U-Statistics for Detecting and Estimating Changes in Weakly Dependent Functional Data

Dissertation
zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

von Lea Wegner
geb. am 06.02.1995 in Köln

genehmigt durch die Fakultät für Mathematik
der Otto-von-Guericke-Universität Magdeburg

Gutachter: Dr. Martin Wendler
Prof. Dr. Claudia Kirch
Assoc. Prof. Michal Pešta

eingereicht am 03.01.2023

Verteidigung am 11.04.2023

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project WE 5988/3 *Analyse funktionaler Daten ohne Dimensionsreduktion*

Abstract

Change-point detection is an important field in time series analysis. As early as the 1950s, it starts with the question whether there is a change in a given time series. Depending on the setting this may for example be a change in the underlying distribution, in mean, variance or correlation. A further step may be to identify the time at which the change occurs. More recently, change-point analysis for high-dimensional and dependent data became important, as more of such data is available and computational power allows for dealing with it.

This thesis consists of two main chapters. In the beginning of the first chapter, the concepts of weakly dependent and functional data are explained. Then, a U-statistic (where U stands for unbiased) for testing the hypothesis of no change is introduced for such data. Subsequently, asymptotic results for this statistic in the presence and absence of a change-point are given. Under the assumption of no change, the limit distribution of the test statistic is established. Since it contains an infinite covariance operator, a bootstrap procedure is introduced for practical applications.

In a simulation study, the performance of the new test statistic is compared to the well established cumulated sum statistic (CUSUM). Special attention is paid to scenarios were the data shows outliers or stems from a heavy tailed distribution.

The first chapter closes with a real world example and a test for a change in environmental data.

The second chapter deals with the task of estimating the time and the direction of change under the assumption of an unknown change-point. A consistent estimator for the time of change is presented. This consistency does not only hold for fixed magnitudes of change but also if the magnitude of the change vanishes in the long run at a certain rate.

Given the consistent estimator of the time of change, an estimator for the direction of change can be constructed, using the spatial median of the data.

The theoretical results are again supported by a simulation study, investigating similar scenarios as in the first chapter. Finally, the example of environmental data of the first chapter is revisited and time and direction of change are estimated for it.

Zusammenfassung

Das Erkennen von Change-Points ist ein wichtiger Teilbereich der Zeitreihenanalyse. Bereits in den 1950er Jahren kamen die ersten Teststatistiken auf, mit deren Hilfe eine Zeitreihe auf Change-Points untersucht werden kann. Je nach Fragestellung kann es an einem Change-Point unter anderem zu einer Veränderung der zugrundeliegenden Verteilung oder einer Veränderung in Erwartungswert, Varianz oder Korrelation kommen. Der nächste Schritt kann dann sein, den Zeitpunkt der Veränderung zu identifizieren. In jüngerer Vergangenheit nimmt die Bedeutung von Change-Point Analysen hochdimensionaler und abhängiger Daten zu, auch weil mehr solcher Daten verfügbar sind und sich mit steigender Rechenleistung untersuchen lassen.

Die vorliegende Arbeit besteht aus zwei Kapiteln. Zu Beginn des ersten Kapitels werden die Konzepte von schwacher Abhängigkeit und funktionalen Daten erklärt. Eine U-Statistik (U steht für unbiased) wird eingeführt, die für solche Daten beim Testen der Hypothese, dass keine Veränderung vorliegt, genutzt werden kann. Anschließend werden asymptotische Ergebnisse für die Statistik unter der Annahme des Vorliegens und Nichtvorliegens eines Change-Points dargestellt. Unter der Annahme, dass kein Change-Point vorliegt, wird die Grenzverteilung der Teststatistik bestimmt. Da diese einen unendlich-dimensionalen Kovarianzoperator enthält, wird ein Bootstrap-Verfahren für die praktische Nutzung eingeführt.
In einer Simulationsstudie wird das Verhalten der neuen Teststatistik mit dem der bewährten Cumulated-Sum-Statistik (CUSUM) verglichen. Insbesondere werden Szenarien betrachtet, in denen die Daten Ausreißer enthalten oder von einer Verteilung mit schweren Rändern stammen. Das erste Kapitel schließt mit einem realen Beispiel von Umweltdaten und einem darauf angewandten Hypothesentest auf einen Change-Point.

Das zweite Kapitel beschäftigt sich unter der Annahme eines vorliegenden Change-Points mit der Aufgabe, den Zeitpunkt und die Richtung der Veränderung zu schätzen. Für den Zeitpunkt wird ein konsistenter Schätzer präsentiert. Diese Konsistenz gilt nicht nur für feste Veränderungsgrößen, sondern auch, wenn die Größe der Veränderung langfristig mit einer bestimmten Rate verschwindet.
Mit Hilfe des konsistenten Schätzers für den Zeitpunkt der Veränderung und der Nutzung des Spatial Median kann ein Schätzer für die Richtung der Veränderung konstruiert werden.
Die theoretischen Ergebnisse werden wiederum durch eine Simulationsstudie gestützt, in der ähnliche Szenarien wie im ersten Kapitel untersucht werden. Schließlich wird das Beispiel der Umweltdaten aus dem ersten Kapitel wieder aufgegriffen, und es werden Zeit und Richtung der Veränderung geschätzt.

Contents

List of Figures

List of Tables

Notation

Let $x \in H$ be an element in a Hilbert space H, $r \in \mathbb{R}$, $u \in \mathbb{R}^n$

$D_{H^2}[0,1]$	Space of cadlag (right-continuous with left limits) functions from $[0,1]$ to H^2
$\|x\|$	$= \|x\|_H$ Hilbert space norm of x
$B_H(0,1)$	$= \{x \in H : \|x\| \leq 1\}$ Unit ball of H around zero
$\lfloor r \rfloor$	Floor function of r, equals greatest integer less than or equal to r
u^T	Transpose of u

Let $A \subset H$ be a subset of a Hilbert space H

$\mathbf{1}_A$	indicator function of set A
A^c	Complement of A

Let X, Z, $\tilde{Z}$, $(X_n)_{n\in\mathbb{N}}$ be random variables in a Hilbert space H, $\tilde{Z}$ independent of X, Y a random variable in $\mathbb{R}$, $f : H \times H \to H$ a function, $p \geq 1$, $p \in \mathbb{R}$

$\mathbb{E}[X]$	Expected value of random variable X
$\mathbb{E}_X[f(X,\tilde{Z})]$	Conditional expectation of $f(X,\tilde{Z})$ given $\tilde{Z} = z$
$\mathrm{Var}(X)$	Variance of X
$\mathrm{Cov}(X,Z)$	Covariance of X and Z
$\mathbb{E}[\|X\|^p]^{1/p}$	$= \|X\|_p$ p-norm / L_p-norm of X
$X_n \xrightarrow{\mathbb{P}} X$	X_n converges to X in probability as $n \to \infty$
$X_n \xrightarrow{\mathcal{D}} X$, $X_n \Rightarrow X$	X_n converges to X in distribution as $n \to \infty$
$X_n \xrightarrow{\text{a.s.}} X$	X_n converges to X almost surely as $n \to \infty$
$X \overset{\mathcal{D}}{=} Z$	X and Z have the same distribution
X, Z i.i.d.	X and Z are independent, identically distributed random variables
a.s.	Almost surely, with probability 1
$Y \sim \mathcal{N}(\mu, \sigma^2)$	Y is normally distributed with mean μ and variance σ^2
$Y \sim t_n$	Y has a Student's t-distribution with n degrees of freedom

Let $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$ with $a_n > 0$ be deterministic sequences in $\mathbb{R}$ and $(X_n)_{n\in\mathbb{N}}$, $(Z_n)_{n\in\mathbb{N}}$ with $X_n > 0$ a.s. be (stochastic) sequences in $\mathbb{R}$

$b_n = \mathcal{o}(a_n)$	$\frac{b_n}{a_n} \xrightarrow{n\to\infty} 0$
$b_n = \mathcal{O}(a_n)$	$\exists\, M > 0$ s.t. $\lvert b_n\rvert \leq M a_n \;\forall\, n \in \mathbb{N}$
$Z_n = \mathcal{o}_P(X_n)$	$\frac{Z_n}{X_n} \xrightarrow{\mathbb{P}} 0$
$Z_n = \mathcal{O}_P(X_n)$	$\forall\, \varepsilon > 0 \;\exists\, M > 0,\ N \in \mathbb{N}$ such that $\mathbb{P}(\lvert Z_n\rvert > M\, X_n) < \varepsilon \;\forall\, n > N$

1 Introduction

Imagine we observe the amount of fine dust particles (also called *particulate matter* or PM_{10}) at Universitätsplatz in Magdeburg every hour of the day over the course of some months. Naturally, the amount of dust varies during the day and also depending on the day of the week. One could think that the amount is higher in the morning and afternoon, when there is a lot of commuter traffic. This may look differently on weekends. We can see our observations as a time series, where each observation (Monday 7 a.m., Monday 8 a.m., ...) resembles a realization of one random variable from our series.
Assume that after some time, we have an external event, having significant impact on the amount of traffic at Universitätsplatz. This may be for example a road closure or more radical a driving ban imposed for the city. After observing the amount of PM_{10} for some more weeks, we are asked whether there is a change in the amount of particulate matter measured at Universitätsplatz.
This is a very typical question from application for time series analysis: Given observations from a fixed time frame, it is asked if there was a change in our observations. The difficulty is that we do observe naturally varying values due to outside effects. The question is, can we nevertheless find a structural change in the data? And furthermore, can we even date the time of the change and its direction?

Taking the example further, imagine that we do not only observe the amount of PM_{10} at Universitätsplatz, but at every major street in Magdeburg. Can we detect a city-wide change in our observations after a driving ban in Magdeburg was imposed? We get an additional problem here: Measuring at different locations over the city means data behaving differently street by street. At some streets we may observe higher values of PM_{10} in general, some may show more variance in the amount of particulate matter over the course of the day than others. However, we still want to evaluate if they share a common change in the amount of fine dust after the driving ban was enforced.
Handling multidimensional observations (this may be really high-dimensional data - maybe we even observe all streets in Magdeburg, not only the major ones) has become a common problem in applications. Therefore, statistical theory for such high-dimensional change-point problems is needed. Various results for real-valued independent multidimensional cases exist. In application though, independence within and between observations can often not be assumed. Think of the streets again: If we have high values of PM_{10} in the morning, they may also be high some hours later. Additionally, it is reasonable to think that the values in neighbouring streets are not independent from each other.
In recent time, theory was extended from independence to data with short range dependence. The underlying assumption is that time points which lie closer together are more likely to influence each other than time points more far away: Values measured today are probably a good indicator of the values on the next day, but presumably not for the same day next year, as too much may happen in between. To further generalise the structure of the time series, we will work with Hilbert

space valued short-range dependent functional data, since this grants a broad range of applications. Functional data consists, as the name says, of functions that in our context contain information of an object over time. Here, the measuring stations in the streets are the objects that collect PM_{10}-data every hour. Multivariate real-valued time series can be expressed as special cases of this construction, but since the functional data is defined to take values in a (potentially) infinite function space, the tools from multivariate time-series analysis cannot be directly applied.
As said before, in practice, data often varies by some outside effects. Even if overall the amount of particulate matter goes down after the driving ban, it may occur that on some days the amount is higher than ever. The challenge is to develop a test procedure that is not influenced (much) by such outlier observations, which is called a robust test.
In general, a change-point test is some sort of hypothesis test, designed to investigate the question whether there is a change in the data at *some* time. It does not necessarily estimate the time of the change as well. Chapter 2 is dedicated to such a hypothesis test. We will generalise a change-point test from the multivariate real space to arbitrary Hilbert spaces, which are general spaces for functionals and include $\mathbb{R}^d$ as well as the often convenient space of square integrable functions. Hilbert space analysis benefits from the generality of the space, which gives a broad range of applications. For example, Hilbert space methods can also be used to analyse real-valued non-linear statistics as the von Mises statistic.
We start by explaining the concepts of short-range dependent and functional data in Section 2.1 and introducing Wilcoxon-type U-statistics (where U stands for unbiased), which are the base for the test statistic, in Section 2.2. Asymptotic results for this test are proven in Section 2.3 and 2.4. Under the hypothesis ("there is no structural change in the data"), the limit distribution of the test statistic is given. On the other hand, it is shown that the test statistic diverges under the alternative ("there is a change at some time point"). To our knowledge, this has not been done before for this theoretical context.
The divergence of the test statistic under the alternative lays the foundation for the use of the test statistic in application. We propose an advancement of the Dependent Wild Bootstrap in Section 2.5, for resampling observations which are in turn used for the test decision. In a simulation study (Section 2.6) we compare our new test to the well-established CUSUM test and observe that ours is indeed more robust in the case of outliers in the data or if the observations stem from some heavy tailed distribution and is not inferior under normality assumptions. We conclude the chapter with a closer look at a real-world example of particulate matter measured all over Germany in the first half of 2020 in Section 2.7.
In Chapter 3, we will focus on estimating the time and direction of a change if our test gives a significant result for the existence of a change-point. The first step is the estimation of the time of change (Section 3.1). While an estimator can be derived naturally from the test statistic, the main result is that this is actually a consistent estimator for the time of change. It is sometimes assumed that the change gets smaller when the sample size grows. One can think that the effect of a change at some time t levels out in the long run. In an extension in Section 3.1.1, we will

prove that we can handle a vanishing difference in the sense that our estimator of the time of change is still consistent.
The consistency of the estimator for the time of change is an essential property required to prove consistency of the estimator of the direction of change in Section 3.2 as well. Here, we encounter a double estimation problem: Even if we knew the true time of change, we would still have to estimate the direction of change. However we only estimated the time of change; so we prove that nevertheless we achieve a suitable estimator for the direction of change if we only know the estimated time of change.
Section 3.3 provides another simulation study, presenting the performance of the estimations of time and direction. Again, we compare our method to the procedure based on CUSUM. As for the hypothesis test, our method is clearly superior in the case of heavy tailed data or data with outliers. We close this thesis by revisiting the real world example of Chapter 2 in Section 3.4.

2 Change-Point Detection

The problem of detecting and quantifying changes in time series data may be called a classical problem in statistics. Change-point analysis offers a wide field for questions, dealing not only with changes in mean but among others with changes in variance or with multiple change-points (see for example Csörgö and Horváth [1997] for a brief overview). Page [1954] used the *cumulative sum (CUSUM)* to detect a change in the mean of a quality parameter in manufacturing.
More recently, the focus shifted to change-point problems of high dimensional data. As more and more of such data is available and computing power increases, there is ambition to develop robust methods useful for various applications. It is often convenient to think of these data as functional observations modelled as random variables taking values in a Hilbert space. We recommend the book of Hörmann and Kokoszka [2012] for an introduction.
There are mainly two approaches for working with functional data. One possibility is to project the data on lower dimensional spaces using functional principal components. Berkes et al. [2009] did this for independent data and Aston and Kirch [2012] for weakly dependent time series. On the other hand, Change-point tests without dimension reduction are also possible, as done by Horváth et al. [2014] under independence and by Sharipov et al. [2016a] and Aue et al. [2018] under dependence.
Using the full information and no dimension reduction leads to some infinite-dimensional covariance operator in the asymptotic distribution, which can either be estimated (see for example Dehling and Fried [2012] for weakly dependent data) or determined by using resampling techniques. In the context of change-point detection for functional time series, the non-overlapping block bootstrap was studied by Sharipov et al. [2016a], the dependent wild bootstrap by Bucchia and Wendler [2017] and the block multiplier bootstrap (for Banachspace-valued times series) by Dette et al. [2020].
In the tradition of Page [1954], these tests are typically based on variants of the CUSUM-test. Since they make use of the sample mean, they are sensitive to outliers. The *Mann-Whitney-Wilcoxon-U-test*, going back to Wilcoxon [1992] and Mann and Whitney [1947], is a more robust option. Change-point tests based on Wilcoxon have been studied before, mainly for real-valued observations, starting with Darkhovsky [1976] and Pettitt [1979]. Yu and Chen [2022] used the maximum of component-wise Wilcoxon-type statistics. As the Mann-Whitney-Wilcoxon-U-statistic is a special case of a two-sample U-statistic, authors like Csörgő and Horváth [1989] studied more general U-statistics for change point detection under independence and Dehling et al. [2015] under dependence.
Leucht and Neumann [2013] have developed a variant of the dependent wild bootstrap (introduced by Shao [2010]) for degenerate U-statistics. As the Wilcoxon-type statistic is non-degenerate, we propose a new version of the dependent wild bootstrap for this Wilcoxon-type of U-statistic in Section 2.5.

