta)}=o(1)\quad a.s.\qquad a.s., \tag{4.8.6}$$

where the last line follows by (3.5.8). Moreover we have

$$\frac{2}{n\tilde n}\sum_{j=1}^{n}X(j)^2\sum_{l=1}^{\tilde n}\frac{c_j^2(l)}{g(l)}\to 2\pi\frac{1+\rho^2}{1-\rho^2}\qquad a.s., \tag{4.8.7}$$

since Theorem C.3 gives $\frac{1}{n}\sum_{j=1}^{n}X(j)^2\to\frac{\sigma^2}{1-\rho^2}\quad a.s.$ and

$$\frac{2}{\tilde n}\sum_{l=1}^{\tilde n}\frac{c_j^2(l)}{g(l)}=\frac{2}{\tilde n}\sum_{l=1}^{\tilde n/2}\frac{1}{\tilde g(l)}=\frac{1}{\pi}\int_0^{\pi}\frac{1}{\tilde g(x)}\,dx+o(1)=\frac{2\pi}{\sigma^2}(1+\rho^2)+o(1).$$

Finally

$$\frac{2}{n\tilde n}\sum_{j=1}^{n}\sum_{i\neq j}X(j)X(i)\sum_{l=1}^{\tilde n}\frac{c_j(l)c_i(l)}{g(l)}\to-2\pi\frac{2\rho^2}{1-\rho^2}\qquad a.s., \tag{4.8.8}$$

since because of $\cos(x)\cos(y)=\frac{1}{2}\left(\cos(x-y)+\cos(x+y)\right)$ it holds

$$\sum_{l=1}^{\tilde n}\frac{c_j(l)c_i(l)}{g(l)}=\frac{2\pi}{\sigma^2}\sum_{l=1}^{\tilde n/2}\cos(2\pi l(j-i)/n)\left(1-2\rho\cos(2\pi l/n)+\rho^2\right)$$
$$=\frac{2\pi}{\sigma^2}(1+\rho^2)\sum_{l=1}^{\tilde n}c_j(l)c_i(l)-\frac{2\pi}{\sigma^2}\rho\sum_{l=1}^{\tilde n/2}\left(\cos(2\pi l(j-i-1)/n)+\cos(2\pi l(j-i+1)/n)\right)$$
$$=\frac{2\pi}{\sigma^2}(1+\rho^2)\sum_{l=1}^{\tilde n}c_j(l)c_i(l)-\frac{2\pi}{\sigma^2}\rho\sum_{l=1}^{\tilde n}c_j(l)c_{i-1}(l)-\frac{2\pi}{\sigma^2}\rho\sum_{l=1}^{\tilde n}c_j(l)c_{i+1}(l).$$

A similar argument as in Remark 4.5.2 thus gives

$$
\begin{aligned}
&\frac{2}{n\tilde{n}}\sum_{j=1}^{n}\sum_{i\neq j}X(j)X(i)\sum_{l=1}^{n}\frac{c_j(l)c_i(l)}{g(l)}\\
&= -\frac{2\pi\rho}{\sigma^2}\left(\frac{1}{n}\sum_{j=2}^{n}X(j)X(j-1)+\frac{1}{n}\sum_{j=1}^{n-1}X(j)X(j+1)\right)+o(1)\\
&= \frac{-4\pi\rho^2}{1-\rho^2}+o(1) \qquad a.s.,
\end{aligned}
$$

where the last line follows because of Theorem C.3. Thus putting together equations (4.8.5) and (4.8.6) - (4.8.8), we get the desired asymptotic. ■

4.8.2. Bootstrapping with the Complete Sequence?

As already mentioned our natural choice for $\alpha(n)$ would be $\alpha(n) = 1$. The above theory, however, does not cover this case. The problem is that the Lindeberg condition is not fulfilled, because the Noether condition is not fulfilled (cf. Theorem E.1) as the following lemma shows.

Lemma 4.8.2. *The Noether condition for $\alpha(n) = 1$ is not fulfilled in this setup, more precisely*

$$
\frac{\sum_{l=1}^{\tilde{n}}\left(d_n(l)-\bar{d}_n\right)^2}{\max_{l=1,\dots,\tilde{n}}\left(d_n(l)-\bar{d}_n\right)^2}=O(1),
$$

where $d_n(l) = \sum_{s=1}^{\lfloor nu\rfloor} c_s(l)$, $u \in (0,1)\setminus\{0.5\}$.

Proof. First of all $\frac{1}{n^2}\sum_{l=1}^{\tilde{n}}\left(d_n(l)-\bar{d}_n\right)^2 = \frac{1}{n^2}\sum_{l=1}^{\tilde{n}}\left(d_n(l)\right)^2-\frac{1}{n}\bar{d}_n^2$. Lemma 4.4.2 shows that the last part converges to 0:

$$
\frac{1}{n^{3/2}}\sum_{l=1}^{\tilde{n}}\sum_{s=1}^{\lfloor nu\rfloor}c_s(l)=O\left(\frac{1}{n^{3/2}}\sum_{s}\max\left(\frac{n}{s},\frac{n}{n-s}\right)\right)=o(1).
$$

Moreover Theorem 4.4.1 gives

$$
\frac{1}{n^2}\sum_{l=1}^{\tilde{n}}\left(\sum_{s=1}^{\lfloor nu\rfloor}c_s(l)\right)^2=\frac{1}{n^2}\sum_{s_1=1}^{\lfloor nu\rfloor}\sum_{s_2=1}^{\lfloor nu\rfloor}\sum_{l=1}^{\tilde{n}}c_{s_1}(l)c_{s_2}(l)\to 1/2u-1/2u^2.
$$

Besides equation 1.342 of Gradshteyn and Ryzhik [38] gives for the maximum

$$
\begin{aligned}
&\frac{1}{n}\max_{l=1,\dots,\tilde{n}}\left|\sum_{s=1}^{\lfloor nu\rfloor}c_s(l)-\bar{d}_n\right|\geqslant\frac{1}{n}\left|\sum_{s=1}^{\lfloor nu\rfloor}\cos(2\pi s/n)\right|-\frac{1}{n}|\bar{d}_n|\\
&=\frac{1}{n}\left|\frac{\cos\left(\frac{(\lfloor nu\rfloor+1)\pi}{n}\right)\sin\left(\frac{\lfloor nu\rfloor\pi}{n}\right)}{\sin\left(\frac{\pi}{n}\right)}\right|+o(1)\to\frac{1}{\pi}|\cos(\pi u)\sin(\pi u)|>0.
\end{aligned}
$$

Putting everything together we arrive at the assertion. ■

Theorem E.1 shows that the Lindeberg condition is equivalent to asymptotic normality, if the Noether condition is fulfilled. This means that in the present setup – even though we know, the Lindeberg condition is not fulfilled – it is still possible to have asymptotic normality. Indeed one would expect that because of the following facts.
We know that any finite number of the Fourier coefficients themselves are already asymptotically normal (cf. Theorem 4.2.1).
Also we have seen that any finite number of the backtransformed bootstrap sample belonging to any frequency ($\neq 0$) is asymptotically normal (cf. Theorem 4.5.2).
Last but not least we also know that using $n/\alpha(n)$ of the backtransformed r.v.´s (instead of a given number of them) we still have asymptotic normality as long as $\alpha(n) \to \infty$, but no matter how slowly (cf. Theorem 4.5.4). It also does not matter, which ones we choose.

As a contrast there is also some thoughts that support the opposite idea, namely that the bootstrap really only holds true for $\alpha(n) \to \infty$. The first one is the asymptotic behavior of the covariance structure if one takes into account the frequency density, confer Remark 4.8.1.
The second thought concerns trigonometric series estimates of densities. Suppose we have observed the i.i.d. sample $Y(1), \dots, Y(n)$. An estimate of the density of $Y(1)$ is given by

$$\frac{1}{2\pi}\left(1 + 2\sum_{j=1}^{m}(\widehat{a}_j \cos(jx) + \widehat{b}_j \sin(jx))\right),$$

where $\widehat{a}_j = \frac{1}{n}\sum_{i=1}^{n}\cos(j\,Y(i))$ and $\widehat{b}_j = \frac{1}{n}\sum_{i=1}^{n}\sin(j\,Y(i))$. In this situation a similar phenomenon happens, precisely Hall [41] shows that in this and some similar situations under certain conditions on the density the mean integrated square error (MISE) is $O(m/n) + o(m^{-r})$ for some $r > 0$ and this rate is exact. Something similar can be found in Anderson and De Figueiredo [1]. Thus we can only use m of the n Fourier coefficients to estimate the density with $m/n \to 0$.

What are the advantages of choosing $\alpha(n) = 1$?

Taking the complete backtransformed bootstrap sequence we can use $\widetilde{Z}_n$ instead of $\widetilde{Z}_n - t\widetilde{Z}_n(1)$. This sum has asymptotically already the same covariance structure as a Brownian bridge as the following Lemma shows. Also note that $\widetilde{Z}_n(1) = 0$ already (confer Remark 4.3.1 – also 1.342 of Gradshteyn and Ryhzik [38] shows $\sum_{s=1}^{n} c_s(l) = 0$ for all $l = 1, \dots, \tilde{n}$).

Lemma 4.8.3. *For $\alpha(n) = 1$ it holds as $n \to \infty$ for all $0 \leqslant u, v \leqslant 1$ for all scores satisfying (4.5.2)*

$$\begin{aligned}
&\mathrm{E}\,\widetilde{Z}_n(u) = 0.\\
&\operatorname{cov}\left(\widetilde{Z}_n(u), \widetilde{Z}_n(v)\right) \to \min(u, v) - uv,
\end{aligned}$$

where

$$\widetilde{Z}_n(u) = \frac{2}{n^{3/2}}\sum_{l=1}^{\tilde{n}}\sum_{s\leqslant nu} c_s(l)\left[\sum_{j=1}^{n} x_j c_j(R_l) - \frac{1}{\tilde{n}}\sum_{k=1}^{\tilde{n}}\sum_{i=1}^{n} x_i c_i(k)\right].$$

Proof. Lemma E.1 gives a closed formula for the covariance. Lemma 4.4.1, equation (4.5.6), and the scores conditions (4.5.2) now give

$$\begin{aligned}
&\operatorname{cov}(\widetilde{Z}_n(u), \widetilde{Z}_n(v)) \\
&= \frac{2}{n^2} \sum_{l=1}^{\tilde{n}} \sum_{s_1=1}^{\lfloor nu \rfloor} c_{s_1}(l) \sum_{s_2=1}^{\lfloor nv \rfloor} c_{s_2}(l) - \frac{2}{n} \left(\frac{1}{\tilde{n}} \sum_{l=1}^{\tilde{n}} \sum_{s_1=1}^{\lfloor nu \rfloor} c_{s_1}(l) \right) \left(\frac{1}{\tilde{n}} \sum_{k=1}^{\tilde{n}} \sum_{s_2=1}^{\lfloor nv \rfloor} c_{s_2}(k) \right) \\
&= \min(u, v) - uv + o(1).
\end{aligned}$$

■

4.8.3. Frequency Bootstrap for Extreme Value Statistics

Another topic of interest is, how one can prove corresponding results for the extreme value statistics. For this one usually needs an approximation with rates, which the central limit theorem cannot give.

4.8.4. Bootstrapping Statistics Based on Periodograms

Picard, Giraitis and Leipus [33, 34] proposed some change point procedures to test whether there occurred a change in the spectral density of a sequence of random variables. Their statistics are already based on the periodogram and they show convergence in $D[0,1]$. The frequency bootstrap should therefore work in their setting. For more details confer Csörgő, Horváth [19], chapter 4.4.2, and references therein.

Part II.

Simulation Study

5. Change Analysis of Stochastic Processes under Strong Invariance

In this chapter we describe the results of a number of simulation studies conducted for the models under strong invariance of Chapter 2. There, we have proven that the permutation principle is asymptotically applicable for processes satisfying (2.1.1) or (2.1.7). The simulation study shows that the permutation method usually gives better results than the asymptotic test if performance is measured by $\alpha-$ and $\beta-$errors respectively. We present some tables giving simulated quantiles of the original statistic under the null as well as permutation quantiles for different alternatives. Although these tables contain a lot of information they are rather difficult to grasp. This is why we also use visual methods below to demonstrate the goodness of the procedures.

The simulations are implemented in the software package R, Version 1.2.3. Computation time is not a problem, even the calculation of the permutation quantiles of the statistics for the gradual test for a sequence of length 200 using 10 000 permutation takes less than 5 seconds on an Intel(R) Pentium 4, 2.66 GHz, 512 MB RAM. The calculation of that statistic is more expensive than for the classical CUSUM statistic in case of the abrupt change.

QQ-Plots

Quantile-quantile plots or QQ-plots are a common tool to compare two distributions. It is a scatterplot, i.e. a plot of the values of one variable against another, of the quantiles of one distribution against the quantiles of another one. If the distributions are similar the plot should be on the diagonal. Even when the data follows the same distribution there are some deviations in the ends, i.e. for very small or very large quantiles. A straight line that is not on the diagonal indicates that the distribution is similar but with a different mean and/or variance. The diagonal is given by the dotted line.

In our case we do the following:

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) *Permutation distribution:* Determine the empirical distribution function of the block permutation statistic based on 10 000 permutations conditioned on the realization of the model from the previous step.

90%	**95%**	**99%**
1.224	1.358	1.628

Table 5.1.1.: Asymptotic quantiles of $M_T^{(1)}$

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

The plot gives a good idea how well the permutation distribution of one specific realization matches the distribution of the original statistic under H_0.

Due to different computational complexities we use different values of t for different models.

It is possible that the performance of the permutation test is good even though the QQ-plot suggests that the match of the distributions is not that good yet. The reason is that the test only compares the permutation quantiles conditioned on one sequence of observations with the value of the original statistic for that exact sequence. Therefore we also use the following type of plots.

SPC-Plots

Size-power-curves or SPC-plots demonstrate the power of a test. They plot the empirical distribution function of the p-values of the statistic for the null hypothesis or a given alternative 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.

In our case we simulate t_1 processes following the given model. For each of these processes we use t_2 permutations to calculate the empirical distribution function of the permutation statistic given the generated process. We can then calculate the p-value of the generated process with respect to the permutational distribution (conditioned on that same process). Finally we plot the empirical distribution function of the so obtained t_1 p-values and obtain a SPC-plot.
Where applicable we also give the SPC-plot of the asymptotic test, i.e. we calculate the p-value with respect to the asymptotic distribution. Of course this is only possible if the distribution function of the limit is known.

Due to different computational complexities t_1 and t_2 differ from model to model.

5.1. Simulations for the Test of an Abrupt Change in the Mean

The following simulations are based on partial sums of normally distributed random variables (with variance 1) (confer Example 2.1.1), and on a Poisson process (confer

N	Partial sums 90%	95%	97.5%	99%	Poisson Process 90%	95%	97.5%	99%
100	1.165	1.295	1.409	1.564	1.156	1.283	1.398	1.554
200	1.190	1.328	1.446	1.603	1.182	1.301	1.426	1.586

Table 5.1.2.: Simulated critical values of $M_T^{(1)}$ (under the null hypothesis)

Example 2.1.2). More specifically, we simulate the increments of the partial sums as i.i.d. random variables, and the increments of the Poisson process are taken at times $1, 2, \ldots$ (instead of $i\frac{T}{N}$, $i = 1, \ldots, N$, since this means only a scaling of the underlying r.v.'s). Other than that, we use the following parameters:

- $N = 100, 200$
- $N^* = \frac{1}{4}N, \frac{1}{2}N, \frac{3}{4}N$

N	N*	d	Partial sums 90%	95%	97.5%	99%	Poisson Process 90%	95%	97.5%	99%
100		0	1.175	1.301	1.423	1.557	1.141	1.273	1.393	1.547
100	25	1	1.167	1.297	1.405	1.532	1.173	1.312	1.420	1.545
100	25	2	1.165	1.290	1.403	1.541	1.172	1.294	1.407	1.532
100	25	3	1.160	1.296	1.421	1.540	1.172	1.311	1.441	1.590
100	25	4	1.166	1.296	1.415	1.551	1.165	1.291	1.406	1.551
100	50	1	1.167	1.301	1.417	1.563	1.173	1.306	1.415	1.542
100	50	2	1.173	1.307	1.426	1.567	1.164	1.290	1.407	1.574
100	50	3	1.181	1.313	1.435	1.590	1.161	1.291	1.402	1.551
100	50	4	1.184	1.324	1.441	1.589	1.160	1.288	1.397	1.557
100	75	1	1.165	1.300	1.412	1.546	1.167	1.284	1.415	1.575
100	75	2	1.170	1.304	1.404	1.547	1.164	1.292	1.408	1.547
100	75	3	1.175	1.295	1.418	1.561	1.174	1.299	1.412	1.534
100	75	4	1.171	1.296	1.422	1.578	1.165	1.299	1.404	1.551
200		0	1.190	1.328	1.455	1.589	1.179	1.311	1.417	1.568
200	50	1	1.185	1.311	1.438	1.579	1.184	1.317	1.434	1.579
200	50	2	1.177	1.306	1.424	1.567	1.180	1.308	1.426	1.555
200	50	3	1.180	1.307	1.429	1.555	1.184	1.313	1.423	1.553
200	50	4	1.180	1.305	1.428	1.555	1.186	1.314	1.427	1.563
200	100	1	1.181	1.315	1.435	1.586	1.196	1.323	1.450	1.583
200	100	2	1.180	1.312	1.423	1.551	1.190	1.316	1.442	1.574
200	100	3	1.174	1.314	1.422	1.545	1.183	1.315	1.431	1.561
200	100	4	1.181	1.311	1.425	1.560	1.182	1.311	1.428	1.569
200	150	1	1.185	1.324	1.442	1.588	1.181	1.310	1.439	1.575
200	150	2	1.185	1.325	1.449	1.583	1.192	1.330	1.449	1.580
200	150	3	1.183	1.317	1.446	1.590	1.181	1.313	1.433	1.588
200	150	4	1.181	1.315	1.441	1.597	1.194	1.330	1.450	1.613

Table 5.1.3.: Simulated critical values of the permutation statistic $M_T^{(1)}(\mathbf{R})$

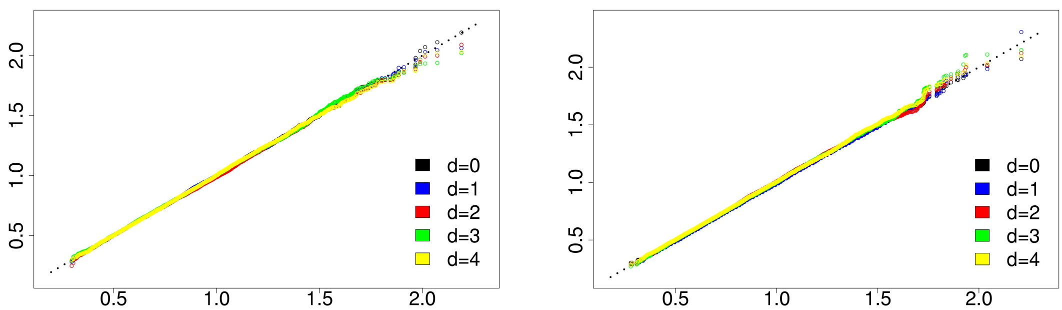
- $d := a^* - a = 0, 0.25, 0.5, 1, 2, 3, 4$

Here N^* is the change-point, and we are in the case of the null hypothesis for $d = 0$.

We generate 10 000 series of increments $\Delta Z_1, \ldots, \Delta Z_N$ corresponding to model (2.1.1) for different parameters under the null hypothesis. The resulting quantiles can be found in Table 5.1.2. The asymptotic critical values are given in Table 5.1.1 for comparison; they are too large. Moreover, the exact quantiles are somewhat larger for the partial sums than for the Poisson process.

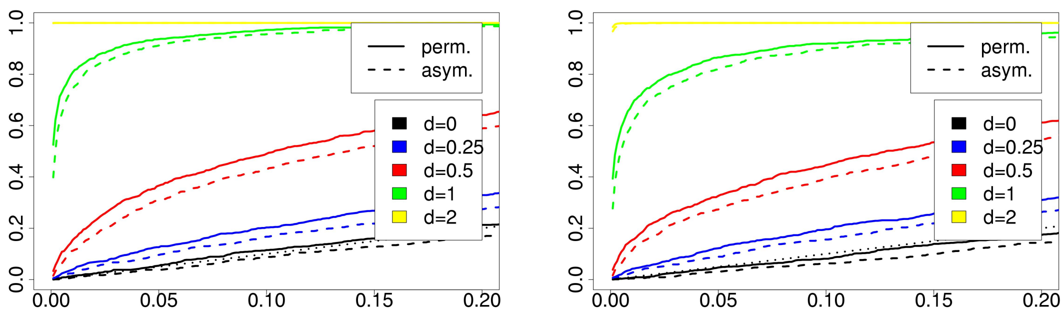
To study the critical values obtained from the permutation method we simulate one realization according to the given model and calculate the permutation quantiles based on 10 000 permutations for this realization. We do this for different realizations but use the same random numbers in each case to get a better idea of the stability of the procedure. The results can be found in Table 5.1.3.
These critical values give better estimates than the asymptotic ones. It also does not seem to be important where exactly the change point is located.

QQ-plots of the simulated null distribution versus different permutation distributions (each conditioned on one realization only) are to be found in Figure 5.1.1.



(1) Partial Sum (2) Poisson Process

Figure 5.1.1.: QQ-plots of $M_T^{(1)}$ (under H_0) against $M_T^{(1)}(\mathbf{R})$ for $N = 100$, $N^* = 75$



(1) Partial Sum (2) Poisson Process

Figure 5.1.2.: Size-power-curves of $M_T^{(1)}(\mathbf{R})$ with respect to the asymptotic distribution and with respect to the permutation distribution for $N = 100$, $N^* = 75$

The permutation distribution fits the null distribution perfectly. Moreover, the result does not depend on the alternative.

Next we are interested in how well the test performs – and also how well it performs in comparison to the asymptotic one. For this reason we create size-power-curves of both methods under the null hypothesis and under alternatives. We use 1 000 realizations of $\{Z(\cdot)\}$ and for each of these 10 000 permutations to calculate the permutation distribution.
The results are presented in Figure 5.1.2.

We have already seen that the asymptotic quantiles are too large by comparing them with the simulated ones of the original statistic under the null hypothesis. This is also confirmed by the size-power-curves which show that the actual level of the asymptotic test is somewhat too small. Even though both methods apparently perform well, we do have a better fit under the permutation method. Under the null hypothesis ($d = a^* - a = 0$), the solid line (representing the permutation method) fits better to the diagonal. Moreover, under alternatives the lines representing the permutation method are also steeper meaning that the power of this test is better than the power of the asymptotic one.

5.2. Simulations for the Test of a Gradual Change in the Mean

The simulations in this section are also based on partial sums of normally distributed r.v. (with variance 1) and on a Poisson process as in the previous section. The following parameters are used:

- $N = 100, 200$
- $N^* = \frac{1}{4}N,\ \frac{1}{2}N,\ \frac{3}{4}N$
- $D = 0, \frac{1}{4}, \frac{1}{2}, 1, 2, 4$

Here, N^* is the change-point, and the null hypothesis is given for $D = 0$. The parameter D has been rescaled in comparison to d in (2.1.7), more precisely, the increments of the change are chosen as $\frac{D}{(1+\gamma)N^\gamma}\left((i - N^*)_+^{1+\gamma} - ((i-1) - N^*)_+^{1+\gamma}\right)$. The latter expression depends on T only through N. The reason is that the magnitude of the parameter D

N	γ	90%	95%	97.5%	99%
100	0.5	1.738	2.150	2.554	3.082
100	1	2.298	2.710	3.114	3.643
100	2	2.130	2.542	2.946	3.474
200	0.5	1.868	2.263	2.649	3.155
200	1	2.353	2.747	3.134	3.640
200	2	2.192	2.586	2.973	3.479

Table 5.2.1.: Asymptotic critical values of $M_T^{(2)}$

is then comparable (as can be seen via the mean value theorem) to the parameter δ in Hušková and Steinebach [48], where $X(i) = \mu + \delta\left(\frac{i-m}{n}\right)_+^\gamma + e(i)$ for an i.i.d. sequence $\{e(\cdot)\}$.

Once again, we generate 10 000 series of increments $\Delta S_1, \ldots, \Delta S_N$ under the null hypothesis for the various choices of parameters. The resulting quantiles can be found in Table 5.2.2. The asymptotic critical values are given in Table 5.2.1. This time the asymptotic quantiles are usually too small.

For comparison, we simulate the critical values obtained through the permutation method as before. Some results can be found in Table 5.2.3. The critical values are quite good but decline as the change becomes more obvious.

We create QQ-plots of the simulated null distribution versus various permutation distributions in order to get an idea how well the approximation fits. The results can be found in Figure 5.2.1.

Here, the matches (and thus the critical values) are quite good but decline, if $\gamma < 1$ as the change becomes more obvious. On the other hand this leads to a greater power of the test, since the critical values are only too small if we are already under an alternative.

Moreover we have some kind of "step behavior" for the Poisson process. Apparently there are several permutations leading to the same maximal value (i.e. the value of the statistic). This, however, does not seem to influence the accuracy of the quantiles as the size-power-curves, below, show. Remember that there are 10 000 points in the plot.

We create size-power-curves of the asymptotic method as well as the permutation method. For $\gamma = 0.25$ we do not know the asymptotic quantiles, since $H_{0.25}$ is not known.
The results can be found in Figure 5.2.2.

First of all the test gives good results for $\gamma = 0.25$, where we do not have the asymptotic test available. Also for $\gamma = 0.5$ the permutation test performs quite well, while the α-errors of the asymptotic one are far too high, e.g. for partial sums we have an actual α-error of 40% for a nominal one of 10%. For $\gamma > 1$ both methods perform well, although the power under the permutation method is always greater than the power under the asymptotic method. The plot on the complete interval $(0, 1)$ also shows, that the asymptotic curve (in contrast to the permutational one) is too high between 0.15

		Partial sums				**Poisson Process**			
N	γ	**90%**	**95%**	**97.5%**	**99%**	**90%**	**95%**	**97.5%**	**99%**
100	0.25	2.589	2.858	3.134	3.435	2.66	2.978	3.249	3.647
100	0.5	2.46	2.742	3.011	3.281	2.496	2.82	3.162	3.543
100	1	2.307	2.613	2.866	3.168	2.336	2.679	3.048	3.454
100	2	2.225	2.527	2.809	3.14	2.226	2.584	2.885	3.385
200	0.25	2.625	2.898	3.154	3.441	2.659	2.968	3.286	3.671
200	0.5	2.481	2.787	3.009	3.319	2.505	2.824	3.167	3.499
200	1	2.355	2.625	2.864	3.215	2.355	2.692	3.042	3.502
200	2	2.247	2.527	2.806	3.09	2.282	2.664	3.008	3.439

Table 5.2.2.: Simulated critical values of $M_T^{(2)}$ (under the null hypothesis)