2.1 Functional Data and Dependence

In functional data analysis, a random variable takes values in a function space. Typical functionals in applications are for example annual curves of environmental data. Instead of considering one (possibly non-stationary) series of observations over several years, the curve is split into annual data, where each curve is one functional observation. Other applications are simultaneously sampled data as for example fMRI-data (functional magnetic resonance imaging) that measures brain activity at several locations in the brain at once. The data collected from one location is one functional observation. Thus, sequentially recorded data can be handled as functional time series, that is a series of random variables, each of it taking values in the function space. Ramsay and Silverman [2005] give a fundamental introduction to functional data and its analysis for the independent case.

Definition 2.1 (Functionals)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, H a Hilbert space and $(\zeta_n)_{n\in\mathbb{Z}}$ a stationary sequence of random variables taking values in a separable measurable space S. The sequence $(X_n)_{n\in\mathbb{Z}}$ is a functional of $(\zeta_n)_{n\in\mathbb{Z}}$, if there exists a measurable function $f : S^{\mathbb{Z}} \to H$, such that $X_n = f((\zeta_{n+k})_{k\in\mathbb{Z}})$.*

A very classic example in time series analysis are stock values. By considering the whole stock market as functional observations, tools of functional time series analysis can be used. The advantage of considering the complete market at once is that systematic characteristics stand out against the individual ones. In Figure 2.1, we see an example where one stock behaves rather differently than the others on the stock market. One could guess that a change-point test on just the single stock gives other results than a test for a systematic change in the overall market.

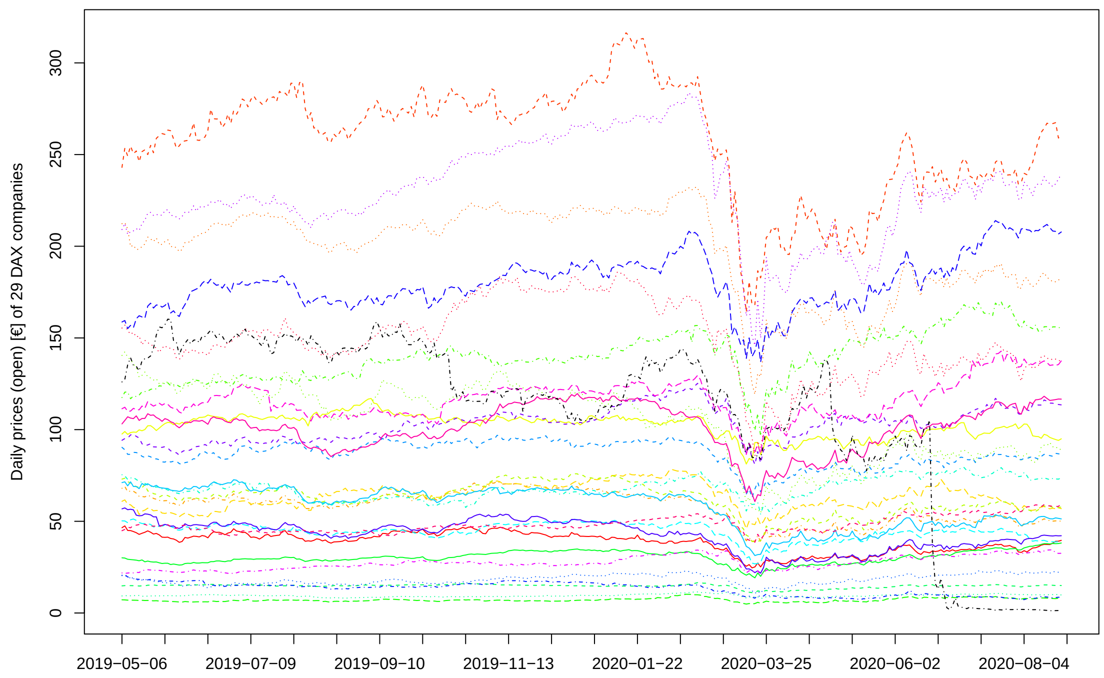


Figure 2.1: Daily open prices in Euro for a single stock (Wirecard) and 29 stocks at the DAX market (black dashed line corresponding to Wirecard) from May 06, 2019 to August 21, 2020. Data extracted from Yahoo!Finance.

Looking at more general examples, we see that the concept of functionals fits for well-known stochastic processes as well.

Example 2.1.

- *Let $(X_t)_{t\in\mathbb{Z}}$ be a d-variate moving average process of order q (MA(q)) with mean zero, i.e.*

$$X_t = \sum_{i=1}^{q} a_i \varepsilon_{t-i} + \varepsilon_t \ \forall t \in \mathbb{Z},$$

where $a_1, ..., a_q \in \mathbb{R}^{d\times d}$ and $(\varepsilon_t)_{t\in\mathbb{Z}}$ a d-variate white noise process (i.e. $(\varepsilon_t)_{t\in\mathbb{Z}}$ is an uncorrelated process with $\mathbb{E}[\varepsilon_t] = 0$, $\mathrm{Cov}(\varepsilon_s, \varepsilon_t) = 0$ if $s \neq t$ and $\mathrm{Cov}(\varepsilon_t, \varepsilon_t) = \Sigma \in \mathbb{R}^{d\times d}$ $\forall t, s \in \mathbb{Z}$).
Clearly, $(X_t)_{t\in\mathbb{Z}}$ is a functional of the white noise process $(\varepsilon_t)_{t\in\mathbb{Z}}$.

- *Let $(X_t)_{t\in\mathbb{Z}}$ be a univariate autoregressive process of order p (AR(p)), i.e.*

$$X_t = \sum_{i=1}^{p} b_i X_{t-i} + \varepsilon_t,$$

where $b_1, ..., b_p \in \mathbb{R}$ and $(\varepsilon_n)_{n\in\mathbb{N}}$ a univariate white noise process.
If all roots of the characteristic polynomial lie outside the unit circle, we can rewrite the process as a univariate moving average process of order ∞:

$$X_t = \sum_{i=1}^{\infty} \tilde{b}_i \varepsilon_{t-i} + \varepsilon_t,$$

where $\tilde{b}_1, \tilde{b}_2, ... \in \mathbb{R}$. Thus, $(X_n)_{n\in\mathbb{N}}$ is again a functional of the white noise process $(\varepsilon_n)_{n\in\mathbb{N}}$.

- *For stock prices, it often does not hold that the data is stationary. So, it is not recommendable to model them as stationary processes like moving average or autoregressive processes. Instead, a model for the log-returns of the stock price can be set up. The log-return at time $t = 1, 2, ...$ is defined as $X_t = \ln(P_t) - \ln(P_{t-1})$, where P_t is the price of the stock at time t. Real-world observations of log-returns show that they are uncorrelated but the squared/absolute log-returns are correlated. This can be captured in a GARCH-process (GARCH stands for* G*eneralised* A*uto*R*egressive with* C*onditional* H*eteroscedasticity). In the univariate setting, $(X_t)_{t\in\mathbb{Z}}$ is called a GARCH(p, q) process if for a univariate white noise process $(\varepsilon_t)_{t\in\mathbb{Z}}$ it fulfils the equation $X_t = \sigma_t \varepsilon_t \ \forall t \in \mathbb{Z}$, where $(\sigma_t)_{t\in\mathbb{Z}}$ is non-negative and the recursion*

$$\sigma_t^2 = a_0 + \sum_{i=1}^{p} a_i X_{t-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-j}^2$$

holds true for real, non-negative parameter $a_0, a_1, ..., a_p$, $b_1, ..., b_q$ *with* $a_p, b_q \neq 0$.
For the special case $p = q = 1$, *the equations*

$$X_t = \sigma_t \varepsilon_t \ \forall t \in \mathbb{Z}$$
$$\sigma_t^2 = a_0 + aX_{t-1}^2 + b\sigma_{t-1}^2 \ \forall t \in \mathbb{Z}$$

have a stationary solution if and only if $\mathbb{E}[\ln(a\varepsilon_1^2 + b)] < 0$. *Especially it holds that*

$$\sigma_t^2 = a_0\Big(1 + \sum_{j=1}^{\infty}\prod_{i=1}^{j}(a\varepsilon_{t-j}^2 + b)\Big) \ \forall t \in \mathbb{Z}$$

(see for example Theorem 14.3 Kreiß and Neuhaus [2006]) and it is thus a functional of the white noise process.

As hinted before, we want to assume some short range dependence in the functional processes. This is a reasonable assumption for many time series applications: Going back to the stock market example, it is likely that a stock value today is dependent on its value yesterday and the day before but not so much on its value one year ago. We want to consider a combination of absolute regularity (introduced by Volkonskii and Rozanov [1959]) and $\mathbb{P}$-near-epoch dependence (introduced by Dehling et al. [2017]). In the following, let H be a seperable Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\|x\| = \sqrt{\langle x, x \rangle}$.

Definition 2.2 (Absolute Regularity)**.** *Let* $(\zeta_n)_{n\in\mathbb{Z}}$ *be a stationary sequence of random variables. We define the mixing coefficients* $(\beta_m)_{m\in\mathbb{Z}}$ *by*

$$\beta_m = \mathbb{E}\Big[\sup_{A\in\mathcal{F}_m^\infty}\big(\mathbb{P}(A|\mathcal{F}_{-\infty}^0) - \mathbb{P}(A)\big)\Big],$$

where $\mathcal{F}_a^b$ *is the* σ*-field generated by* $\zeta_a, \dots, \zeta_b$, *and call the sequence* $(\zeta_n)_{n\in\mathbb{Z}}$ *absolutely regular if* $\beta_m \to 0$ *as* $m \to \infty$.

Mixing coefficients are some sort of measure for dependence. The request $\beta_m \to 0$ as $m \to \infty$ describes the weakening dependence in the data with growing distance, as $\beta_m = 0$ if $(\zeta_n)_{n\in\mathbb{Z}}$ is an independent sequence.
Several other mixing conditions can be found in the literature, the book of Doukhan [1994] gives a broad introduction to mixing theory. See also Bradley [2005] for some basic properties of different mixing coefficients and their connection.

Definition 2.3 (L_p-NED)**.** *Let* $(\zeta_n)_{n\in\mathbb{Z}}$ *be a stationary sequence of random variables.* $(X_n)_{n\in\mathbb{Z}}$ *is called* L_p*-near-epoch-dependent (*L_p*-NED) for* $p \geq 1$ *on* $(\zeta_n)_{n\in\mathbb{Z}}$ *if there exists a sequence of approximation constants* $(a_{k,p})_{k\in\mathbb{N}}$ *with* $a_{k,p} \xrightarrow{k\to\infty} 0$ *and*

$$\mathbb{E}[\|X_0 - \mathbb{E}[X_0|\mathcal{F}_{-k}^k]\|^p]^{\frac{1}{p}} \leq a_{k,p},$$

where $\mathcal{F}_a^b$ *is the* σ*-field generated by* $\zeta_a, \dots, \zeta_b$.

Definition 2.4 ($\mathbb{P}$-NED). *Let $(\zeta_n)_{n\in\mathbb{Z}}$ be a stationary sequence of random variables. $(X_n)_{n\in\mathbb{Z}}$ is called near-epoch-dependent in probability ($\mathbb{P}$-NED) on $(\zeta_n)_{n\in\mathbb{Z}}$ if there exist sequences $(a_k)_{k\in\mathbb{N}}$ with $a_k \xrightarrow{k\to\infty} 0$ and $(f_k)_{k\in\mathbb{Z}}$ and a nonincreasing function $\Phi : (0,\infty) \to (0,\infty)$ such that*

$$\mathbb{P}(\|X_0 - f_k(\zeta_{-k}, ..., \zeta_k)\|_H > \epsilon) \leq a_k \Phi(\epsilon) \;\; \forall k \in \mathbb{N},\, \epsilon > 0.$$

The concept of L_p-NED was introduced by Ibragimov [1962] presenting the idea of weakly dependent data. It is known that $ARMA(p,q)$ and $GARCH(l,q)$ models are L_p-NED (see for example Qiu and Lin [2011]). The concept of $\mathbb{P}$-NED is more general, as it does not require finite moments, which enlarges its application for example to heavy tailed data. Dehling et al. [2017] already proved that any $(X_n)_{n\in\mathbb{Z}}$ being L_p-NED on $(\zeta_n)_{n\in\mathbb{Z}}$ is also $\mathbb{P}$-NED on $(\zeta_n)_{n\in\mathbb{Z}}$. If $(X_n)_{n\in\mathbb{Z}}$ is bounded, the inverse holds true as well.
For a series $(X_n)_{n\in\mathbb{Z}}$ that is $\mathbb{P}$-NED on $(\zeta_n)_{n\in\mathbb{Z}}$, we introduce the following notation for the "truncated" functional

$$X_{n,k} := f_k(\zeta_{n-k}, ..., \zeta_{n+k}) \quad n, k \in \mathbb{N}.$$

2.2 Wilcoxon-Type U-statistic for Change-Point Analysis

Originally introduced by Hoeffding [1948], U-statistics are a class of *unbiased statistics* from estimation theory. In the context of two-sample problems, they can be used to test for differences between the two samples, e.g. in location or distribution. The *Mann-Whitney-Wilcoxon-U-statistic* for two real-valued samples $X_1, ..., X_{n_1} \in \mathbb{R}$ and $Z_1, ..., Z_{n_2} \in \mathbb{R}$ can be written as

$$U(X_1, ..., X_{n_1}, Z_1, ..., Z_{n_2}) = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \operatorname{sgn}(X_i - Z_j) = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \frac{X_i - Z_j}{|X_i - Z_j|}$$

and can be used to test if the two samples have the same location. Chakraborty and Chaudhuri [2017] have generalised this test statistic to Hilbert spaces by replacing the sign by the so called spatial sign. For $X_1, ..., X_{n_1}, Z_1, ..., Z_{n_2} \in H$ it reads

$$U(X_1, ..., X_{n_1}, Z_1, ..., Z_{n_2}) = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \frac{X_i - Z_j}{\|X_i - Z_j\|}$$

They have shown weak convergence against a Gaussian distribution for independent random variables. This two-sample test statistic can be of use for change-point detection. Assuming we have a sample $X_1, ..., X_n \in H$ and a potential change-point $1 < k < n$, we want to compare the X_i before and after k to decide whether k is a change-point or not. A natural approach is to split the sample at k and insert it in the two-sample spatial sign statistic

$$U(X_1, ..., X_k, X_{k+1}, ..., X_n) = \frac{1}{k(n-k)} \sum_{i=1}^{k} \sum_{j=k+1}^{n} \frac{X_i - X_j}{\|X_i - X_j\|}$$

to compare the location before and after k.
In practice, one encounters several problems. First of all, the change-point is typically unknown, so it is not known where to split the sequence of observations into two samples. To tackle this problem, we will maximise the test statistic over all possible splitting points.
We will treat the CUSUM statistic and the Wilcoxon-type statistic as two special cases of a general class of change-point statistics based on two-sample U-statistics. Let $h : H^2 \to H$ be a kernel function. We define

$$U_{n,k}(X) := \sum_{i=1}^{k} \sum_{j=k+1}^{n} h(X_i, X_j).$$

Example 2.2 (CUSUM)**.** *For $h(x, y) = x - y$, we can construct the CUSUM-statistic for functional data by taking the maximum of the norm of $U_{n,k}$ combined with a suitable factor:*

$$\max_{1 \leq k < n} \frac{1}{n^{3/2}} \|U_{n,k}(X)\| = \max_{1 \leq k < n} \frac{1}{n^{3/2}} \Big\| \sum_{i=1}^{k} \sum_{j=k+1}^{n} (X_i - X_j) \Big\|$$

$$= \max_{1\le k<n} \frac{1}{n^{3/2}} \big\|(n-k)\sum_{i=1}^{k} X_i - k\sum_{j=k+1}^{n} X_j\big\|$$

$$= \max_{1\le k<n} \frac{1}{n^{3/2}} \big\|(n-k)X_1 - \sum_{j=k+1}^{n} X_j + ... + (n-k)X_k - \sum_{j=k+1}^{n} X_j \pm k\sum_{j=1}^{k} X_j\big\|$$

$$= \max_{1\le k<n} \frac{1}{n^{3/2}} \big\|nX_1 - \sum_{j=1}^{n} X_j + ... + nX_k - \sum_{j=1}^{n} X_j\big\|$$

$$= \max_{1\le k<n} \frac{1}{n^{3/2}} \big\|n\sum_{i=1}^{k} \big(X_i - \frac{1}{n}\sum_{j=1}^{n} X_j\big)\big\| = \max_{1\le k<n} \frac{1}{n^{1/2}} \big\|\sum_{i=1}^{k} \big(X_i - \bar{X}\big)\big\|$$

Example 2.3 (Wilcoxon-type)**.** *For $h(x,y) = (x-y)/\|x-y\|$, we can construct the Wilcoxon-type changepoint statistic for functional data similar to Example 2.2. The factor $\frac{1}{n^{3/2}}$ is chosen to achieve the desired asymptotic behaviour:*

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

2.2.1 The Kernel h

We will prove a limit theorem for a general class of change-point statistics given by $\max_{1\le k<n} \frac{1}{n^{3/2}}\|U_{n,k}\|$. Nevertheless, we have to make some assumptions on the kernel function h.