				Partial sums				Poisson process						Partial sums				Poisson process			
γ	**D**	**N**	**N***	**90%**	**95%**	**97.5%**	**99%**	**90%**	**95%**	**97.5%**	**99%**	**N**	**N***	**90%**	**95%**	**97.5%**	**99%**	**90%**	**95%**	**97.5%**	**99%**
0.25	0	100		2.667	2.905	3.127	3.397	2.584	2.974	3.302	4.015	200		2.606	2.859	3.114	3.366	2.575	2.863	3.117	3.574
0.25	0.25	100	25	2.67	2.94	3.17	3.426	2.447	2.774	3.169	3.758	200	50	2.608	2.864	3.094	3.371	2.557	2.836	3.1	3.606
0.25	0.5	100	25	2.712	2.943	3.186	3.454	2.452	2.779	3.175	3.646	200	50	2.595	2.882	3.071	3.382	2.53	2.82	3.048	3.408
0.25	1	100	25	2.677	2.891	3.13	3.408	2.421	2.729	3.067	3.287	200	50	2.518	2.763	2.985	3.267	2.427	2.7	2.916	3.18
0.25	2	100	25	2.362	2.61	2.778	3.027	2.181	2.381	2.593	2.859	200	50	2.203	2.417	2.611	2.839	2.123	2.31	2.497	2.714
0.25	4	100	25	1.65	1.774	1.924	2.095	1.551	1.726	1.876	2.043	200	50	1.539	1.688	1.84	1.996	1.484	1.635	1.782	1.924
0.25	0.25	100	50	2.666	2.931	3.153	3.42	2.436	2.703	3.073	3.435	200	100	2.592	2.866	3.066	3.354	2.421	2.662	2.953	3.307
0.25	0.5	100	50	2.662	2.927	3.16	3.397	2.432	2.704	3.078	3.358	200	100	2.585	2.85	3.043	3.334	2.364	2.632	2.905	3.166
0.25	1	100	50	2.585	2.845	3.097	3.336	2.339	2.635	3.002	3.115	200	100	2.492	2.754	2.936	3.225	2.247	2.501	2.682	2.98
0.25	2	100	50	2.264	2.484	2.718	2.941	2.092	2.292	2.504	2.745	200	100	2.158	2.38	2.543	2.759	1.928	2.124	2.272	2.527
0.25	4	100	50	1.567	1.72	1.859	2.016	1.491	1.642	1.786	1.938	200	100	1.506	1.655	1.793	1.927	1.378	1.512	1.631	1.782
0.25	0.25	100	75	2.693	2.912	3.131	3.406	2.397	2.652	3.032	3.407	200	150	2.619	2.888	3.093	3.37	2.604	2.898	3.186	3.597
0.25	0.5	100	75	2.691	2.899	3.134	3.392	2.372	2.645	2.998	3.347	200	150	2.632	2.892	3.093	3.377	2.596	2.889	3.173	3.577
0.25	1	100	75	2.642	2.864	3.09	3.344	2.331	2.598	2.93	3.168	200	150	2.577	2.849	3.115	3.368	2.536	2.82	3.084	3.464
0.25	2	100	75	2.441	2.65	2.866	3.094	2.151	2.409	2.654	2.842	200	150	2.407	2.687	2.921	3.21	2.337	2.598	2.836	3.031
0.25	4	100	75	1.879	2.066	2.241	2.445	1.725	1.894	2.087	2.295	200	150	1.923	2.142	2.351	2.563	1.842	2.037	2.19	2.384
0.5	0	100		2.57	2.747	2.994	3.249	2.394	2.834	3.159	4.015	200		2.456	2.734	2.978	3.261	2.413	2.704	2.974	3.465
0.5	0.25	100	25	2.602	2.776	3.021	3.285	2.275	2.584	2.977	3.792	200	50	2.495	2.779	2.982	3.244	2.39	2.777	2.952	3.452
0.5	0.5	100	25	2.611	2.792	3.04	3.3	2.289	2.608	2.998	3.726	200	50	2.471	2.772	2.943	3.221	2.378	2.75	2.94	3.472
0.5	1	100	25	2.556	2.764	3.003	3.266	2.307	2.593	2.963	3.473	200	50	2.417	2.681	2.922	3.171	2.316	2.677	2.855	3.135
0.5	2	100	25	2.298	2.565	2.748	3.006	2.158	2.414	2.647	2.915	200	50	2.184	2.414	2.618	2.834	2.107	2.307	2.505	2.757
0.5	4	100	25	1.679	1.824	1.994	2.181	1.621	1.799	1.953	2.166	200	50	1.601	1.767	1.892	2.068	1.54	1.692	1.849	2.024
0.5	0.25	100	50	2.581	2.774	3.01	3.274	2.269	2.579	2.921	3.455	200	100	2.489	2.774	2.991	3.235	2.274	2.586	2.835	3.28
0.5	0.5	100	50	2.568	2.792	3.022	3.275	2.274	2.55	2.924	3.419	200	100	2.473	2.764	2.949	3.206	2.242	2.576	2.79	3.21
0.5	1	100	50	2.538	2.776	2.999	3.249	2.256	2.526	2.888	3.269	200	100	2.42	2.703	2.927	3.178	2.159	2.479	2.675	2.938
0.5	2	100	50	2.334	2.571	2.795	3.072	2.14	2.389	2.697	2.862	200	100	2.22	2.518	2.688	2.911	1.958	2.224	2.378	2.618
0.5	4	100	50	1.786	1.988	2.162	2.345	1.706	1.907	2.027	2.246	200	100	1.727	1.908	2.07	2.243	1.527	1.688	1.858	2.014
0.5	0.25	100	75	2.557	2.758	3	3.251	2.234	2.569	2.876	3.429	200	150	2.491	2.745	2.99	3.255	2.44	2.752	3.025	3.493
0.5	0.5	100	75	2.542	2.772	3.008	3.261	2.232	2.546	2.869	3.404	200	150	2.503	2.751	2.953	3.243	2.444	2.742	3.024	3.466
0.5	1	100	75	2.532	2.766	2.997	3.246	2.217	2.491	2.859	3.328	200	150	2.492	2.749	3.004	3.244	2.417	2.706	2.99	3.423
0.5	2	100	75	2.46	2.704	2.912	3.194	2.155	2.43	2.765	3.09	200	150	2.414	2.696	2.944	3.242	2.349	2.603	2.883	3.275
0.5	4	100	75	2.225	2.413	2.625	2.855	1.952	2.208	2.424	2.63	200	150	2.171	2.43	2.675	3.088	2.086	2.353	2.588	2.785
1	0	100		2.452	2.667	2.857	3.135	2.246	2.658	3.033	4.015	200		2.351	2.618	2.837	3.165	2.253	2.572	2.903	3.331
1	0.25	100	25	2.469	2.685	2.877	3.146	2.128	2.434	2.79	3.818	200	50	2.347	2.624	2.858	3.165	2.236	2.627	2.888	3.379
1	0.5	100	25	2.449	2.679	2.888	3.155	2.169	2.447	2.824	3.795	200	50	2.336	2.622	2.894	3.151	2.237	2.604	2.905	3.362
1	1	100	25	2.471	2.64	2.872	3.147	2.199	2.466	2.849	3.664	200	50	2.289	2.609	2.817	3.126	2.217	2.577	2.902	3.289
1	2	100	25	2.261	2.491	2.703	2.964	2.109	2.434	2.739	3.127	200	50	2.139	2.447	2.62	2.909	2.106	2.374	2.666	2.828
1	4	100	25	1.78	1.965	2.14	2.344	1.721	1.939	2.145	2.31	200	50	1.705	1.897	2.06	2.263	1.655	1.864	2.011	2.202
1	0.25	100	50	2.468	2.698	2.869	3.138	2.099	2.431	2.756	3.468	200	100	2.347	2.629	2.857	3.169	2.142	2.484	2.705	3.175
1	0.5	100	50	2.473	2.713	2.882	3.137	2.103	2.456	2.76	3.46	200	100	2.351	2.631	2.888	3.158	2.128	2.458	2.691	3.143
1	1	100	50	2.443	2.72	2.882	3.151	2.115	2.476	2.758	3.415	200	100	2.327	2.641	2.861	3.14	2.086	2.414	2.687	3.096
1	2	100	50	2.385	2.658	2.832	3.102	2.142	2.361	2.736	3.214	200	100	2.244	2.57	2.774	3.081	1.992	2.282	2.551	2.751
1	4	100	50	2.13	2.342	2.541	2.762	1.94	2.217	2.495	2.655	200	100	1.992	2.247	2.505	2.85	1.77	1.966	2.137	2.474
1	0.25	100	75	2.456	2.681	2.86	3.142	2.073	2.421	2.72	3.442	200	150	2.35	2.621	2.843	3.173	2.292	2.609	2.875	3.342
1	0.5	100	75	2.463	2.691	2.859	3.143	2.07	2.426	2.723	3.438	200	150	2.347	2.626	2.853	3.178	2.291	2.604	2.873	3.344
1	1	100	75	2.469	2.699	2.868	3.139	2.071	2.431	2.726	3.425	200	150	2.344	2.627	2.873	3.167	2.284	2.596	2.866	3.327
1	2	100	75	2.463	2.696	2.854	3.121	2.07	2.428	2.722	3.38	200	150	2.341	2.637	2.859	3.175	2.275	2.554	2.858	3.32
1	4	100	75	2.345	2.663	2.823	3.081	2.084	2.394	2.688	3.224	200	150	2.294	2.581	2.833	3.109	2.226	2.514	2.769	3.17

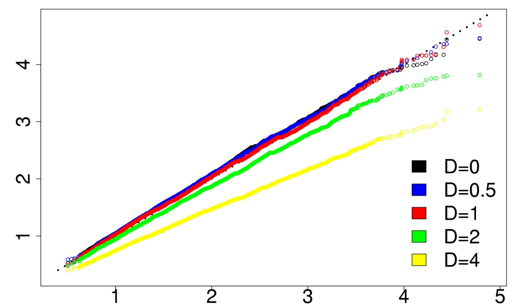
Table 5.2.3.: Simulated critical values of the permutation statistic $M_T^{(2)}(\mathbf{R})$

and 1. However, this is not a problem for the test, since one would hardly choose any critical value in that range.

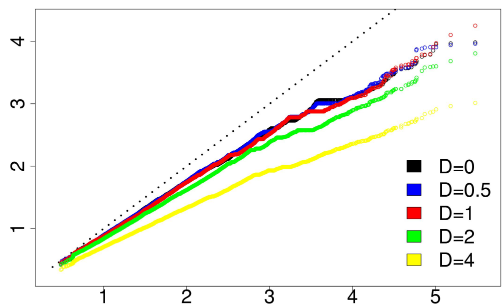
Furthermore the power declines with increasing γ; for $\gamma = 2$ it is almost impossible to distinguish between any alternatives. However, this is not surprising, since for $\gamma = 2$ (and $N^* = \frac{3}{4}N$) we have an effective mean difference of the increments of approximately $\frac{D}{16}$, which is not very much.

When we use d without rescaling i.e. as in (2.1.7) and $T = N$ (which changes d slightly), the critical values decrease significantly. Nevertheless, this does not seem to affect the permutation method at all – apparently the permutation quantiles are still smaller than the value of the test statistic for the unpermuted observations. With the asymptotic method, however, we only obtain good β-errors for smaller d's, but observe a sudden jump in the β-errors (up to 100%) as soon as d gets larger. For example this jump occurs at $d = 2$ for the 90%-quantile with $\gamma = 0.5$, $N = 100, 200$.

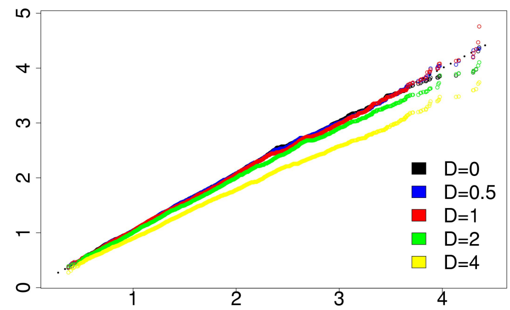
Note that here (in contrast to the i.i.d case) the consistency of the test is not guaranteed, since the estimator for b is unbounded under the alternative (which violates condition (2.4) of Steinebach [79]).



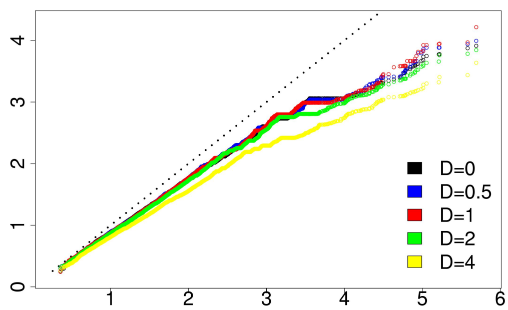
(1) Partial Sum, $\gamma = 0.25$



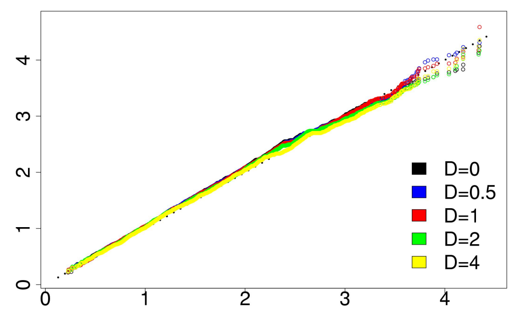
(2) Poisson Process, $\gamma = 0.25$



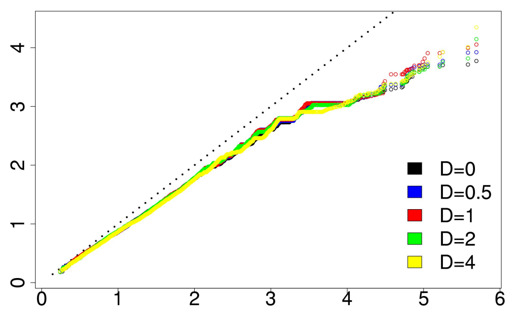
(3) Partial Sum, $\gamma = 0.5$



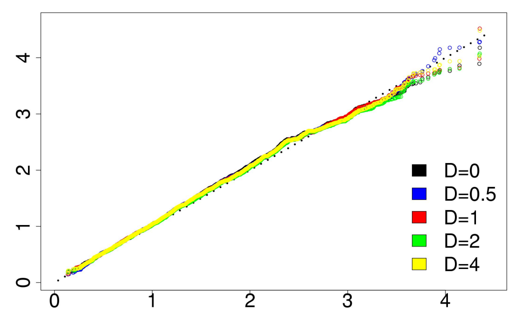
(4) Poisson Process, $\gamma = 0.5$



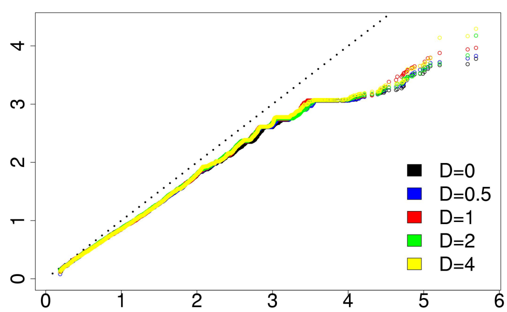
(5) Partial Sum, $\gamma = 1$



(6) Poisson Process, $\gamma = 1$

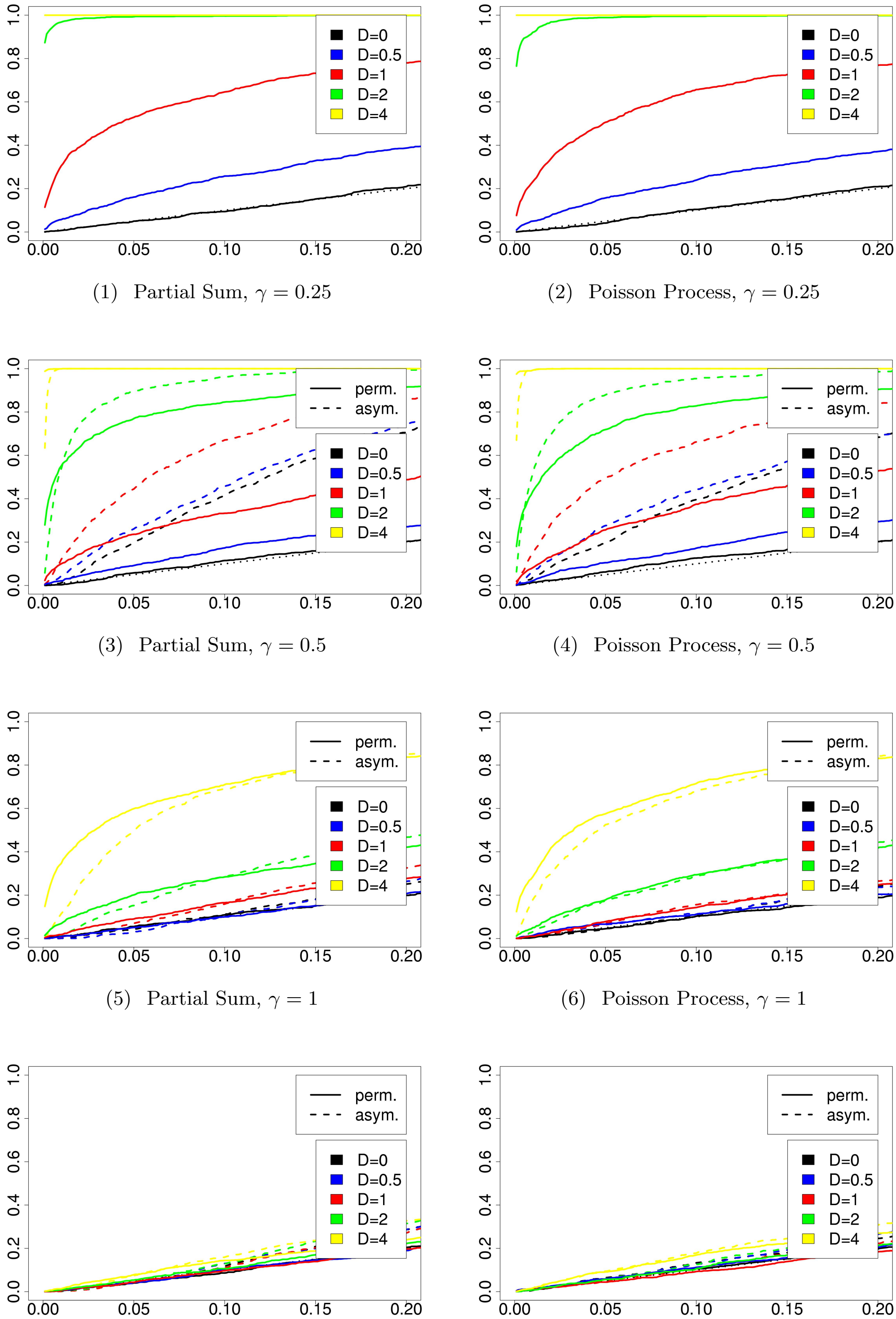


(7) Partial Sum, $\gamma = 2$



(8) Poisson Process, $\gamma = 2$

Figure 5.2.1.: QQ-plots of $M_T^{(2)}$ (under H_0) against $M_T^{(2)}(\mathbf{R})$ for $N = 100$, $N^* = 75$



(1) Partial Sum, $\gamma = 0.25$

(2) Poisson Process, $\gamma = 0.25$

(3) Partial Sum, $\gamma = 0.5$

(4) Poisson Process, $\gamma = 0.5$

(5) Partial Sum, $\gamma = 1$

(6) Poisson Process, $\gamma = 1$

(7) Partial Sum, $\gamma = 2$

(8) Poisson Process, $\gamma = 2$

Figure 5.2.2.: Size-power-curves of $M_T^{(2)}(\mathbf{R})$ with respect to the asymptotic distribution and with respect to the permutation distribution for $N = 100$, $N^* = 75$

6. Change Analysis of the Location Model with an AR(1)-Error-Sequence

In this chapter we describe the results of a simulation study concerning the block permutation method of Chapter 3 and the frequency permutation method of Chapter 4. So far we have only proven that they are asymptotically applicable. The purpose of the simulation study below is to determine how well they are actually working for small sample sizes n.
The linear process we implement is a causal AR(1) sequence. It turns out that the permutation methods usually behave better than the asymptotic method if performance is measure by $\alpha-$ and $\beta-$errors, respectively. Moreover the performance of the block permutation method is better than the frequency permutation as long as we choose an appropriate block length. Otherwise the goodness of the block permutation method drops significantly.
Due to limitation of space we will only give a small part of the simulation study, yet the results are typical. We present some tables giving simulated quantiles of the original statistic under the null as well as permutation quantiles for different alternatives, but we focus on QQ-plots and SPC-plots (for a short introduction confer Chapter 5) since they provide the information in a much more comprehensible way.

The simulations are implemented in the software package R, Version 1.8.1. Computation time is not a problem, even the calculation of the permutation quantiles of the frequency statistics for a sequence of length 80 using 10 000 permutation takes less than 5 seconds on an Intel(R) Pentium 4, 2.66 GHz, 512 MB RAM. Because of the Fourier transforms the calculation of that statistic is more expensive than of the block permutation statistic.

In a first section we summarize the outcome of the simulation study for all three tests – asymptotic , block permutation and frequency permutation test. We focus our attention on a comparison of the three methods.
The following two sections then give more details of the simulation study for the block permutation test respectively the frequency permutation test. We concentrate on the q-weighted CUSUM statistic and only give a small portion of the simulation results of the other three statistics.

6.1. Comparison of the Results of the Three Methods

Asymptotically all three methods are equivalent, yet there are huge differences in their small sample behavior.

The asymptotic critical values have the obvious advantage that they are easily obtained for the extreme value statistics and in some cases for the q-weighted CUSUM as well as

for sum statistics. In many other cases, however, they are not known theoretically. Furthermore a disadvantage of the asymptotic critical values is the rather slow convergence rates.
The main problem is to get a good estimator for $\sigma^2 \left(\sum w_s\right)^2$ which is needed to get the correct asymptotic for the original statistic. It is not easy to obtain such an estimator and the performance is usually rather bad. As a result the performance of the test also drops.
This can be seen by comparing for example the SPC-plots of the asymptotic test in Section 6.2 with the ones in Section 6.3 for statistic $T_n^{(3)}(q_1)$, $q_1 \equiv 1$. The only difference is that in the curves for the block bootstrap we use an estimator (more precisely the Bartlett log window estimator (confer (6.2.1))) for the above term, whereas in the frequency case we use the exact value.
The test performs much worse if we use the estimator instead of the real value. Moreover the quantiles are rather too conservative (high) if we use the correct value, but far too small if we use the estimator instead. In real life situations the exact value is usually not known.
Similarly the frequency permutation test also needs an estimator for $\sigma^2 \left(\sum w_s\right)^2$. The simulation study suggests that, when using the correct value, the test performs quite well even somewhat better than the asymptotic method. However, just like with the latter there will be a problem using an estimator. Since the quantiles of the frequency method are not as conservative as the asymptotic ones, the quantiles with estimator will be even worse.
So both methods should only be applied in real life situations when one has some additional information which leads to an improvement of the estimator. Then the frequency test usually performs better than the asymptotic one.

As a contrast the block bootstrap method is independent of an estimator. The reason is that the variance of the rank statistic, which is to be used in the bootstrap, is invariant under block permutations. On the other hand Lemma 3.5.1 shows that it is a valid estimator for $\sigma^2 \left(\sum w_s\right)^2$. Thus it can also be used in the original statistic. Because it is invariant under block permutations the value of the original statistic for the given sample is divided by the same value as the permutation statistic for each permutation. Since we are interested only in a comparison of the value of the original statistic for the given sample with some quantile of the block permutation statistic, this comparison is independent of the actual value of that estimator for the given observations. So the performance of the test is indeed independent of the performance of the estimator, which is a huge advantage.

Another drawback of both, the frequency method as well as the asymptotic method, is the independence of the critical values from the correlation of the underlying time series. For the asymptotic quantiles this is obvious, for the frequency method it is suggested by the simulation study. As a contrast the simulated correct quantiles do depend strongly on the underlying correlation. However, the QQ-plots as well as size-power-curves suggest that the frequency bootstrap performs slightly better than the asymptotic method in this regard.

The block bootstrap on the other hand takes the correlation much better into account. Yet it only does so, if we choose our block length appropriately. The simulations show that the block bootstrap depends crucially on a good choice of block length K. If chosen incorrectly it behaves worse than the asymptotic and the frequency permutation test.

However, taking a large block length is a safe choice if one either assumes the data to be dependent or is not sure about it. Even for independent data a large K gives good α–errors, yet the β–errors slightly increase.

In real life situations it seems best to use the block bootstrap with a large block length if uncertain. In case one has a better knowledge of the data (and thus has a good estimator for $\sigma^2 \left(\sum w_s\right)^2$) it may also be appropriate to use the frequency bootstrap.

For the MOSUM statistic it is always best to use the asymptotic quantiles, since they are known and easy to obtain. The block bootstrap works usually not so good for this statistic. The frequency bootstrap works better than the block method yet not as good as the asymptotic one.

6.2. Simulation of the Block Permutation Method

The simulations show that the results for the weighted CUSUM statistic $T^{(1)}$, the q-weighted CUSUM statistic $T^{(3)}(q)$, and the sum-statistic $T^{(4)}(r)$ are very similar. The block permutation test does behave better than the asymptotic method for an appropriately chosen block length K.

The asymptotic test, however, does perform better with the MOSUM statistic $T^{(2)}(G)$. The block permutation 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 of 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.

Following the advice of Antoch et al. (confer [4], Remark 3) we use the following Bartlett log window estimator as variance estimator in the asymptotic case:

$$\widetilde{\tau}_n^2(\Lambda) = \widehat{R}(0) + 2\sum_{k=1}^{\Lambda}\left(1 - \frac{k}{\Lambda}\right)\widehat{R}(k), \tag{6.2.1}$$

where

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

$\widehat{m} = \min\{\arg\max\{|S_k| : k = 1, \ldots, n\}\}$ and $\overline{X}^*_{\widehat{m}} = \frac{1}{n-\widehat{m}}\sum_{i=\widehat{m}+1}^{n} X_i$. For $n = 80$ we choose $\Lambda = 8$ ($\Lambda = 12$ for $n = 120$) in accordance with the results obtained by Antoch et al. [4].

For the block permutation test, however, we need to use estimator (3.5.1) in order to have the correct asymptotic behavior – conditioned on the observations. Since it is invariant under permutations, the original as well as permutation statistic is divided by the same value so that the actual value is irrelevant. This is an advantage, because the goodness of the test does not depend on the goodness of the estimator.

In the simulation study we use the model of Section 3.3, where $\{e(i) : i \geqslant 1\}$ forms an AR(1) sequence with autoregressive coefficient $\rho \in \{0, \pm 0.1, \pm 0.2, \pm 0.3, \pm 0.5, 0.7\}$ and $\{\epsilon(j) : -\infty < j < \infty\}$ are i.i.d. $N(0,1)$, hence $\tau = \frac{1}{1-\rho}$. We create the AR(1)-sequences recursively (i.e. $e(i) = \rho\, e(i-1) + \epsilon(i)$, where we throw the first 50 away). 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 $d \in \{0, 0.25, 0.5, 1, 2, 4\}$.

We discuss the q-weighted CUSUM and MOSUM statistics in more detail but only give a small yet typical selection of plots for the weighted CUSUM and Sum statistics, since the results are very similar. For more details confer Kirch [51].

6.2.1. Variance Estimation

In this subsection we investigate the estimator (3.5.1)

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

for $\tau^2 = \sigma^2 \left(\sum_{s \geqslant 0} w_s \right)^2$.

This is not exactly the same estimator as in Section 3.5, yet it is asymptotically equivalent (both under the null hypothesis and alternatives) with fast enough convergence rates. Moreover it is unbiased for i.i.d. errors and – as some simulations show – does behave better for all error sequences. Confer also Remark 3.5.1.

The purpose of the first simulations is to illustrate the behavior of the above estimator for different values of the block length K. Therefore we simulate AR(1) sequences with the parameters as above. Table 6.2.1.1 shows the mean of $\widehat{\tau}_{LK}$ of 10 000 repetitions (under the null hypothesis).

		n = 80		n = 120		n = 210				n = 80		n = 120		n = 210	
ρ	τ	K	$\widehat{\tau}_{LK}$	K	$\widehat{\tau}_{LK}$	K	$\widehat{\tau}_{LK}$	ρ	τ	K	$\widehat{\tau}_{LK}$	K	$\widehat{\tau}_{LK}$	K	$\widehat{\tau}_{LK}$
-0.5	0.67	1	1.15	1	1.16	1	1.15	0.1	1.11	1	1	1	1	1	1
-0.5	0.67	2	0.81	2	0.82	3	0.82	0.1	1.11	2	1.05	2	1.05	3	1.07
-0.5	0.67	4	0.76	4	0.76	5	0.75	0.1	1.11	4	1.07	4	1.07	5	1.08
-0.5	0.67	10	0.69	12	0.68	15	0.68	0.1	1.11	10	1.06	12	1.07	15	1.08
-0.3	0.77	1	1.05	1	1.05	1	1.05	0.2	1.25	1	1.01	1	1.02	1	1.02
-0.3	0.77	2	0.87	2	0.88	3	0.85	0.2	1.25	2	1.11	2	1.11	3	1.16
-0.3	0.77	4	0.82	4	0.83	5	0.81	0.2	1.25	4	1.16	4	1.17	5	1.19
-0.3	0.77	10	0.77	12	0.77	15	0.77	0.2	1.25	10	1.18	12	1.19	15	1.21
-0.2	0.83	1	1.02	1	1.02	1	1.02	0.3	1.43	1	1.04	1	1.04	1	1.05
-0.2	0.83	2	0.91	2	0.91	3	0.89	0.3	1.43	2	1.18	2	1.19	3	1.26
-0.2	0.83	4	0.87	4	0.87	5	0.86	0.3	1.43	4	1.28	4	1.3	5	1.32
-0.2	0.83	10	0.82	12	0.82	15	0.83	0.3	1.43	10	1.32	12	1.34	15	1.37
-0.1	0.91	1	1.00	1	1.01	1	1.00	0.5	2	1	1.13	1	1.14	1	1.15
-0.1	0.91	2	0.95	2	0.95	3	0.94	0.5	2	2	1.39	2	1.39	3	1.55
-0.1	0.91	4	0.92	4	0.93	5	0.92	0.5	2	4	1.2	4	1.63	5	1.7
-0.1	0.91	10	0.89	12	0.89	15	0.9	0.5	2	10	1.78	12	1.82	15	1.87
0	1	1	1	1	1	1	1	0.7	3.33	1	1.35	1	1.37	1	1.38
0	1	2	0.99	2	1	3	1	0.7	3.33	2	1.75	2	1.78	3	2.07
0	1	4	0.99	4	0.99	5	0.99	0.7	3.33	4	2.2	4	2.24	5	2.41
0	1	10	0.96	12	0.97	15	0.98	0.7	3.33	10	2.68	12	2.8	15	2.93

Table 6.2.1.1.: Mean values of $\widehat{\tau}_{LK}$ under the null hypothesis

Under strong correlation large values of K give best results. Also the greater n the better are the results, which is consistent with the asymptotic given in Lemma 3.5.1. However the convergence is rather slow.

6.2.2. q-weighted CUSUM Statistic

In this subsection we examine the behavior of the statistic $T_n^{(3)}(q)$ and its block permutation counterparts. We concentrate on $q_1 \equiv 1$, yet the simulations for $q_2 = (\mathrm{id}(1-\mathrm{id}))^{1/4}$ give very similar results. As far as we know the asymptotic quantiles are theoretically only known for $q_1 \equiv 1$. For all other choices we have to run a simulation to use the asymptotic method at all. The block permutation method, however, is applicable no matter which weight function q we choose.

We compute the quantiles of the null hypothesis from 10 000 simulated values of $T_n^{(3)}(q)/\widehat{\tau}_{LK}$ under the null hypothesis. We use the same set of random numbers for each combination of variables (as long as n remains the same) to get better comparable results. The distribution of the null hypothesis depends on the block length K, since we use a different variance estimator for different K. Table 6.2.2.3 gives the results of this simulation for q_1.

In the same way we then compute the quantiles of $T_n^{(3)}(q)/\tau$, i.e. we divide the statistic by τ, instead of an estimator for it. The results can be found in Table 6.2.2.2 for $q_1 = 1$.

The simulations show that the quantiles for this statistic also depend strongly on the correlation ρ between the observations. This is more obvious if we use the real value of the variance, since the estimators (3.5.1) used in Table 6.2.2.3 compensate for the difference somewhat. Not surprisingly the results of Tables 6.2.2.2 and 6.2.2.3 are closest for large K, which is the best estimator (cf. Table 6.2.1.1).

90%	**95%**	**97.5%**
1.224	1.358	1.480

Table 6.2.2.1.: Asymptotic quantiles (CUSUM statistic, $q_1 \equiv 1$)

	n = 80			**n = 120**			**n = 210**		
ρ	**90%**	**95%**	**97.5%**	**90%**	**95%**	**97.5%**	**90%**	**95%**	**97.5%**
-0.5	1.24	1.373	1.49	1.234	1.36	1.485	1.222	1.354	1.477
-0.3	1.205	1.337	1.461	1.206	1.329	1.449	1.2	1.332	1.453
-0.2	1.192	1.32	1.45	1.192	1.317	1.439	1.193	1.324	1.445
-0.1	1.178	1.307	1.436	1.182	1.308	1.434	1.184	1.317	1.435
0	1.163	1.295	1.425	1.173	1.298	1.421	1.177	1.309	1.427
0.1	1.151	1.279	1.412	1.159	1.284	1.412	1.17	1.302	1.416
0.2	1.136	1.268	1.394	1.146	1.273	1.398	1.163	1.295	1.405
0.3	1.119	1.25	1.377	1.135	1.26	1.383	1.154	1.285	1.394
0.5	1.079	1.21	1.34	1.105	1.233	1.35	1.128	1.256	1.369
0.7	1.012	1.144	1.262	1.048	1.187	1.296	1.088	1.217	1.333

Table 6.2.2.2.: Simulated critical values of $T_n^{(3)}(q_1)/\tau$

The asymptotic quantiles (confer Table 6.2.2.1 for $q_1 \equiv 1$) on the other hand are the same no matter which model the errors follow. This means that they are wrong in some cases. We can also see that the difference between the largest and the smallest value of $T_n^{(3)}(q)/\tau$ (i.e. for $\rho = -0.5$ and $\rho = 0.7$) diminishes for a greater number of observations, which is consistent with the asymptotic results. However the convergence is rather slow. The above observations are true for both q_1 and q_2, they are even more obvious for q_2 than q_1.

To study the critical values obtained from the block permutation method, we simulate one realization according to the given model and calculate the block permutation quantiles based on 10 000 permutations for this realization. We do this for different realizations, but use the same random numbers in each case to get a better idea of the stability of the procedure. A selection of these quantiles can be found in Table 6.2.2.4 for q_1.
The results depend on the correlation about as much as the ones for the null hypothesis.

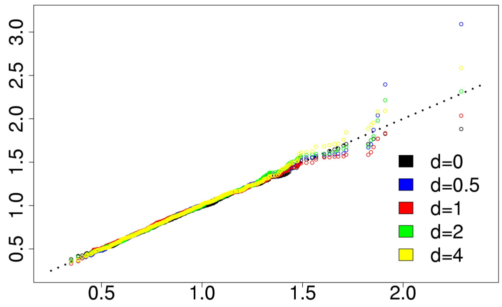
Under positive correlation they are quite stable for different alternatives. Under negative

		n = 80				**n = 120**				**n = 210**		
ρ	**K**	**90%**	**95%**	**97.5%**	**K**	**90%**	**95%**	**97.5%**	**K**	**90%**	**95%**	**97.5%**
−0.5	1	0.718	0.804	0.876	1	0.714	0.792	0.861	1	0.715	0.795	0.867
−0.5	2	1.012	1.117	1.218	2	1.021	1.137	1.221	3	1.018	1.12	1.221
−0.5	4	1.092	1.202	1.304	4	1.087	1.198	1.304	5	1.096	1.211	1.309
−0.5	10	1.216	1.336	1.456	12	1.196	1.296	1.39	15	1.187	1.294	1.39
−0.3	1	0.887	0.992	1.077	1	0.882	0.986	1.088	1	0.887	0.985	1.081
−0.3	2	1.044	1.162	1.254	2	1.048	1.166	1.271	3	1.096	1.211	1.314
−0.3	4	1.113	1.229	1.313	4	1.12	1.231	1.328	5	1.123	1.238	1.342
−0.3	10	1.193	1.306	1.411	12	1.188	1.295	1.394	15	1.178	1.284	1.381
−0.2	1	0.961	1.071	1.173	1	0.973	1.085	1.178	1	0.974	1.077	1.186
−0.2	2	1.089	1.204	1.299	2	1.079	1.2	1.3	3	1.107	1.246	1.351
−0.2	4	1.123	1.241	1.334	4	1.12	1.241	1.348	5	1.156	1.272	1.379
−0.2	10	1.182	1.294	1.392	12	1.18	1.283	1.361	15	1.178	1.292	1.379
−0.1	1	1.059	1.172	1.284	1	1.071	1.184	1.286	1	1.075	1.201	1.309
−0.1	2	1.114	1.234	1.341	2	1.118	1.244	1.359	3	1.142	1.269	1.38
−0.1	4	1.134	1.242	1.349	4	1.143	1.263	1.366	5	1.162	1.274	1.382
−0.1	10	1.188	1.291	1.39	12	1.169	1.279	1.368	15	1.179	1.291	1.388
0	1	1.154	1.281	1.41	1	1.162	1.29	1.403	1	1.181	1.321	1.425
0	2	1.152	1.268	1.37	2	1.166	1.293	1.421	3	1.183	1.314	1.425
0	4	1.141	1.258	1.35	4	1.157	1.276	1.392	5	1.172	1.289	1.407
0	10	1.17	1.275	1.367	12	1.169	1.266	1.361	15	1.174	1.282	1.374
0.1	1	1.255	1.389	1.517	1	1.278	1.424	1.543	1	1.298	1.437	1.564
0.1	2	1.19	1.315	1.425	2	1.217	1.348	1.462	3	1.213	1.341	1.456
0.1	4	1.157	1.275	1.371	4	1.166	1.282	1.4	5	1.193	1.311	1.421
0.1	10	1.166	1.271	1.367	12	1.168	1.272	1.36	15	1.172	1.272	1.358
0.2	1	1.371	1.523	1.656	1	1.388	1.536	1.67	1	1.431	1.592	1.751
0.2	2	1.233	1.357	1.477	2	1.277	1.406	1.529	3	1.243	1.384	1.501
0.2	4	1.173	1.279	1.384	4	1.207	1.322	1.42	5	1.209	1.341	1.46
0.2	10	1.153	1.246	1.327	12	1.159	1.261	1.342	15	1.169	1.273	1.36
0.3	1	1.494	1.65	1.776	1	1.534	1.707	1.861	1	1.579	1.754	1.919
0.3	2	1.302	1.438	1.562	2	1.352	1.504	1.627	3	1.296	1.437	1.567
0.3	4	1.2	1.314	1.414	4	1.216	1.346	1.444	5	1.228	1.357	1.467
0.3	10	1.137	1.231	1.316	12	1.159	1.25	1.331	15	1.171	1.288	1.376
0.5	1	1.824	2.02	2.19	1	1.88	2.098	2.273	1	1.946	2.168	2.367
0.5	2	1.489	1.631	1.761	2	1.544	1.698	1.838	3	1.423	1.562	1.697
0.5	4	1.26	1.386	1.473	4	1.304	1.434	1.546	5	1.294	1.428	1.549
0.5	10	1.128	1.215	1.29	12	1.144	1.234	1.311	15	1.169	1.279	1.366
0.7	1	2.323	2.542	2.73	1	2.473	2.71	2.928	1	2.564	2.83	3.055
0.7	2	1.78	1.928	2.081	2	1.87	2.052	2.224	3	1.703	1.887	2.042
0.7	4	1.394	1.512	1.6	4	1.48	1.614	1.726	5	1.452	1.605	1.734
0.7	10	1.116	1.192	1.252	12	1.142	1.223	1.289	15	1.174	1.274	1.35

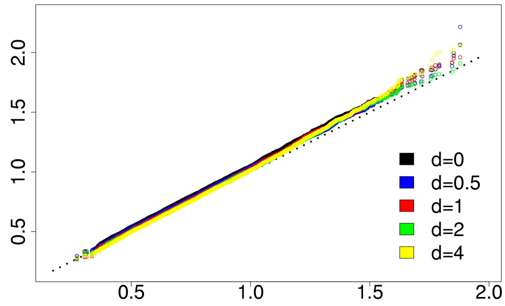
Table 6.2.2.3.: Simulated quantiles under null hypothesis (CUSUM statistic, $q_1 \equiv 1$)

		n = 80					n = 210				
ρ	d	K	m	90%	95%	97.5%	K	m	90%	95%	97.5%
−0.5		2		1.278	1.399	1.513	3		1.209	1.329	1.45
−0.5	0.5	2	20	1.239	1.359	1.463	3	52	1.199	1.326	1.437
−0.5	1	2	20	1.19	1.309	1.416	3	52	1.192	1.313	1.431
−0.5	2	2	20	1.148	1.269	1.384	3	52	1.179	1.313	1.426
−0.5	4	2	20	1.13	1.26	1.375	3	52	1.173	1.305	1.422
−0.5	0.5	2	40	1.217	1.342	1.45	3	105	1.202	1.322	1.434
−0.5	1	2	40	1.178	1.304	1.413	3	105	1.185	1.319	1.427
−0.5	2	2	40	1.152	1.28	1.381	3	105	1.173	1.313	1.419
−0.5	4	2	40	1.142	1.265	1.376	3	105	1.171	1.299	1.428
−0.5		4		1.21	1.317	1.405	5		1.221	1.338	1.451
−0.5	0.5	4	20	1.147	1.257	1.357	5	52	1.194	1.319	1.433
−0.5	1	4	20	1.11	1.232	1.335	5	52	1.169	1.29	1.387
−0.5	2	4	20	1.097	1.218	1.34	5	52	1.148	1.275	1.389
−0.5	4	4	20	1.09	1.215	1.342	5	52	1.145	1.271	1.389
−0.5	0.5	4	40	1.139	1.246	1.342	5	105	1.191	1.31	1.416
−0.5	1	4	40	1.108	1.225	1.321	5	105	1.163	1.294	1.405
−0.5	2	4	40	1.094	1.219	1.327	5	105	1.146	1.282	1.395
−0.5	4	4	40	1.099	1.224	1.328	5	105	1.143	1.28	1.396
−0.5		10		1.327	1.397	1.449	15		1.194	1.282	1.352
−0.5	0.5	10	20	1.1	1.171	1.225	15	52	1.126	1.232	1.314
−0.5	1	10	20	1.035	1.105	1.151	15	52	1.081	1.203	1.296
−0.5	2	10	20	1.002	1.134	1.143	15	52	1.066	1.18	1.271
−0.5	4	10	20	0.974	1.143	1.143	15	52	1.058	1.177	1.265
−0.5	0.5	10	40	1.074	1.179	1.263	15	105	1.109	1.211	1.311
−0.5	1	10	40	1.035	1.085	1.302	15	105	1.075	1.187	1.29
−0.5	2	10	40	1.007	1.046	1.317	15	105	1.062	1.188	1.295
−0.5	4	10	40	1.001	1.022	1.321	15	105	1.049	1.233	1.293
0		2		1.167	1.291	1.393	3		1.18	1.306	1.408
0	0.5	2	20	1.153	1.279	1.384	3	52	1.175	1.304	1.413
0	1	2	20	1.144	1.268	1.365	3	52	1.179	1.303	1.405
0	2	2	20	1.128	1.251	1.367	3	52	1.175	1.302	1.408
0	4	2	20	1.127	1.256	1.368	3	52	1.174	1.303	1.405
0	0.5	2	40	1.154	1.277	1.384	3	105	1.174	1.299	1.416
0	1	2	40	1.144	1.266	1.377	3	105	1.172	1.299	1.415
0	2	2	40	1.136	1.26	1.371	3	105	1.168	1.3	1.418
0	4	2	40	1.139	1.265	1.374	3	105	1.167	1.291	1.423
0		4		1.12	1.225	1.317	5		1.176	1.302	1.412
0	0.5	4	20	1.102	1.218	1.312	5	52	1.17	1.292	1.397
0	1	4	20	1.093	1.208	1.308	5	52	1.158	1.28	1.377
0	2	4	20	1.087	1.21	1.315	5	52	1.146	1.266	1.376
0	4	4	20	1.085	1.212	1.339	5	52	1.139	1.271	1.375
0	0.5	4	40	1.095	1.21	1.305	5	105	1.172	1.289	1.39
0	1	4	40	1.09	1.203	1.306	5	105	1.161	1.274	1.379
0	2	4	40	1.095	1.212	1.321	5	105	1.147	1.276	1.384
0	4	4	40	1.099	1.219	1.331	5	105	1.142	1.282	1.395
0		10		1.34	1.426	1.509	15		1.113	1.189	1.244
0	0.5	10	20	1.105	1.168	1.237	15	52	1.099	1.198	1.277
0	1	10	20	1.066	1.117	1.142	15	52	1.072	1.187	1.262
0	2	10	20	1.016	1.125	1.127	15	52	1.055	1.181	1.269
0	4	10	20	0.985	1.14	1.14	15	52	1.055	1.173	1.26
0	0.5	10	40	1.089	1.205	1.268	15	105	1.077	1.181	1.263
0	1	10	40	0.999	1.118	1.282	15	105	1.066	1.175	1.279
0	2	10	40	0.99	1.06	1.311	15	105	1.064	1.18	1.274
0	4	10	40	0.993	1.03	1.32	15	105	1.057	1.192	1.286
0.5		2		1.13	1.26	1.365	3		1.15	1.276	1.384
0.5	0.5	2	20	1.125	1.257	1.366	3	52	1.164	1.272	1.388
0.5	1	2	20	1.122	1.251	1.375	3	52	1.162	1.278	1.391
0.5	2	2	20	1.122	1.243	1.351	3	52	1.168	1.294	1.4
0.5	4	2	20	1.117	1.243	1.351	3	52	1.17	1.298	1.404
0.5	0.5	2	40	1.133	1.261	1.379	3	105	1.152	1.282	1.385
0.5	1	2	40	1.132	1.258	1.361	3	105	1.155	1.281	1.399
0.5	2	2	40	1.13	1.257	1.357	3	105	1.162	1.289	1.406
0.5	4	2	40	1.134	1.258	1.369	3	105	1.166	1.298	1.422
0.5		4		1.089	1.198	1.293	5		1.15	1.273	1.373
0.5	0.5	4	20	1.088	1.204	1.3	5	52	1.149	1.27	1.383
0.5	1	4	20	1.09	1.202	1.293	5	52	1.148	1.261	1.385
0.5	2	4	20	1.081	1.195	1.297	5	52	1.142	1.265	1.369
0.5	4	4	20	1.087	1.211	1.309	5	52	1.139	1.263	1.377
0.5	0.5	4	40	1.09	1.198	1.302	5	105	1.153	1.272	1.377
0.5	1	4	40	1.091	1.199	1.292	5	105	1.145	1.264	1.369
0.5	2	4	40	1.087	1.202	1.305	5	105	1.143	1.262	1.367
0.5	4	4	40	1.097	1.21	1.327	5	105	1.142	1.275	1.379
0.5		10		1.158	1.241	1.412	15		1.078	1.152	1.211
0.5	0.5	10	20	1.132	1.214	1.281	15	52	1.084	1.181	1.249
0.5	1	10	20	1.028	1.114	1.143	15	52	1.07	1.174	1.251
0.5	2	10	20	1.048	1.075	1.091	15	52	1.061	1.17	1.254
0.5	4	10	20	1.015	1.121	1.121	15	52	1.054	1.175	1.263
0.5	0.5	10	40	1.099	1.194	1.26	15	105	1.063	1.154	1.232
0.5	1	10	40	1.089	1.145	1.189	15	105	1.051	1.144	1.225
0.5	2	10	40	0.977	1.109	1.258	15	105	1.052	1.159	1.251
0.5	4	10	40	0.996	1.067	1.304	15	105	1.058	1.184	1.278

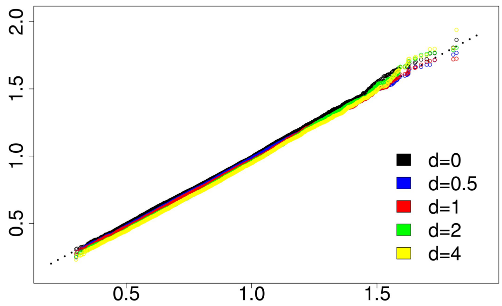
Table 6.2.2.4.: Permutation quantiles (CUSUM statistic, $q_1 \equiv 1$)



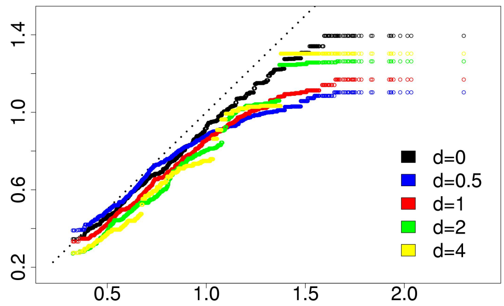
(1) $\rho = -0.5,\ K = 10$



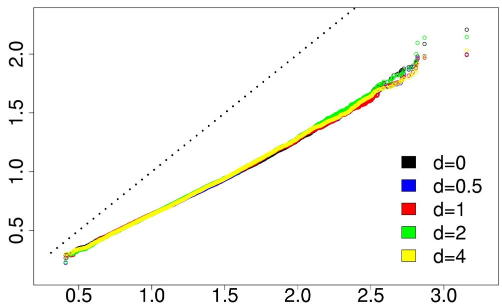
(2) $\rho = -0.1,\ K = 2$



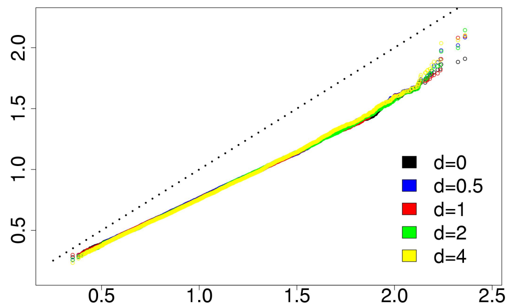
(3) $\rho = -0.1,\ K = 4$



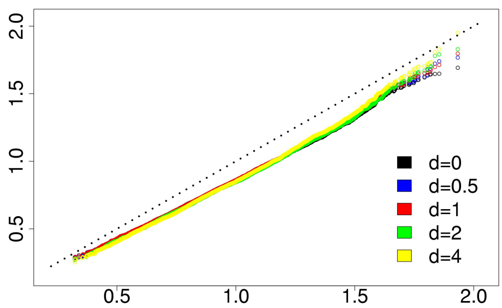
(4) $\rho = -0.1,\ K = 10$



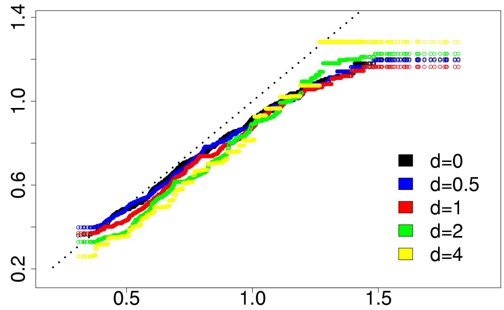
(5) $\rho = 0.5,\ K = 1$



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

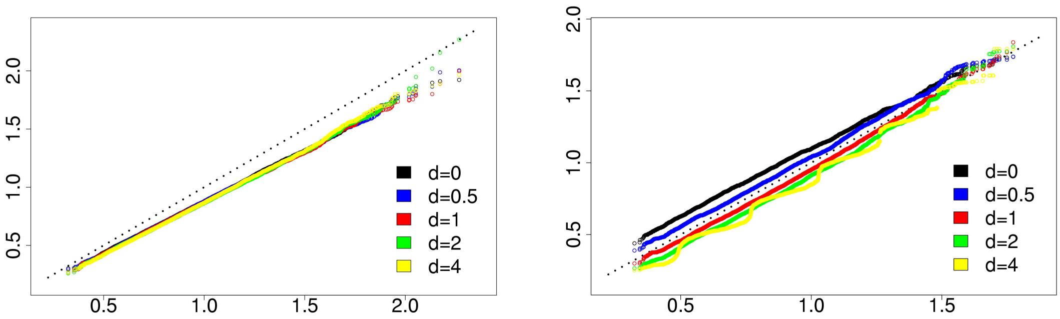


(7) $\rho = 0.5,\ K = 4$



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

Figure 6.2.2.1(i).: QQ-plots of $T_n^{(3)}(q_1)/\widehat{\tau}_{LK}$ (under H_0) against $T_{L,K}^{(3)}(q_1)(\mathbf{R})/\widehat{\tau}_{LK}$ 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 6.2.2.1(ii).: QQ-plots of $T_n^{(3)}(q_1)/\widehat{\tau}_{LK}$ (under H_0) against $T_{L,K}^{(3)}(q_1)(\mathbf{R})/\widehat{\tau}_{LK}$ for $n = 210$, $m = 105$ and different values for ρ and K

correlation, however, 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 (see e.g. $\rho = -0.5$, $K = 10$, 15, $n = 80$, 210). The dependence on the mean difference d is even greater for q_2.

Comparing the values with those of Table 6.2.2.3 shows that there is a good match for an appropriately chosen value of K. For stronger correlated error sequences the match between null hypothesis and permutation quantiles is the better the longer the block length K is. This is not very surprising since in this case the dependence structure is much better preserved. In the independent case we have a better match for a shorter block length. Yet for $\rho = -0.5$ i.e. negatively correlated error sequences the quantiles match better under $K = 4$ than $K = 10$.

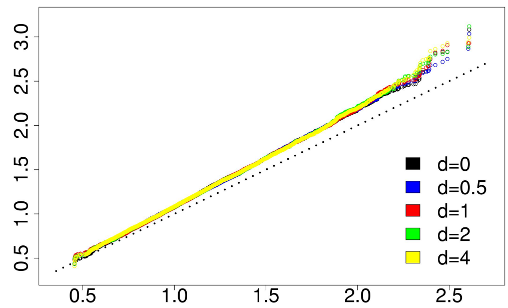
Figures 6.2.2.1(i), 6.2.2.1(ii), and 6.2.2.2 show some typical QQ-plots of the distribution of the original statistic under the null versus the block permutation distribution conditioned on just one realization. Here we use 10 000 permutations.
For n=80 we get best matches for $K = 4$. This is even true for $\rho = \pm 0.5$. 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 6.2.2.1(ii), which gives the QQ-plots for $n = 210$ and $K = 15$). Even for small n, this phenomenon does not seem to influence the goodness 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 (remember that there are 10 000 points for each alternative in the plot). The simulated quantiles of the block permutation statistic for different values of K also confirm this. Even the 97.5%–quantiles are good approximations of the null quantiles (also for $K = 10$).

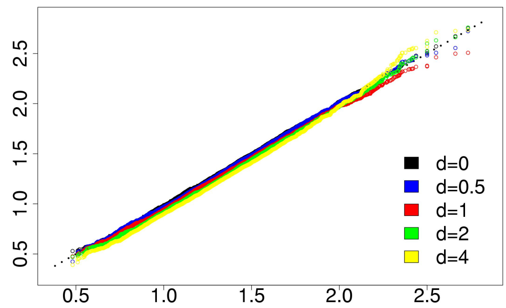
The plots displayed below have a change at $n/2$, but we get essentially the same picture for different choices of m. Moreover the distribution seems to be independent of d (i.e. the mean difference before and after the change) for positively correlated observations. There is again only the exception of $K = 10$, but fortunately the values get smaller if there is a more obvious change. This 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.

Next we are interested in how well the test performs – and also how well it performs in comparison to the asymptotic one. For this reason we create size-power-curves of both methods under the null hypothesis and under alternatives. To have a fair comparison we use $\widetilde{\tau}_n(8)$ (cf. (6.2.1)) as variance estimator for the asymptotic test. For the block permutation test we use the statistics as proposed in Section 3.5. The asymptotic method does not depend on K at all. The curves may look slightly different, since we use different sample paths in each plot namely the same as for the block permutation method. We use 10 000 random permutations to obtain the distribution of the block permutation statistic conditioned on one realization, but we only use 1 000 realizations to create the size-power-curves.

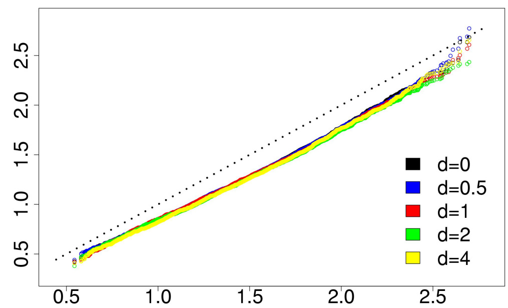
Some results are presented in Figures 6.2.2.3 and 6.2.2.4. The asymptotic test does not perform too well, except in the case of $\rho = -0.5$. The α–errors in all other cases are too high, e.g. for $\rho = 0.5$ we even have an actual α–error of 20% for a nominal one of 10%. Using double exponentially distributed innovations we get essentially the same pictures.



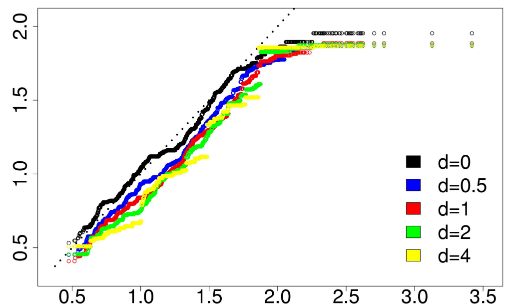
(1) $\rho = -0.1$, $K = 1$



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

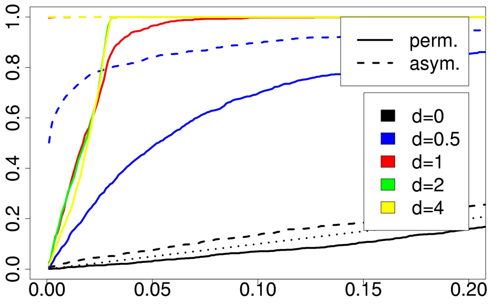


(3) $\rho = 0.5$, $K = 4$

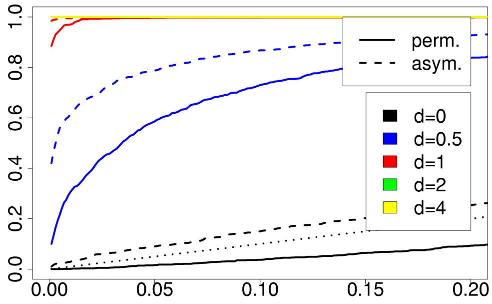


(4) $\rho = 0.5$, $K = 10$

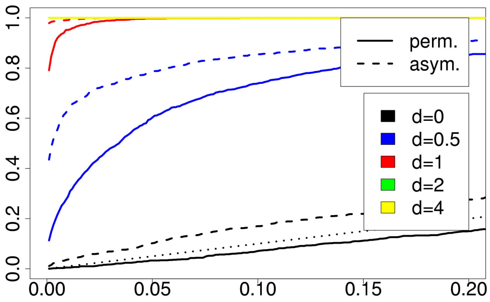
Figure 6.2.2.2.: QQ-plots of $T_n^{(3)}(q_2)/\widehat{\tau}_{LK}$ (under H_0) against $T_{L,K}^{(3)}(q_2)(\mathbf{R})/\widehat{\tau}_{LK}$ for $n = 80$, $m = 40$ and different values for ρ and K



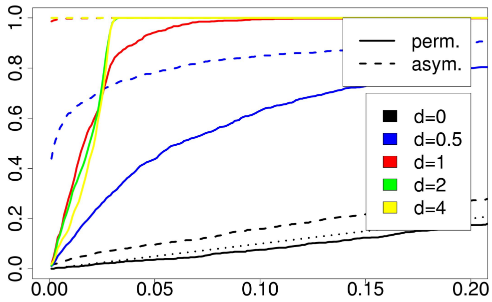
(1) $\rho = -0.5,\ K = 10$



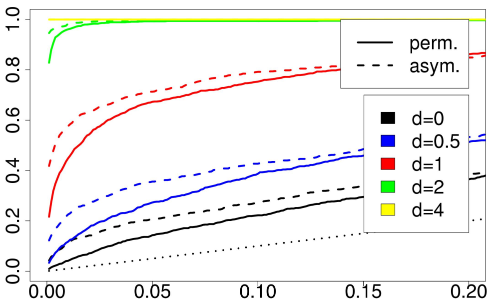
(2) $\rho = -0.3,\ K = 2$



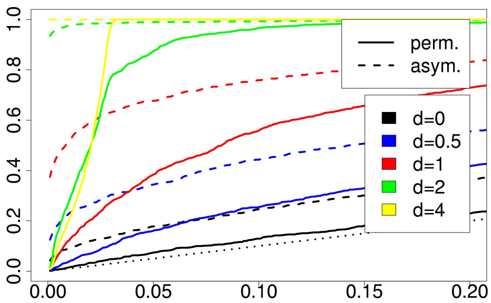
(3) $\rho = -0.3,\ K = 4$



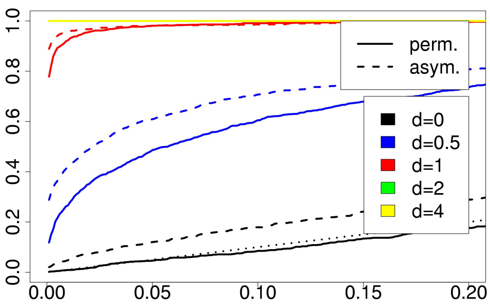
(4) $\rho = -0.3,\ K = 10$



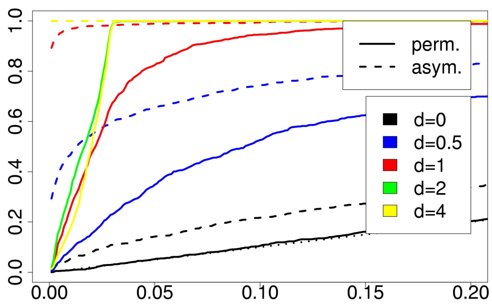
(5) $\rho = 0.5,\ K = 4$



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



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



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

Figure 6.2.2.3.: Size-power-curves of $T_n^{(3)}(q_1)/\widetilde{\tau}_n(8)$ with respect to the asymptotic distribution and of $T_{L,K}^{(3)}(\mathbf{R}, q_1)/\widehat{\tau}_{LK}$ with respect to the permutation distribution for $n = 80$, $m = 40$ and different values for ρ and K

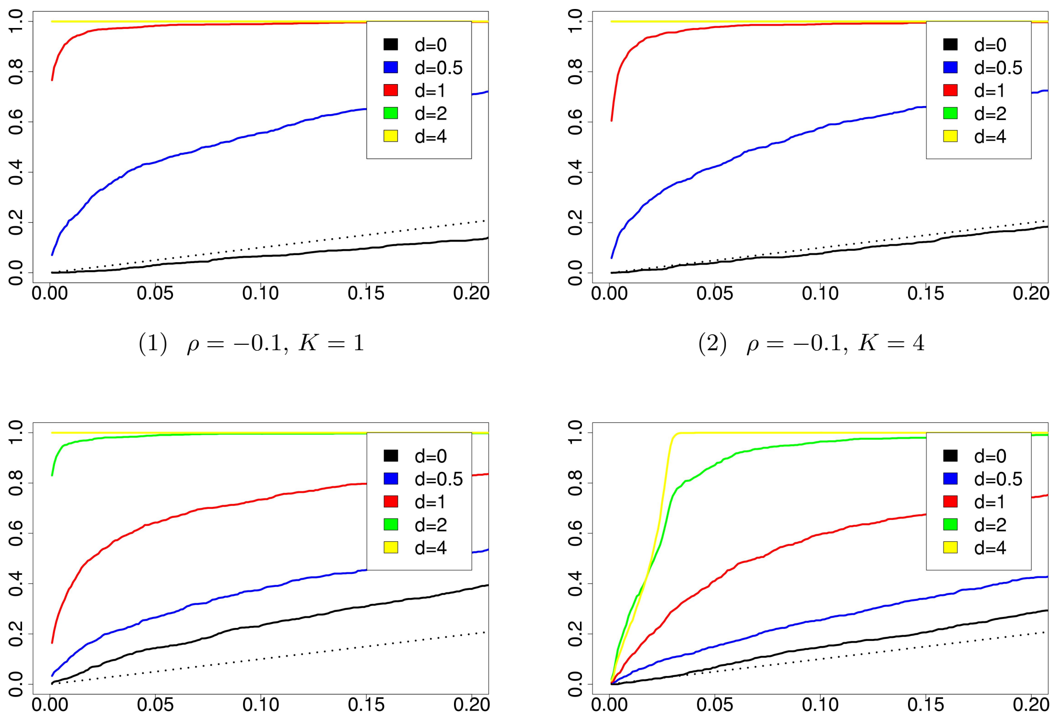
For the block permutation test it is apparently very important to choose an appropriate K. For an independent error-sequence the α–error is good for all choices of K, however, the β–errors increase (i.e. the y-values decrease under alternatives) somewhat for higher K. The best choice here would indeed be $K = 1$. On the other hand a choice of $K = 10$ is best under strong dependence.