Definition 2.5 (Antisymmetry)**.** *A kernel $h : H^2 \to H$ is called antisymmetric, if for all $x, y \in H$*

$$h(x,y) = -h(y,x)$$

Antisymmetric kernels are natural candidates for comparing two distributions, because if X and $\tilde{X}$ are two independent H-valued random variables with the same distribution and h is antisymmetric, we have

$$\mathbb{E}[h(X,\tilde{X})] = \mathbb{E}[h(\tilde{X},X)] = \mathbb{E}[-h(X,\tilde{X})] = -\mathbb{E}[h(X,\tilde{X})]$$

meaning $E[h(X,\tilde{X})] = 0$, so our test statistic should have values close to 0 if there is no change in the data.

Definition 2.6 (Uniform Moments)**.** *Let $(X_n)_{n\in\mathbb{Z}}$ be a $\mathbb{P}$-NED functional of $(\zeta_n)_{n\in\mathbb{Z}}$. If there exists $M > 0$ such that for all $k, n \in \mathbb{N}$*

$$\mathbb{E}[\|h\big(X_{0,k}, X_{n,k}\big)\|^m] = \mathbb{E}[\|h\big(f_k(\zeta_{-k}, ..., \zeta_k), f_k(\zeta_{n-k}, ..., \zeta_{n+k})\big)\|^m] \le M,$$

$$\mathbb{E}[\|h\big(X_0, X_{n,k}\big)\|^m] = \mathbb{E}[\|h\big(X_0, f_k(\zeta_{n-k}, ..., \zeta_{n+k})\big)\|^m] \le M,$$

$$\mathbb{E}[\|h\big(X_0, X_n\big)\|^m] \le M,$$

we say that the kernel has uniform m-th moments under approximation.

Furthermore, we need the following mild continuity condition on the kernel, introduced by Denker and Keller [1986].

Definition 2.7 (Variation condition). *Let X, Z be two random variables in H. The kernel h fulfils the variation condition, if there exist $L > 0$, $\epsilon_0 > 0$ such that for every $\epsilon \in (0, \epsilon_0)$ it holds that*

$$\mathbb{E}\left[\Big(\sup_{\substack{\|x-X\|\leq\epsilon \\ \|y-Z\|\leq\epsilon}} \|h(x,y) - h(X,Z)\|\Big)^2\right] \leq L\epsilon \quad x, y \in H.$$

It can easily be seen that $h(x,y) = x - y$ fulfils the condition. The kernel $h(x,y) = (x-y)/\|x-y\|$ will fulfil the condition, as long as there exists a constant C such that $P(\|X_1 - x\| \leq \epsilon) \leq C\epsilon$ for all $x \in H$ and $\epsilon > 0$. This can be proved along the lines of Remark 2 in Dehling et al. [2022].
Finally, we will need Hoeffding's decomposition of the kernel to be able to define the limit distribution:

Definition 2.8 (Hoeffding's decomposition). *Let $h : H \times H \to H$ be an antisymmetric kernel. Let X, Z be two independent random variables in H. Hoeffding's decomposition of h with respect to X, Z is defined as*

$$h(x,y) = h_1^{(Z)}(x) - h_1^{(X)}(y) + h_2^{(X,Z)}(x,y),$$

where

$$h_1^{(X)}(y) = \mathbb{E}[h(y,X)], \ h_1^{(Z)}(x) = \mathbb{E}[h(x,Z)]$$
$$h_2^{(X,Z)}(x,y) = h(x,y) - h_1^{(Z)}(x) + h_1^{(X)}(y)$$

with $x, y \in H$.

We get a special case if X and Z follow the same distribution, which is explained in the next remark.

Remark 2.1. *For two independent random variables X, X' following the same distribution, Hoeffding's decomposition with respect to X, X' reads*

$$h(x,y) = h_1^{(X)}(x) - h_1^{(X)}(y) + h_2^{(X)}(x,y),$$

where

$$h_1^{(X)}(x) = \mathbb{E}[h(x,X)]$$

and

$$h_2^{(X)}(x,y) = h(x,y) - h_1^{(X)}(x) + h_1^{(X)}(y)$$

with $x, y \in H$.

In our context, we want to use Hoeffdings's decomposition with respect to samples from a sequence which are not necessarily independent. To handle this, we introduce independent copies:

Remark 2.2. *If X and Z are not independent, we construct Hoeffding's decomposition of the kernel h with respect to X, Z by introducing independent copies $\tilde{X}$ and $\tilde{Z}$ of X resp. Z and say that*

$$h(x, y) = h_1^{(Z)}(x) - h_1^{(X)}(y) + h_2^{(X,Z)}(x, y),$$

where

$$h_1^{(X)}(y) = \mathbb{E}[h(y, \tilde{X})],\ h_1^{(Z)}(x) = \mathbb{E}[h(x, \tilde{Z})]$$
$$h_2^{(X,Z)}(x, y) = h(x, y) - h_1^{(Z)}(x) + h_1^{(X)}(y)$$

for $x, y \in H$.
This ensures independence such that the expectation $\mathbb{E}_X[h_1^{(Z)}(X)] = \mathbb{E}[h(X, \tilde{Z})]$, but note that $\mathbb{E}_X[h(y, X)] = \mathbb{E}_{\tilde{X}}[h(y, \tilde{X})]$ since $\tilde{X}$ is a copy of X.

In the special case described in Remark 2.1, we get some important properties for the expectation of $h_1^{(X)}(X)$ and $h_2^{(X)}(X, y)$:

Remark 2.3. *Let X be a random variable in H, and $\tilde{X}$ and independent copy of X. For the two types of functions in Hoeffding's decomposition, the following properties hold:*

$$\begin{aligned}\mathbb{E}[h_1^{(X)}(X)] &= \mathbb{E}_X\left[\mathbb{E}_{\tilde{X}}[h(X, \tilde{X})]\right] = -\mathbb{E}_X\left[\mathbb{E}_{\tilde{X}}[h(\tilde{X}, X)]\right] = -\mathbb{E}_X\left[\mathbb{E}_{\tilde{X}}[h(X, \tilde{X})]\right] \\ &= -\mathbb{E}[h_1^{(X)}(X)].\end{aligned}$$

Meaning that $\mathbb{E}[h_1^{(X)}(X)] = 0$. Here we first used the antisymmetry of h and the fact that X and $\tilde{X}$ follow the same distribution. Furthermore, it follows:

$$\begin{aligned}\mathbb{E}[h_2^{(X)}(X, y)] &= \mathbb{E}_X[h(X, y)] - \mathbb{E}_X[h_1^{(X)}(X)] + h_1^{(X)}(y) \\ &= \mathbb{E}_X[h(X, y)] + \mathbb{E}_{\tilde{X}}[h(y, \tilde{X})] \\ &= -\mathbb{E}_X[h(y, X)] + \mathbb{E}_{\tilde{X}}[h(y, \tilde{X})] \\ &= 0\end{aligned}$$

again using the antisymmetry of h and the fact that X and $\tilde{X}$ follow the same distribution. Applying the same arguments, it holds that $\mathbb{E}[h_2^{(X)}(y, X)] = 0$ as well. $h_2^{(X)}$ with this property is called degenerate.

It is important to note that uniform moments and variation condition transfer to the components of Hoeffding's decomposition.

Proposition 2.1.

i) The variation condition holds for $h_1^{(X)}$, $h_1^{(Z)}$ and $h_2^{(X,Z)}$.

ii) $h_2^{(X)}$ has uniform m-th moments under approximation, if h has it.

Proof.
i) Variation condition
By Hoeffding's decomposition for h it holds that for all $x, x' \in H$

$$\|h_1^{(X)}(x) - h_1^{(X)}(x')\| = \|\mathbb{E}[h(x,\tilde{X})] - \mathbb{E}[h(x',\tilde{X})]\|.$$

Let X, Z be two random variables in H. Then by Jensen's inequality for conditional expectations and the variation condition for h it is

$$\begin{aligned} &\mathbb{E}\Big[\Big(\sup_{\|y-Z\|\le\epsilon} \|h_1^{(X)}(y) - h_1^{(X)}(Z)\|\Big)^2\Big] \qquad (1)\\ &= \mathbb{E}\Big[\Big(\sup_{\|y-Z\|\le\epsilon} E\big[\|h(y,\tilde{X}) - h(Z,\tilde{X})\big|Z\big]\|\Big)^2\Big]\\ &\le \mathbb{E}\Big[\Big(\sup_{\|y-Z\|\le\epsilon} \|h(y,\tilde{X}) - h(Z,\tilde{X})\|\Big)^2\Big]\\ &\le \mathbb{E}\Big[\Big(\sup_{\substack{\|y-Z\|\le\epsilon\\ \|x-\tilde{X}\|\le\epsilon}} \|h(y,x) - h(Z,\tilde{X})\|\Big)^2\Big] \le L\epsilon. \end{aligned}$$

For $h_1^{(Z)}$ it holds the same. Therefore it follows for $h_2^{(X,Z)}$, using Hoeffding's decomposition:

$$\begin{aligned} &\mathbb{E}\Big[\Big(\sup_{\substack{\|x-X\|\le\epsilon\\ \|y-Z\|\le\epsilon}} \|h_2^{(X,Z)}(x,y) - h_2^{(X,Z)}(X,Z)\|\Big)^2\Big]^{1/2}\\ &\le \mathbb{E}\Big[\Big(\sup_{\substack{\|x-X\|\le\epsilon\\ \|y-Z\|\le\epsilon}} \|h(x,y) - h(X,Z)\|\Big)^2\Big]^{1/2} + \mathbb{E}\Big[\Big(\sup_{\|x-X\|\le\epsilon} \|h_1^{(Z)}(x) - h_1^{(Z)}(X)\|\Big)^2\Big]^{1/2}\\ &+ \mathbb{E}\Big[\Big(\sup_{\|y-Z\|\le\epsilon} \|h_1^{(X)}(y) - h_1^{(X)}(Z)\|\Big)^2\Big]^{1/2} \le 3(L\epsilon)^{1/2}. \end{aligned}$$

Taking the square gives the result.
ii) Uniform moments
This can again be shown by using Hoeffding's decomposition and then Jensen's inequality. We will show this exemplary for $\mathbb{E}[\|h_2^{(X)}(X_0, X_n)\|^m]$:

$$\begin{aligned} &\mathbb{E}[\|h_2^{(X)}(X_0,X_n)\|^m]^{1/m}\\ &\le \mathbb{E}[\|h(X_0,X_n)\|^m]^{1/m} + \mathbb{E}\big[\|\mathbb{E}_{\tilde{X}}[h(X_0,\tilde{X})]\|^m\big]^{1/m} + \mathbb{E}\big[\|\mathbb{E}_{\tilde{X}}[h(\tilde{X},X_n)]\|^m\big]^{1/m}\\ &\le \mathbb{E}[\|h(X_0,X_n)\|^m]^{1/m} + \mathbb{E}\big[\mathbb{E}_{\tilde{X}}[\|h(X_0,\tilde{X})\|^m]\big]^{1/m} + \mathbb{E}\big[\mathbb{E}_{\tilde{X}}[\|h(\tilde{X},X_n)\|^m]\big]^{1/m}\\ &\le 3M^{1/m} \end{aligned}$$

by uniform moments of h. □

2.3 Main Result Under the Hypothesis

A change-point test can be seen as a hypothesis test on a sample $X_1, ..., X_n$ with the hypothesis of no structural change in the data

$$H_0: \quad X_1 \stackrel{\mathcal{D}}{=} X_2 \stackrel{\mathcal{D}}{=} ... \stackrel{\mathcal{D}}{=} X_n$$

against the alternative of at most one change

$$H_1: \quad \exists\, 1 < k < n \text{ such that } \; X_1 \stackrel{\mathcal{D}}{=} ... \stackrel{\mathcal{D}}{=} X_k \stackrel{\mathcal{D}}{\neq} X_{k+1} \stackrel{\mathcal{D}}{=} ... \stackrel{\mathcal{D}}{=} X_n$$

Given the definitions and concepts introduced above, we can state our first theorem on the asymptotic distribution of our test statistic $\max_{1\leq k<n} \frac{1}{n^{3/2}} \|U_{n,k}\|$ under the hypothesis, that is the assumption of no change.

Theorem 2.1. *Let $(X_n)_{n\in\mathbb{Z}}$ be $\mathbb{P}-$NED on an absolutely regular sequence $(\zeta_n)_{n\in\mathbb{Z}}$ such that $a_k \Phi(k^{-8\frac{\delta+3}{\delta}}) = \mathcal{O}(k^{-8\frac{(\delta+3)(\delta+2)}{\delta^2}})$ and $\sum_{k=1}^{\infty} k^2 \beta_k^{\frac{\delta}{4+\delta}} < \infty$ for some $\delta > 0$. Assume that $h : H^2 \to H$ is an antisymmetric kernel that fulfils the variation condition and is either bounded or has uniform $(4+\delta)$-moments under approximation. Then it holds that*

$$\max_{1\leq k<n} \frac{1}{n^{3/2}} \|U_{n,k}(X)\| = \max_{1\leq k<n} \frac{1}{n^{3/2}} \Big\| \sum_{i=1}^{k} \sum_{j=k+1}^{n} h(X_i, X_j) \Big\| \xrightarrow{\mathcal{D}} \sup_{\lambda\in[0,1]} \|W(\lambda) - \lambda W(1)\|,$$

where W is an H-valued Brownian motion and the covariance operator S of $W(1)$ is given by

$$\langle S(x), y\rangle = \sum_{i=-\infty}^{\infty} \mathrm{Cov}\left(\langle h_1(X_0), x\rangle, \langle h_1(X_i), y\rangle\right).$$

Remark 2.4. *Since in the definition of absolute regularity it is assumed that $(\zeta_n)_{n\in\mathbb{Z}}$ is a stationary process, this implies that the functional $(X_n)_{n\in\mathbb{Z}}$ of $(\zeta_n)_{n\in\mathbb{Z}}$ is stationary as well.*

For the kernel $h(x, y) = x - y$, we obtain as a special case a limit theorem for the functional CUSUM-statistic similar to Corollary 1 of Sharipov et al. [2016a]. Before we start proving this theorem, already note that the limit distribution depends on an infinite dimensional covariance operator, which is unknown in practice. We propose to use a new version of the dependent wild bootstrap for resampling. The actual procedure and asymptotic validity of the method is presented in Chapter 2.5.
The proofs will make use of Hoeffding's decomposition of the kernel h, which allows us to prove asymptotic results each for the linear and degenerate part of the test statistic. In this chapter, we will omit the superscript of $h_1^{(X)}$ and $h_2^{(X)}$ for notational simplicity.