This shows that the block permutation test performs very well. This is especially important if we want to choose $q \not\equiv 1$, because it saves us an extensive simulation of the asymptotic critical values.

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 as already mentioned.

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 block permutation test are always better than those of the asymptotic test (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.



(1) $\rho = -0.1$, $K = 1$ (2) $\rho = -0.1$, $K = 4$

(3) $\rho = 0.5$, $K = 4$ (4) $\rho = 0.5$, $K = 10$

Figure 6.2.2.4.: Size-power-curves of $T^{(3)}_{L,K}(\mathbf{R}, q_2)/\widehat{\tau}_{LK}$ with respect to the permutation distribution for $n = 80$, $m = 40$ and different values for ρ and K

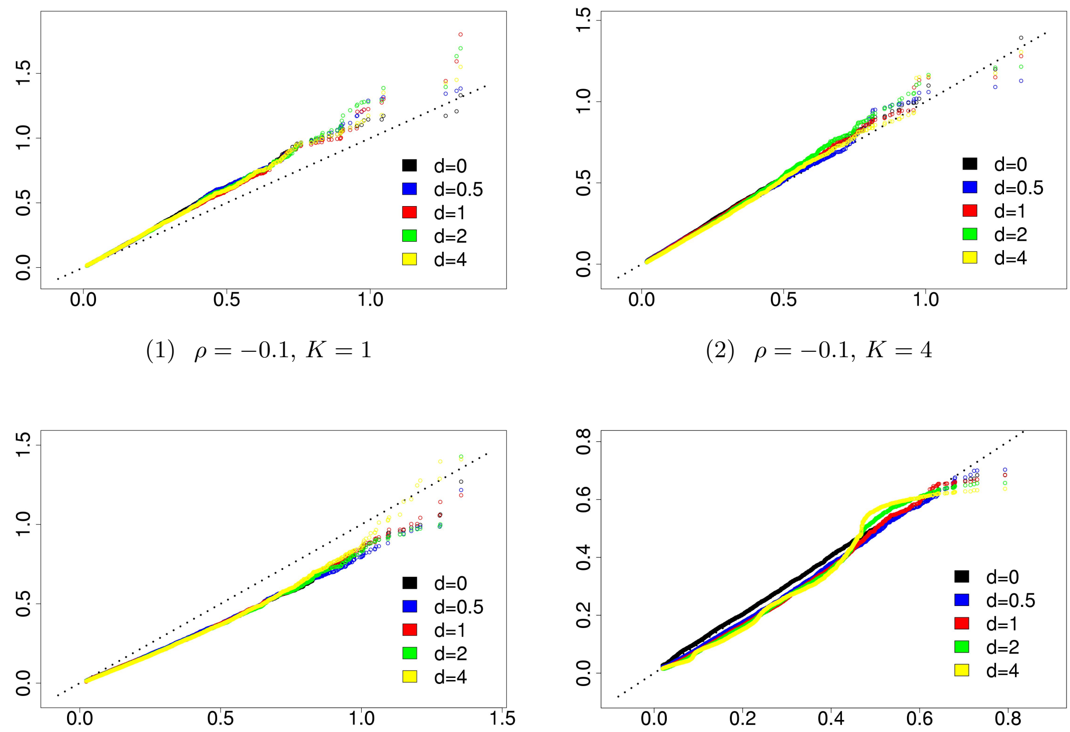
6.2.3. Sum Statistic

In this section we examine the behavior of the Sum statistic $T_n^{(4)}(r)$ and its block permutation counterparts. Here we use $r_1 \equiv 1$ and $r_2 = (\mathrm{id}(1-\mathrm{id}))^{3/2}$, which fulfill conditions (3.2.6).

As far as we know the asymptotic quantiles are only known for $r_1 \equiv 1$. The distribution function is given in Kiefer [49], Table 3, $B_1(x)$ (confer also Antoch et. al. [3], p. 15). For all other choices we have to run a simulation to use the asymptotic method at all. The block permutation method, however, is applicable no matter which weight function r we choose.

Since the results are very similar to the ones for the q-weighted CUSUM statistic we refrain from giving them in detail and only show a selection of QQ-plots and size-power-curves without comment. The only real difference is that the QQ-plots even for large values of the block length K do not show the step behavior that is typical for the q-weighted CUSUM statistics. This is because we are here looking at sums and not maxima.

The QQ-plots can be found in Figures 6.2.3.1, 6.2.3.2(i) and 6.2.3.2(ii). For r_2 the QQ-plots for a large correlation e.g. $\rho = 0.5$ and long block length (e.g. $K = 10$) are different for the null hypothesis and alternatives. This difference, however, diminishes for more observations. SPC-plots can be found in Figures 6.2.3.3 and 6.2.3.4.



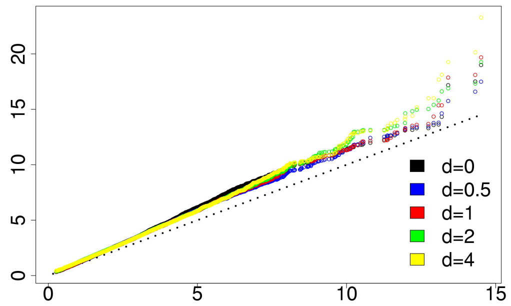
(1) $\rho = -0.1$, $K = 1$

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

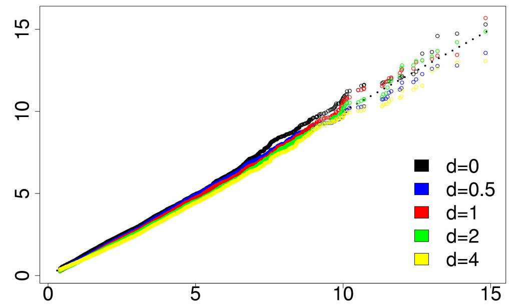
(3) $\rho = 0.5$, $K = 4$

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

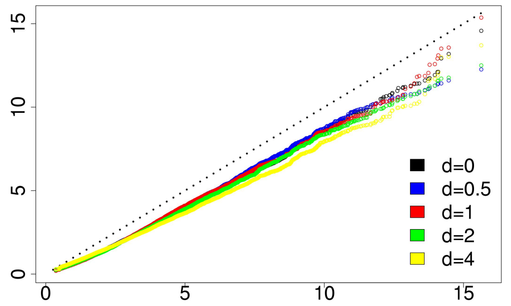
Figure 6.2.3.1.: QQ-plots of $T_n^{(4)}(r_1)/\hat{\tau}_{LK}^2$ (under H_0) against $T_{L,K}^{(4)}(r_1)(\mathbf{R})/\hat{\tau}_{LK}^2$ for $n = 80$, $m = 40$ and different values for ρ and K



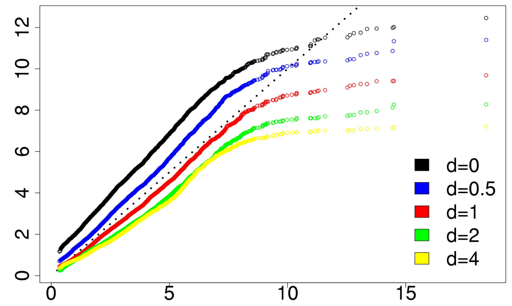
(1) $\rho = -0.1,\ K = 1$



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

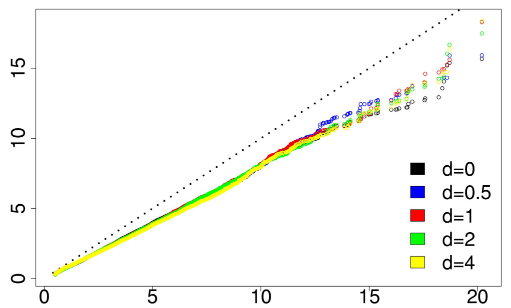


(3) $\rho = 0.5,\ K = 4$

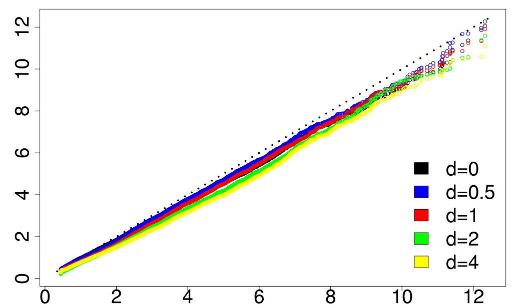


(4) $\rho = 0.5,\ K = 10$

Figure 6.2.3.2(i).: QQ-plots of $T_n^{(4)}(r_2)/\hat{\tau}_{LK}^2$ (under H_0) against $T_{L,K}^{(4)}(r_2)(\mathbf{R})/\hat{\tau}_{LK}^2$ for $n = 80$, $m = 40$ and different values for ρ and K

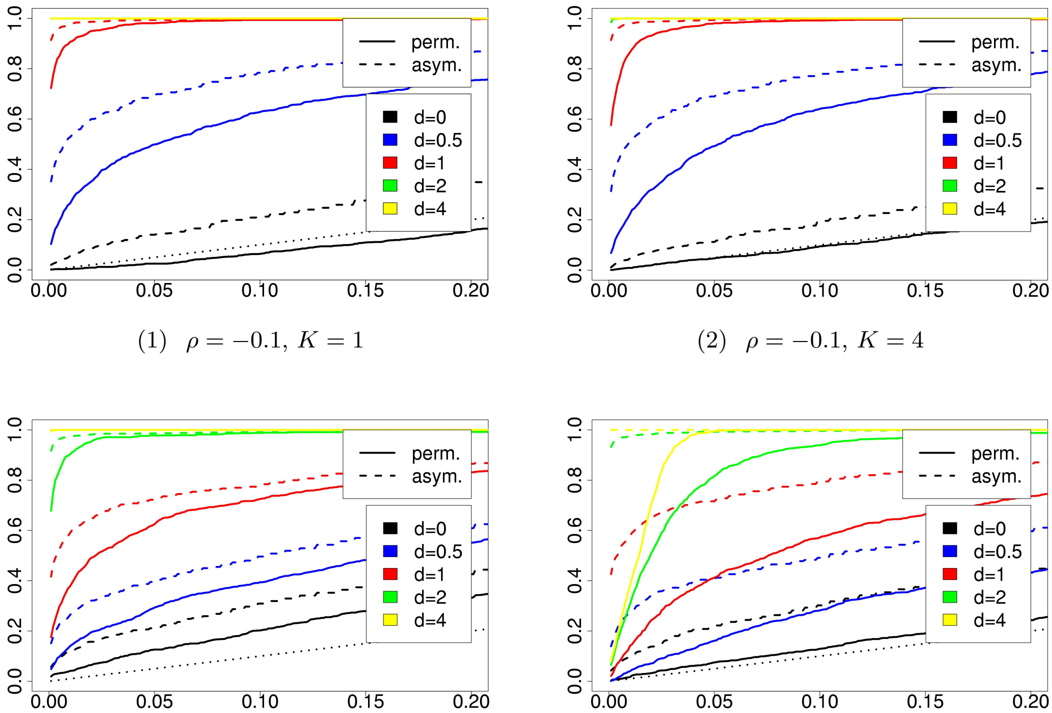


(5) $\rho = 0.5,\ K = 5$



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

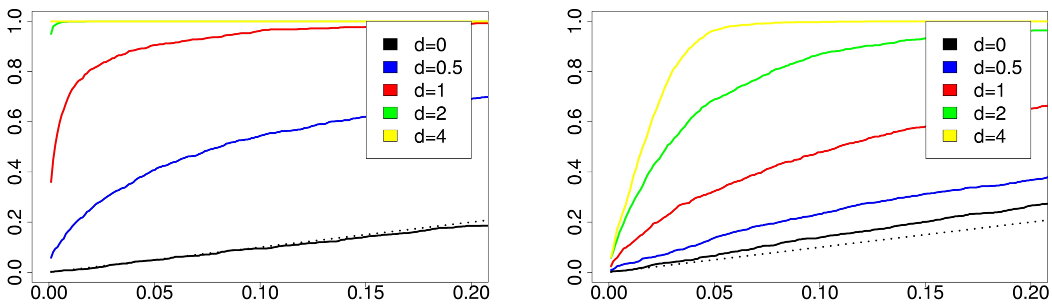
Figure 6.2.3.2(ii).: QQ-plots of $T_n^{(4)}(r_2)/\hat{\tau}_{LK}^2$ (under H_0) against $T_{L,K}^{(4)}(r_2)(\mathbf{R})/\hat{\tau}_{LK}^2$ for $n = 210$, $m = 105$ and different values for ρ and K



(1) $\rho = -0.1,\ K = 1$ (2) $\rho = -0.1,\ K = 4$

(3) $\rho = 0.5,\ K = 4$ (4) $\rho = 0.5,\ K = 10$

Figure 6.2.3.3.: Size-power-curves of $T_n^{(4)}(r_1)/\widetilde{\tau}_n^2(8)$ with respect to the asymptotic distribution and of $T_{L,K}^{(4)}(\mathbf{R}, r_1)/\widehat{\tau}_{LK}^2$ with respect to the permutation distribution for $n = 80$, $m = 40$ and different values for ρ and K

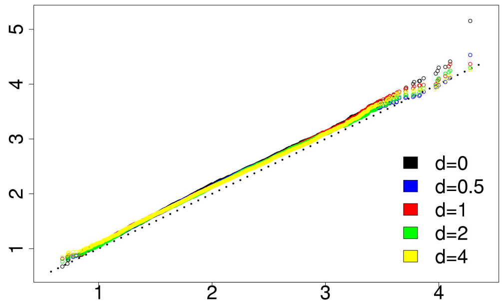


(1) $\rho = -0.1,\ K = 4$ (2) $\rho = 0.5,\ K = 10$

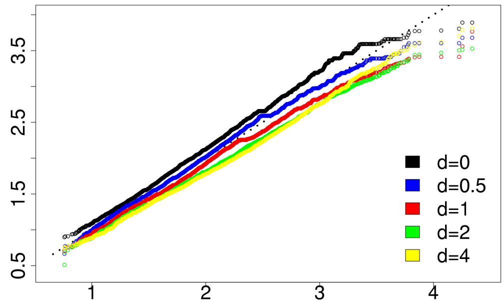
Figure 6.2.3.4.: Size-power-curves of $T_{L,K}^{(4)}(\mathbf{R}, r_2)/\widehat{\tau}_{LK}^2$ with respect to the permutation distribution for $n = 80$, $m = 40$ and different values for ρ and K

6.2.4. Weighted CUSUM Statistic

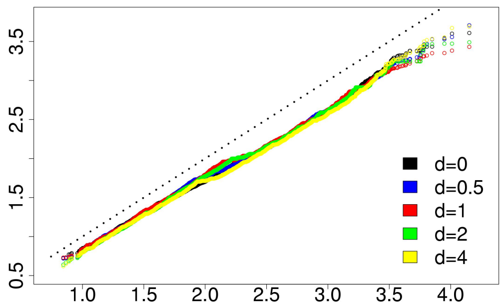
In this subsection we examine the behavior of the weighted CUSUM statistic $T_n^{(1)}$ and its block permutation counterpart. Since the results are very similar to the q-weighted CUSUM statistic we only give some QQ- and SPC-plots in Figures 6.2.4.1 resp. 6.2.4.2.



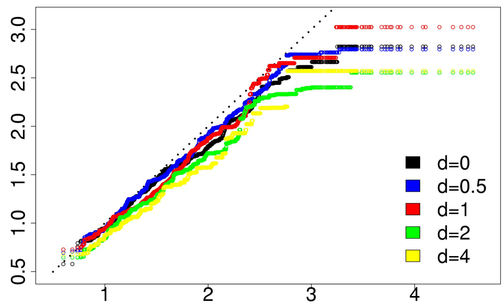
(1) $\rho = -0.1$, $K = 1$



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

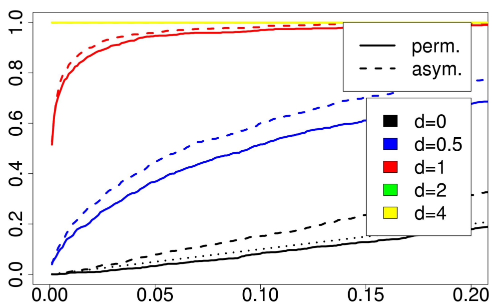


(3) $\rho = 0.5$, $K = 4$

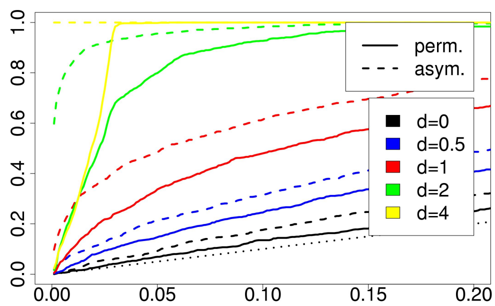


(4) $\rho = 0.5$, $K = 10$

Figure 6.2.4.1.: QQ-plots of $T_{L,K}^{(1)}/\widehat{\tau}_{LK}$ (under H_0) against $T_{L,K}^{(1)}(\mathbf{R})/\widehat{\tau}_{LK}$ for $n = 80$, $m = 40$ and different values for ρ and K



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

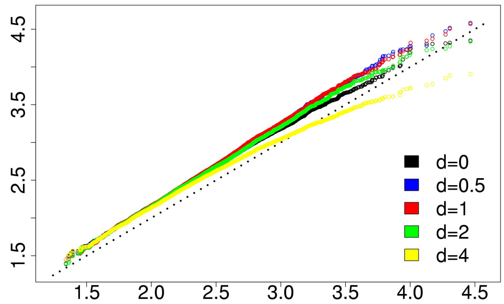


(2) $\rho = 0.5$, $K = 10$

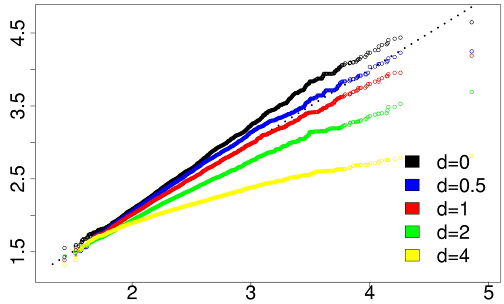
Figure 6.2.4.2.: Size-power-curves of $T_n^{(1)}/\widetilde{\tau}_n(8)$ with respect to the asymptotic distribution and of $T_{L,K}^{(1)}(\mathbf{R})/\widehat{\tau}_{LK}$ with respect to the permutation distribution for $n = 80$, $m = 40$ and different values for ρ and K

6.2.5. MOSUM Statistic

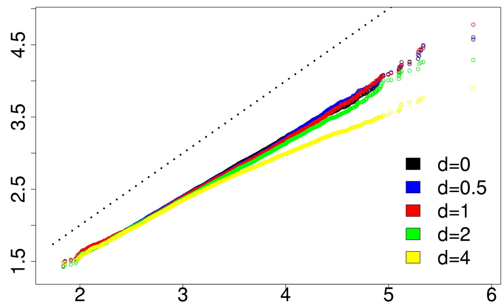
As already mentioned the block permutation method is not very well suited for the MOSUM statistic. We are looking at a generally very small (e.g. $G = 0.05\,n$ or $G = 0.1\,n$) window of data. So K has to be small in comparison to G, otherwise the maximum is taken of the same numbers for many permutations.



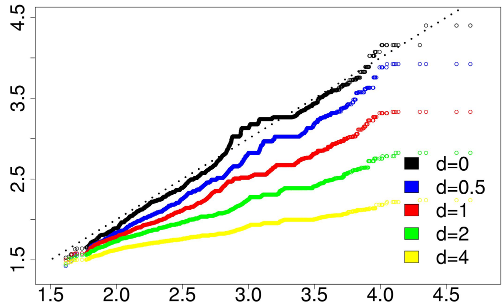
(1) $\rho = -0.1,\ K = 1$



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

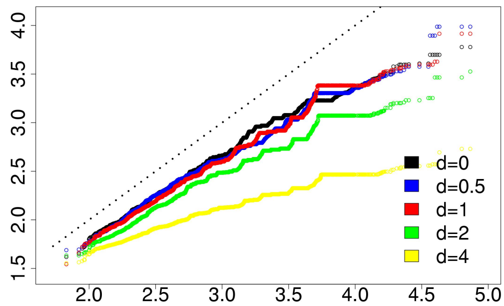


(3) $\rho = 0.3,\ K = 1$

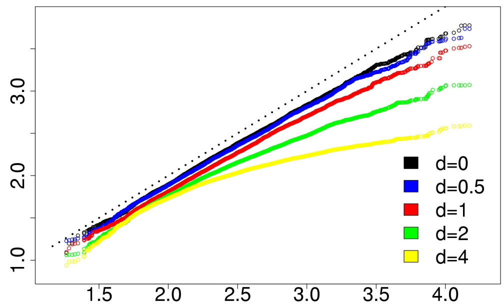


(4) $\rho = 0.3,\ K = 3$

Figure 6.2.5.1(i).: QQ-plots of $T_n^{(2)}(G_1)/\widehat{\tau}_{LK}$ (under H_0) against $T_{L,K}^{(2)}(G_1)(\mathbf{R})/\widehat{\tau}_{LK}$ for $n = 120$, $m = 60$ and different values for ρ and K



(5) $\rho = 0.3,\ K = 3$

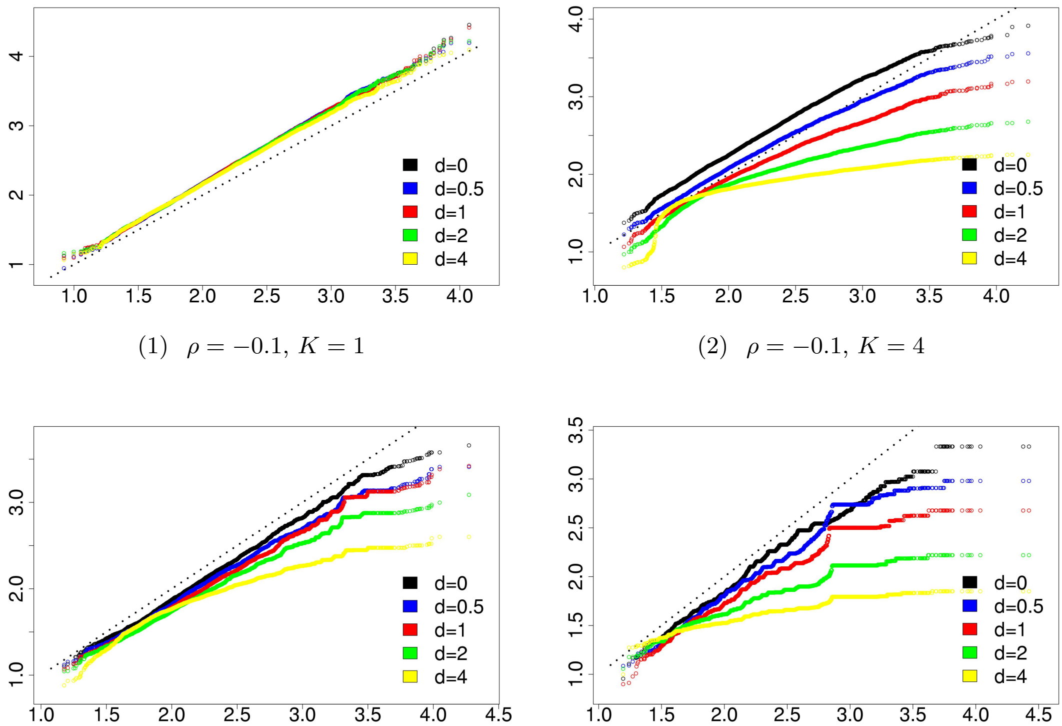


(6) $\rho = 0.3,\ K = 6$

Figure 6.2.5.1(ii).: QQ-plots of $T_n^{(2)}(G_1)/\widehat{\tau}_{LK}$ (under H_0) against $T_{L,K}^{(2)}(G_1)(\mathbf{R})/\widehat{\tau}_{LK}$ for $n = 210$, $m = 105$ and different values for ρ and K

We give some QQ-plots and size-power-curves to demonstrate the fact.

Here, we use $n = 120$ and $n = 210$ so that G is not all too small. Moreover we use $G_1 \approx 0.05\,n$ (more precisely $G_1(120) = 6$, $G_1(210) = 10$) and $G_2 \approx 0.1\,n$ (more precisely $G_2(120) = 12$, $G_2(210) = 20$). The simulations show that the block permutation method is indeed working better for G_2 than G_1, probably because the "window" is somewhat larger.



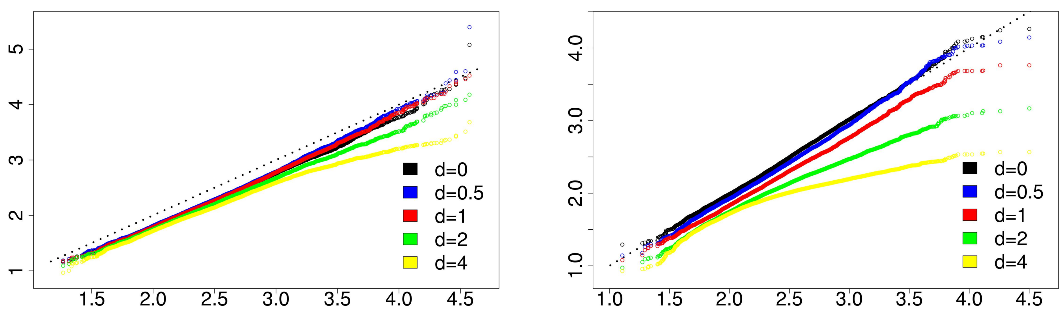
(1) $\rho = -0.1$, $K = 1$

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

(3) $\rho = 0.3$, $K = 4$

(4) $\rho = 0.3$, $K = 6$

Figure 6.2.5.2(i).: QQ-plots of $T_n^{(2)}(G_2)/\widehat{\tau}_{LK}$ (under H_0) against $T_{L,K}^{(2)}(G_2)(\mathbf{R})/\widehat{\tau}_{LK}$ for $n = 120$, $m = 60$ and different values for ρ and K



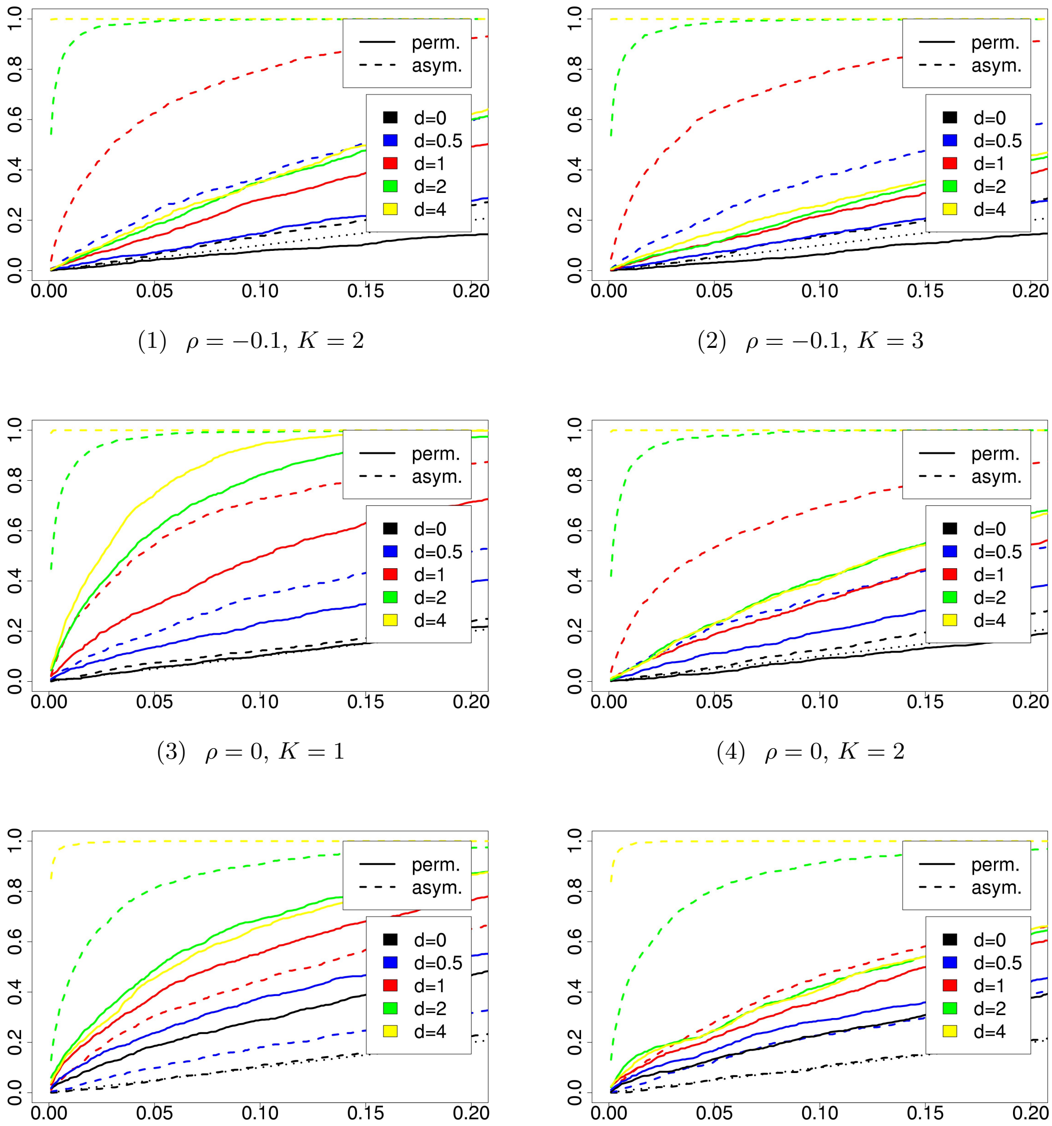
(5) $\rho = 0.3$, $K = 3$

(6) $\rho = 0.3$, $K = 6$

Figure 6.2.5.2(ii).: QQ-plots of $T_n^{(2)}(G_2)/\widehat{\tau}_{LK}$ (under H_0) against $T_{L,K}^{(2)}(G_2)(\mathbf{R})/\widehat{\tau}_{LK}$ for $n = 210$, $m = 105$ and different values for ρ and K

For $K = 1$ the QQ-plots only fit the null distribution if we have independent error sequences or maybe error sequences under weak dependence ($\rho = \pm 0.1$). For $K > 1$ the fit under the null hypothesis becomes better (at least if K is not too large), but the curves under alternatives are not a good match.