Lemma 2.1 (Hoeffding's decomposition of $U_{n,k}$)**.** *Let $h : H \times H \to H$ be an anti-symmetric kernel.*
Under Hoeffding's decomposition it holds for the test statistic that

$$U_{n,k}(X) = \sum_{i=1}^{k} \sum_{j=k+1}^{n} h(X_i, X_j) = \underbrace{n \sum_{i=1}^{k} (h_1(X_i) - \overline{h_1(X)})}_{\text{linear part}} + \underbrace{\sum_{i=1}^{k} \sum_{j=k+1}^{n} h_2(X_i, X_j)}_{\text{degenerate part}},$$

where $\overline{h_1(X)} = \frac{1}{n} \sum_{j=1}^{n} h_1(X_j)$.

Proof. After using Hoeffding's decomposition for h, some calculations lead to:

$$\begin{aligned}
U_{n,k}(X) &= \sum_{i=1}^{k} \sum_{j=k+1}^{n} h(X_i, X_j) = \sum_{i=1}^{k} \sum_{j=k+1}^{n} [h_1(X_i) - h_1(X_j) + h_2(X_i, X_j)] \\
&= \sum_{i=1}^{k} \sum_{j=k+1}^{n} [h_1(X_i) - h_1(X_j)] + \sum_{i=1}^{k} \sum_{j=k+1}^{n} h_2(X_i, X_j) \\
&= (n-k) h_1(X_1) - \sum_{j=k+1}^{n} h_1(X_j) + \ldots + (n-k) h_1(X_k) \\
&\quad - \sum_{j=k+1}^{n} h_1(X_j) + \sum_{i=1}^{k} \sum_{j=k+1}^{n} h_2(X_i, X_j) \\
&= n h_1(X_1) - \sum_{j=1}^{n} h_1(X_j) + \ldots + n h_1(X_k) - \sum_{j=1}^{n} h_1(X_j) \\
&\quad + \sum_{i=1}^{k} \sum_{j=k+1}^{n} h_2(X_i, X_j) \\
&= n \Big(\sum_{i=1}^{k} [h_1(X_i) - \frac{1}{n} \sum_{j=1}^{n} h_1(X_j)] \Big) + \sum_{i=1}^{k} \sum_{j=k+1}^{n} h_2(X_i, X_j) \\
&= n \sum_{i=1}^{k} \Big(h_1(X_i) - \overline{h_1(X)} \Big) + \sum_{i=1}^{k} \sum_{j=k+1}^{n} h_2(X_i, X_j).
\end{aligned}$$

□

The idea of proof for Theorem 2.1 is to handle the linear and degenerate part of $\max_{1 \leq k < n} \frac{1}{n^{3/2}} \|U_{n,k}(X)\|$ separately. For the linear part, convergence results of Sharipov et al. [2016a] for partial sums can be used, given some properties of $(h_1(X_n))_{n \in \mathbb{Z}}$ that need to be checked. In fact, the linear part already converges in distribution to the process given in Theorem 2.1. As a consequence, the second (and bigger) task is to show that the degenerate part converges to zero in probability.
To use existing results about partial sums for the linear part, it is first checked that $(h_1(X_n))_{n \in \mathbb{Z}}$ is L_2-NED:

Lemma 2.2. *Under the assumptions of Theorem 2.1, $(h_1(X_n))_{n\in\mathbb{Z}}$ is L_2-NED with approximation constants $a_{k,2} = \mathcal{O}(k^{-4\frac{\delta+3}{\delta}})$.*

Proof. We recall the following notation $X_{n,k} = f_k(\zeta_{n-k}, ..., \zeta_{n+k})$ and let $\tilde{X}_{n,k}$ be an independent copy of this random variable. Now, we can find the approximation constants of $(h_1(X_n))_{n\in\mathbb{Z}}$ by using its variation condition (1) stated in Corollary 2.1 and some further inequalities.

$$
\begin{aligned}
&\mathbb{E}[\|h_1(X_0) - \mathbb{E}[h_1(X_0)|\mathcal{F}_{-k}^k]\|^2] \leq \mathbb{E}[\|h_1(X_0) - h_1(X_{0,k})\|^2] \\
&= \mathbb{E}[\|h_1(X_0) - h_1(X_{0,k})\|^2 \mathbf{1}_{\{\|X_0 - X_{0,k}\| > s_k\}}] \\
&\qquad + \mathbb{E}[\|h_1(X_0) - h_1(X_{0,k})\|^2 \mathbf{1}_{\{\|X_0 - X_{0,k}\| \leq s_k\}}] \\
&\leq \mathbb{E}[\|h_1(X_0) - h_1(X_{0,k})\|^2 \mathbf{1}_{\{\|X_0 - X_{0,k}\| > s_k\}}] \\
&\qquad + \underbrace{\mathbb{E}\left[\left(\sup_{\|X_0 - X_{0,k}\| \leq s_k} \|h_1(X_0) - h_1(X_{0,k})\|\right)^2\right]}_{\overset{(1)}{\leq} Ls_k} \\
&\leq \left\|\|h_1(X_0) - h_1(X_0,k)\|^2\right\|_{\frac{2+\delta}{2}} + \left\|\mathbf{1}_{\{\|X_0 - X_{0,k}\| > s_k\}}\right\|_{\frac{2+\delta}{\delta}} + Ls_k \\
&= \left\|\|h_1(X_0) - h_1(X_0,k)\|^2\right\|_{\frac{2+\delta}{2}} + \mathbb{P}(\|X_0 - X_{0,k}\| > s_k)^{\frac{\delta}{2+\delta}} + Ls_k \\
&\leq \mathbb{E}[\|h_1(X_0) - h_1(X_{0,k})\|^{2+\delta}]^{\frac{2}{2+\delta}} + (a_k\Phi(s_k))^{\frac{\delta}{2+\delta}} + Ls_k
\end{aligned}
$$

by using Hölder's inequality and the fact that $(X_n)_n$ is $\mathbb{P}$-NED. Now, also using Jensen's and Minkowski's inequality and the uniform moment condition, it holds that

$$
\begin{aligned}
&\mathbb{E}[\|h_1(X_0) - h_1(X_{0,k})\|^{2+\delta}]^{\frac{2}{2+\delta}} + (a_k\Phi(s_k))^{\frac{\delta}{2+\delta}} + Ls_k \\
&= \mathbb{E}\left[\left\|\mathbb{E}[h(X_0, \tilde{X}_0)|X_0, X_{0,k}] - \mathbb{E}[h(X_{0,k}, \tilde{X}_{0,k})|X_0, X_{0,k}]\right\|^{2+\delta}\right]^{\frac{2}{2+\delta}} (a_k\Phi(s_k))^{\frac{\delta}{2+\delta}} \\
&\qquad + Ls_k \\
&\leq \mathbb{E}\left[\mathbb{E}[\|h(X_0, \tilde{X}_0) - h(X_{0,k}, \tilde{X}_{0,k})\|^{2+\delta}|X_0, X_{0,k}]\right]^{\frac{2}{2+\delta}} (a_k\Phi(s_k))^{\frac{\delta}{2+\delta}} + Ls_k \\
&= \left(\mathbb{E}[\|h(X_0, \tilde{X}_0) - h(X_{0,k}, \tilde{X}_{0,k})\|^{2+\delta}]^{\frac{1}{2+\delta}}\right)^2 (a_k\Phi(s_k))^{\frac{\delta}{2+\delta}} + Ls_k \\
&\leq \left(\mathbb{E}[\|h(X_0, \tilde{X}_0)\|^{2+\delta}]^{\frac{1}{2+\delta}} + \mathbb{E}[\|h(X_{0,k}, \tilde{X}_{0,k})\|^{2+\delta}]^{\frac{1}{2+\delta}}\right)^2 (a_k\Phi(s_k))^{\frac{\delta}{2+\delta}} + Ls_k \\
&\leq (M^{\frac{1}{2+\delta}} + M^{\frac{1}{2+\delta}})^2 (a_k\Phi(s_k))^{\frac{\delta}{2+\delta}} + Ls_k.
\end{aligned}
$$

Let $s_k = k^{-8\frac{3+\delta}{\delta}}$ and recall the assumptions on the $\mathbb{P}$-NED coefficients to get

$$
\begin{aligned}
&(M^{\frac{1}{2+\delta}} + M^{\frac{1}{2+\delta}})^2 (a_k\Phi(s_k))^{\frac{\delta}{2+\delta}} + Ls_k \\
&\leq C(k^{-8\frac{(3+\delta)(2+\delta)}{\delta^2}})^{\frac{\delta}{2+\delta}} + Lk^{-8\frac{3+\delta}{\delta}}.
\end{aligned}
$$

By taking the square root, we get the result:

$$
\left(\mathbb{E}[\|h_1(X_0) - \mathbb{E}[h_1(X_0)|\mathcal{F}_{-k}^k]\|^2]\right)^{\frac{1}{2}} \leq Ck^{-4\frac{3+\delta}{\delta}} =: a_{k,2}.
$$

Since it holds that $a_{k,2} \xrightarrow{k\to\infty} 0$, $(X_n)_{n\in\mathbb{Z}}$ is L_2-NED. □

Proposition 2.2. *Under the assumptions of Theorem 2.1 it holds:*

$$\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor n\lambda\rfloor} h_1(X_i)\Big)_{\lambda\in[0,1]} \xrightarrow{\mathcal{D}} (W(\lambda))_{\lambda\in[0,1]},$$

where $(W(\lambda))_{\lambda\in[0,1]}$ *is a Brownian motion with covariance operator as defined in Theorem 2.1.*

Proof. We want to use Theorem 1 Sharipov et al. [2016a] for $(h_1(X_n))_{n\in\mathbb{Z}}$, so we have to check its assumptions:
Assumption 1: $(h_1(X_n))_{n\in\mathbb{Z}}$ is L_1-NED.
We know by Lemma 2.2 that $(h_1(X_n))_{n\in\mathbb{Z}}$ is L_2-NED. Thus, L_1-NED follows by Jensen's inequality:

$$\mathbb{E}[\|h_1(X_0) - \mathbb{E}[h_1(X_0)|\mathcal{F}_{-k}^{k}]\|] \le \mathbb{E}[\|h_1(X_0) - \mathbb{E}[h_1(X_0)|\mathcal{F}_{-k}^{k}]\|^2]^{\frac{1}{2}} \le a_{k,2}.$$

So, $(h_1(X_n))_{n\in\mathbb{Z}}$ is L_1-NED with constants $a_{k,1} = a_{k,2} = Ck^{-4\frac{3+\delta}{\delta}}$.
Assumption 2: Existing $(4+\delta)$-moments.
This follows from the assumption of uniform moments under approximation and Jensen's inequality:

$$\begin{aligned}\mathbb{E}[\|h_1(X_0)\|^{4+\delta}] &= \mathbb{E}\big[\|\mathbb{E}[h(X_0,\tilde{X}_0)|X_0]\|^{4+\delta}\big]\\ &\le \mathbb{E}\big[\mathbb{E}[\|h(X_0,\tilde{X}_0)\|^{4+\delta}|X_0]\big]\\ &= \mathbb{E}[\|h(X_0,\tilde{X}_0)\|^{4+\delta}] \le M < \infty.\end{aligned}$$

In the case that h is bounded, h_1 is bounded as well and thus moments exist.
Assumption 3: $\sum_{m=1}^{\infty} m^2 a_{m,1}^{\frac{\delta}{3+\delta}} < \infty$.
A short calculations leads to:

$$\sum_{m=1}^{\infty} m^2 a_{m,1}^{\frac{\delta}{3+\delta}} = C\sum_{m=1}^{\infty} m^2 (m^{-4\frac{3+\delta}{\delta}})^{\frac{\delta}{3+\delta}} = C\sum_{m=1}^{\infty} m^2 m^{-4} = C\sum_{m=1}^{\infty} m^{-2} < \infty.$$

Assumption 4: $\sum_{m=1}^{\infty} m^2 \beta_m^{\frac{\delta}{4+\delta}} < \infty$.
This holds directly by the assumed rate on the coefficients β_m.
We have checked that all assumptions for Theorem 1 Sharipov et al. [2016a] are fulfilled and since $\mathbb{E}[h_1(X_0)] = 0$, the statement of the proposition follows directly by the use of the said theorem. □

With this result, we can already show the convergence of the linear part of $\max_{1\le k<n} \frac{1}{n^{3/2}}\|U_{n,k}(X)\|$. To complete the proof of Theorem 2.1, we will prove that the degenerate part converges to zero in probability. To do so, we first state and prove four technical lemmas, needed to handle the L_2-norm of the degenerate part. This is done by approximating h_2 by some truncated counterparts that can in turn be suitably bounded and are not too far away from h_2.

Lemma 2.3. *Under the assumptions of Theorem 2.1, there exists a universal constant $C > 0$ such that for every $i, k, l \in \mathbb{N}$, $\epsilon > 0$ it holds that*

$$\mathbb{E}[\|h_2(X_i, X_{i+k+2l}) - h_2(X_{i,l}, X_{i+k+2l,l})\|^2]^{\frac{1}{2}} \leq C\big(\sqrt{\epsilon} + \beta_k^{\frac{\delta}{2(2+\delta)}} + (a_l \Phi(\epsilon))^{\frac{\delta}{2(2+\delta)}}\big),$$

where $X_{i,l} = f_l(\zeta_{i-l}, ..., \zeta_{i+l})$.

Proof. By Lemma D1 Dehling et al. [2017] there exist independent copies $(\zeta'_n)_{n\in\mathbb{Z}}$, $(\zeta''_n)_{n\in\mathbb{Z}}$ of $(\zeta_n)_{n\in\mathbb{Z}}$ satisfying

$$\mathbb{P}((\zeta'_n)_{n\geq i+k+l} = (\zeta_n)_{n\geq i+k+l}) = 1 - \beta_k \quad \text{and} \quad \mathbb{P}((\zeta''_n)_{n\leq i+l} = (\zeta_n)_{n\leq i+l}) = 1 - \beta_k. \tag{2}$$

Define

$$\begin{aligned} X'_i &= f((\zeta'_{i+n})_{n\in\mathbb{Z}}) \ , \quad X''_i = f((\zeta''_{i+n})_{n\in\mathbb{Z}}) \\ X'_{i,l} &= f_l(\zeta'_{i-l}, ..., \zeta'_{i+l}) \ , \quad X''_{i,l} = f_l(\zeta''_{i-l}, ..., \zeta''_{i+l}). \end{aligned}$$

With the help of these, we can write

$$\begin{aligned} &\mathbb{E}[\|h_2(X_i, X_{i+k+2l}) - h_2(X_{i,l}, X_{i+k+2l,l})\|^2]^{\frac{1}{2}} \\ \leq\ &\mathbb{E}[\|h_2(X_i, X_{i+k+2l}) - h_2(X''_i, X'_{i+k+2l})\|^2]^{\frac{1}{2}} && (3) \\ &+ \mathbb{E}[\|h_2(X''_i, X'_{i+k+2l}) - h_2(X''_{i,l}, X'_{i+k+2l,l})\|^2]^{\frac{1}{2}} && (4) \\ &+ \mathbb{E}[\|h_2(X''_{i,l}, X'_{i+k+2l,l}) - h_2(X_{i,l}, X_{i+k+2l,l})\|^2]^{\frac{1}{2}} && (5) \end{aligned}$$

by using the triangle inequality. We will look at the three summands separately. For abbreviation, define

$$\begin{aligned} B &= \{(\zeta'_n)_{n\geq i+k+l} = (\zeta_n)_{n\geq i+k+l}, \ (\zeta''_n)_{n\leq i+l} = (\zeta_n)_{n\leq i+l}\} \\ B^c &= \{(\zeta'_n)_{n\geq i+k+l} \neq (\zeta_n)_{n\geq i+k+l} \text{ or } (\zeta''_n)_{n\leq i+l} \neq (\zeta_n)_{n\leq i+l}\} \end{aligned}$$

and write

$$\begin{aligned} (3) &= \mathbb{E}[\|h_2(X_i, X_{i+k+2l}) - h_2(X''_i, X'_{i+k+2l})\|^2]^{\frac{1}{2}} \\ &\leq \mathbb{E}[\|h_2(X_i, X_{i+k+2l}) - h_2(X''_i, X'_{i+k+2l})\|^2 \mathbf{1}_{B^c}]^{\frac{1}{2}} && (3.\text{A}) \\ &+ \mathbb{E}[\|h_2(X_i, X_{i+k+2l}) - h_2(X''_i, X'_{i+k+2l})\|^2 \mathbf{1}_{B}]^{\frac{1}{2}}. && (3.\text{B}) \end{aligned}$$

For (3.A), we use Hölder's inequality together with our assumptions on uniform moments under approximation and get

$$\begin{aligned} (3.A) &\leq \mathbb{E}[\|h_2(X_i, X_{i+k+2l}) - h_2(X''_i, X'_{i+k+2l})\|^{\frac{2(2+\delta)}{2}}]^{\frac{2}{2(2+\delta)}} \mathbb{P}(B^c)^{\frac{\delta}{2(2+\delta)}} \\ &\leq \Big(\mathbb{E}[\|h_2(X_i, X_{i+k+2l})\|^{2+\delta}]^{\frac{1}{2+\delta}} + \mathbb{E}[\|h_2(X''_i, X'_{i+k+2l})\|^{2+\delta}]^{\frac{1}{2+\delta}}\Big) \\ &\quad \cdot \Big(\mathbb{P}(\{(\zeta'_n)_{n\geq i+k+l} \neq (\zeta_n)_{n\geq i+k+l}\}) + \mathbb{P}(\{(\zeta''_n)_{n\leq i+l} \neq (\zeta_n)_{n\leq i+l}\})\Big)^{\frac{\delta}{2(2+\delta)}} \end{aligned}$$

$$
\begin{aligned}
&\le 2M^{\frac{1}{2+\delta}}(2\beta_k^{\frac{\delta}{2(2+\delta)}}) \\
&\le C\beta_k^{\frac{\delta}{2(2+\delta)}},
\end{aligned}
$$

where we used property (2) of the copied series $(\zeta'_n)_{n\in\mathbb{Z}}$, $(\zeta''_n)_{n\in\mathbb{Z}}$ for the second to last inequality. For (3.B), we split up again:

$$
\begin{aligned}
(3.B) \le\ & \mathbb{E}\Big[\|h_2(X_i, X_{i+k+2l}) - h_2(X''_i, X'_{i+k+2l})\|^2 \mathbf{1}_B \\
&\qquad\qquad \cdot \mathbf{1}_{\{\|X_i - X''_i\| \le 2\epsilon,\, \|X_{i+k+2l} - X'_{i+k+2l}\| \le 2\epsilon\}}\Big]^{\frac{1}{2}} \\
&+ \mathbb{E}\Big[\|h_2(X_i, X_{i+k+2l}) - h_2(X''_i, X'_{i+k+2l})\|^2 \mathbf{1}_B \\
&\qquad\qquad \cdot \mathbf{1}_{\{\|X_i - X''_i\| > 2\epsilon \text{ or } \|X_{i+k+2l} - X'_{i+k+2l}\| > 2\epsilon\}}\Big]^{\frac{1}{2}}
\end{aligned}
$$

For the first summand, we use the variation condition. For the second, notice that on B:

$$
\|X_i - X''_i\| \le \|X_i - X_{i,l}\| + \|X_{i,l} - X''_i\| = \|X_i - X_{i,l}\| + \|X''_{i,l} - X''_i\|
$$

and

$$
\begin{aligned}
\|X_{i+k+2l} - X'_{i+k+2l}\| &\le \|X_{i+k+2l} - X_{i+k+2l,l}\| + \|X_{i+k+2l,l} - X'_{i+k+2l}\| \\
&= \|X_{i+k+2l} - X_{i+k+2l,l}\| + \|X'_{i+k+2l,l} - X'_{i+k+2l}\|.
\end{aligned}
$$

So, by the moment assumptions and Hölder's inequality

$$
\begin{aligned}
(3.B) &\le \sqrt{L2\epsilon} \\
&\quad + \mathbb{E}[\|h_2(X_i, X_{i+k+2l}) - h_2(X''_i, X'_{i+k+2l})\|^2 \mathbf{1}_{\{\|X_i - X_{i,l}\| > \epsilon\}}]^{\frac{1}{2}} \\
&\quad + \mathbb{E}[\|h_2(X_i, X_{i+k+2l}) - h_2(X''_i, X'_{i+k+2l})\|^2 \mathbf{1}_{\{\|X''_i - X''_{i,l}\| > \epsilon\}}]^{\frac{1}{2}} \\
&\quad + \mathbb{E}[\|h_2(X_i, X_{i+k+2l}) - h_2(X''_i, X'_{i+k+2l})\|^2 \mathbf{1}_{\{\|X_{i+k+2l} - X_{i+k+2l,l}\| > \epsilon\}}]^{\frac{1}{2}} \\
&\quad + \mathbb{E}[\|h_2(X_i, X_{i+k+2l}) - h_2(X''_i, X'_{i+k+2l})\|^2 \mathbf{1}_{\{\|X'_{i+k+2l} - X'_{i+k+2l,l}\| > \epsilon\}}]^{\frac{1}{2}} \\
&\le \sqrt{L2\epsilon} + 4 \cdot 2M^{\frac{1}{2+\delta}} (\mathbb{P}(\|X_i - X_{i,l}\| > \epsilon))^{\frac{\delta}{2(2+\delta)}} \\
&\le \sqrt{L2\epsilon} + 4 \cdot 2M^{\frac{1}{2+\delta}} (a_l \Phi(\epsilon))^{\frac{\delta}{2(2+\delta)}} \\
&\le C\left(\sqrt{\epsilon} + (a_l\Phi(\epsilon))^{\frac{\delta}{2(2+\delta)}}\right),
\end{aligned}
$$

using that $(X_n)_{n\in\mathbb{Z}}$ is $\mathbb{P}$-NED. Combining the results for (3.A) and (3.B) we get

$$
(3) \le (3.A) + (3.B) \le C\left(\beta_k^{\frac{\delta}{2(2+\delta)}} + \sqrt{\epsilon} + (a_l\Phi(\epsilon))^{\frac{\delta}{2(2+\delta)}}\right).
$$

We can now look at (4). Again, split the term into two summands (similar as for (3)), and use the variation condition for the first and Hölder's inequality for the second summand:

$$
(4) = \mathbb{E}[\|h_2(X''_i, X'_{i+k+2l}) - h_2(X''_{i,l}, X'_{i+k+2l,l})\|^2]^{\frac{1}{2}}
$$

$$
\begin{aligned}
&\leq \mathbb{E}\Big[\|h_2(X_i'', X_{i+k+2l}') - h_2(X_{i,l}'', X_{i+k+2l,l}')\|^2 \\
&\qquad\qquad \cdot \mathbf{1}_{\{\|X_i''-X_{i,l}''\|\leq\epsilon,\, \|X_{i+k+2l}'-X_{i+k+2l,l}'\|\leq\epsilon\}}\Big]^{\frac{1}{2}} \\
&\quad + \mathbb{E}\Big[\|h_2(X_i'', X_{i+k+2l}') - h_2(X_{i,l}'', X_{i+k+2l,l}')\|^2 \\
&\qquad\qquad \cdot \mathbf{1}_{\{\|X_i''-X_{i,l}''\|>\epsilon \text{ or } \|X_{i+k+2l}'-X_{i+k+2l,l}'\|>\epsilon\}}\Big]^{\frac{1}{2}} \\
&\leq \sqrt{L\epsilon} + \Big(\mathbb{E}[\|h_2(X_i'', X_{i+k+2l}')\|^{2+\delta}]^{\frac{1}{2+\delta}} + \mathbb{E}[\|h_2(X_{i,l}'', X_{i+k+2l,l}')\|^{2+\delta}]^{\frac{1}{2+\delta}}\Big) \\
&\qquad \cdot \Big(\mathbb{P}(\|X_i'' - X_{i,l}''\| > \epsilon) + \mathbb{P}(\|X_{i+k+2l}' - X_{i+k+2l,l}'\| > \epsilon)\Big)^{\frac{\delta}{2(2+\delta)}} \\
&\leq \sqrt{L\epsilon} + 2M^{\frac{1}{2+\delta}}(2a_l\Phi(\epsilon))^{\frac{\delta}{2(2+\delta)}} \\
&\leq C\left(\sqrt{\epsilon} + (a_l\Phi(\epsilon))^{\frac{\delta}{2(2+\delta)}}\right),
\end{aligned}
$$

since $(X_n)_{n\in\mathbb{Z}}$ is $\mathbb{P}$-NED. Lastly, split up (5) as well and note that since on B it is $X_{i+k+2l,l} = X_{i+k+2l,l}'$ and $X_{i,l} = X_{i,l}''$, the second summand equals zero. For the first summand, use Hölder's inequality again:

$$
\begin{aligned}
(5) &= \mathbb{E}[\|h_2(X_{i,l}'', X_{i+k+2l,l}') - h_2(X_{i,l}, X_{i+k+2l,l})\|^2]^{\frac{1}{2}} \\
&\leq \mathbb{E}[\|h_2(X_{i,l}'', X_{i+k+2l,l}') - h_2(X_{i,l}, X_{i+k+2l,l})\|^2 \mathbf{1}_{B^c}]^{\frac{1}{2}} \\
&\quad + \mathbb{E}[\|h_2(X_{i,l}'', X_{i+k+2l,l}') - h_2(X_{i,l}, X_{i+k+2l,l})\|^2 \mathbf{1}_{B}]^{\frac{1}{2}} \\
&\leq 2M^{\frac{1}{2+\delta}}\Big(\mathbb{P}(\{(\zeta_n')_{n\geq i+k+l} \neq (\zeta_n)_{n\geq i+k+l}\}) + \mathbb{P}(\{(\zeta_n'')_{n\leq i+l} \neq (\zeta_n)_{n\leq i+l}\})\Big)^{\frac{\delta}{2(2+\delta)}} \\
&\overset{(2)}{\leq} 2M^{\frac{1}{2+\delta}}(2\beta_k)^{\frac{\delta}{2(2+\delta)}} \leq C\beta_k^{\frac{\delta}{2(2+\delta)}}.
\end{aligned}
$$

Finally, we can put everything together, which completes the proof:

$$
\begin{aligned}
&\mathbb{E}[\|h_2(X_i, X_{i+k+2l}) - h_2(X_{i,l}, X_{i+k+2l,l})\|^2]^{\frac{1}{2}} \leq (3) + (4) + (5) \\
&\leq C\left(\beta_k^{\frac{\delta}{2(2+\delta)}} + \sqrt{\epsilon} + (a_l\Phi(\epsilon))^{\frac{\delta}{2(2+\delta)}}\right) + C\left(\sqrt{\epsilon} + (a_l\Phi(\epsilon))^{\frac{\delta}{2(2+\delta)}}\right) + C\beta_k^{\frac{\delta}{2(2+\delta)}} \\
&\leq C\left(\sqrt{\epsilon} + \beta_k^{\frac{\delta}{2(2+\delta)}} + (a_l\Phi(\epsilon))^{\frac{\delta}{2(2+\delta)}}\right).
\end{aligned}
$$

□

Lemma 2.4. *Under the assumptions of Theorem 2.1 it holds for any $n_1 < n_2 < n_3 < n_4 \leq n$ and $l = \left\lfloor n^{\frac{3}{16}} \right\rfloor$:*

$$
\mathbb{E}\Bigg[\Big(\sum_{n_1\leq i\leq n_2}\sum_{n_3\leq j\leq n_4}\|h_2(X_i, X_j) - h_2(X_{i,l}, X_{j,l})\|\Big)^2\Bigg]^{\frac{1}{2}} \leq C(n_4 - n_3)n^{\frac{1}{4}}.
$$

Proof. The important step of the proof is to bound the left hand side expectation from above by a sum of $\mathbb{E}[\|h_2(X_i, X_j) - h_2(X_{i,l}, X_{j,l})\|^2]^{1/2}$ terms. We can then use

Lemma 2.3 to achieve the stated approximation. First note that

$$\mathbb{E}\Big[\Big(\sum_{n_1\le i\le n_2}\sum_{n_3\le j\le n_4}\|h_2(X_i,X_j)-h_2(X_{i,l},X_{j,l})\|\Big)^2\Big]^{\frac{1}{2}}$$
$$\le \mathbb{E}\Big[\Big(\sum_{1\le i<j}\sum_{n_3\le j\le n_4}\|h_2(X_i,X_j)-h_2(X_{i,l},X_{j,l})\|\Big)^2\Big]^{\frac{1}{2}}.$$

For any fixed $n_3 \le j \le n_4$ it is

$$\mathbb{E}\Big[\sum_{1\le i<j}\|h_2(X_i,X_j)\|\Big]=\mathbb{E}\Big[\sum_{k=1}^{j-1}\|h_2(X_{j-k},X_j)\|\Big]\le \mathbb{E}\Big[\sum_{k=1}^{n_4}\|h_2(X_{j-k},X_j)\|\Big].$$

And for j there are at most (n_4-n_3) possibilities. So,

$$\mathbb{E}\Big[\sum_{n_3\le j\le n_4}\sum_{1\le i<j}\|h_2(X_i,X_j)\|\Big]\le (n_4-n_3)\mathbb{E}\Big[\sum_{k=1}^{n_4}\|h_2(X_{j-k},X_j)\|\Big].$$

The analogous holds for $h_2(X_{i,l},X_{j,l})$. Thus,

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\sum_{1\le i<j}\sum_{n_3\le j\le n_4}\|h_2(X_i,X_j)-h_2(X_{i,l},X_{j,l})\|\Big)^2\Big]^{\frac{1}{2}}\\
\le\ &\sum_{n_3\le j\le n_4}\sum_{1\le i<j}\mathbb{E}[\|h_2(X_i,X_j)-h_2(X_{i,l},X_{j,l})\|^2]^{\frac{1}{2}}\\
\le\ &(n_4-n_3)\sum_{k=1}^{n_4}\mathbb{E}[\|h_2(X_{j-k},X_j)-h_2(X_{j-k,l},X_{j,l})\|^2]^{\frac{1}{2}}\\
\le\ &(n_4-n_3)\sum_{k=1}^{n_4}C\left(\sqrt{\epsilon}+\beta_{k-2l}^{\frac{\delta}{2(2+\delta)}}+(a_l\Phi(\epsilon))^{\frac{\delta}{2(2+\delta)}}\right) \qquad (6)
\end{aligned}$$

by Lemma 2.3.
Now set $\epsilon = l^{-8\frac{3+\delta}{\delta}}$ and define $\beta_k=1$ if $k<0$. Then by the assumptions on the approximation constants and the mixing coefficients

$$\begin{aligned}
(6) &= C(n_4-n_3)\sum_{k=1}^{n_4}\left(l^{-8\frac{3+\delta}{\delta}\frac{1}{2}}+\beta_{k-2l}^{\frac{\delta}{2(2+\delta)}}+(a_l\Phi(l^{-8\frac{3+\delta}{\delta}}))^{\frac{\delta}{2(2+\delta)}}\right)\\
&\le C(n_4-n_3)\sum_{k=1}^{n_4}\left(l^{-4\frac{3+\delta}{\delta}}+\beta_{k-2l}^{\frac{\delta}{2(2+\delta)}}+l^{-4\frac{3+\delta}{\delta}}\right)\\
&\le C(n_4-n_3)\Big(\sum_{k=1}^{n_4}l^{-4}+\sum_{k=1}^{2l-1}\underbrace{\beta_{k-2l}^{\frac{\delta}{4+\delta}}}_{=1}+\sum_{k=2l}^{n_4}\beta_{k-2l}^{\frac{\delta}{4+\delta}}\Big)\\
&\le C(n_4-n_3)\Big(n_4l^{-4}+2l+\underbrace{\sum_{k=2l}^{n_4}(k-2l)^2\beta_{k-2l}^{\frac{\delta}{4+\delta}}}_{<\infty}\Big)
\end{aligned}$$

$$\leq C(n_4 - n_3)\big(nl^{-4} + 2l\big)$$
$$\leq C(n_4 - n_3)n^{\frac{1}{4}}.$$

So the statement of the lemma is proven. □

Lemma 2.5. *Under the assumptions of Theorem 2.1, it holds for any $n_1 < n_2 < n_3 < n_4 \leq n$ and $l = \left\lfloor n^{\frac{3}{16}} \right\rfloor$:*

$$\mathbb{E}\Bigg[\Big(\sum_{n_1 \leq i \leq n_2} \sum_{n_3 \leq j \leq n_4} \|h_{2,l}(X_{i,l}, X_{j,l}) - h_2(X_{i,l}, X_{j,l})\|\Big)^2\Bigg]^{\frac{1}{2}} \leq C(n_4 - n_3)n^{\frac{1}{4}}.$$

where $h_{2,l}(x, y) = h(x, y) - \mathbb{E}[h(x, \tilde{X}_{0,l})] - \mathbb{E}[h(\tilde{X}_{0,l}, y)]$ $\forall\ x, y \in H$ and $\tilde{X}_{n,k} = f_k(\tilde{\zeta}_{n-k}, ..., \tilde{\zeta}_{n+k})$, where $(\tilde{\zeta}_n)_{n\in\mathbb{Z}}$ is an independent copy of $(\zeta_n)_{n\in\mathbb{Z}}$.