As already mentioned earlier the curves for larger G, i.e. Figures 6.2.5.2(i) and 6.2.5.2(ii), are better than those for smaller G, i.e. Figures 6.2.5.1(i) and 6.2.5.1(ii). The reason is that we can choose a somewhat larger K to better fit the correlated data, but the match is still not very good.



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

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

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

(4) $\rho = 0,\ K = 2$

(5) $\rho = 0.3,\ K = 2$

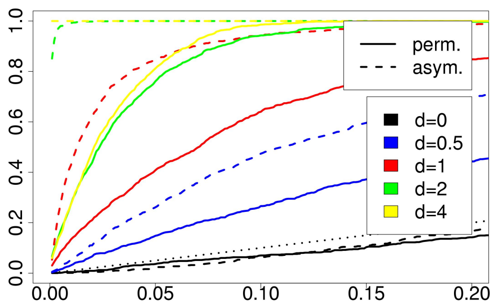
(6) $\rho = 0.3,\ K = 3$

Figure 6.2.5.3.: Size-power-curves of $T_n^{(2)}(G_1)/\widetilde{\tau}_n(8)$ with respect to the asymptotic distribution and of $T_{L,K}^{(2)}(G_1, \mathbf{R})/\widehat{\tau}_{LK}$ with respect to the permutation distribution for $n = 120$, $m = 60$ and different values for ρ and K

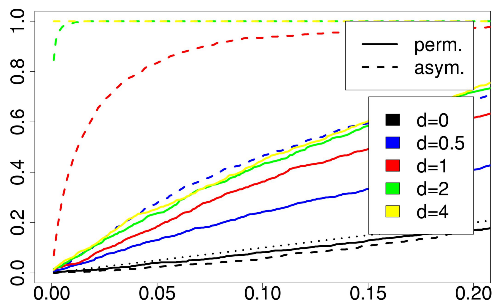
Secondly the plots get slightly better for larger n (confer Figures 6.2.5.1(ii) respectively 6.2.5.2(ii)), but apparently the convergence is rather slow.

To create size-power-curves we only use 1 000 permutations for the empirical distribution function of each realization to save some computational time. Some results are presented in Figures 6.2.5.3 and 6.2.5.4.

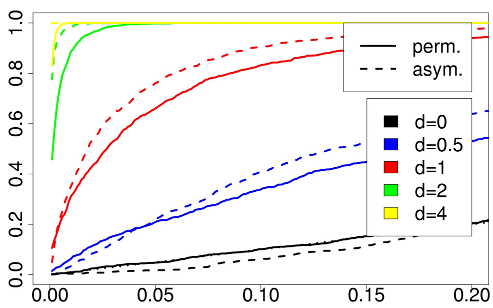
As expected the asymptotic test performs quite well, although in some cases it performs better than in others. This is due to the fact that the asymptotic quantiles do not depend on the correlation, but the null quantiles do.



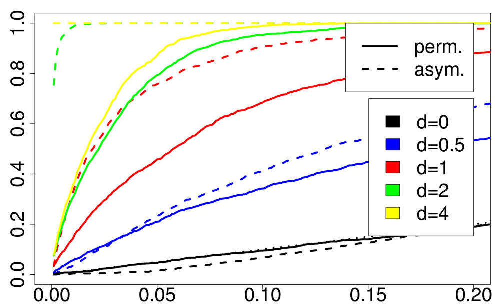
(1) $\rho = -0.1$, $K = 2$



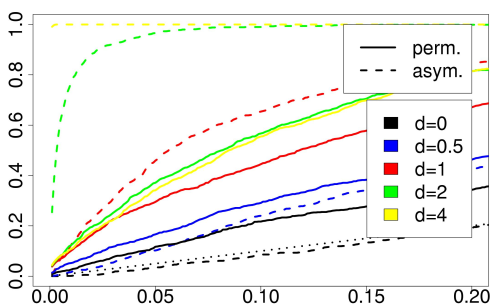
(2) $\rho = -0.1$, $K = 4$



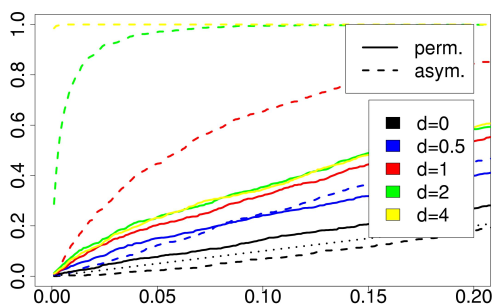
(3) $\rho = 0$, $K = 1$



(4) $\rho = 0$, $K = 2$



(5) $\rho = 0.3$, $K = 4$



(6) $\rho = 0.3$, $K = 6$

Figure 6.2.5.4.: Size-power-curves of $T_n^{(2)}(G_2)/\widetilde{\tau}_n(8)$ with respect to the asymptotic distribution and of $T_{L,K}^{(2)}(G_2, \mathbf{R})/\widehat{\tau}_{LK}$ with respect to the permutation distribution for $n = 120$, $m = 60$ and different values for ρ and K

The block permutation test performs quite well for independent error sequences and $K = 1$, confer Figure 6.2.5.3, (3), respectively 6.2.5.4, (3). For G_1 the α-errors are slightly better than the asymptotic α-errors, however the β-errors are higher. For G_2 the α-errors for the permutation method are also better and the β-errors about comparable.

In general the block permutation test does not perform very well. For $K = 1$ it can distinguish the alternatives, but the α–errors are completely wrong, especially for stronger correlated data. A larger K puts the α–error closer to where they are supposed to be, but also leads to a worse distinction of the alternatives.

Furthermore the block permutation method performs better for G_2 than G_1. All of the effects described above are weaker for a larger G. Then the permutation method even has a good fit under weak alternatives, i.e. $\rho = \pm 0.1$, and $K = 2$. This is in accordance with the asymptotic behavior of the permutation statistic as proven in Theorem 3.5.1.

6.3. Simulation of the Frequency Permutation Method

In this section we investigate the performance of the frequency permutation method in more detail.

Again we use an AR(1) sequence (throwing away the first 50 data points) with parameter $\rho \in \{0, \pm 0.1, \pm 0.3, \pm 0.5, 0.7\}$, standard normally as well as double-exponentially distributed innovations (mean 0, variance 1) and length $n = 80$, change-point $m = 40$. Then we use the algorithm as described in Section 4.2 with 60 respectively 70 bootstrap r.v.´s, i.e. $\frac{n}{\alpha(n)} = 60$ respectively $= 70$. Here, we take the first 60 respectively first 70. Moreover we use all 80 r.v.´s, which corresponds to $\alpha(n) = 1$.

We do not use any variance estimators in the fourth step but rather use the actual values to get better comparable results. Indeed if we compare the asymptotic results here with the ones we have obtained in the previous section, we see that they perform much better here. The only difference is that we have used estimators there. For practical purposes the goodness of the procedure of the frequency test as well as of the asymptotic one depends crucially on the goodness of the estimator for $\sigma^2 \left(\sum w_s\right)^2$. That is the disadvantage over the block permutation test. Here, however, we are interested in the performance of the algorithm itself, so to get an objective picture we use the (usually unknown) exact value of $\sigma^2 \left(\sum w_s\right)^2$.

The simulations for n odd are slower than the ones for n even, because of the nature of the FFT-Algorithm. This is why one should use the algorithm for n even.

Again we give the details of the simulation study only for the q-weighted CUSUM statistic. For the other statistics we present a small selection of (representative) plots due to limitations of space. The results are, however, very similar. Details can be found in Kirch [51].

6.3.1. q-weighted CUSUM Statistic

In this subsection we investigate the statistics $T_n^{(3)}(q)$ for $q_1 \equiv 1$ as well as $q_2 := (\mathrm{id}(1-\mathrm{id}))^{1/4}$. We concentrate on the discussion of q_1, since we can then also compare

					Normal r.v.´s			Double exp. r.v.´s		
n	**ρ**	**N**	**m**	**d**	**90%**	**95%**	**97.5%**	**90%**	**95%**	**97.5%**
80	-0.5	60			1.175	1.356	1.527	1.115	1.283	1.397
80	-0.5	70			1.2	1.346	1.545	1.133	1.329	1.451
80	-0.5	80			1.165	1.344	1.531	1.104	1.266	1.422
80	-0.5	60	40	0.5	1.14	1.287	1.439	1.171	1.261	1.442
80	-0.5	70	40	0.5	1.128	1.302	1.463	1.176	1.322	1.444
80	-0.5	80	40	0.5	1.133	1.255	1.385	1.169	1.317	1.455
80	-0.5	60	40	1	1.189	1.326	1.434	1.121	1.293	1.441
80	-0.5	70	40	1	1.215	1.372	1.562	1.139	1.291	1.429
80	-0.5	80	40	1	1.185	1.379	1.518	1.138	1.291	1.408
80	-0.5	60	40	2	1.169	1.318	1.51	1.12	1.279	1.403
80	-0.5	70	40	2	1.21	1.384	1.546	1.148	1.277	1.391
80	-0.5	80	40	2	1.186	1.336	1.559	1.146	1.269	1.364
80	-0.5	60	40	4	1.193	1.347	1.519	1.119	1.265	1.376
80	-0.5	70	40	4	1.182	1.408	1.617	1.122	1.262	1.39
80	-0.5	80	40	4	1.194	1.388	1.55	1.123	1.276	1.37
80	0	60			1.155	1.287	1.402	1.12	1.226	1.298
80	0	70			1.172	1.328	1.441	1.127	1.217	1.302
80	0	80			1.174	1.307	1.412	1.104	1.211	1.28
80	0	60	40	0.5	1.156	1.297	1.375	1.158	1.261	1.357
80	0	70	40	0.5	1.164	1.318	1.414	1.14	1.262	1.336
80	0	80	40	0.5	1.149	1.3	1.386	1.126	1.221	1.307
80	0	60	40	1	1.163	1.291	1.381	1.141	1.26	1.378
80	0	70	40	1	1.209	1.322	1.449	1.145	1.254	1.338
80	0	80	40	1	1.192	1.306	1.43	1.141	1.257	1.36
80	0	60	40	2	1.155	1.306	1.46	1.135	1.207	1.285
80	0	70	40	2	1.204	1.339	1.476	1.127	1.215	1.308
80	0	80	40	2	1.19	1.344	1.427	1.117	1.224	1.306
80	0	60	40	4	1.16	1.308	1.45	1.123	1.242	1.351
80	0	70	40	4	1.195	1.333	1.473	1.154	1.258	1.364
80	0	80	40	4	1.182	1.35	1.46	1.157	1.254	1.37
80	0.3	60			1.151	1.303	1.422	1.128	1.232	1.34
80	0.3	70			1.134	1.287	1.477	1.132	1.247	1.318
80	0.3	80			1.114	1.293	1.456	1.133	1.247	1.296
80	0.3	60	40	0.5	1.128	1.274	1.413	1.17	1.274	1.369
80	0.3	70	40	0.5	1.155	1.298	1.471	1.161	1.272	1.383
80	0.3	80	40	0.5	1.159	1.313	1.459	1.148	1.265	1.361
80	0.3	60	40	1	1.146	1.305	1.433	1.169	1.299	1.424
80	0.3	70	40	1	1.169	1.359	1.541	1.182	1.303	1.422
80	0.3	80	40	1	1.143	1.356	1.567	1.184	1.303	1.401
80	0.3	60	40	2	1.179	1.296	1.398	1.135	1.241	1.372
80	0.3	70	40	2	1.177	1.333	1.46	1.149	1.271	1.378
80	0.3	80	40	2	1.177	1.34	1.476	1.145	1.288	1.363
80	0.3	60	40	4	1.145	1.268	1.385	1.136	1.263	1.411
80	0.3	70	40	4	1.126	1.296	1.419	1.167	1.296	1.432
80	0.3	80	40	4	1.124	1.239	1.385	1.166	1.322	1.413
80	0.5	60			1.149	1.358	1.539	1.146	1.281	1.392
80	0.5	70			1.128	1.332	1.628	1.164	1.277	1.391
80	0.5	80			1.138	1.329	1.56	1.15	1.27	1.387
80	0.5	60	40	0.5	1.109	1.245	1.444	1.158	1.28	1.364
80	0.5	70	40	0.5	1.143	1.302	1.509	1.159	1.261	1.361
80	0.5	80	40	0.5	1.137	1.305	1.509	1.136	1.264	1.364
80	0.5	60	40	1	1.15	1.33	1.517	1.173	1.373	1.475
80	0.5	70	40	1	1.174	1.344	1.568	1.173	1.363	1.551
80	0.5	80	40	1	1.174	1.34	1.587	1.209	1.335	1.464
80	0.5	60	40	2	1.163	1.342	1.491	1.137	1.255	1.355
80	0.5	70	40	2	1.146	1.402	1.606	1.186	1.293	1.412
80	0.5	80	40	2	1.153	1.39	1.608	1.166	1.306	1.428
80	0.5	60	40	4	1.11	1.26	1.461	1.14	1.285	1.468
80	0.5	70	40	4	1.107	1.336	1.475	1.15	1.345	1.478
80	0.5	80	40	4	1.092	1.294	1.466	1.176	1.308	1.481
80	0.7	60			1.2	1.467	1.698	1.187	1.355	1.543
80	0.7	70			1.183	1.549	1.769	1.231	1.404	1.512
80	0.7	80			1.166	1.481	1.754	1.224	1.376	1.479
80	0.7	60	40	0.5	1.107	1.335	1.506	1.183	1.333	1.423
80	0.7	70	40	0.5	1.118	1.385	1.618	1.18	1.311	1.456
80	0.7	80	40	0.5	1.127	1.38	1.701	1.154	1.334	1.484
80	0.7	60	40	1	1.109	1.319	1.553	1.194	1.362	1.488
80	0.7	70	40	1	1.129	1.436	1.669	1.206	1.397	1.561
80	0.7	80	40	1	1.134	1.371	1.682	1.206	1.383	1.536
80	0.7	60	40	2	1.192	1.385	1.49	1.156	1.309	1.469
80	0.7	70	40	2	1.184	1.361	1.57	1.148	1.368	1.583
80	0.7	80	40	2	1.179	1.349	1.603	1.138	1.35	1.586
80	0.7	60	40	4	1.141	1.315	1.484	1.15	1.359	1.601
80	0.7	70	40	4	1.133	1.379	1.529	1.151	1.439	1.646
80	0.7	80	40	4	1.123	1.311	1.54	1.171	1.353	1.623

Table 6.3.1.1.: Simulated permutation quantiles of statistic $T_n^{(3f)}(\mathbf{R}, q_1)$ for a fixed innovation sequence based on 1 000 permutations

the results with the asymptotic test. We only give a short extract of the simulations for q_2 to demonstrate that the results are very similar.

First, we simulate the critical values of the statistics under the null, where we use the correct (rather than an estimated) variance. The critical values are based on 10 000 simulated values, the results can be found in Table 6.3.1.2.

To get an impression of the permutation quantiles for one fixed underlying error sequence we calculate them by means of the empirical distribution function of the frequency permutation statistic based on 1 000 permutations. We use the same underlying error sequence for all choices of parameters and also the same 1 000 permutations. This way we get a better idea, how much the obtained quantiles depend on the alternative. Some results can be found in Table 6.3.1.1.

The frequency permutation quantiles are stable for different alternatives as well as different values of N. However, unlike the null quantiles they are also stable for different correlations. This is due to the fact that by permuting the coefficients they loose their variance structure, which in the end means the permuted sequence looses its covariance structure and thus is close to the independent case. The asymptotic quantiles also do not depend on the covariance structure of the sequence, so the same problem arises.

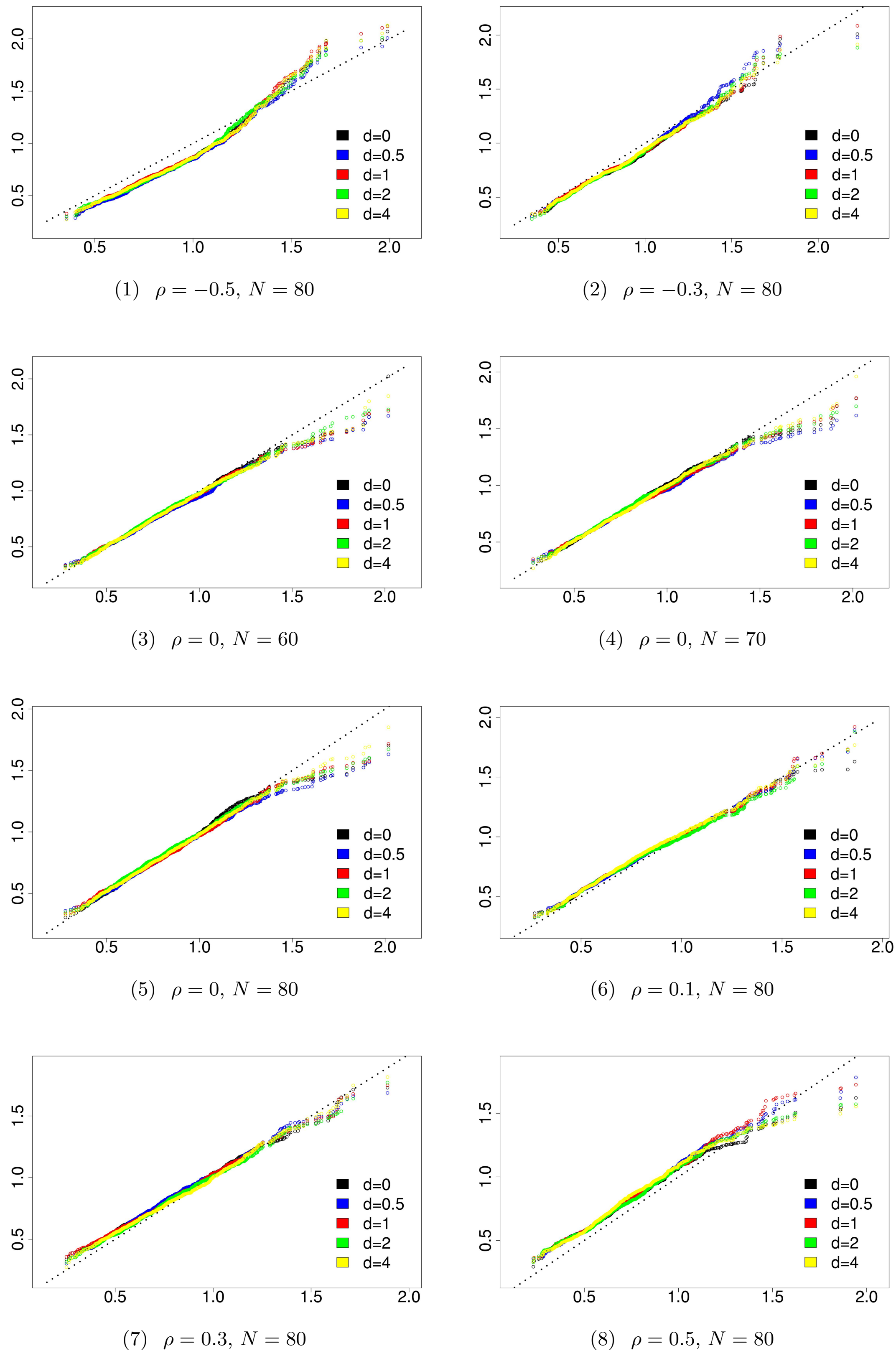
Again QQ-plots will better illustrate the similarity of the distribution of the frequency permutation statistic conditioned on a given sequence and the distribution of the original statistic under the null. Here, we use 1 000 simulated time series respectively 1 000 permutations to obtain the QQ-plot. The permutation quantiles are based on one fixed AR(1)-process for different mean changes.

The plots look very similar, no matter how many data points we use (the value of N is not important). This remains true for a different selection function β. To illustrate this point we give all three plots for $\rho = 0$ (yet this is true for every other choice of ρ as well). For a selection of other values of ρ we then only give the plot where we use all n values. The results can be found in Figures 6.3.1.1(i), 6.3.1.1(ii) respectively 6.3.1.2(i), 6.3.1.2(ii).

The plots for independent or only slightly correlated data look very good, whereas for a strong correlation there seems to be a slight difference in distribution.

	Normal r.v.´s			**Double exp. r.v.´s**		
ρ	**90%**	**95%**	**97.5%**	**90%**	**95%**	**97.5%**
-0.5	1.24	1.373	1.49	1.255	1.396	1.512
-0.3	1.205	1.337	1.461	1.208	1.348	1.466
-0.2	1.192	1.32	1.45	1.19	1.33	1.446
-0.1	1.178	1.307	1.436	1.175	1.312	1.43
0	1.163	1.295	1.425	1.161	1.295	1.411
0.1	1.151	1.279	1.412	1.144	1.282	1.401
0.2	1.136	1.268	1.394	1.131	1.271	1.388
0.3	1.119	1.25	1.377	1.118	1.256	1.376
0.5	1.079	1.21	1.34	1.084	1.214	1.336
0.7	1.012	1.144	1.262	1.019	1.151	1.267

Table 6.3.1.2.: Simulated quantiles of the statistic $\tilde{T}_n^{(3)}(q_1)$ under the null, $n = 80$



(1) $\rho = -0.5,\ N = 80$

(2) $\rho = -0.3,\ N = 80$

(3) $\rho = 0,\ N = 60$

(4) $\rho = 0,\ N = 70$

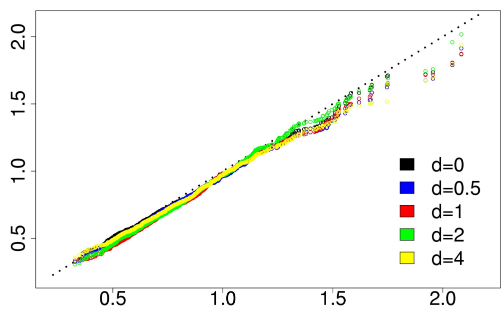
(5) $\rho = 0,\ N = 80$

(6) $\rho = 0.1,\ N = 80$

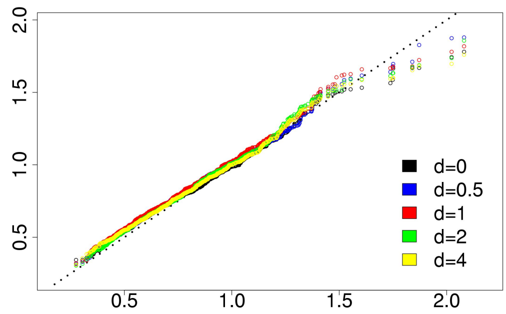
(7) $\rho = 0.3,\ N = 80$

(8) $\rho = 0.5,\ N = 80$

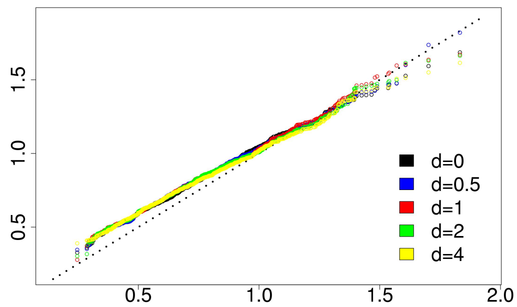
Figure 6.3.1.1(i).: QQ-plots of $T_n^{(3)}(q_1)$ (under H_0) against $T_n^{(3f)}(\mathbf{R}, q_1)$ for standard normally distributed innovations, $n = 80$, $m = 40$ and different values for ρ and N



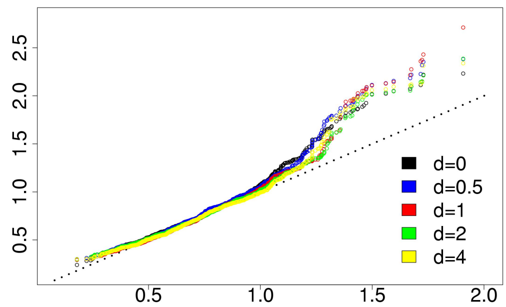
(9) $\rho = -0.3$, $N = 80$



(10) $\rho = 0.1$, $N = 80$



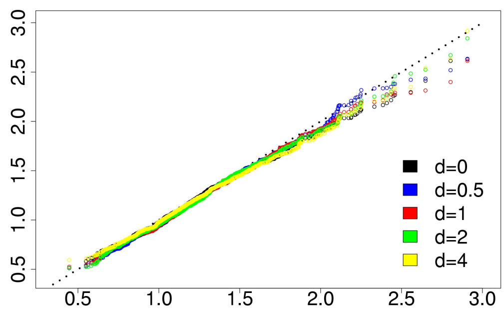
(11) $\rho = 0.3$, $N = 80$



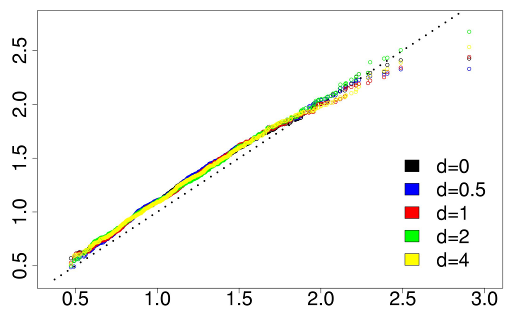
(12) $\rho = 0.5$, $N = 80$

Figure 6.3.1.1(ii).: QQ-plots of $T_n^{(3)}(q_1)$ (under H_0) against $T_n^{(3f)}(\mathbf{R}, q_1)$ for double exponentially distributed innovations, $n = 80$, $m = 40$ and different values for ρ and N

Finally we also give a selection of size-power-curves in Figures 6.3.1.3(i), 6.3.1.3(ii) respectively 6.3.1.4. Here again the results are very similar for all three choices of N. To create the SPC-plots we use 10 000 time series according to the model (null hypothesis as well as alternatives) and for each of these 1 000 permutations. When we only use

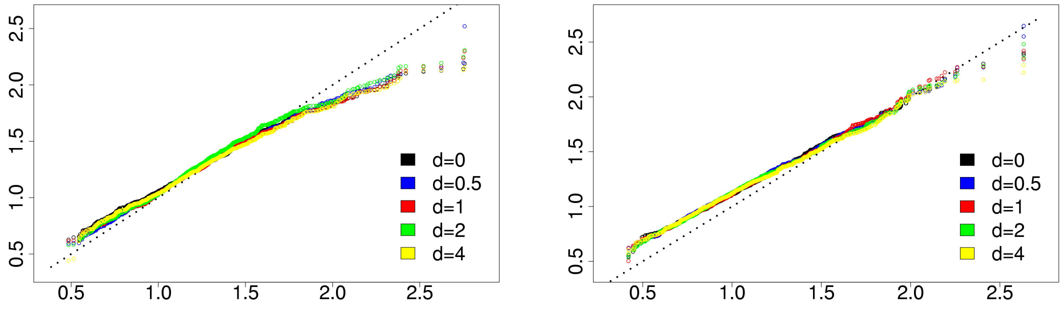


(1) $\rho = -0.1$, $N = 80$



(2) $\rho = 0.3$, $N = 80$

Figure 6.3.1.2(i).: QQ-plots of $T_n^{(3)}(q_2)$ (under H_0) against $T_n^{(3f)}(\mathbf{R}, q_2)$ for standard normally distributed innovations, $n = 80$, $m = 40$ and different values for ρ and N

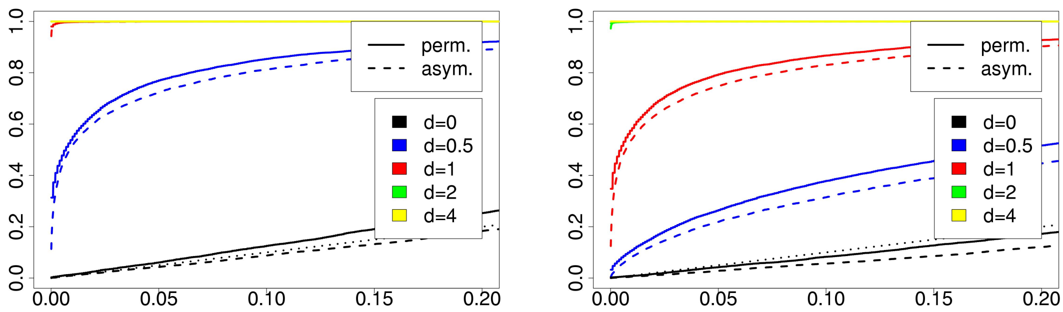


(3) $\rho = -0.1$, $N = 80$ (4) $\rho = 0.3$, $N = 80$

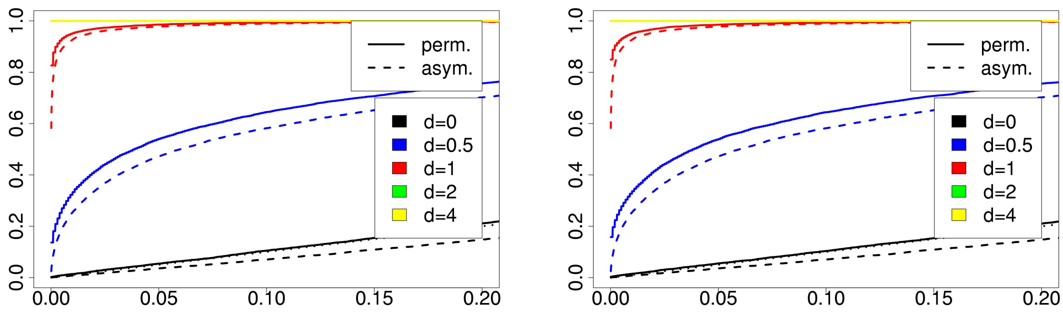
Figure 6.3.1.2(ii).: QQ-plots of $T_n^{(3)}(q_2)$ (under H_0) against $T_n^{(3f)}(\mathbf{R}, q_2)$ for double exponentially distributed innovations, $n = 80$, $m = 40$ and different values for ρ and N

1 000 time series, the plots depend on the seed of the random generator. Using 10 000 gives almost identical pictures and thus the plots can be considered to be correct.

The match is usually better than the asymptotic quantiles (if known) with the exception of negatively correlated errors, where the asymptotic quantiles match best.