Remark 2.5. *The "truncated" function $h_{2,l}$ is degenerated as well, meaning that for all $i \in \mathbb{N}$*

$$\mathbb{E}[h_{2,l}(X_{i,l}, y)] = \mathbb{E}_X[h(X_{i,l}, y)] - \underbrace{\mathbb{E}_X[\mathbb{E}_{\tilde{X}}[h(X_{i,l}, \tilde{X}_{i,l})]]}_{=0 \text{ by antisymmetry of } h} - \mathbb{E}_X[\mathbb{E}_{\tilde{X}}[h(\tilde{X}_{i,l}, y)]]$$
$$= \mathbb{E}_X[h(X_{i,l}, y)] - \mathbb{E}_{\tilde{X}}[h(\tilde{X}_{i,l}, y)]$$
$$= 0,$$

since $\tilde{X}_{i,l}$ is a copy of $X_{i,l}$. And similarly it is $\mathbb{E}[h_{2,l}(y, X_{i,l})] = 0$.

Proof of Lemma 2.5. For $(\tilde{\zeta}_n)_{n\in\mathbb{Z}}$ an independent copy of $(\zeta_n)_{n\in\mathbb{Z}}$, write $\tilde{X}_n = f((\tilde{\zeta}_{n+k})_{k\in\mathbb{Z}})$. So $(\tilde{X}_n)_{n\in\mathbb{Z}}$ is an independent copy of $(X_n)_{n\in\mathbb{Z}}$. We will use Hoeffding's decomposition and rewrite h_2 as $h_2(x, y) = h(x, y) - \mathbb{E}[h(x, \tilde{X}_j)] - \mathbb{E}[h(\tilde{X}_i, y)]$ and similarly for $h_{2,l}$. By doing so, we obtain

$$\begin{aligned}
&\mathbb{E}[\|h_{2,l}(X_{i,l}, X_{j,l}) - h_2(X_{i,l}, X_{j,l})\|^2]^{\frac{1}{2}} \\
=\ &\mathbb{E}\Big[\|\ h(X_{i,l}, X_{j,l}) - \mathbb{E}_{\tilde{X}}[h(X_{i,l}, \tilde{X}_{j,l})] - \mathbb{E}_{\tilde{X}}[h(\tilde{X}_{i,l}, X_{j,l})] \\
&\quad - h(X_{i,l}, X_{j,l}) + \mathbb{E}_{\tilde{X}}[h(X_{i,l}, \tilde{X}_j)] + \mathbb{E}_{\tilde{X}}[h(\tilde{X}_i, X_{j,l})]\|^2\Big]^{\frac{1}{2}} \\
\leq\ &\mathbb{E}\big[\|h(X_{i,l}, \tilde{X}_{j,l}) - h(X_{i,l}, \tilde{X}_j)\|^2\big]^{\frac{1}{2}} && (7)\\
&\quad + \mathbb{E}\big[\|h(\tilde{X}_{i,l}, X_{j,l}) - h(\tilde{X}_i, X_{j,l})\|^2\big]^{\frac{1}{2}}. && (8)
\end{aligned}$$

We bound the two terms separately, starting with (8):

$$\begin{aligned}
&\mathbb{E}\big[\|h(\tilde{X}_{i,l}, X_{j,l}) - h(\tilde{X}_i, X_{j,l})\|^2\big]^{\frac{1}{2}} \\
\leq\ &\mathbb{E}\big[\|h(\tilde{X}_{i,l}, X_{j,l}) - h(\tilde{X}_i, X_j)\|^2\big]^{\frac{1}{2}} && (8.A)\\
&\quad + \mathbb{E}\big[\|h(\tilde{X}_i, X_{j,l}) - h(\tilde{X}_i, X_j)\|^2\big]^{\frac{1}{2}}. && (8.B)
\end{aligned}$$

Now, decompose the first summand and use the variation condition and Hölder's inequality:

$$
\begin{aligned}
(8.A) &= \mathbb{E}\left[\|h(\tilde{X}_{i,l}, X_{j,l}) - h(\tilde{X}_i, X_j)\|^2 \mathbf{1}_{\{\|\tilde{X}_i - \tilde{X}_{i,l}\| \leq \epsilon,\ \|X_j - X_{j,l}\| \leq \epsilon\}}\right]^{\frac{1}{2}} \\
&\quad + \mathbb{E}\left[\|h(\tilde{X}_{i,l}, X_{j,l}) - h(\tilde{X}_i, X_j)\|^2 \mathbf{1}_{\{\|\tilde{X}_i - \tilde{X}_{i,l}\| > \epsilon \text{ or } \|X_j - X_{j,l}\| > \epsilon\}}\right]^{\frac{1}{2}} \\
&\leq \sqrt{L\epsilon} + \mathbb{E}\left[\|h(\tilde{X}_{i,l}, X_{j,l}) - h(\tilde{X}_i, X_j)\|^{2+\delta}\right]^{\frac{1}{2+\delta}} \\
&\qquad \cdot \Big(\mathbb{P}(\|\tilde{X}_i - \tilde{X}_{i,l}\| > \epsilon) + \mathbb{P}(\|X_j - X_{j,l}\| > \epsilon)\Big)^{\frac{\delta}{2(2+\delta)}} \\
&\leq \sqrt{L\epsilon} + 2M^{\frac{1}{2+\delta}} (2a_l \Phi(\epsilon))^{\frac{\delta}{2(2+\delta)}} \\
&\leq C\left(\sqrt{\epsilon} + (2a_l\Phi(\epsilon))^{\frac{\delta}{2(2+\delta)}}\right)
\end{aligned}
$$

by the moment and $\mathbb{P}$-NED assumptions.
For $(8.B)$ we use similar arguments:

$$
\begin{aligned}
(8.B) &\leq \mathbb{E}\left[\|h(\tilde{X}_i, X_{j,l}) - h(\tilde{X}_i, X_j)\|^2 \mathbf{1}_{\{\|X_j - X_{j,l}\| > \epsilon\}}\right]^{\frac{1}{2}} \\
&\quad + \mathbb{E}\left[\|h(\tilde{X}_i, X_{j,l}) - h(\tilde{X}_i, X_j)\|^2 \mathbf{1}_{\{\|X_j - X_{j,l}\| \leq \epsilon\}}\right]^{\frac{1}{2}} \\
&\leq \mathbb{E}\left[\|h(\tilde{X}_i, X_{j,l}) - h(\tilde{X}_i, X_j)\|^{2+\delta}\right]^{\frac{1}{2+\delta}} \cdot \mathbb{P}(\|X_j - X_{j,l}\| > \epsilon)^{\frac{\delta}{2(2+\delta)}} + \sqrt{L\epsilon} \\
&\leq 2M^{\frac{1}{2+\delta}} (a_l\Phi(\epsilon))^{\frac{\delta}{2(2+\delta)}} + \sqrt{L\epsilon} \\
&\leq C\left(\sqrt{\epsilon} + (a_l\Phi(\epsilon))^{\frac{\delta}{2(2+\delta)}}\right).
\end{aligned}
$$

Putting these two terms together, we get

$$
(8) \leq C\left(\sqrt{\epsilon} + (a_l\Phi(\epsilon))^{\frac{\delta}{2(2+\delta)}}\right).
$$

Bounding (7) works completely analogous, just with i and j interchanged, so

$$
(7) \leq \left(\sqrt{\epsilon} + (a_l\Phi(\epsilon))^{\frac{\delta}{2(2+\delta)}}\right).
$$

All together this yields

$$
\mathbb{E}[\|h_{2,l}(X_{i,l}, X_{j,l}) - h_2(X_{i,l}, X_{j,l})\|^2]^{\frac{1}{2}} \leq (7) + (8) \leq C\left(\sqrt{\epsilon} + (a_l\Phi(\epsilon))^{\frac{\delta}{2(2+\delta)}}\right).
$$

So we finally get that

$$
\begin{aligned}
&\mathbb{E}\left[\Big(\sum_{n_1 \leq i \leq n_2} \sum_{n_3 \leq j \leq n_4} \|h_{2,l}(X_{i,l}, X_{j,l}) - h_2(X_{i,l}, X_{j,l})\|\Big)^2\right]^{\frac{1}{2}} \\
&\leq \mathbb{E}\left[\Big(\sum_{1 \leq i < j} \sum_{n_3 \leq j \leq n_4} \|h_{2,l}(X_{i,l}, X_{j,l}) - h_2(X_{i,l}, X_{j,l})\|\Big)^2\right]^{\frac{1}{2}}
\end{aligned}
$$

$$
\begin{aligned}
&\leq \sum_{1\leq i<j} \sum_{n_3\leq j\leq n_4} \mathbb{E}\big[\|h_{2,l}(X_{i,l},X_{j,l})-h_2(X_{i,l},X_{j,l})\|^2\big]^{\frac{1}{2}} \\
&\leq \sum_{1\leq i<j} \sum_{n_3\leq j\leq n_4} C\left(\sqrt{\epsilon}+(a_l\Phi(\epsilon))^{\frac{\delta}{2(2+\delta)}}\right) \\
&\leq C(n_4-n_3)\sum_{k=1}^{n_4}\left(\sqrt{\epsilon}+(a_l\Phi(\epsilon))^{\frac{\delta}{2(2+\delta)}}\right) \\
&\leq C(n_4-n_3)n^{\frac{1}{4}},
\end{aligned}
$$

where the last line is achieved by setting $\epsilon = l^{-8\frac{3+\delta}{\delta}}$ and similar calculations as in Lemma 2.4. □

Remark 2.6. *If the kernel h is not antisymmetric, we get an additional expectation in h_2 and $h_{2,l}$ respectively. More precisely we get*

$$h_2(X_{i,l},X_{j,l}) = h(X_{i,l},X_{j,l}) - \mathbb{E}_{\tilde{X}}[h(X_{i,l},\tilde{X}_j)] - \mathbb{E}_{\tilde{X}}[h(\tilde{X}_i,X_{j,l})] + \mathbb{E}_{X,\tilde{X}}[h(\tilde{X}_i,X_j)]$$

and

$$h_{2,l}(X_{i,l},X_{j,l}) = h(X_{i,l},X_{j,l}) - \mathbb{E}_{\tilde{X}}[h(X_{i,l},\tilde{X}_{j,l})] - \mathbb{E}_{\tilde{X}}[h(\tilde{X}_{i,l},X_{j,l})] + \mathbb{E}_{X,\tilde{X}}[h(\tilde{X}_{i,l},X_{j,l})].$$

This means we have to bound the additional term $\mathbb{E}[\|h(\tilde{X}_{i,l},X_{j,l})-h(\tilde{X}_i,X_j)\|^2]^{\frac{1}{2}}$ in Lemma 2.5. But notice that this equals the term (8.A), which we have bounded above.

Lemma 2.6. *Under the assumptions of Theorem 2.1, it holds for any $n_1,n_2,n_3,n_4 \leq n$ with $n_1<n_2$, $n_3<n_4$ and $l=\left\lfloor n^{\frac{3}{16}}\right\rfloor$:*

$$\mathbb{E}\left[\left(\|\sum_{n_1\leq i\leq n_2}\sum_{n_3\leq j\leq n_4} h_{2,l}(X_{i,l},X_{j,l})\|\right)^2\right] \leq C(n_4-n_3)(n_2-n_1)n^{\frac{3}{8}}.$$

For the definition of $h_{2,l}$, see Lemma 2.5.

Proof. In this proof, we will use Lemma 1 Yoshihara [1976], which is the following: *Let $g(x_1,...,x_k)$ be a Borel function. For any $0\leq j\leq k-1$ with*

$$\mathbb{E}[|g(X_{I,l},X'_{I^C,l})|^{1+\tilde{\delta}}] \leq M \tag{$\Diamond$}$$

for some $\tilde{\delta}>0$, where $I=\{i_1,...,i_j\}$, $I^C=\{i_{j+1},...,i_k\}$ and X' an independent copy of X, it holds that

$$\left|\mathbb{E}[g(X_{i_1,l},...,X_{i_k,l})]-\mathbb{E}[g(X_{I,l},X'_{I^C,l})]\right| \leq 4M^{1/(1+\tilde{\delta})}\beta^{\tilde{\delta}/(1+\tilde{\delta})}_{(i_{j+1}-i_j)-2l}. \tag{Y}$$

Now, for the proof of the lemma, first observe that we can rewrite the squared norm as the scalar product and thus:

$$
\begin{aligned}
&\mathbb{E}\Big[\| \sum_{n_1\le i\le n_2}\sum_{n_3\le j\le n_4} h_{2,l}(X_{i,l},X_{j,l})\|^2\Big] \\
&= \mathbb{E}\big[\langle \sum_{n_1\le i\le n_2}\sum_{n_3\le j\le n_4} h_{2,l}(X_{i,l},X_{j,l}), \sum_{n_1\le i\le n_2}\sum_{n_3\le j\le n_4} h_{2,l}(X_{i,l},X_{j,l})\rangle\big] \\
&= \underset{(i_1\neq i_2)\text{ or }(j_1\neq j_2)\text{ or both}}{\sum_{n_1\le i_1\le n_2}\sum_{n_3\le j_1\le n_4}\sum_{n_1\le i_2\le n_2}\sum_{n_3\le j_2\le n_4}} \mathbb{E}[\langle h_{2,l}(X_{i_1,l},X_{j_1,l}), h_{2,l}(X_{i_2,l},X_{j_2,l})\rangle] \qquad (9)\\
&\quad + \sum_{n_1\le i\le n_2}\sum_{n_3\le j\le n_4} \mathbb{E}[\langle h_{2,l}(X_{i,l},X_{j,l}), h_{2,l}(X_{i,l},X_{j,l})\rangle]. \qquad (10)
\end{aligned}
$$

We know by the uniform moments under approximation that (10) is bounded by the following:

$$
\begin{aligned}
(10) &= \sum_{n_1\le i\le n_2}\sum_{n_3\le j\le n_4} \mathbb{E}\big[\|h_{2,l}(X_{i,l},X_{j,l})\|^2\big] \le (n_2-n_1)(n_4-n_3)M \\
&\le M(n_2-n_1)(n_4-n_3)n^{\frac{3}{8}}.
\end{aligned}
$$

For (9) we use the above mentioned lemma of Yoshihara [1976]. Note that by the double summation, we have three different cases to analyse: $(i_1\neq i_2)$ or $(j_1\neq j_2)$ or both. Universal, let $m=\max(j_1-i_1, j_2-i_2)$, first assume that $m=j_1-i_1$ and let $\tilde{\delta}=\delta/2>0$.

<u>First case:</u> $i_1\neq i_2$ and $j_1\neq j_2$.

Define the function $g(x_1,x_2,x_3,x_4) := \langle h_{2,l}(x_1,x_2), h_{2,l}(x_3,x_4)\rangle$ and check that $(\Diamond)$ holds true for $I=\{i_1\}$ and $I^C=\{j_1,i_2,j_2\}$:

$$
\begin{aligned}
&\mathbb{E}\big[|g(X_{i_1,l},X'_{j_1,l},X'_{i_2,l},X'_{j_2,l})|^{1+\tilde{\delta}}\big] \le \mathbb{E}\big[\|h_{2,l}(X_{i_1,l},X'_{j_1,l})\|^{1+\tilde{\delta}}\|h_{2,l}(X'_{i_2,l},X'_{j_2,l})\|^{1+\tilde{\delta}}\big] \\
&\le \mathbb{E}\big[\|h_{2,l}(X_{i_1,l},X'_{j_1,l})\|^{2(1+\tilde{\delta})}\big]^{1/2}\mathbb{E}\big[\|h_{2,l}(X'_{i_2,l},X'_{j_2,l})\|^{2(1+\tilde{\delta})}\big]^{1/2} \le M
\end{aligned}
$$

by the moment assumptions and $\delta=\tilde{\delta}/2$. Here, we first use the Cauchy-Schwarz inequality and then Hölder's inequality. Now (Y) states that

$$
\big|\mathbb{E}[g(X_{i_1,l},X_{j_1,l},X_{i_2,l},X_{j_2,l})] - \mathbb{E}[g(X_{i_1,l},X'_{j_1,l},X'_{i_2,l},X'_{j_2,l})]\big| \le C\beta_{m-2l}^{\tilde{\delta}/(1+\tilde{\delta})}. \qquad (11)
$$

The second expectation equals 0, which can be seen by using the law of the iterated expectation:

$$
\begin{aligned}
&\mathbb{E}[g(X_{i_1,l},X'_{j_1,l},X'_{i_2,l},X'_{j_2,l})] = \mathbb{E}\big[\mathbb{E}[g(X_{i_1,l},X'_{j_1,l},X'_{i_2,l},X'_{j_2,l})|X'_{j_1,l},X'_{i_2,l},X'_{j_2,l}]\big] \\
&= \mathbb{E}\big[\mathbb{E}[\langle h_{2,l}(X_{i_1,l},X'_{j_1,l}), h_{2,l}(X'_{i_2,l},X'_{j_2,l})\rangle|X'_{j_1,l},X'_{i_2,l},X'_{j_2,l}]\big] \\
&= \mathbb{E}\big[\langle \mathbb{E}[h_{2,l}(X_{i_1,l},X'_{j_1,l})|X'_{j_1,l},X'_{i_2,l},X'_{j_2,l}], h_{2,l}(X'_{i_2,l},X'_{j_2,l})\rangle\big] \qquad (12)
\end{aligned}
$$

since $h_{2,l}(X'_{i_2,l},X'_{j_2,l})$ is measurable with respect to the inner (conditional) expectation. In general it holds for random variables X,Y that $\mathbb{E}[\langle Y,X\rangle|\mathfrak{B}] = \langle Y,\mathbb{E}[X|\mathfrak{B}]\rangle$ if Y is measurable with respect to $\mathfrak{B}$. So,

$$
(12) = \mathbb{E}\big[\underbrace{\langle \mathbb{E}[h_{2,l}(X_{i_1,l},X'_{j_1,l})|X'_{j_1,l},X'_{i_2,l},X'_{j_2,l}]}_{=0\text{ because }h_{2,l}\text{ is degenerated}}, h_{2,l}(X'_{i_2,l},X'_{j_2,l})\rangle\big] = 0.
$$

Plugging this into (11), we get that

$$\mathbb{E}[\langle h_{2,l}(X_{i_1,l}, X_{j_1,l}), h_{2,l}(X_{i_2,l}, X_{j_2,l})\rangle] \le \big|\mathbb{E}[g(X_{i_1,l}, X_{j_1,l}, X_{i_2,l}, X_{j_2,l})]\big| \le C\beta_{m-2l}^{\tilde{\delta}/(1+\tilde{\delta})}.$$

We repeat the above argumentation for the other two cases:
<u>Second case:</u> $i_1 \neq i_2$ but $j_1 = j_2$.
Define the function $g(x_1, x_2, x_3) := \langle h_{2,l}(x_1, x_2), h_{2,l}(x_3, x_2)\rangle$ and check that ($\Diamond$) holds true for $I = \{i_1\}$ and $I^C = \{j_1, i_2\}$:

$$\begin{aligned} &\mathbb{E}\big[|g(X_{i_1,l}, X'_{j_1,l}, X'_{j_2,l})|^{1+\tilde{\delta}}\big] \\ \le\ &\mathbb{E}\big[\|h_{2,l}(X_{i_1,l}, X'_{j_1,l})\|^{2(1+\tilde{\delta})}\big]^{1/2}\mathbb{E}\big[\|h_{2,l}(X'_{i_2,l}, X'_{j_1,l})\|^{2(1+\tilde{\delta})}\big]^{1/2} \le M. \end{aligned}$$

Here, (Y) states that

$$\big|\mathbb{E}[g(X_{i_1,l}, X_{j_1,l}, X_{i_2,l})] - \mathbb{E}[g(X_{i_1,l}, X'_{j_1,l}, X'_{i_2,l})]\big| \le C\beta_{m-2l}^{\tilde{\delta}/(1+\tilde{\delta})}. \tag{13}$$

Again, the second expectation equals zero:

$$\begin{aligned} &\mathbb{E}[g(X_{i_1,l}, X'_{j_1,l}, X'_{i_2,l})] \\ =\ &\mathbb{E}\big[\mathbb{E}[\langle h_{2,l}(X_{i_1,l}, X'_{j_1,l}), h_{2,l}(X'_{i_2,l}, X'_{j_1,l})\rangle | X'_{i_2,l}, X'_{j_1,l}]\big] \\ =\ &\mathbb{E}\big[\langle \underbrace{\mathbb{E}[h_{2,l}(X_{i_1,l}, X'_{j_1,l}) | X'_{i_2,l}, X'_{j_1,l}]}_{=0}, h_{2,l}(X'_{i_2,l}, X'_{j_1,l})\rangle\big] \\ =\ &0. \end{aligned}$$

Plugging this into (13), we get that

$$\mathbb{E}[\langle h_{2,l}(X_{i_1,l}, X_{j_1,l}), h_{2,l}(X_{i_2,l}, X_{j_1,l})\rangle] \le |\mathbb{E}[g(X_{i_1,l}, X_{j_1,l}, X_{i_2,l})]| \le C\beta_{m-2l}^{\tilde{\delta}/(1+\tilde{\delta})}.$$

<u>Third case:</u> $j_1 \neq j_2$ but $i_1 = i_2$.
Define the function $g(x_1, x_2, x_3) := \langle h_{2,l}(x_1, x_2), h_{2,l}(x_1, x_3)\rangle$. Checking that ($\Diamond$) holds true for $I = \{i_1\}$ and $I^C = \{j_1, j_2\}$ works completely similar to the second case. Noting that we have to condition on $X_{i_1,l}, X'_{j_2,l}$ in this case, yields:

$$\mathbb{E}[\langle h_{2,l}(X_{i_1,l}, X_{j_1,l}), h_{2,l}(X_{i_1,l}, X_{j_2,l})\rangle] \le \big|\mathbb{E}[g(X_{i_1,l}, X_{j_1,l}, X_{j_2,l})]\big| \le C\beta_{m-2l}^{\tilde{\delta}/(1+\tilde{\delta})}.$$

We can conclude for the quadratic term

$$\begin{aligned} &\mathbb{E}\Big[\big\|\sum_{n_1 \le i \le n_2}\sum_{n_3 \le j \le n_4} h_{2,l}(X_{i,l}, X_{j,l})\big\|^2\Big] \\ =\ &\sum_{n_1 \le i_1 \le n_2}\sum_{n_3 \le j_1 \le n_4}\underset{(i_1 \neq i_2)\text{ or }(j_1 \neq j_2)\text{ or both}}{\sum_{n_1 \le i_2 \le n_2}\sum_{n_3 \le j_2 \le n_4}} C\beta_{m-2l}^{\tilde{\delta}/(1+\tilde{\delta})} + (n_4 - n_3)(n_2 - n_1)M. \end{aligned} \tag{14}$$

For a fixed m we have the following possibilities to choose:
Since we assumed $m = j_1 - i_1$, there are

- at most $n_2 - n_1$ possibilities for i_1, so only 1 possibility for j_1.
- at most $(n_4 - n_3)$ possibilities for j_2, so at most m possibilities for i_2, since by the definition of m the value $j_2 - i_2$ is smaller (or equal) than m.

So, recalling that $\delta = \tilde{\delta}/2$, we have for $l > \frac{1}{2}$:

$$\sum_{n_1 \le i_1 \le n_2} \sum_{n_3 \le j_1 \le n_4} \sum_{\substack{n_1 \le i_2 \le n_2, \\ (i_1 \ne i_2) \text{ or } (j_1 \ne j_2) \text{ or both}}} \sum_{n_3 \le j_2 \le n_4} C\beta_{m-2l}^{\tilde{\delta}/(1+\tilde{\delta})}$$

$$\begin{aligned}
&\le C(n_4 - n_3)(n_2 - n_1) \sum_{m=1}^{n} m\beta_{m-2l}^{\frac{\delta}{2+\delta}} \\
&= C(n_4 - n_3)(n_2 - n_1) \left(\sum_{m=1}^{2l-1} m \underbrace{\beta_{m-2l}^{\frac{\delta}{2+\delta}}}_{=1} + \sum_{m=2l}^{n} \beta_{m-2l}^{\frac{\delta}{2+\delta}} \right) \\
&\le C(n_4 - n_3)(n_2 - n_1) \left(\sum_{m=1}^{2l-1} m + \sum_{m=2l}^{n} (m-2l)\beta_{m-2l}^{\frac{\delta}{2+\delta}} + \sum_{m=2l}^{n} 2l\beta_{m-2l}^{\frac{\delta}{2+\delta}} \right) \\
&\le C(n_4 - n_3)(n_2 - n_1) \left((2l)^2 + \sum_{m=2l}^{n} (m-2l)\beta_{m-2l}^{\frac{\delta}{2+\delta}} + 2l \sum_{m=2l}^{n} (m-2l)\beta_{m-2l}^{\frac{\delta}{2+\delta}} \right) \\
&= C(n_4 - n_3)(n_2 - n_1) \left(l^2 + (1+2l) \sum_{m=2l}^{n} (m-2l)\beta_{m-2l}^{\frac{\delta}{2+\delta}} \right) \\
&\le C(n_4 - n_3)(n_2 - n_1) \left(l^2 + (2l)^2 \sum_{m=2l}^{n} (m-2l)\beta_{m-2l}^{\frac{\delta}{2+\delta}} \right) \\
&\le C(n_4 - n_3)(n_2 - n_1) \Big(l^2 + l^2 \underbrace{\sum_{m=2l}^{n} (m-2l)^2\beta_{m-2l}^{\frac{\delta}{2+\delta}} \Big)}_{<\infty} \\
&\le C(n_4 - n_3)(n_2 - n_1) l^2 \\
&\le C(n_4 - n_3)(n_2 - n_1) n^{\frac{3}{8}}.
\end{aligned}$$

Thus, it follows

$$(14) \le C(n_4 - n_3)(n_2 - n_1)n^{\frac{3}{8}}.$$

If $m = j_2 - i_2$, it works very similar. Just a few comments on what changes: We get in the first case $I = \{i_1, j_1, j_2\}$, $I^C = \{j_2\}$, which leads to defining the function $g(X_{i_1,l}, X_{j_1,l}, X_{i_2,l}, X'_{j_2,l}) := \langle h_{2,l}(X_{i_1,l}, X_{j_1,l}), h_{2,l}(X_{i_2,l}, X'_{j_2,l}) \rangle$ and conditioning on $X_{i_1,l}, X_{j_1,l}, X_{i_2,l}$. For the second case it is $I = \{i_1, i_2\}$, $I^C = \{j_2\}$. We define $g(X_{i_1,l}, X'_{j_2,l}, X_{i_2,l}) := \langle h_{2,l}(X_{i_1,l}, X'_{j_2,l}), h_{2,l}(X_{i_2,l}, X'_{j_2,l}) \rangle$ and condition on $X_{i_2,l}, X'_{j_2,l}$. In the third case it is $I = \{i_1, j_1\}$, $I^C = \{j_2\}$, define
$g(X_{i_1,l}, X_{j_1,l}, X'_{j_2,l}) := \langle h_{2,l}(X_{i_1,l}, X_{j_1,l}), h_{2,l}(X_{i_1,l}, X'_{j_2,l}) \rangle$ and condition on $X_{i_1,l}, X_{j_1,l}$. This proves the lemma. $\square$

We can now state the asymptotic behaviour of the degenerate part of the U-statistic.

Proposition 2.3. *Under the assumptions of Theorem 2.1, it holds that*

a)

$$\mathbb{E}\Big[\Big(\max_{1\le n_1<n}\Big\|\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n}h_2(X_i,X_j)\Big\|\Big)^2\Big]^{\frac{1}{2}}\le Cs^2 2^{\frac{5s}{4}}$$

for s large enough that $n\le 2^s$.

b)

$$\max_{1\le n_1<n}\frac{1}{n^{3/2}}\Big\|\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n}h_2(X_i,X_j)\Big\|\xrightarrow{a.s.}0\quad \text{for } n\to\infty.$$

Proof.
Part a) We split the expectation with the help of the triangle inequality into three parts:

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\max_{1\le n_1<n}\Big\|\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n}h_2(X_i,X_j)\Big\|\Big)^2\Big]^{\frac{1}{2}}\\
\le\ &\mathbb{E}\Big[\Big(\max_{1\le n_1<n}\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n}\|h_2(X_i,X_j)-h_2(X_{i,l},X_{j,l})\|\Big)^2\Big]^{\frac{1}{2}} &(15)\\
&+\mathbb{E}\Big[\Big(\max_{1\le n_1<n}\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n}\|h_{2,l}(X_{i,l},X_{j,l})-h_2(X_{i,l},X_{j,l})\|\Big)^2\Big]^{\frac{1}{2}} &(16)\\
&+\mathbb{E}\Big[\Big(\max_{1\le n_1<n}\Big\|\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n}h_{2,l}(X_{i,l},X_{j,l})\Big\|\Big)^2\Big]^{\frac{1}{2}} &(17)
\end{aligned}$$

We will use Lemmas 2.4 to 2.6 to bound the three terms. The idea for all three terms is to use a suitable partition to rewrite the double sum inside of the expectation. By Lemma 2.4,

$$\begin{aligned}
(15)&\le C\mathbb{E}\Big[\Big(\sum_{d=0}^{s}\max_{\tilde{j}=1,\dots,2^{s-d}}\sum_{i=1}^{(\tilde{j}-1)2^d-1}\sum_{j=(\tilde{j}-1)2^d}^{\tilde{j}2^d}\|h_2(X_i,X_j)-h_2(X_{i,l},X_{j,l})\|\Big)^2\Big]^{\frac{1}{2}}\\
&\le Cs\sum_{d=0}^{s}\sum_{\tilde{j}=1}^{2^{s-d}}\mathbb{E}\Big[\Big(\sum_{i=1}^{(\tilde{j}-1)2^d-1}\sum_{j=(\tilde{j}-1)2^d}^{\tilde{j}2^d}\|h_2(X_i,X_j)-h_2(X_{i,l},X_{j,l})\|\Big)^2\Big]^{\frac{1}{2}}\\
&\le Cs\sum_{d=0}^{s}\sum_{\tilde{j}=1}^{2^{s-d}}(\tilde{j}2^d-(\tilde{j}-1)2^d)2^{\frac{s}{4}}\\
&\le C2^{\frac{s}{4}}s\sum_{d=0}^{s}\sum_{\tilde{j}=1}^{2^{s-d}}2^d=C2^{\frac{s}{4}}s\sum_{d=0}^{s}2^{s-d}2^d
\end{aligned}$$

$$= C2^{\frac{5s}{4}} s^2.$$

For (16) it works similar:

$$(16) \leq Cs \sum_{d=0}^{s} \sum_{\tilde{j}=1}^{2^{s-d}} \mathbb{E}\Big[\Big(\sum_{i=1}^{(\tilde{j}-1)2^d-1} \sum_{j=(\tilde{j}-1)2^d}^{\tilde{j}2^d} \|h_{2,l}(X_{i,l}, X_{j,l}) - h_2(X_{i,l}, X_{j,l})\|\Big)^2\Big]^{\frac{1}{2}}$$

$$\leq Cs \sum_{d=0}^{s} \sum_{\tilde{j}=1}^{2^{s-d}} (\tilde{j}2^d - (\tilde{j}-1)2^d) 2^{\frac{s}{4}}$$