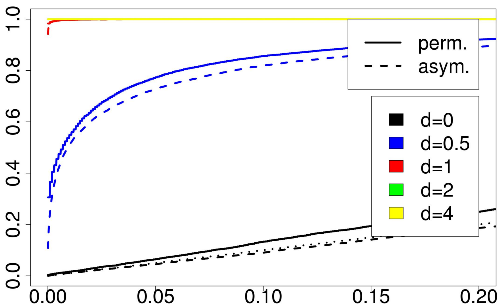


(1) $\rho = -0.3$, $N = 80$ (2) $\rho = 0.3$, $N = 80$

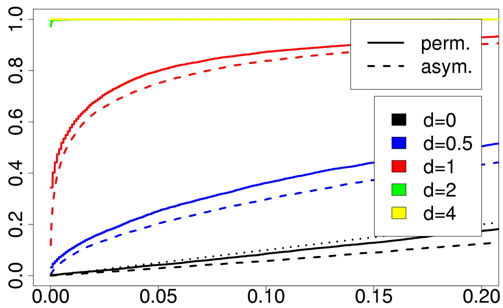


(3) $\rho = 0$, $N = 60$ (4) $\rho = 0$, $N = 80$

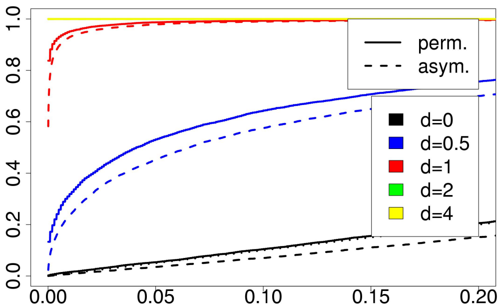
Figure 6.3.1.3(i).: SPC-plots for $T_n^{(3f)}(\mathbf{R}, q_1)$ for standard normally distributed innovations, $n = 80$, $m = 40$ and different values for ρ and N



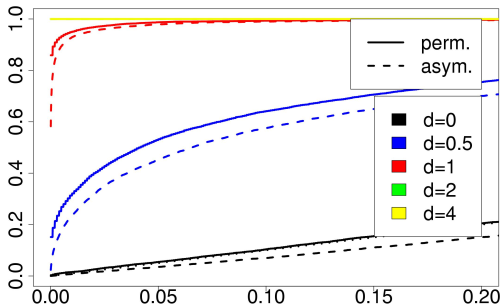
(5) $\rho = -0.3$, $N = 80$



(6) $\rho = 0.3$, $N = 80$

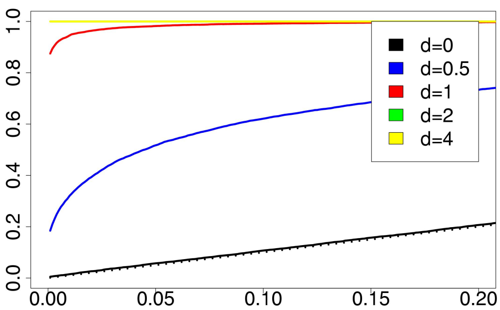


(7) $\rho = 0$, $N = 60$

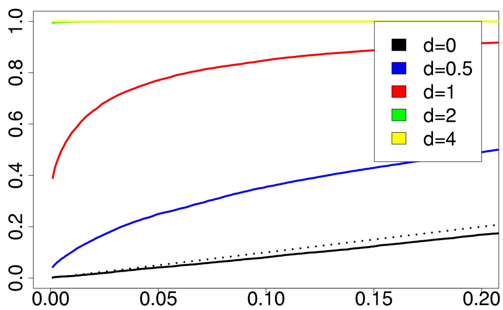


(8) $\rho = 0$, $N = 80$

Figure 6.3.1.3(ii).: SPC-plots for $T_n^{(3f)}(\mathbf{R}, q_1)$ for double exponentially distributed innovations, $n = 80$, $m = 40$ and different values for ρ and N



(1) $\rho = 0$, $N = 80$



(2) $\rho = 0.3$, $N = 80$

Figure 6.3.1.4.: SPC-plots for $T_n^{(3f)}(\mathbf{R}, q_2)$ for standard normally distributed innovations, $n = 80$, $m = 40$ and different values for ρ and N

6.3.2. Sum Statistic

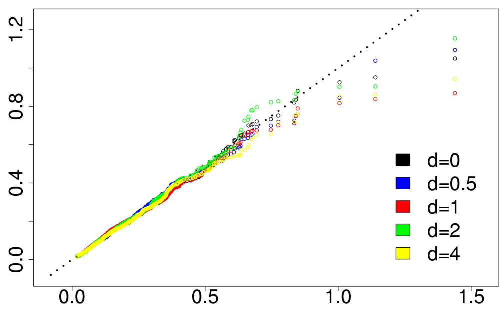
In this subsection we investigate the sum statistics $T_n^{(4)}(r)$ for $r_1 \equiv 1$ as well as $r_2 := (\mathrm{id}(1-\mathrm{id}))^{\frac{3}{2}}$. $\tilde{T}_n^{(4f)}(r)$ denotes the corresponding frequency permutation statistic, it is essentially statistic $T_n^{(4f)}(r)$ from Corollary 4.6.1, where we approximated the integral by a sum. Precisely

$$\tilde{T}_n^{(4f)}(r) = \frac{1}{N} \sum_{m=1}^{N-1} \frac{1}{r(m/N)} \left[\widetilde{Z}_n^{\mathbf{X}}(m/N, \mathbf{R}) - \frac{m}{N} \widetilde{Z}_n^{\mathbf{X}}(1, \mathbf{R}) \right]^2$$

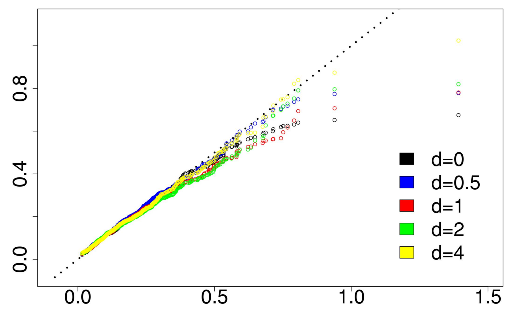
with $\{\widetilde{Z}_n^{\mathbf{X}}(u, \mathbf{R})\}$ as in Theorem 4.6.2. We only give a short extract from the simulation, yet results are similar to the previous sections. We concentrate on r_1, since then we also know the limit distribution.

Some QQ-plots can be found in Figures 6.3.2.1(i) and 6.3.2.1(ii), some size-power-curves in Figures 6.3.2.2(i), 6.3.2.2(ii) respectively 6.3.2.3(i), 6.3.2.3(ii).

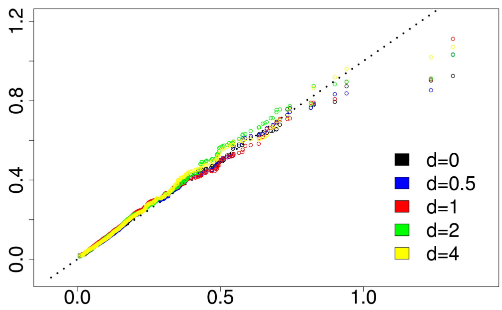
For r_1 we have very good plots for both the frequency permutation test as well as the asymptotic one. For r_2 the plot is best for independent errors in case of normal innovations and for positively correlated errors in the case of double exponentially distributed errors.



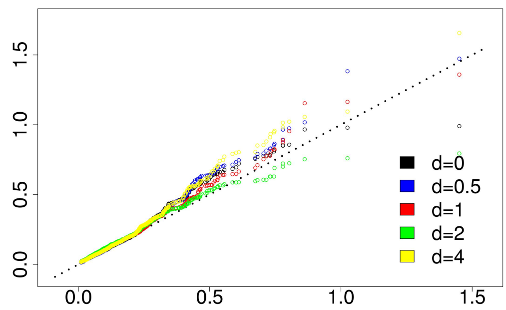
(1) $\rho = -0.3$, $N = 80$



(2) $\rho = 0$, $N = 80$

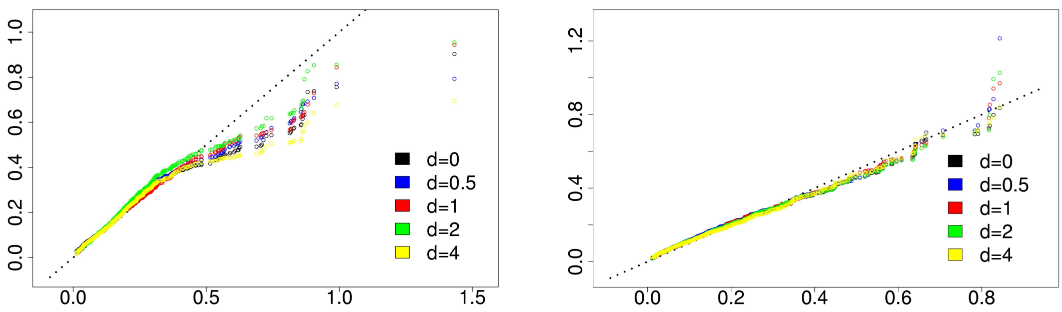


(3) $\rho = 0.3$, $N = 80$



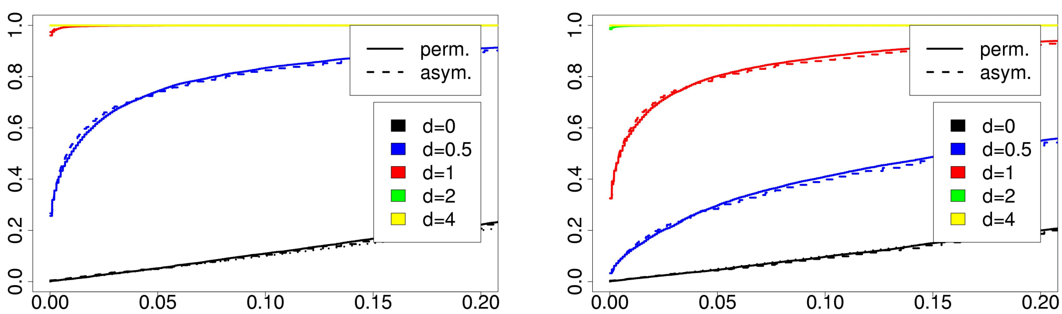
(4) $\rho = 0.5$, $N = 80$

Figure 6.3.2.1(i).: QQ-plots of $T_n^{(4)}(r_1)$ (under H_0) against $\tilde{T}_n^{(4f)}(\mathbf{R}, r_1)$ for standard normally distributed innovations, $n = 80$, $m = 40$ and different values for ρ and N



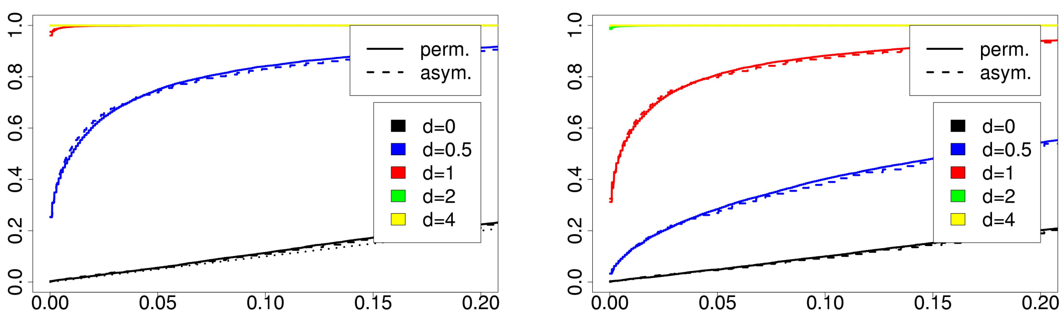
(5) $\rho = -0.1$, $N = 80$ (6) $\rho = 0.3$, $N = 80$

Figure 6.3.2.1(ii).: QQ-plots of $T_n^{(4)}(r_1)$ (under H_0) against $\tilde{T}_n^{(4f)}(\mathbf{R}, r_1)$ for double exponentially distributed innovations, $n = 80$, $m = 40$ and different values for ρ and N



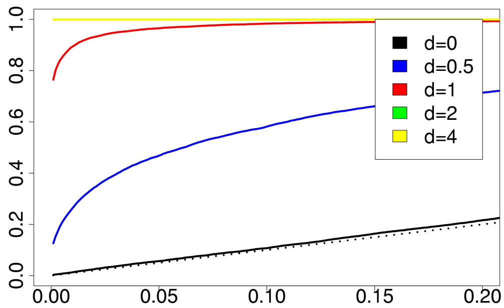
(1) $\rho = -0.3$, $N = 80$ (2) $\rho = 0.3$, $N = 80$

Figure 6.3.2.2(i).: SPC-plots for $\tilde{T}_n^{(4f)}(\mathbf{R}, r_1)$ for standard normally distributed innovations, $n = 80$, $m = 40$ and different values for ρ and N

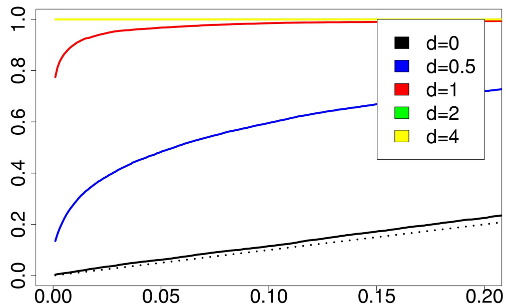


(3) $\rho = -0.3$, $N = 80$ (4) $\rho = 0.3$, $N = 80$

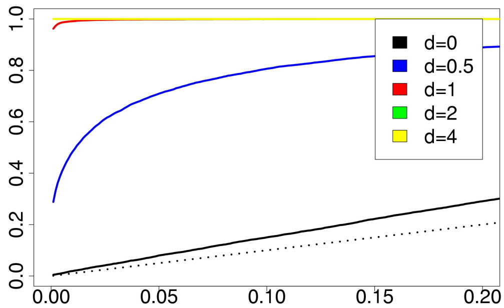
Figure 6.3.2.2(ii).: SPC-plots for $\tilde{T}_n^{(4f)}(\mathbf{R}, r_1)$ for double exponentially distributed innovations, $n = 80$, $m = 40$ and different values for ρ and N



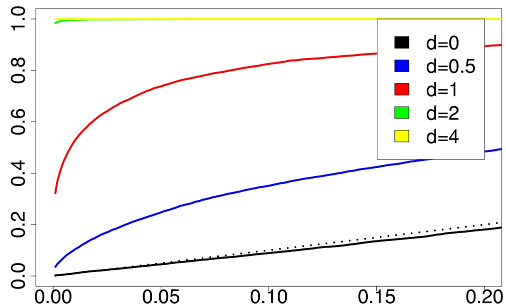
(1) $\rho = 0,\ N = 60$



(2) $\rho = 0,\ N = 80$

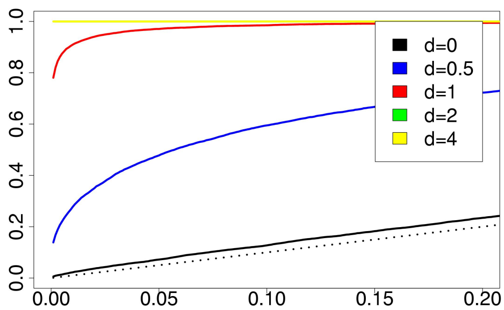


(3) $\rho = -0.3,\ N = 80$

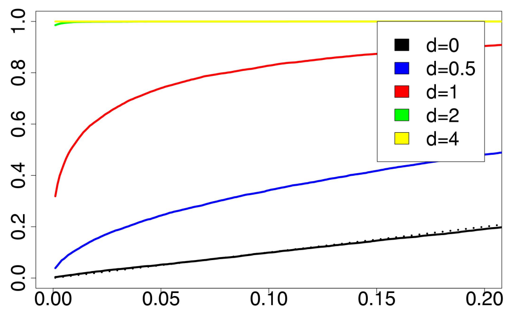


(4) $\rho = 0.3,\ N = 80$

Figure 6.3.2.3(i).: SPC-plots for $\tilde{T}_n^{(4f)}(\mathbf{R}, r_2)$ for standard normally distributed innovations, $n = 80$, $m = 40$ and different values for ρ and N



(5) $\rho = 0,\ N = 80$



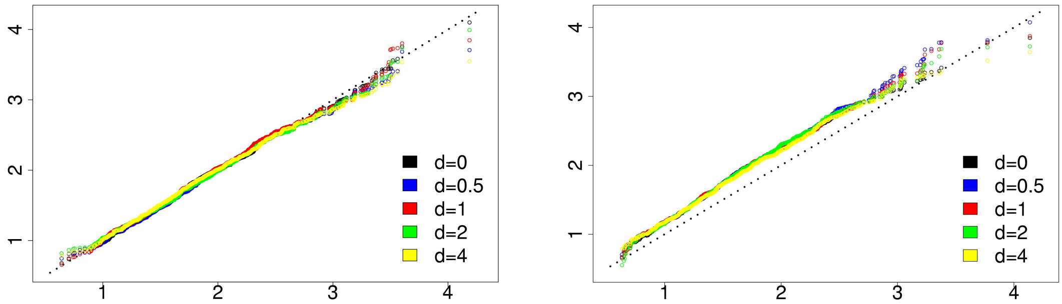
(6) $\rho = 0.3,\ N = 80$

Figure 6.3.2.3(ii).: SPC-plots for $\tilde{T}_n^{(4f)}(\mathbf{R}, r_2)$ for double exponentially distributed innovations, $n = 80$, $m = 40$ and different values for ρ and N

6.3.3. Weighted CUSUM Statistic

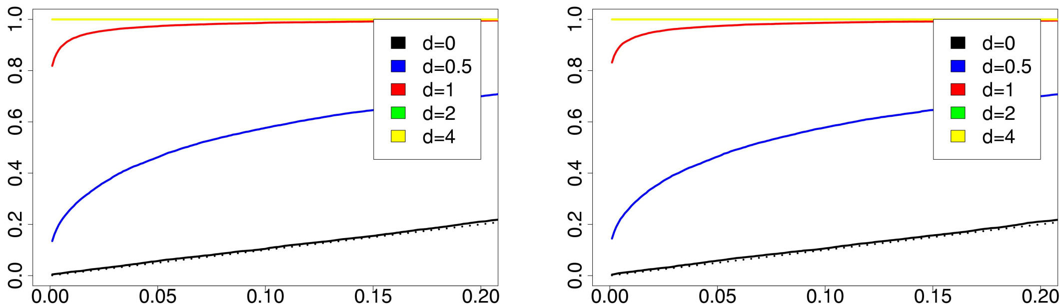
Now we investigate the trimmed version of the weighted CUSUM statistic ($\epsilon = 0.1$)

$$\tilde{T}_n^{(1)} := \max_{0.1n \leqslant m < 0.9n} \left(\sqrt{\frac{n}{m(n-m)}} |S_m| \right).$$

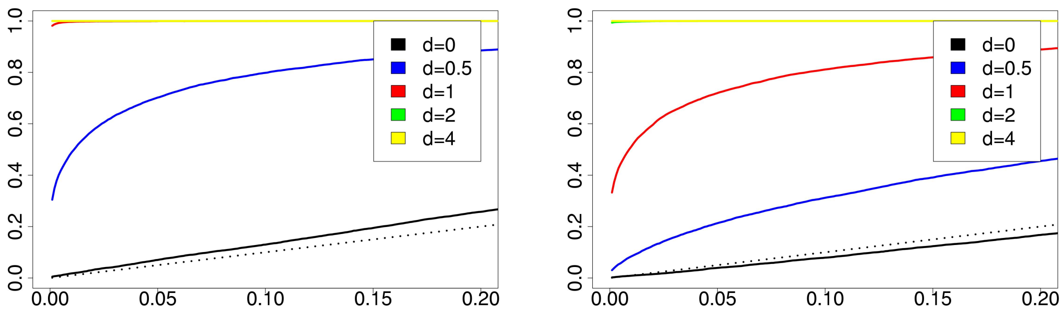


(1) $\rho = -0.1$, $N = 80$ (2) $\rho = 0.5$, $N = 80$

Figure 6.3.3.1.: QQ-plots of $\tilde{T}_n^{(1)}$ (under H_0) against $T_n^{(1f)}(\mathbf{R})$ for standard normally distributed innovations, $n = 80$, $m = 40$ and different values for ρ and N



(1) $\rho = 0$, $N = 60$ (2) $\rho = 0$, $N = 80$



(3) $\rho = -0.3$, $N = 80$ (4) $\rho = 0.3$, $N = 80$

Figure 6.3.3.2.: SPC-plots for $T_n^{(1f)}(\mathbf{R})$ for standard normally distributed innovations, $n = 80$, $m = 40$ and different values for ρ and N

Since it is very similar to the results for the q-weighted CUSUM statistic we only give some QQ- and SPC-plots in Figures 6.3.3.1 and 6.3.3.2 for normally distributed errors, yet the plots for double exponentially distributed innovation are very similar.

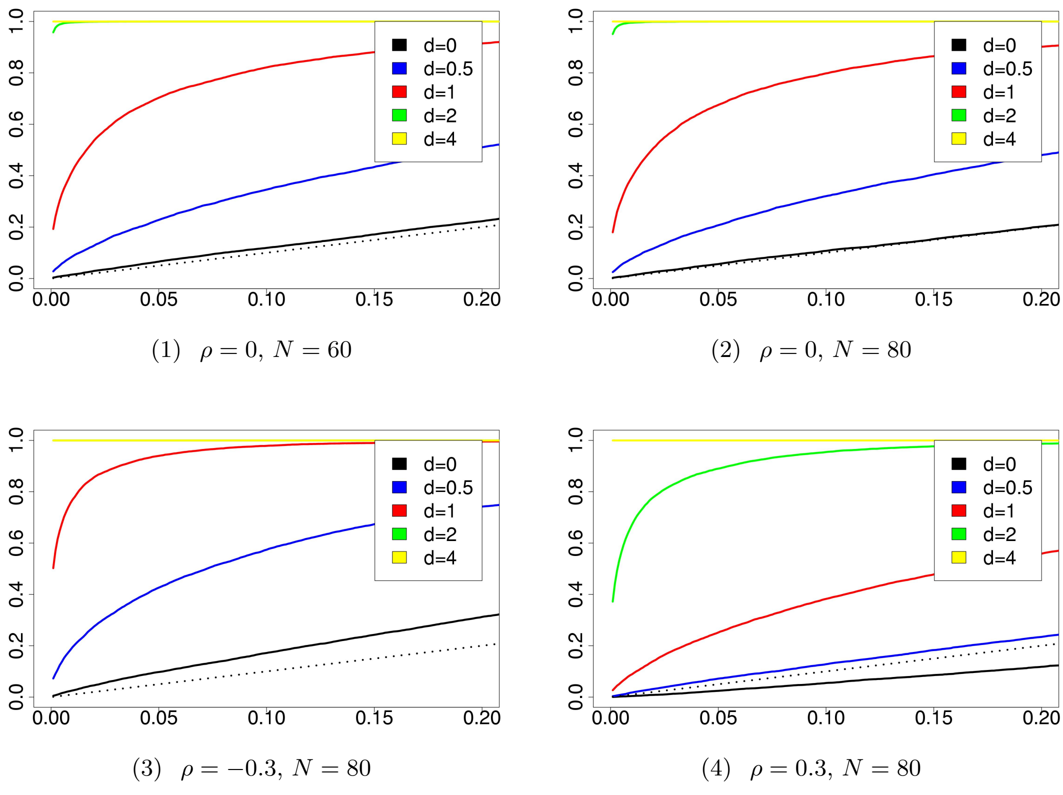
6.3.4. MOSUM Statistic

Finally we investigate the trimmed version of the MOSUM statistic ($\epsilon = 0.1$)

$$\tilde{T}_n^{(2)} := \max_{0.1n \leqslant m < 0.9n} \frac{1}{\sqrt{n}} |S_m - S_{m-0.1n}|.$$

It corresponds to $G = 0.1n$.

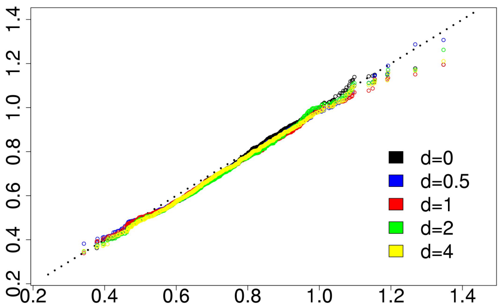
Just as the block permutation test the frequency permutation test only works fine for the MOSUM statistic if we have independent errors. The results for the double exponentially distributed innovations and normal ones are very similar, so we only give some QQ-plots and size-power-curves for normal innovations in Figures 6.3.4.2 and 6.3.4.1.



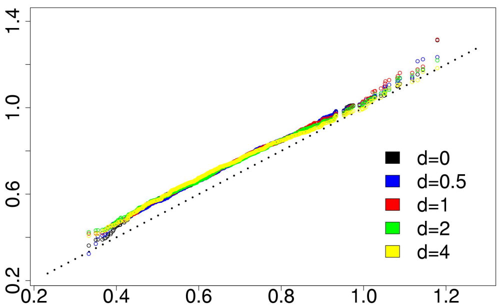
(1) $\rho = 0$, $N = 60$ (2) $\rho = 0$, $N = 80$

(3) $\rho = -0.3$, $N = 80$ (4) $\rho = 0.3$, $N = 80$

Figure 6.3.4.1.: SPC-plots for $T_n^{(2f)}(\mathbf{R})$ for standard normally distributed innovations, $n = 80$, $m = 40$ and different values for ρ and N



(1) $\rho = -0.1$, $N = 80$



(2) $\rho = 0.3$, $N = 80$

Figure 6.3.4.2.: QQ-plots of $\tilde{T}_n^{(2)}$ (under H_0) against $T_n^{(2f)}(\mathbf{R})$ for standard normally distributed innovations, $n = 80$, $m = 40$ and different values for ρ and N

Part III.

Appendix

A. Landau Symbols in Stochastics

In this appendix we give definitions as well as some easy rules of calculus for the deterministic as well as stochastic Landau symbols. We use them quite frequently throughout this work without further comment. Those symbols are short expressions for convergence respectively boundedness and are very convenient in longer calculations. We do not explicitely give the proofs for the rules of calculus, but they are easily obtained by definition in case of deterministic or *a.s.* Landau symbols. The P-stochastic rules are a little more difficult to obtain, for details confer e.g. van der Vaart [81], Section 2.2.

Definition A.1 (Deterministic Landau Symbols). *Let $\{X_n\}$, $\{U_n\}$ be deterministic sequences with $U_n > 0$, then as $n \to \infty$*

$$X_n = o(U_n) \qquad :\Longleftrightarrow \qquad \frac{X_n}{U_n} \to 0 \quad \text{as } n \to \infty;$$

$$X_n = O(U_n) \qquad :\Longleftrightarrow \qquad |X_n| \leqslant C\, U_n \quad \text{for some } C > 0, \text{for all } n.$$

Note that even though it is standard to use the equality sign, this is in fact not a symmetric relation.

We also use the following notation for $O(\cdot)$, because it is more convenient in longer inequalities. It reminds of an inequality sign and ultimately it is an inequality sign that also contains constants.

Definition A.2. *Let $\{X_n\}; \{U_n\}$ be sequences, $\{U_n\}$ additionally positive, then*

$$X_n \ll U_n \qquad :\Longleftrightarrow \qquad X_n = O(U_n).$$

We now obtain immediately some rules of calculus for the Landau Symbols.

Lemma A.1. *Let $\{X_n\}$, $\{Y_n\}$ be deterministic sequences, $\{U_n\}$, $\{V_n\}$ additionally positive, then it holds:*

(i) $X_n = o(U_n) \quad \Longrightarrow \quad X_n = O(U_n).$

(ii) $X_n = O(U_n),\ Y_n = O(V_n) \quad \Longrightarrow \quad X_n \pm Y_n = O(\max(U_n, V_n)),\ X_nY_n = O(U_nV_n).$

(iii) $X_n = o(U_n),\ Y_n = o(V_n) \quad \Longrightarrow \quad X_n \pm Y_n = o(\max(U_n, V_n)).$

(iv) $X_n = O(U_n),\ Y_n = o(V_n) \quad \Longrightarrow \quad X_nY_n = o(U_nV_n).$

(v) $X_n = O(U_n),\ U_n = O(V_n) \quad \Longrightarrow \quad X_n = O(V_n),$
if we replace at least one of the O by o we obtain $X_n = o(V_n)$.

(vi) *If* $X_n \to X$ *as* $n \to \infty$, *it holds* $X_n = O(1)$.

(vii) $X_n = O(1) \quad \Longrightarrow \quad \max_{1 \leqslant i \leqslant n} X_i = O(1)$.

Proof. The assertions are easily obtained from the definitions, which is why they are omitted. ■

Now we turn to stochastic Landau symbols.
Note that they reduce to the deterministic definitions for deterministic sequences.

Definition A.3 (Stochastic Landau Symbols). *Let* $\{X_n\}$, $\{U_n\}$ *be stochastic sequences with* $U_n > 0$ *almost surely, then as* $n \to \infty$

$$X_n = o_P(U_n) \qquad :\Longleftrightarrow \qquad \frac{X_n}{U_n} \xrightarrow{P} 0 \quad \textit{as } n \to \infty;$$

$$X_n = o(U_n) \quad a.s. \qquad :\Longleftrightarrow \qquad \frac{X_n}{U_n} \to 0 \quad a.s. \quad \textit{as } n \to \infty;$$

$$X_n = O_P(U_n) \qquad :\Longleftrightarrow \qquad \forall \epsilon \ \exists C : P\left(|X_n| > C\, U_n\right) \leqslant \epsilon \quad \forall n.$$

$$X_n = O(U_n) \quad a.s. \qquad :\Longleftrightarrow \qquad \exists C : |X_n| \leqslant C\, U_n \quad a.s. \qquad \forall n,$$

where in the last line C *might depend on the realization* ω. *In the almost sure case we again also use* $X_n \ll U_n$ *a.s. instead of* $X_n = O(U_n)$ *a.s.*

We also have the following rules of calculus:

Lemma A.2. *Let now* $\{X_n\}$, $\{Y_n\}$ *be stochastic sequences,* $\{U_n\}$, $\{V_n\}$ *additionally almost surely positive, then it holds:*

a) If we replace the deterministic Landau symbols with the corresponding almost sure symbols, assertions (i)-(vii) of Lemma A.1 remain true.

b) If we replace o, O *with* o_P *respectively* O_P, *assertions (i)-(v) of Lemma A.1 remain true. In assertion (vi) we get* $X_n = O_P(1)$ *for* $X_n \xrightarrow{P} X$ *for some random variable* X, *particularly it holds for any random variable* $X = O_P(1)$. *Assertion (vii) is in general not fulfilled.*

c) $X_n = O(Y_n) \quad a.s. \quad \Longrightarrow \quad X_n = O_P(1)$
and $X_n = o(Y_n) \quad a.s. \quad \Longrightarrow \quad X_n = o_P(1)$.

Proof. For the almost sure convergence this follows immediately from the definitions and Lemma A.1. The results for o_P and O_P are somewhat more difficult. Some details can be found in van der Vaart [81], Section 2.2. ■

B. Some Useful Inequalities

In this appendix we state some inequalities we frequently use throughout this work. A first section deals with Hájek–Rényi–type inequalities not only for the i.i.d. case but also for dependent random variables. In a second section we give moment inequalities for higher order moments of sums of possibly dependent random variables. This also leads to upper bounds for higher order moments of the maximum of partial sums. Finally one can use these moment inequalities to obtain a strong law of large numbers, which is essential for the proof of Theorem 3.5.1, and even give a convergence rate.

B.1. Hájek–Rényi–Type Inequalities

The first lemma shows that the Hájek–Rényi inequality remains true for linear processes. That result can be obtained from the inequality for i.i.d. random variables by an application of the BN decomposition in Lemma C.1.

Lemma B.1. *Let $\{b_k\}$ be a non-increasing positive sequence of constants, i.e. $b_1 \geqslant b_2 \geqslant \ldots \geqslant b_n > 0$. If*

$$e(i) = \sum_{s \geqslant 0} w_s \, \epsilon(i-s), \qquad i = 1, 2, \ldots$$

is a linear process with i.i.d. innovations with existing second moment and

$$\sum_{s \geqslant 0} \sqrt{s} \, |w_s| < \infty,$$

the following inequalities hold:

$$a) \qquad P\left(\max_{k_0 \leqslant k \leqslant n} b_k \left| \sum_{i=1}^{k} e(i) \right| > \alpha \right) \leqslant \frac{C}{\alpha^2} \left(k_0 b_{k_0}^2 + \sum_{i=k_0+1}^{n} b_i^2 \right).$$

$$b) \qquad P\left(\max_{k_0 \leqslant k \leqslant n-l} b_{n-k} \left| \sum_{i=k+1}^{n} e(i) \right| > \alpha \right) \leqslant \frac{C}{\alpha^2} \left(l b_l^2 + \sum_{i=l+1}^{n-k_0} b_i^2 \right),$$

where $C < \infty$ is a constant only depending on the weights and the variance of the linear process.

Proof. The proof of a) can be found in Bai [6], Proposition 1. The assumptions on the weights are somewhat stronger there. Lemma C.1 and Remark C.1, however, show that

the BN decomposition remains true under the above assumptions, i.e. there exists a stationary process $\widetilde{e}(\cdot)$ with finite variance such that $e(i) = \epsilon(i)\sum_{j\geqslant 0} w_j - \widetilde{e}(i) + \widetilde{e}(i-1)$. This gives $\sum_{i=k+1}^{n} e(i) = \sum_{i=k+1}^{n} \epsilon(i)\sum_{j\geqslant 0} w_j + \widetilde{e}(k) - \widetilde{e}(n)$.
The proof of b) is analogous, yet we give it for the sake of completeness. It is also based on the above BN decomposition. Because $\{b_k\}$ is non-increasing it holds $b_{n-k} \leqslant b_l$, $k \leqslant n-l$, thus the Chebyshev inequality and the Hájek–Rényi inequality for independent errors give

$$\begin{aligned}
&P\left(\max_{k_0\leqslant k\leqslant n-l} b_{n-k}\left|\sum_{j=k+1}^{n} e(j)\right| > \alpha\right)\\
&\leqslant P\left(\left|\sum_{j\geqslant 0} w_j\right| \max_{k_0\leqslant k\leqslant n-l} b_{n-k}\left|\sum_{j=k+1}^{n} \epsilon(j)\right| > \alpha/3\right)\\
&\quad + \sum_{k=k_0}^{n-l} P\left(b_{n-k}|\widetilde{e}(k)| > \alpha/3\right) + P\left(b_l|\widetilde{e}(n)| > \alpha/3\right)\\
&\leqslant P\left(\left|\sum_{j\geqslant 0} w_j\right| \max_{l\leqslant k\leqslant n-k_0} b_k\left|\sum_{j=1}^{k} \epsilon(n-j+1)\right| > \alpha/3\right) + \frac{\widetilde{C}}{\alpha^2}\left(\sum_{k=l}^{n-k_0} b_k^2 + b_l^2\right)\\
&\leqslant \frac{C}{\alpha^2}\left(l b_l^2 + \sum_{j=l+1}^{n-k_0} b_j^2\right).
\end{aligned}$$

■

We turn now to a Hájek–Rényi–type inequality for possibly dependent random variables, where the only assumption is one on the higher order moments of their partial sums. That assumption is fulfilled for a large class of random variables, not surprisingly the result is somewhat weaker.

Móricz et al. [65] , Theorem 3.1., give an approximation of higher order moments for the maximum of partial sums. Going along the lines of their proof it is possible to obtain a first result for moments of maxima of weighted partial sums as the following theorem shows. For $\gamma = 2$ it was proven by Lavielle and Moulines [58].

Theorem B.1. *Let $\{Y(i) : i \in \mathbb{N}\}$ be a sequence of random variables satisfying*

$$\mathrm{E}\,|S_{i,j}|^{\gamma} \leqslant C|j-i+1|^{\varphi}$$

for some $\gamma \geqslant 1$, $\varphi > 1$ and some constant $C > 0$, where $S_{i,j} = \sum_{l=i}^{j} Y(l)$.

Then for any positive and non-increasing sequence $b_1 \geqslant b_2 \geqslant \ldots \geqslant b_n > 0$, there exists a constant $A(\varphi,\gamma) \geqslant 1$ (only depending on φ and γ) with

$$\mathrm{E}\left[\max_{k=1,\ldots,n}(b_k|S_{1,k}|)\right]^{\gamma} \leqslant C A(\varphi,\gamma) n^{\varphi-1}\sum_{k=1}^{n} b_k^{\gamma},$$

where C is as above.

Proof. The proof follows closely that of Theorem 3.1. in Móricz et al. [65].
For $n = 1$ the assertion is obviously right for any $A(\varphi,\gamma) \geqslant 1$. We prove the assertion

for $n > 1$ by induction. Let $m = \lfloor \frac{n}{2} \rfloor + 1$, which gives

$$m - 1 \leqslant \frac{n}{2}, \qquad n - m \leqslant \frac{n}{2}.$$

Moreover let $M_{i,j} := \max_{k=i,\ldots,j} b_k |S_{i,k}|$. Since for $k > m$ it holds $b_k|S_{1,k}| \leqslant b_k(|S_{1,m}| + |S_{(m+1),k}|) \leqslant b_m|S_{1,m}| + b_k|S_{(m+1),k}|$, we get

$$\begin{aligned} M_{1,n} &= \max[M_{1,(m-1)}\,,\, b_m|S_{1,m}|\,,\, \max_{k=m+1,\ldots,n} b_k|S_{1,k}|] \\ &\leqslant b_m|S_{1,m}| + \left(M^{\gamma}_{1,(m-1)} + M^{\gamma}_{(m+1),n}\right)^{\frac{1}{\gamma}}. \end{aligned}$$

Note that $m b_m^{\gamma} \leqslant \sum_{i=1}^{m} b_i^{\gamma} \leqslant \sum_{i=1}^{n} b_i^{\gamma}$, which together with the Minkowski inequality and the induction hypothesis gives

$$\begin{aligned} &(\mathrm{E}\, M^{\gamma}_{1,n})^{\frac{1}{\gamma}} \\ &\leqslant b_m(\mathrm{E}\,|S_{1,m}|^{\gamma})^{\frac{1}{\gamma}} + \left(\mathrm{E}(M^{\gamma}_{1,(m-1)}) + \mathrm{E}(M^{\gamma}_{(m+1),n})\right)^{\frac{1}{\gamma}} \\ &\leqslant b_m(Cm^{\varphi})^{\frac{1}{\gamma}} + \left(A(\varphi,\gamma)C\left[(m-1)^{\varphi-1}\sum_{k=1}^{m-1} b_k^{\gamma} + (n-m)^{\varphi-1}\sum_{k=m+1}^{n} b_k^{\gamma}\right]\right)^{\frac{1}{\gamma}} \\ &\leqslant \left(Cm^{\varphi-1}\sum_{i=1}^{n} b_i^{\gamma}\right)^{\frac{1}{\gamma}} + \left(A(\varphi,\gamma)C\left(\frac{n}{2}\right)^{\varphi-1}\left[\sum_{k=1}^{m-1} b_k^{\gamma} + \sum_{k=m+1}^{n} b_k^{\gamma}\right]\right)^{\frac{1}{\gamma}} \\ &\leqslant \left(Cn^{\varphi-1}\sum_{k=1}^{n} b_k^{\gamma}\right)^{\frac{1}{\gamma}}\left[1 + \left(\frac{A(\varphi,\gamma)}{2^{\varphi-1}}\right)^{\frac{1}{\gamma}}\right]. \end{aligned}$$

Now we only need to find a constant $A(\varphi,\gamma) \geqslant 1$ such that

$$1 + \left(\frac{A(\varphi,\gamma)}{2^{\varphi-1}}\right)^{\frac{1}{\gamma}} \leqslant A(\varphi,\gamma)^{\frac{1}{\gamma}}. \tag{B.1}$$

By the assumption that $\varphi > 1$ it holds $\frac{1}{2^{\varphi-1}} < 1$, which shows that equation (B.1) is fulfilled for any constant

$$A(\varphi,\gamma) \geqslant \left(1 - \frac{1}{2^{\frac{\varphi-1}{\gamma}}}\right)^{-\gamma} \geqslant 1.$$

This gives the assertion. ■

However, one gets an improved version if one uses the result by Móricz et al. [65] (i.e. the above result for $b_j \equiv 1$) and the following theorem by Fazekas and Klesov [29]:

Theorem B.2. *Let $\alpha_1, \ldots, \alpha_n$ be non-negative numbers, $\{Y(i) : i \in \mathbb{N}\}$ a sequence of random variables satisfying for each $1 \leqslant m \leqslant n$*

$$\mathrm{E} \max_{1 \leqslant j \leqslant m} |S_j|^{\gamma} \leqslant \sum_{j=1}^{m} \alpha_l$$

for some $\gamma \geqslant 0$, where $S_j = \sum_{l=1}^{j} Y(l)$.

Then for any positive and non-increasing sequence $b_1 \geqslant b_2 \geqslant \ldots \geqslant b_n > 0$ it holds

$$\mathrm{E}\left[\max_{k=1,\ldots,n}(b_k|S_k|)\right]^\gamma \leqslant 4\sum_{k=1}^n b_k^\gamma \alpha_k.$$

Proof. Confer Theorem 1.1 in Fazekas and Klesov [29]. ■

We now get an improved version of Theorem B.1.

Theorem B.3. *Let $\{Y(i) : i \in \mathbb{N}\}$ be a sequence of random variables satisfying*

$$\mathrm{E}\,|S_{i,j}|^\gamma \leqslant C|j-i+1|^\varphi$$

for some $\gamma \geqslant 1$, $\varphi > 1$ and some constant $C > 0$, where $S_{i,j} = \sum_{l=i}^j Y(l)$.

Then for any positive and non-increasing sequence $b_1 \geqslant b_2 \geqslant \ldots \geqslant b_n > 0$, there exists a constant $A(\varphi,\gamma) \geqslant 4$ (only depending on φ and γ) with

$$\mathrm{E}\left[\max_{k=1,\ldots,n}(b_k|S_{1,k}|)\right]^\gamma \leqslant CA(\varphi,\gamma)\sum_{k=1}^n b_k^\gamma k^{\varphi-1},$$

where C is as above.

Proof. First of all Theorem B.1 with $b_j \equiv 1$ gives

$$\mathrm{E}\left[\max_{k=1,\ldots,m}|S_{1,k}|\right]^\gamma \leqslant C\widetilde{A}(\varphi,\gamma)m^\varphi \leqslant C(\varphi\widetilde{A}(\varphi,\gamma))\sum_{k=1}^m k^{\varphi-1},$$

since $\sum_{l=1}^m l^{\varphi-1} \geqslant \int_0^m x^{\varphi-1}\,dx = \frac{1}{\varphi}m^\varphi$. Then Theorem B.2 gives

$$\mathrm{E}\left[\max_{k=1,\ldots,n}(b_k|S_{1,k}|)\right]^\gamma \leqslant C(4\varphi\widetilde{A}(\varphi,\gamma))\sum_{k=1}^n b_k^\gamma k^{\varphi-1},$$

which is the desired assertion. ■

A corresponding result for $\varphi = 1$ is given in the following theorem.

Theorem B.4. *Let $\{Y(i) : i \in \mathbb{N}\}$ be a sequence of random variables satisfying*

$$\mathrm{E}\,|S_{i,j}|^\gamma \leqslant C|j-i+1|$$

for some $\gamma > 1$ and some constant $C > 0$, where $S_{i,j} = \sum_{l=i}^j Y(l)$ and $b_1 \geqslant b_2 \geqslant \ldots \geqslant b_n > 0$ a positive and non-increasing sequence. Then

$$\mathrm{E}\left[\max_{k=1,\ldots,n} b_k|S_{1,k}|\right]^\gamma \leqslant 4C(\log(2n))^\gamma \sum_{j=1}^n b_j^\gamma$$

where C is as above.

Proof. Theorem 3 in Móricz [64] gives the result for $b_j \equiv 1$. An application of Theorem B.2 then gives the desired assertion. ■

B.2. Moment Inequalities and Strong Law of Large Numbers

In this section we summarize some results on moment inequalities for strong-mixing random sequences. Further we use these results to obtain a strong law of large numbers for triangular arrays of dependent random variables.

First we give the definition for strong-mixing.

Definition B.1 (strong-mixing). *Given a random sequence $\{Y(i) : i \geqslant 1\}$, let $\mathcal{A}_n^m$ be the σ-Algebra generated by $\{Y(i) : n \leqslant i \leqslant m\}$, and define the corresponding α-mixing sequence by*

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

where A and B vary over the σ-fields $\mathcal{A}_{-\infty}^n$ and $\mathcal{A}_{n+k}^\infty$, respectively. We call $\alpha_Y(k)$ the mixing coefficient. Note that in case the sequence $\{Y(\cdot)\}$ is strictly stationary, the $\sup_n$ in this definition becomes redundant. The sequence $\{Y(\cdot)\}$ is called α-mixing or strong-mixing if $\alpha_Y(k) \to 0$ as $k \to \infty$.

The following theorem gives a moment inequality for α-mixing random variables, which are not necessarily stationary.

Theorem B.5. *Let $\{Y(i) : i \in \mathbb{Z}\}$ be a random sequence with $\mathrm{E}\, Y(i) = 0$, $i \in \mathbb{Z}$. Assume there is a $\delta, \Delta > 0$, $2l \leqslant \delta \leqslant 2(l+1)$, $l = 0, 1, 2 \ldots$, with*

$$\mathrm{E}\, |Y(i)|^{2+\delta+\Delta} \leqslant D_1 \quad \textit{for all } i \in \mathbb{Z} \tag{B.1}$$

and

$$\sum_{k=0}^{\infty} (k+1)^{2l+2} \alpha_Y(k)^{\Delta/(2l+4+\Delta)} \leqslant D_2(\delta, \Delta), \tag{B.2}$$

where α_Y is the corresponding mixing coefficient. Then it holds

$$\mathrm{E} \left| \sum_{i=1}^{n} Y(i) \right|^{2+\delta} \leqslant \Gamma(D_1, D_2, \delta, \Delta)\, n^{(2+\delta)/2},$$

where $\Gamma(D_1, D_2, \delta, \Delta)$ is a constant just depending on D_1, D_2, δ, and Δ.

Proof. Analogous to Politis et al. [73], Corollary A.0.1, p. 319. ■

The following theorem gives the corresponding result for stationary sequences. Here the conditions on the α-mixing sequence is somewhat weaker.

Theorem B.6. *Let $\{Y(i) : i \in \mathbb{Z}\}$ be a strictly stationary sequence with $\mathrm{E}\, Y(i) = 0$, $i \in \mathbb{Z}$. Assume there are $\delta, \Delta > 0$ with*

$$\mathrm{E}\, |Y(i)|^{2+\delta+\Delta} \leqslant D_1 \quad \textit{for all } i \in \mathbb{Z} \tag{B.3}$$

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

$$\sum_{k=0}^{\infty} (k+1)^{\delta/2} \alpha(k)^{\Delta/(2+\delta+\Delta)} \leqslant D_2(\delta, \Delta), \tag{B.4}$$

where α_Y is the corresponding mixing coefficient. Then it holds

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

where $\Gamma(D_1,\alpha,\delta,\Delta)$ is a constant just depending on D_1, $\alpha(k)$, $k \in \mathbb{N}$, δ, and Δ.

Proof. Confer Yokoyama [83], Theorem 1. ■

Under some stronger assumptions, e.g. for martingale difference sequences, certain stationary and Φ-mixing sequences, certain stationary aperiodic Markov sequences, we do not need the existence of a higher moment. Here, we give the result for independent sequences; for the above examples and more details confer Stout [80], Theorem 3.7.8.

Theorem B.7. *Let $\{Y(i) : i \geqslant 1\}$ be a sequence of independent random variables with $\mathrm{E}\,Y(i) = 0$, $i \geqslant 1$, satisfying $\mathrm{E}\,|Y(i)|^{2+\delta} \leqslant C$, $i \geqslant 1$ for some $\delta > 0$. Then, there is a constant D such that*

$$\mathrm{E}\left|\sum_{i=1}^{n} Y(i)\right|^{2+\delta} \leqslant D\; n^{(2+\delta)/2},$$

where D only depends on C and δ.

Remark B.1. Again one can derive a corresponding result for linear processes using the BN decomposition. Let $\{e(\cdot)\}$ be a linear process fulfilling the assumptions of Lemma C.1, particularly the innovations have a νth moment ($\nu > 2$). Lemma C.1 then gives for $0 < \delta < \nu - 2$

$$\mathrm{E}\left|\sum_{i=1}^{n} e(i)\right|^{2+\delta} \leqslant D_1\, n^{(2+\delta)/2} + D_2\,\mathrm{E}\,|\widetilde{e}(0)|^{2+\delta} \leqslant \widetilde{D} n^{(2+\delta)/2}.$$

Serfling [74] obtained moment inequalities for the maximum of partial sums as well as a convergence rate in the strong law of large numbers in cases where moment inequalities of the partial sums as above are fulfilled.

Theorem B.8. *Under the conditions of Theorem B.5, B.6, B.7, or Remark B.1 it holds:*

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

where D only depends on δ and the joint distribution of the $Y(i)$.

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

Proof. Confer Lemma B respectively Theorem 3.1 in Serfling [74]. ■

Remark B.2. Assertion b) of Theorem B.8 remains true for a triangular array that fulfills uniformly the assertion in a).

Remark B.3. If we have a stationary time series with $\mathrm{E}\,|Y(i)|^{2+\delta} \leqslant D$, $i \geqslant 0$, for some $\delta, D > 0$, and if $\alpha_Y(k) = O(k^{-2})$, then the invariance principle by Kuelbs and Philipp in [56] and the law of iterated logarithm give

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

The following strong law of large numbers is a consequence of Theorem B.8 b) and Remark B.2. It can also be easily derived using a result as in Theorems B.5, B.6, or B.7 and the Markov inequality.

Corollary B.1. *Let $\{Y_n(i) : 1 \leqslant i \leqslant n\}_n$ be a triangular array, which fulfills uniformly the conditions of Theorem B.5, B.6, B.7, or Remark B.1 then as $n \to \infty$ we have:*

$$\frac{1}{n}\sum_{i=1}^{n} Y_n(i) \to 0 \quad a.s.$$

Proof. The Markov inequality gives

$$P\left(\frac{1}{n}\left|\sum_{i=1}^{n} Y_n(i)\right| \geqslant \epsilon\right) \leqslant \frac{1}{n^{2+\delta}\epsilon^{2+\delta}}\,\mathrm{E}\left|\sum_{i=1}^{n} Y_n(i)\right|^{2+\delta} \ll n^{-(1+\delta/2)}.$$

The assertion follows, since $\sum_{n\geqslant 1} \frac{1}{n^{1+\delta/2}} < \infty$. ∎

Finally we give a weak law of large numbers for triangular arrays of row-wise independent random variables, which we need to prove the validity of the bootstrap with replacement in Chapter 2.

Lemma B.2. *Let $\{Y_n(i) : 1 \leqslant i \leqslant n\}_n$ be a triangular array of row-wise independent random variables. Let $b_n > 0$ with $b_n \to \infty$, and let $\widetilde{Y}_n(i) = Y_n(i) 1_{\{|Y_n(i)| \leqslant b_n\}}$. Suppose that as $n \to \infty$*

$$(i) \qquad P\Big(\max_{i=1,\ldots,n} |Y_n(i)| > b_n\Big) \to 0, \quad \textit{and}$$

$$(ii) \qquad b_n^{-2}\sum_{i=1}^{n} \mathrm{E}\,\widetilde{Y}_n(i)^2 \to 0.$$

Then

$$\frac{1}{b_n}\sum_{i=1}^{n}(Y_n(i) - \mathrm{E}\,\widetilde{Y}_n(i)) \xrightarrow{P} 0.$$

Proof. Confer Durrett [23], Chapter 1 (5.5). The proof remains true if we replace condition (i) there with the somewhat weaker condition above. ∎

C. Beveridge-Nelson Decomposition of a Linear Process

In this appendix we introduce a method of deriving asymptotics for linear processes, which we use frequently in the present work. The key to the approach is an algebraic decomposition of the linear filter into long-run and transitory elements. In the econometric literature it is known as the Beveridge-Nelson or BN decomposition (confer Beveridge and Nelson [8]). The method offers a simple unified approach to strong laws and central limit theory for linear processes.

Phillips and Solo [71] put the method on a mathematical foundation and prove strong laws of large numbers, central limit theorems and even a law of iterated logarithm by it. We will now give a short description of it using the example of the strong law of large numbers and state some more of the results by Phillips and Solo that are important for this thesis, for details confer their paper [71].

Lemma C.1 (BN decomposition). *Let $\{\epsilon(\cdot)\}$ be a sequence of i.i.d. random variables with zero mean and $\mathrm{E}\,|\epsilon(0)|^\nu < \infty$ for some $\nu > 2$. Let $\{w_s : s \geqslant 0\}$ be weights satisfying*

$$\sum_{s\geqslant 0} \sqrt{s}|w_s| < \infty.$$

Then the following decomposition holds:

$$e(i) := \sum_{s\geqslant 0} w_s\,\epsilon(i-s) = \epsilon(i)\sum_{s\geqslant 0} w_s + \widetilde{e}(i-1) - \widetilde{e}(i),$$

where $\{\widetilde{e}(\cdot)\}$ is a stationary process satisfying $\mathrm{E}\,|\widetilde{e}(0)|^p < \infty$ for any $p < \nu$.

Remark C.1. If we only assume that $\{\epsilon(\cdot)\}$ has a finite second moment, then we still get the above decomposition. But we also only have a finite second moment of $\{\widetilde{e}(\cdot)\}$. It is also possible to use somewhat different assumptions on the weights. For details confer Phillips and Solo [71].

Proof. Let

$$\widetilde{w}_s := \sum_{j>s} w_j.$$

Then $\sum_{s\geqslant 0} \widetilde{w}_s^2 < \infty$, since

$$\sum_{s\geqslant 0}\left(\sum_{j>s} w_j\right)^2 = \sum_{s\geqslant 0}\sum_{k>s}\sum_{l>s} w_k w_l = \sum_{k\geqslant 1}\sum_{l\geqslant 1} w_k w_l \min(k,l)$$

$$\leqslant \left(\sum_{k\geqslant 1}\sqrt{k}|w_k|\right)^2 < \infty.$$

Thus an application of Kolmogorov´s three series theorem gives that $\sum_{s\geqslant 0}\widetilde{w}_s\epsilon(i-s)$ converges almost surely. Hence

$$\widetilde{e}(i) := \sum_{s\geqslant 0}\widetilde{w}_s\epsilon(i-s)$$

is a stationary sequence. Now a simple calculation yields

$$e(i) = \epsilon(i)\sum_{s\geqslant 0} w_s + \widetilde{e}(i-1) - \widetilde{e}(i).$$

For a proof of $\mathrm{E}\,|\widetilde{e}(0)|^p < \infty$ for all $p < \nu$ confer the proof of Theorem 3.3 in Phillips and Solo [71]. ■

We illustrate how the above lemma can be used to obtain results such as strong laws of large numbers on the following Theorem:

Theorem C.1 (SLLN). *Under the assumptions of Lemma C.1 it holds as $n\to\infty$*

$$\frac{1}{n}\sum_{i=1}^{n} e(i) \to 0 \quad a.s.$$

Proof. As a consequence of the BN decomposition, Lemma C.1, it holds

$$\frac{1}{n}\sum_{i=1}^{n} e(i) = \sum_{s\geqslant 0} w_s\,\frac{1}{n}\sum_{i=1}^{n}\epsilon(i) + \frac{1}{n}(\widetilde{e}(0) - \widetilde{e}(n)).$$

This shows that the strong law of large numbers for $\{e(\cdot)\}$ follows directly from the strong law of large numbers for $\{\epsilon(\cdot)\}$ as long as

$$n^{-1}\widetilde{e}(0) \to 0, \quad n^{-1}\widetilde{e}(n) \to 0 \qquad a.s.$$

The Markov inequality shows that this is fulfilled because

$$\sum_{n\geqslant 1} P\left(n^{-1}|\widetilde{e}(0)| \geqslant \delta\right) \leqslant \frac{1}{\delta^2}\,\mathrm{E}\,\widetilde{e}(0)^2\sum_{n\geqslant 1} n^{-2} < \infty.$$

An analogous argument yields the almost sure convergence of $n^{-1}\widetilde{e}(n)$. ■

We give now two more results of Phillips and Solo [71] that we need in this work. The proofs are also based on the BN decomposition, but we will not give the details here.

Theorem C.2 (LIL). *Under the assumptions of Lemma C.1 it holds as $n\to\infty$*

$$\limsup_{n\to\infty}\sqrt{\frac{1}{2\tau^2 n\log\log n}}\sum_{i=1}^{n} e(i) = 1 \qquad a.s.$$

$$\textit{and}\quad \liminf_{n\to\infty}\sqrt{\frac{1}{2\tau^2 n\log\log n}}\sum_{i=1}^{n} e(i) = 1 \qquad a.s.,$$

where $\tau^2 = \left(\sum_{s\geqslant 0} w_s\right)^2 \operatorname{var}\epsilon(0)$.

Proof. Confer Phillips and Solo [71], Theorem 3.3. ■

Theorem C.3 (SLLN for covariances). *Under the assumptions of Lemma C.1 it holds for $h \geqslant 0$ as $n \to \infty$*

$$\frac{1}{n}\sum_{i=1}^{n} e(i)e(i+h) \to \operatorname{var}\epsilon(0)\sum_{s\geqslant 0} w_s w_{s+h} \qquad a.s.$$

Particularly we have for the variance as $n \to \infty$

$$\frac{1}{n}\sum_{i=1}^{n} e(i)^2 \to \operatorname{var}\epsilon(0)\sum_{s\geqslant 0} w_s^2 \qquad a.s.$$

Proof. Confer Theorem 3.7 and Remark 3.9 of Phillips and Solo [71]. ■

D. A Weak Approximation to Permutation and Exchangeable Processes

In this chapter we state an approximation by a Brownian bridge to permutation and exchangeable processes. The proof is based on the Skorokhod embedding for martingales and goes back to Einmahl and Mason [28], Theorem 1. It allows us to derive the asymptotics for rank statistics as well as score processes from that of the statistics for i.i.d. normal variables.

Theorem D.1. *Let $Z_n = (Z_n(1), \ldots, Z_n(n))$, $n = 1, 2, \ldots$, be a sequence of n-dimensional random vectors satisfying $\sum_{i=1}^n Z_n(i) = 0$,*

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[|Z_n(i)|^\nu\right] \leqslant D, \tag{D.1}$$

for some $D > 0$, $\nu > 2$ and

$$\frac{1}{n}\sum_{i=1}^{n}(Z_n(i))^2 = 1 + O_P\left(n^{-2s}\right), \tag{D.2}$$

where $s := \min\left(\frac{\nu-2}{2\nu}, \frac{1}{4}\right)$. Then, on a rich enough probability space, there exists a sequence of stochastic processes $\{\Pi_n(k) : 1 \leqslant k \leqslant n\}$ $(n = 1, 2, \ldots)$ with

$$\{\Pi_n(k) : 1 \leqslant k \leqslant n\} \stackrel{\mathcal{D}}{=} \left\{\sum_{i=1}^{k} Z_n(R_i) : 1 \leqslant k \leqslant n\right\},$$

where $(R_1, \ldots, R_n)$ is a random permutation of $(1, 2, \ldots, n)$, independent of Z_n, and there is a fixed Brownian bridge $\{B(t) : 0 \leqslant t \leqslant 1\}$ such that, for $0 \leqslant \mu < s$,

$$\max_{1\leqslant k<n}\left(\frac{k(n-k)}{n}\right)^{\mu}\frac{n}{\sqrt{k(n-k)}}\left|\frac{1}{\sqrt{n}}\Pi_n(k) - B(k/n)\right| = O_P(1).$$

The proof of the above theorem goes along the lines of Theorem 1 of Einmahl and Mason [28], which gives the result for $\nu = 4$. We only need to replace the Hájek-Rényi inequality (cf. [28], p. 110) respectively Lemma 3 there with the following lemmas. The details for Corollary D.1, i.e. in the case where Z_n is a vector of scores, can be found in Kirch [50], Theorem 5.1.1. It is also possible to derive the result from a generalization by Häusler and Mason [42], Theorem 2.

Lemma D.1. *Let $M(0) = 0$, $M(1), \ldots, M(m)$, $m \geqslant 1$, be a mean 0, square-integrable martingale, and $a(1) \geqslant \ldots \geqslant a(m) \geqslant 0$ be constants. Then, for $1 < \gamma \leqslant 2$ and $\lambda > 0$,*

$$P\left(\max_{1\leqslant i\leqslant m} a(i)|M(i)| > \lambda\right) \leqslant 2^{\gamma-1}\frac{1}{\lambda^\gamma}\sum_{i=1}^{m} a(i)^\gamma\,\mathrm{E}\left|M(i) - M(i-1)\right|^\gamma.$$

Proof. Confer Lemma 1 in Häusler and Mason [42], or Lemma 5.1.2 in Kirch [50] together with Einmahl [27]. ■

Lemma D.2. *Let $Z_n = (Z_n(1), \ldots, Z_n(n))$ be random vectors with $\sum_{i=1}^n Z_n(i) = 0$, and $(R_1, \ldots, R_n)$ be a random permutation independent of Z_n as in Theorem D.2. Then, for $1 \leqslant i \leqslant n$ and $1 \leqslant \gamma \leqslant 2$,*

$$\mathrm{E}\left|\sum_{j=1}^{i} Z_n(R_j)\right|^{\gamma} \leqslant 2\min(i, n-i)\,\frac{1}{n}\sum_{j=1}^{n}\mathrm{E}\left|Z_n(j)\right|^{\gamma}.$$

Proof. Confer Mason [63], the proof for rank statistics is given in Lemma 5.1.3 in Kirch [50]. It remains valid in our situation: The case $n = 1$ is obvious, so let $n \geqslant 2$. Moreover let $U_n(1), \ldots, U_n(n)$ be i.i.d. random variables on $\{1, \ldots, n\}$ with

$$P(U_n(l) = i) = \frac{1}{n}, \quad i, l = 1, \ldots, n.$$

Let $1 \leqslant i \leqslant n/2$. Hoeffding´s inequality (cf. Shorack and Wellner [75], p. 878) gives for all $1 \leqslant s \leqslant 2$

$$E\left(\left|\sum_{k=1}^{i} Z_n(R_n(k))\right|^{\gamma} \,\middle|\, Z\right) \leqslant E\left(\left|\sum_{k=1}^{i} Z_n(U_n(k))\right|^{\gamma} \,\middle|\, Z\right).$$

The von Bahr-Esseen inequality (cf. [75], p. 858) now gives

$$E\left(\left|\sum_{k=1}^{i} Z_n(U_n(k))\right|^{\gamma} \,\middle|\, Z\right) \leqslant 2\sum_{k=1}^{i} E\left(\left|Z_n(U_n(k))\right|^{\gamma} \,\middle|\, Z\right) = \frac{2i}{n}\sum_{j=1}^{n}\left|Z_n(j)\right|^{\gamma}.$$

Taking expectations on both sides now gives the assertion.

For $n/2 < i < n$ we also get the assertion, since

$$\sum_{k=1}^{i} Z_n(R_n(k)) = -\sum_{k=i+1}^{n} Z_n(R_n(k)).$$

The case $i = n$ is trivial, since the left side is equal to 0. ■

In this work we need two special cases of the above theorem, which we state in two corollaries. The first one is for rank statistics and an immediate consequence of the theorem. We need it to obtain the rank asymptotics which in turn help prove the permutation statistic results.

Corollary D.1. *Let $b_n(1), \ldots, b_n(n)$ be scores satisfying*

$$\sigma_n^2(\mathbf{b}) := \frac{1}{n}\sum_{i=1}^{n}\left(b_n(i) - \bar{b}_n\right)^2 \geqslant D_1, \tag{D.3}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\left|b_n(i) - \bar{b}_n\right|^{\nu} \leqslant D_2, \tag{D.4}$$

Then, on a rich enough probability space, there is a sequence of stochastic processes $\{\Pi_n(k) : 1 \leqslant k \leqslant n\}$ $(n = 1, 2, \ldots)$ *with*

$$\{\Pi_n(k) : 1 \leqslant k \leqslant n\} \stackrel{\mathcal{D}}{=} \left\{ \frac{1}{\sqrt{\sigma_n^2(\mathbf{b})}} \sum_{i=1}^{k} \left(b_n(R_i) - \bar{b}_n\right) : 1 \leqslant k \leqslant n \right\},$$

where $(R_1, \ldots, R_n)$ *is a random permutation of* $(1, 2, \ldots, n)$, $\bar{b}_n = \frac{1}{n}\sum_{i=1}^{n} b_n(i)$, *and there is a fixed Brownian bridge* $\{B(t) : 0 \leqslant t \leqslant 1\}$ *such that, for* $0 \leqslant \mu < \min\left(\frac{\nu-2}{2\nu}, \frac{1}{4}\right)$,

$$\max_{1 \leqslant k < n} \left(\frac{k(n-k)}{n}\right)^{\mu} \frac{n}{\sqrt{k(n-k)}} \left| \frac{1}{\sqrt{n}} \Pi_n(k) - B(k/n) \right| = O_P(1).$$

Proof. Follows immediately from Theorem D.1 by choosing $Z_n(i) := \frac{1}{\sigma_n(\mathbf{b})}(b_n(i) - \bar{b}_n)$.
■

The second corollary is also a consequence of the above theorem. We need it to prove the bootstrapping results with replacement.

Corollary D.2. *Let* $b_n(1), \ldots, b_n(n)$ *be scores satisfying (D.3) and (D.4). Then, on a rich enough probability space, there is a sequence of stochastic processes* $\{\mathcal{E}_n(k) : 1 \leqslant k \leqslant n\}$ $(n = 1, 2, \ldots)$ *with*

$$\{\mathcal{E}_n(k) : 1 \leqslant k \leqslant n\} \stackrel{\mathcal{D}}{=} \left\{ \frac{1}{\sqrt{\sigma_n^2(\mathbf{b})}} \sum_{i=1}^{k} \left(b_n(U_i) - \bar{b}_{U,n}\right) : 1 \leqslant k \leqslant n \right\},$$

where $\{U_i : i = 1, \ldots, n\}$ *is a triangular array of rowwise i.i.d. random variables uniformly distributed on* $\{1, 2, \ldots, n\}$, $\bar{b}_n = \frac{1}{n}\sum_{i=1}^{n} b_n(i)$, $\bar{b}_{U,n} = \frac{1}{n}\sum_{i=1}^{n} b_n(U_i)$, *and there is a fixed Brownian bridge* $\{B(t) : 0 \leqslant t \leqslant 1\}$ *such that, for* $0 \leqslant \mu < \min\left(\frac{\nu-2}{2\nu}, \frac{1}{4}\right)$,

$$\max_{1 \leqslant k < n} \left(\frac{k(n-k)}{n}\right)^{\mu} \frac{n}{\sqrt{k(n-k)}} \left| \frac{1}{\sqrt{n}} \mathcal{E}_n(k) - B(k/n) \right| = O_P(1).$$

Proof. We show that the assumptions of Theorem D.1 are fulfilled for $Z_n(i) = \frac{1}{\sigma_n(\mathbf{b})}(b_n(U_i) - \bar{b}_{U,n})$. This yields the assertion because

$$(Z_n(1), \ldots, Z_n(n)) \stackrel{\mathcal{D}}{=} (Z_n(R_1), \ldots, Z_n(R_n)),$$

where $(R_1, \ldots, R_n)$ is a random permutation of $(1, \ldots, n)$ independent of $\{U_i : i = 1, \ldots, n\}$.

Let $s := \min\left(\frac{\nu-2}{2\nu}, \frac{1}{4}\right)$ and note that

$$\bar{b}_n - \bar{b}_{U,n} = \frac{1}{n}\sum_{i=1}^{n}(b_n(U_i) - \bar{b}_n) = O_P(n^{-s}), \tag{D.5}$$

because the Markov inequality gives

$$P\left(|\bar{b}_n - \bar{b}_{U,n}| \geqslant C/n^s\right) \leqslant \frac{n^{2s}}{C^2} \operatorname{var}\left(\frac{1}{n}\sum_{i=1}^{n}(b_n(U_i) - \bar{b}_n)\right) \ll \frac{n^{2s-1}}{C^2} \to 0.$$

Furthermore the Minkowski inequality gives

$$\mathrm{E}\,|\bar{b}_n - \bar{b}_{U,n}|^\nu \leqslant \mathrm{E}\,|b_n(U_1) - \bar{b}_n|^\nu \leqslant D_2.$$

Thus we get

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\,|Z_n(i)|^\nu \ll \frac{1}{|\sigma_n(\mathbf{b})|^\nu} \left(\mathrm{E}\,|b_n(U_1) - \bar{b}_n|^\nu + \mathrm{E}\,|\bar{b}_n - \bar{b}_{U,n}|^\nu\right) \ll 1.$$

Moreover the Markov inequality additionally to the von Bahr-Esseen inequality (cf. e.g. Shorack and Wellner [75], p. 858) gives ($r := \min(\nu, 4)$)

$$\begin{aligned}
&P\left(\left|\sum_{i=1}^{n}\Big[(b_n(U_i) - \bar{b}_n)^2 - \frac{1}{n}\sum_{j=1}^{n}(b_n(j) - \bar{b}_n)^2\Big]\right| \geqslant C/n^{2s-1}\right) \\
&\leqslant \frac{n^{sr-r/2}}{C^{r/2}}\,\mathrm{E}\left|\sum_{i=1}^{n}\Big[(b_n(U_i) - \bar{b}_n)^2 - \frac{1}{n}\sum_{j=1}^{n}(b_n(j) - \bar{b}_n)^2\Big]\right|^{r/2} \\
&\leqslant \frac{1}{C^{r/2}}\,2^{r/2-1}\,\mathrm{E}\left|(b_n(U_i) - \bar{b}_n)^2 - \frac{1}{n}\sum_{j=1}^{n}(b_n(j) - \bar{b}_n)^2\right|^{r/2} \\
&\ll \frac{1}{C^{r/2}}\left(\frac{1}{n}\sum_{i=1}^{n}|b_n(i) - \bar{b}_n|^r + (1 + D_2)^{r/2}\right) \ll \frac{1}{C^{r/2}},
\end{aligned}$$

note that $sr - r/2 + 1 = 0$. This now gives together with (D.5)

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sigma_n(\mathbf{b})}(b_n(U_i) - \bar{b}_{U,n})\right)^2 = 1 + O_P(n^{-2s}), \tag{D.6}$$

because

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sigma_n(\mathbf{b})}(b_n(U_i) - \bar{b}_{U,n})\right)^2 - 1 \\
&= \frac{1}{\sigma_n^2(\mathbf{b})}\frac{1}{n}\sum_{i=1}^{n}\Big[(b_n(U_i) - \bar{b}_n)^2 - \frac{1}{n}\sum_{j=1}^{n}(b_n(j) - \bar{b}_n)^2\Big] - \frac{1}{\sigma_n^2(\mathbf{b})}(\bar{b}_{U,n} - \bar{b}_n)^2 \\
&= O_P(n^{-2s}).
\end{aligned}$$

■

E. Some Results on Rank Statistics

In this chapter we give some simple linear rank statistic results that we need in the proofs of Chapter 4. For a detailed discussion of the subject confer the book by Hájek et al. [40].

The first lemma gives closed formulas for some moments of linear rank statistics.

Lemma E.1. *Consider the following two simple linear rank statistics:*

$$S_1 = \sum_{i=1}^{n} c(i)a(R_i), \qquad S_2 = \sum_{i=1}^{n} d(i)b(R_i),$$

where $c(\cdot)$, $a(\cdot)$, $d(\cdot)$ and $b(\cdot)$ are some vectors and $(R_1, \ldots, R_n)$ is a random permutation of $(1, \ldots, n)$. Then it holds

$$a) \qquad \operatorname{var} S_1 = \frac{1}{n-1} \sum_{i=1}^{n} (a(i) - \bar{a})^2 \sum_{j=1}^{n} (c(j) - \bar{c})^2,$$

$$b) \qquad \operatorname{cov}(S_1, S_2) = \frac{1}{n-1} \sum_{i=1}^{n} (a(i) - \bar{a})(b(i) - \bar{b}) \sum_{j=1}^{n} (c(j) - \bar{c})(d(j) - \bar{d}),$$

$$c) \qquad \mathrm{E}(S_1 - \mathrm{E}\, S_1)^4 = 3\frac{(n-1)^3}{n+1}(\sigma_a^2)^2(\sigma_c^2)^2 + \frac{(n-1)(n-2)(n-3)}{n(n+1)} k_{4a}k_{4c},$$

where $\bar{a} = \frac{1}{n}\sum_{i=1}^{n} a(i)$, $\sigma_a^2 = \frac{1}{n-1}\sum_{i=1}^{n}(a(i) - \bar{a})^2$ and

$$k_{4a} = \frac{1}{(n-1)(n-2)(n-3)} \left[n(n+1) \sum_{i=1}^{n} (a(i) - \bar{a})^4 - 3(n-1)^3(\sigma_a^2)^2 \right]$$

and analogous expressions for b, c, d.

Proof. Confer Hájek et al. [40], Theorems 3.3.3 and 3.3.4 and problem 25 in Section 3.3. ■

The next theorem gives necessary and sufficient conditions for asymptotic normality of simple linear rank statistics.

Theorem E.1. *The simple linear rank statistic $S_n = \sum_{i=1}^{n} c(i)a(R_i)$ is asymptotically normal, if the Lindeberg condition*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{|\delta_{nij}|>\tau} \delta_{nij}^2 = 0 \qquad \text{for any } \tau > 0$$

with

$$\delta_{nij} = (c(i) - \bar{c})(a(j) - \bar{a}) \left[\frac{1}{n} \sum_{i=1}^{n} (c(i) - \bar{c})^2 \sum_{j=1}^{n} (a(j) - \bar{a})^2 \right]^{-1/2}$$

is fulfilled.
If additionally the Noether conditions

$$\frac{\sum_{i=1}^{n} (a(i) - \bar{a})^2}{\max_{i=1,\dots,n} (a(i) - \bar{a})^2} \to \infty, \qquad \frac{\sum_{i=1}^{n} (c(i) - \bar{c})^2}{\max_{i=1,\dots,n} (c(i) - \bar{c})^2} \to \infty$$

hold, then the Lindeberg condition is also necessary.
Furthermore the Lindeberg condition implies the Noether condition.

Proof. Confer Hájek et al. [40] Section 6.1., problems 2 and 3, p. 241 and Hájek [39]. Hájek [39], Theorem 5.1, shows that the Lindeberg condition implies the Noether condition. ■

F. Some Results from Change-Point Analysis

In this appendix we state some well-known facts from change-point analysis that we use throughout this work.

The first lemma gives conditions under which the limit distribution remains the same even if one only considers the maximum over fewer elements.

Lemma F.1. *Let α_n, β_n be sequences of positive numbers, $\beta_n \to \infty$, and A_n, B_n random variables satisfying $A_n = o(\beta_n/\alpha_n)$, then it holds for all $y \in \mathbb{R}$*

$$|P(\alpha_n \max(A_n, B_n) - \beta_n \leqslant y) - P(\alpha_n B_n - \beta_n \leqslant y)| \to 0$$

as $n \to \infty$.

Proof. First it holds

$$P(\alpha_n \max(A_n, B_n) \leqslant y + \beta_n) \leqslant P(\alpha_n B_n \leqslant y + \beta_n),$$

moreover for all y there exists $N(y)$ with $y/\beta_n + 1 \geqslant 1/2$ for all $n \geqslant N(y)$. Thus

$$\begin{aligned}
P(\alpha_n \max(A_n, B_n) \leqslant y + \beta_n) &\geqslant P(\alpha_n B_n \leqslant y + \beta_n) + P\left(\frac{\alpha_n}{\beta_n} A_n \leqslant \frac{1}{2}\right) - 1 \\
&\geqslant P(\alpha_n B_n \leqslant y + \beta_n) - \varepsilon
\end{aligned}$$

for n large enough. The proof can already be found in Lemma 4.1.1 in Kirch [50]. ■

Now we state some results involving the weight functions $q(\cdot)$ we use for the q-weighted CUSUM statistics and $r(\cdot)$ we use for the sum statistics (confer Section 3.2):

Lemma F.2. *Let $q \in FC_0^1$ and*

$$\int_0^1 \frac{t(1-t)}{q^4(t)}\, dt < \infty \quad or$$

$$\int_0^1 \frac{1}{q^2(t)}\, dt.$$

Then $I^*(q, c) < \infty$ for all $c > 0$.

Proof. The proof for squared q can be found in Csörgő and Horváth [18], Corollary 1.3. The proof for the first integral is analogous to the proof given there. Note that for all $c > 0$ and $x \geqslant 0$

$$x^2 \exp(-cx) \leqslant \frac{4}{c^2 e^2},$$

which gives

$$\begin{aligned}
\int_0^1 \frac{1}{t(1-t)} \exp\left(-c\frac{q^2(t)}{t(1-t)}\right) dt &= \int_0^1 \frac{t(1-t)}{q^4(t)} \frac{q^4(t)}{t^2(1-t)^2} \exp\left(-c\frac{q^2(t)}{t(1-t)}\right) dt \\
&\leqslant \frac{4}{c^2e^2} \int_0^1 \frac{t(1-t)}{q^4(t)}\, dt < \infty
\end{aligned}$$

■

Lemma F.3. *a) If $q \in Q_{0,1}$ and $I^*(q,c) < \infty$ for some $c > 0$, it holds*

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

b) If $q \in Q_{0,1}$ and $I^*(q,c) < \infty$ for all $c > 0$, it holds

$$\lim_{t\to 0} \frac{|B(t)|}{q(t)} = 0 = \lim_{t\to 1} \frac{|B(t)|}{q(t)} \qquad a.s.$$

c) If $r(\cdot)$ is positive on $(0,1)$ and $0 \leqslant p < \infty$, then the following statements are equivalent:

$$\int_0^1 \frac{(t(1-t))^{p/2}}{r(t)}\, dt < \infty$$

$$\int_0^1 \frac{|B(t)|^p}{r(t)}\, dt < \infty \qquad a.s.,$$

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

Proof. For a) confer Csörgő and Horváth [18], Chapter 4, Corollary 1.1, for b) Corollary 1.2 and for c) Chapter 5, Lemma 3.2. ■

Acknowledgements

I would like to express my thanks to my supervisor Prof. Dr. Josef Steinebach for always listening to my problems. Moreover I am deeply grateful to Prof. Dr. Marie Hušková from Charles University, Prague, whose deep insights in rank statistics as well as the field of change-point analysis helped solve many problems. Furthermore it was her who suggested the possibility of bootstrapping in the frequency domain. Last but not least my thanks go to Prof. Dr. David Mason from the University of Delaware for pointing out to me that the frequency bootstrap is essentially just another linear rank statistic problem.

Bibliography

[1] Anderson, L.G., and Figueiredo, R.J.P. An adaptive orthogonal-series estimator for probability density functions. *Ann. Statist.*, 8:347–376, 1980.

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

[3] Antoch, J., Hušková, M., and Jarušková, D. Change point detection. pages 1–75, Voorburg, 2000. ISI. Lecture Notes of the 5th IASC Summer School.

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

[5] Aue, A., Berkes, I., and Horváth, L. Strong approximation for the sums of squares of augmented garch sequences. *Bernoulli*, 2006. To appear.

[6] J. Bai. Least squares estimation of a shift in linear processes. *J. Time Ser. Anal.*, 15:435–472, 1994.

[7] 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.

[8] Beveridge, S., and Nelson, C.R. A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the ´business cycle´. *Journal of Monetary Economics*, 7:151–174, 1981.

[9] Billingsley, P. *Convergence of Probability Measures.* Wiley, New York, 1968.

[10] Bose, A. Edgeworth correction by bootstrap in autoregression. *Ann. Statist.*, 16:1709–1722, 1988.

[11] Braun, W.J., and Kulperger, R.J. Properties of a fourier bootstrap method for time series. *Comm. Statist. Theory Methods*, 26:1329–1336, 1997.

[12] Brillinger, D.R. *Time Series, Data Analysis and Theory.* SIAM, San Francisco, 1981.

[13] Brockwell, P.J., and Davis, R.A. *Time Series: Theory and Methods.* Springer, New York, second edition, 1991.

[14] Bronstein, I.N., Semendjajew, K.A., Musiol, G., and Mühlig, H. *Taschenbuch der Mathematik.* Harri Deutsch, Frankfurt am Main, 4th edition, 1999.

[15] Carlstein, E. The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist.*, 14:1171–1179, 1986.

[16] Chen, J., and Gupta, A.K. *Parametric Statistical Change Point Analysis.* Birkhäuser, Boston, 2000.

[17] Chernoff, H., and Zacks, S. Estimating the current mean of a normal distribution which is subjected to changes in time. *Ann. Math. Stat.*, 35:999–1018, 1964.

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

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

[20] Csörgő, M., Horváth, L., and Steinebach, J. Invariance principles for renewal processes. *Ann. Probab.*, 15:1441–1460, 1987.

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

[22] Dahlhaus, R., and Janas, D. A frequence domain bootstrap for ratio statistics in time series analysis. *Ann. Statist.*, 24:1934–1963, 1996.

[23] Durrett, R. *Probability: Theory and Examples.* Brooks/Cole, Belmont, third edition, 2005.

[24] Eberlein, E. Strong approximation of very weak bernoulli processes. *Z. Wahrsch. verw. Geb.*, 62:17–37, 1983.

[25] Efron, B. Bootstrap methods: Another look at the jackknife. *Ann. Statist.*, 7:1–26, 1979.

[26] Einmahl, U. Extensions of results of komlós, major, and tusnády to the multivariate case. *J. Multivariate Anal.*, 28:20–68, 1989.

[27] Einmahl, U. Personal communication, 2003.

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

[29] Fazekas, I., and Klesov, O. A general approach to the strong law of large numbers. *Theory Probab. Appl.*, 45:436–448, 2000.

[30] Franke, J., and Härdle, W. On bootstrapping kernel spectral estimates. *Ann. Statist.*, 20:121–145, 1992.

[31] Freedman, D.A. Bootstrapping regression models. *Ann. Statist.*, 9:1218–1228, 1981.

[32] Freedman, D.A. On bootstrapping two-stage least-squares estimates in stationary linear models. *Ann. Statist.*, 12:827–842, 1984.

[33] Giraitis, L., and Leipus, R. Functional clt for nonparametric estimates of the spectrum and change-point problem for a spectral function. *Liet. Mat. Rink.*, 30:674–697, 1990.

[34] Giraitis, L., and Leipus, R. Testing and estimating in the change-point problem of the spectral function. *Liet. Mat. Rink.*, 32:15–29, 1992.

[35] Good, P. *Permutation Tests*. Springer, New York, second edition, 2000.

[36] Good, P. *Permutation, Parametric, and Bootstrap Tests of Hypothesis*. Springer, New York, third edition, 2005.

[37] Gorodetskii, V.V. On the strong mixing property for linear sequences. *Theory Probab. Appl.*, 22:411–413, 1977.

[38] Gradshteyn, I.S., and Ryzhik, I.M. *Table of Integrals, Series, and Products*. Academic Press, New York, corrected and enlarged edition, 1980.

[39] Hájek, J. Some extensions of the wald-wolfowitz-noether theorem. *Ann. Math. Statist.*, 23:506–523, 1961.

[40] Hájek, J., Šidák, Z., and Sen, P.K. *Theory of Rank Tests*. Academic Press, San Diego, second edition, 1999.

[41] Hall, P. On trigonometric series estimates of densities. *Ann. Statist.*, 9:683–685, 1981.

[42] Häusler, E., and Mason, D.M. Weighted approximation to continuous time martingales with applications. *Scand. J. Statist.*, 26:281–295, 1999.

[43] Hawkins, D.M. Testing a sequence of observations for a shift in location. *J. Amer. Statist. Assoc.*, 72:180–186, 1977.

[44] Hidalgo, J. An alternative bootstrap to moving blocks for time series regression models. *J. Econometrics*, 117:369–399, 2003.

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

[46] Horváth, L., and Steinebach, J. Testing for changes in the mean or variance of a stochastic process under weak invariance. *J. Statist. Plann. Inference*, 91:365–376, 2000.

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

[48] Hušková, M., and Steinebach, J. Limit theorems for a class of tests of gradual changes. *J. Statist. Plann. Inference*, 89:57–77, 2000.

[49] Kiefer, J. K-sample analogues of the kolmogorov-smirnov and cramér-v. mises tests. *Ann. Math. Statist.*, 30:420–447, 1960.

[50] Kirch, C. Permutationsprinzipien in der Changepoint-Analyse. Diploma thesis, Philipps-Universität Marburg, 2003.

[51] 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.

[52] 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.

[53] Kokoszka, P., and Leipus, R. Change-point in the mean of dependent observations. *Statist. Probab. Lett.*, 40:385–393, 1998.

[54] Komlós, J., Major, P., and Tusnády, G. An approximation of partial sums of independent rv´s and the sample df. i. *Z. Wahrsch. verw. Geb.*, 32:111–131, 1975.

[55] Komlós, J., Major, P., and Tusnády, G. An approximation of partial sums of independent rv´s and the sample df. ii. *Z. Wahrsch. verw. Geb.*, 34:33–58, 1976.

[56] Kuelbs, J., and Philipp, W. Almost sure invariance principles for partial sums of mixing *b*-valued random variables. *Ann. Probab.*, 8:1003–1036, 1980.

[57] Künsch, H.R. The jackknife and the bootstrap for general stationary observations. *Ann. Statist.*, 17:1217–1241, 1989.

[58] Lavielle, M., and Moulines, E. Least-squares estimation of an unknown number of shifts in a time series. *J. Time Ser. Anal.*, 21:33–59, 2000.

[59] Leadbetter, M.R., Lindgren, L., and Rootzén, H. *Extremes and Related Properties of Random Sequences and Processes.* Springer, New York, 1983.

[60] Liu, R.Y., and Singh, K. Moving blocks jackknife and bootstrap capture weak dependence. In LePage, R. and Billard, L., editors, *Exploring the Limits of Bootstrap*, pages 225–248, New York, 1992. Wiley.

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

[62] Major, P. An approximation of partial sums of independent rv´s. *Z. Wahrsch. verw. Geb.*, 35:213–220, 1976.

[63] Mason, D.M. Personal communication, 2003.

[64] Móricz, F. Moment inequalities and the strong laws of large numbers. *Z. Wahrsch. verw. Geb.*, 35:299–314, 1976.

[65] Móricz, F., Serfling, R., and Stout, W. Moment and probability bounds with quasi-superadditive structure for the maximum partial sum. *Ann. Probab.*, 10:1032–1040, 1982.

[66] Neuhaus, G., and Kreiss, J.P. *Einführung in die Zeitreihenanalyse.* Gesellschaft zur Förderung der mathematischen Statistik, Münster, 1985.

[67] Page, E.S. Continuous inspection schemes. *Biometrika*, 41:100–115, 1954.

[68] Page, E.S. Control charts with warning lines. *Biometrika*, 42:243–257, 1955.

[69] Page, E.S. A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42:523–527, 1955.

[70] Philipp, W. Invariance principles for independent and weakly dependent random variables. In Eberlein, E. and Taqqu, M.S., editors, *Dependence in Probability and Statistics*, pages 225–268, Boston, 1986. Birkhäuser.

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

[72] 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.

[73] Politis, D.N., Romano, J.P., and Wolf, M. *Subsampling*. Springer, New York, 1999.

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

[75] Shorack, G.R., and Wellner, J.A. *Empirical Processes with Applications to Statistics*. Wiley, New York, 1986.

[76] Singh, K. On the asymptotic accuracy of efron´s bootstrap. *Ann. Statist.*, 9:1187–1195, 1981.

[77] Slabý, A. Limit theorems for rank statistics detecting gradual changes. *Comment. Math. Univ. Carolin.*, 42:591–600, 2001.

[78] Steinebach, J. Change-point and jump estimates in an amoc renewal model. In Mandl,P. and Hušková, M., editors, *Asymptotic Statistics*, pages 447–457, Heidelberg, 1994. Physica-Verlag.

[79] Steinebach, J. Some remarks on the testing of smooth changes in the linear drift of a stochastic process. *Theory Probab. Math. Statist.*, 61:164–175, 1999.

[80] Stout, W.F. *Almost Sure Convergence*. Academic Press, New York, 1974.

[81] van der Vaart, A.W. *Asymptotic Statistics*. Cambridge University Press, Cambridge, 1998.

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

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

Notation

Real and Complex Numbers

$\lfloor x \rfloor$	integer part of x, largest integer $n \leqslant x$
$\lceil x \rceil$	smallest integer $n \geqslant x$
x_+	positive part, $\max(x, 0)$
$x \vee y$	maximum of x and y
$x \wedge y$	minimum of x and y
o, O	Landau symbols, confer Appendix A
$a_n \ll b_n$	$a_n = O(b_n)$, confer Appendix A
$\langle\ ,\ \rangle$	standard scalar product on $\mathbb{R}^n$
1_A	indicator function, $= 1$ on A, $= 0$ else
$\overline{a+ib}$	conjugated complex, $= a - ib$
$\mathrm{Re}(z)$	real part of z
$\mathrm{Im}(z)$	imaginary part of z

Probability

$a.s.$	almost surely
i.i.d.	independent and identically distributed
r.v.	random variable
$\mathrm{E}\,X$	expectation of X
$\mathrm{var}\,X$	variance of X
$\mathrm{cov}(X, Y)$	covariance of X and Y
$\mathrm{E}(X \vert Y)$	conditional expectation of X given Y (the σ-algebra generated by Y)
$P(A \vert Y)$	conditional probability, $= E(1_A \vert Y)$
o_P, O_P	Landau symbols, confer Appendix A
$a_n \ll b_n$ a.s.	$a_n = O(b_n)$ a.s., confer Appendix A
$\xrightarrow{P}$	converges in probability
$\xrightarrow{\mathcal{D}}$	converges weakly, in distribution
$X \stackrel{\mathcal{D}}{=} Y$	X and Y have the same distribution
$X_n \stackrel{\mathcal{D}}{=} Y_n + o_P(\gamma_n)$	there exist $Z_n^{(j)} = o_P(\gamma_n)$, $j = 1, 2$, such that $X_n + Z_n^{(1)} \stackrel{\mathcal{D}}{=} Y_n + Z_n^{(2)}$; this means in particular that X_n/γ_n has the same distributional asymptotics as Y_n/γ_n

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Josef Steinebach betreut worden.

Köln, im März 2006

(Claudia Kirch)

Teilpublikationen:

[1] Permutation principles for the change analysis of stochastic processes under strong invariance. *J. Comput. Appl. Math*, 186:64-88, 2006 (mit J. Steinebach).

[2] Block permutation principles for the change analysis of dependent data. *J. Statist. Plann. Inference*, 2006. Eingereicht.

[3] Resampling in the frequency domain of time series to determine critical values for change-point tests. 2006. In Vorbereitung.

Lebenslauf

Persönliche Daten

Name:	Claudia Kirch
Nationalität:	deutsch
Geburtsdatum:	30.12.1977
Geburtsort:	Saarbrücken
Familienstand:	ledig

Schulbildung

1984 - 1988	Grundschule Viktoria, Püttlingen
1988 - 1997	Marie-Luise-Kaschnitz-Gymnasium, Völklingen
07.1997	Abitur

Studium

10.1998 - 08.2003	Mathematik (Diplom) mit Nebenfach Informatik an der Philipps-Universität, Marburg
10.2000	Vordiplom
08.2003	Diplom

Berufstätigkeit

07.1997 - 08.1998	au pair Aufenthalt in Washington D.C., USA
seit 10.2003	Wissenschaftliche Mitarbeiterin am Mathematischen Institut der Universität zu Köln