$$\leq Cs^2 2^{\frac{5s}{4}}$$

by Lemma 2.5.
For (17), first look at its squared value and observe

$$(17)^2 = \mathbb{E}\Big[\Big(\max_{1 \leq n_1 < n} \| \sum_{i=1}^{n_1} \sum_{j=n_1+1}^{n} h_{2,l}(X_{i,l}, X_{j,l})\|\Big)^2\Big]$$

$$\leq \mathbb{E}\Big[\Big(\max_{1 \leq n_1 < n} \max_{n_1 < n_2 \leq n} \| \sum_{i=1}^{n_1} \sum_{j=n_2}^{n} h_{2,l}(X_{i,l}, X_{j,l})\|\Big)^2\Big]$$

$$\leq C\mathbb{E}\Big[\Big(\sum_{d=0}^{s} \max_{\tilde{j}=1,\dots,2^{s-d}} \max_{1 \leq n_1 < n} \| \sum_{i=1}^{n_1} \sum_{j=(\tilde{j}-1)2^d}^{\tilde{j}2^d} h_{2,l}(X_{i,l}, X_{j,l})\|\Big)^2\Big]$$

$$\leq Cs^2 \sum_{d=0}^{s} \sum_{\tilde{j}=1}^{2^{s-d}} \mathbb{E}\Big[\Big(\max_{1 \leq n_1 < n} \| \sum_{i=1}^{n_1} \sum_{j=(\tilde{j}-1)2^d}^{\tilde{j}2^d} h_{2,l}(X_{i,l}, X_{j,l})\|\Big)^2\Big]$$

$$= Cs^2 \sum_{d=0}^{s} \sum_{\tilde{j}=1}^{2^{s-d}} \mathbb{E}\Big[\Big(\max_{1 \leq n_1 < n} \| \sum_{i=1}^{n_1} \hat{X}_{i,l}\|\Big)^2\Big],$$

where $\hat{X}_{i,l} = \sum_{j=(\tilde{j}-1)2^d}^{\tilde{j}2^d} h_{2,l}(X_{i,l}, X_{j,l})$.
We will use Theorem 1 of Móricz [1976] to bound the expectation of the maximum.
We know by Lemma 2.6 that

$$\mathbb{E}\Big[\| \sum_{i=b+1}^{b+n} \hat{X}_{i,l}\|^2\Big] = \mathbb{E}\Big[\| \sum_{i=b+1}^{b+n} \sum_{j=(\tilde{j}-1)2^d}^{\tilde{j}2^d} h_{2,l}(X_{i,l}, X_{j,l})\|^2\Big]$$

$$\leq C(b+n-b)(\tilde{j}2^d - (\tilde{j}-1)2^d)(n+b)^{3/8} \leq Cn^{9/8} 2^d (n+b)^{3/8} \quad \forall b \geq 0.$$

In the notation of Móricz [1976], let $g(F_{b,n}) = Cn2^{d8/9}(n+b)^{3/9}$ and $\alpha = 9/8$. We need to check that $g(F_{b,n})$ fulfils

$$g(F_{b,k}) + g(F_{b+k,l}) \leq g(F_{b,k+l})$$

for all $b \geq 0$, $1 \leq k < k+l$.

So we calculate

$$
\begin{aligned}
g(F_{b,k}) &= k2^{d8/9}(b+k)^{3/9} \leq k2^{d8/9}(b+k+l)^{3/9} \\
g(F_{b+k,l}) &= l2^{d8/9}(b+k+l)^{3/9} \\
g(F_{b,k+l}) &= (k+l)2^{d8/9}(b+k+l)^{3/9}
\end{aligned}
$$

and see that the assumption on $g(F_{b,n})$ is fulfilled. Now, Theorem 1 of Móricz [1976] states that

$$
\mathbb{E}\Big[\Big(\max_{1\leq n_1 \leq n} \Big\|\sum_{i=b+1}^{b+n} \hat{X}_{i,l}\Big\|\Big)^2\Big] \leq C g^{\alpha}(F_{b,n}) \quad \forall b, n \geq 0.
$$

Using this for $b = 0$, we get that

$$
\mathbb{E}\Big[\Big(\max_{1\leq n_1 \leq n} \Big\|\sum_{i=1}^{n} \hat{X}_{i,l}\Big\|\Big)^2\Big] \leq C n^{9/8} 2^d n^{3/8} = C n^{3/2} 2^d.
$$

So,

$$
\begin{aligned}
(17)^2 &\leq C s^2 n^{3/2} \sum_{d=0}^{s} \sum_{\tilde{j}=1}^{2^{s-d}} 2^d = C s^2 n^{3/2} \sum_{d=0}^{s} 2^{s-d} 2^d = C s^2 n^{3/2} \sum_{d=0}^{s} 2^s = C s^2 n^{3/2} s 2^s \\
&\leq C s^3 2^{\frac{5s}{2}}
\end{aligned}
$$

recalling that $n \leq 2^s$.
Taking the square root yields

$$
\mathbb{E}\Big[\Big(\max_{1\leq n_1 < n} \Big\|\sum_{i=1}^{n_1} \sum_{j=n_1+1}^{n} h_{2,l}(X_{i,l}, X_{j,l})\Big\|\Big)^2\Big]^{1/2} \leq C s^{3/2} 2^{\frac{5s}{4}} \leq C s^2 2^{\frac{5s}{4}}.
$$

Combining all three parts gives us the stated result:

$$
\mathbb{E}\Big[\Big(\max_{1\leq n_1 < n} \Big\|\sum_{i=1}^{n_1} \sum_{j=n_1+1}^{n} h_2(X_i, X_j)\Big\|\Big)^2\Big]^{\frac{1}{2}} \leq C s^2 2^{\frac{5s}{4}}
$$

Part b) Recall that s is chosen such that $n \leq 2^s$ and thus $n^{\frac{3}{2}} \leq 2^{\frac{3s}{2}}$. To prove almost sure convergence, it is enough to prove that for any $\epsilon > 0$

$$
\sum_{s=1}^{\infty} \mathbb{P}\Big(2^{-\frac{3s}{2}} \max_{1\leq n_1 < n} \Big\|\sum_{s=1}^{n_1} \sum_{j=n_1+1}^{n} h_2(X_i, X_j)\Big\| > \epsilon\Big) < \infty.
$$

We do this by using Markov's inequality and our result from part a):

$$
\sum_{s=1}^{\infty} \mathbb{P}\Big(2^{-\frac{3s}{2}} \max_{1\leq n_1 < n} \Big\|\sum_{s=1}^{n_1} \sum_{j=n_1+1}^{n} h_2(X_i, X_j)\Big\| > \epsilon\Big)
$$

$$
\begin{aligned}
&\leq \frac{1}{\epsilon^2}\sum_{s=1}^{\infty}\mathbb{E}\Bigg[\Big(2^{-\frac{3s}{2}}\max_{1\leq n_1<n}\big\|\sum_{s=1}^{n_1}\sum_{j=n_1+1}^{n}h_2(X_i,X_j)\big\|\Big)^2\Bigg]\\
&=\frac{1}{\epsilon^2}\sum_{s=1}^{\infty}2^{-3s}\mathbb{E}\Bigg[\Big(\max_{1\leq n_1<n}\big\|\sum_{s=1}^{n_1}\sum_{j=n_1+1}^{n}h_2(X_i,X_j)\big\|\Big)^2\Bigg]\\
&\leq \frac{1}{\epsilon^2}\sum_{s=1}^{\infty}2^{-3s}(Cs^2 2^{\frac{5s}{4}})^2\\
&=\frac{C}{\epsilon^2}\sum_{s=1}^{\infty}s^4 2^{-\frac{s}{2}}<\infty.
\end{aligned}
$$

By the lemma of Borel-Cantelli almost sure convergence follows

$$
\max_{1\leq n_1<n}\frac{1}{n^{3/2}}\big\|\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n}h_2(X_i,X_j)\big\|\xrightarrow{\text{a.s.}}0\quad\text{for } n\to\infty.
$$

□

To prove Theorem 2.1, it is left to combine the results for the linear and degenerate part of the U-statistic.

Proof of Theorem 2.1. We will bound the maximum from above by the sum of the degenerate and the linear part, using Hoeffding's decomposition, as shown in Lemma 2.1 and using the triangle inequality afterwards:

$$
\begin{aligned}
\max_{1\leq k<n}\frac{1}{n^{3/2}}\|U_{n,k}\| &= \max_{1\leq k<n}\frac{1}{n^{3/2}}\big\|n\sum_{i=1}^{k}(h_1(X_i)-\overline{h_1(X)})+\sum_{i=1}^{k}\sum_{j=k+1}^{n}h_2(X_i,X_j)\big\|\\
\leq \max_{1\leq k<n}\frac{1}{n^{3/2}}\big\|n\sum_{i=1}^{k}(h_1(X_i)-\overline{h_1(X)})\big\| &+ \max_{1\leq k<n}\frac{1}{n^{3/2}}\big\|\sum_{i=1}^{k}\sum_{j=k+1}^{n}h_2(X_i,X_j)\big\|
\end{aligned}
$$

For the degenerate part, use the convergence to zero from Proposition 2.3:

$$
\max_{1\leq k<n}\frac{1}{n^{3/2}}\big\|\sum_{i=1}^{k}\sum_{j=k+1}^{n}h_2(X_i,X_j)\big\|\xrightarrow{\mathbb{P}}0,
$$

since convergence in probability follows from almost sure convergence.
Now observe that we can write the linear part for n large enough as

$$
\begin{aligned}
\max_{1\leq k<n}\frac{1}{n^{3/2}}\big\|n\sum_{i=1}^{k}\big(h_1(X_i)-\overline{h_1(X)}\big)\big\| &= \max_{\lambda\in[0,1]}\frac{1}{n^{3/2}}\big\|n\sum_{i=1}^{\lfloor n\lambda\rfloor}\big(h_1(X_i)-\overline{h_1(X)}\big)\big\|\\
=\max_{\lambda\in[0,1]}\frac{1}{n^{3/2}}\big\|n\sum_{i=1}^{\lfloor n\lambda\rfloor}h_1(X_i)-n\lfloor n\lambda\rfloor\frac{1}{n}\sum_{j=1}^{n}h_1(X_j)\big\|
\end{aligned}
$$

$$
\begin{aligned}
&= \max_{\lambda\in[0,1]} \Big\| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor n\lambda\rfloor} h_1(X_i) - \frac{\lfloor n\lambda\rfloor}{n^{3/2}} \sum_{j=1}^{n} h_1(X_j) \Big\| \\
&\approx \sup_{\lambda\in[0,1]} \Big\| \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor n\lambda\rfloor} h_1(X_i)}_{=:x(\lambda)} - \frac{\lambda}{\sqrt{n}} \sum_{j=1}^{n} h_1(X_i) \Big\| \\
&= \sup_{\lambda\in[0,1]} \|x(\lambda) - \lambda x(1)\|.
\end{aligned}
$$

We know by Proposition 2.2 that

$$
(x(\lambda))_{\lambda\in[0,1]} \xrightarrow{\mathcal{D}} (W(\lambda))_{\lambda\in[0,1]}.
$$

By the continuous mapping theorem it follows that

$$
(x(\lambda) - \lambda x(1))_{\lambda\in[0,1]} \xrightarrow{\mathcal{D}} (W(\lambda) - \lambda W(1))_{\lambda\in[0,1]}.
$$

And thus we can finally conclude

$$
\max_{1\leq k<n} \frac{1}{n^{3/2}} \|U_{n,k}\| \xrightarrow{\mathcal{D}} \sup_{\lambda\in[0,1]} \|W(\lambda) - \lambda W(1)\|.
$$

□

2.4 Validity Under the Alternative

The next theorem will show that the test statistic converges to infinity in probability under the alternative, so a test based on it is consistent.
For this, we consider the following model: We have a stationary, $H \otimes H$-valued sequence $(X_n, Z_n)_{n\in\mathbb{Z}}$ and we observe $Y_1, ..., Y_n$ with

$$Y_i = \begin{cases} X_i & \text{for } i \leq \lfloor n\lambda^\star \rfloor = k^\star \\ Z_i & \text{for } i > \lfloor n\lambda^\star \rfloor = k^\star \end{cases},$$

so $\lambda^\star \in (0,1)$ is the proportion of observations after which the change happens. The change is detectable, if $\|\mathbb{E}[h(X_1, \tilde{Z}_1)]\| \neq 0$ for an independent copy $\tilde{Z}_1$ of Z_1.

Example 2.4 (Detectability)**.** *Let $H = \mathbb{R}^2$ and sequences of i.i.d. random variables $(X_i)_{i\leq n}$ and $(Z_i)_{i\leq n}$ in H. Define $Y_i = (Y_{i,1}, Y_{i,2})^T = X_i - Z_i$ for $i = 1, ..., n$ and assume that $Y_{i,1}$ and $Y_{i,2}$ are independent and have a density function f_1 resp. f_2 which is symmetric around $\mathbb{E}[Y_{i,1}]$ resp. $\mathbb{E}[Y_{i,2}]$ for all $i = 1, ..., n$. For example, let $X_{i,1}, X_{i,2}$ and $Z_{i,1}, Z_{i,2}$ be independent and normal distributed, then $Y_{i,1}, Y_{i,2}$ are independent and normal distributed as well with a density function symmetric around its expected value.*
For $h(x,y) = x - y$ and $h(x,y) = (x-y)/\|x-y\|$ a change is detectable if

$$\mathbb{E}[X_1] \neq \mathbb{E}[Z_1] \Leftrightarrow \mathbb{E}[Y_1] \neq 0.$$

Or equivalently, a change is not detectable if

$$\mathbb{E}[X_1] = \mathbb{E}[Z_1] \Leftrightarrow \mathbb{E}[Y_1] = 0.$$

For $h(x,y) = x - y$ this can quickly be seen by

$$\|\mathbb{E}[h(X_1, Z_1)]\| = \|\mathbb{E}[X_1] - \mathbb{E}[Z_1]\| = 0 \Leftrightarrow \mathbb{E}[X_1] = \mathbb{E}[Z_1].$$

For $h(x,y) = (x-y)/\|x-y\|$, first rewrite the expectation:

$$\begin{aligned} &\|\mathbb{E}[h(X_1, Z_2)]\| = \|\mathbb{E}[(X_1 - Z_1)/\|X_1 - Z_1\|]\| = \|\mathbb{E}[Y_1/\|Y_1\|]\| \\ &= \|(\mathbb{E}[Y_{1,1}/\sqrt{Y_{1,1}^2 + Y_{1,2}^2}], \mathbb{E}[Y_{1,2}/\sqrt{Y_{1,1}^2 + Y_{1,2}^2}])^T\| \end{aligned}$$

Now, by independence and the existence of a density function, it holds that

$$\begin{aligned} \mathbb{E}[Y_{1,1}/\sqrt{Y_{1,1}^2 + Y_{1,2}^2}] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{y_1}{\sqrt{y_1^2 + y_2^2}} f_1(y_1) f_2(y_2) dy_1 dy_2 \\ &= \int_{-\infty}^{\infty} \Big[\int_{-\infty}^{\infty} \frac{y_1}{\sqrt{y_1^2 + y_2^2}} f_1(y_1) dy_1\Big] f_2(y_2) dy_2. \end{aligned}$$

Rewriting the inner integral yields:

$$\begin{aligned}
&\int_{-\infty}^{\infty} \frac{y_1}{\sqrt{y_1^2+y_2^2}} f_1(y_1)dy_1 \\
&= \int_{-\infty}^{0} \frac{y_1}{\sqrt{y_1^2+y_2^2}} f_1(y_1)dy_1 + \int_{0}^{\infty} \frac{y_1}{\sqrt{y_1^2+y_2^2}} f_1(y_1)dy_1 \\
&= -\int_{0}^{-\infty} \frac{y_1}{\sqrt{y_1^2+y_2^2}} f_1(y_1)dy_1 + \int_{0}^{\infty} \frac{y_1}{\sqrt{y_1^2+y_2^2}} f_1(y_1)dy_1 \\
&= -\int_{0}^{\infty} \frac{-y_1}{\sqrt{y_1^2+y_2^2}} f_1(-y_1)(-1)dy_1 + \int_{0}^{\infty} \frac{y_1}{\sqrt{y_1^2+y_2^2}} f_1(y_1)dy_1 \\
&= \int_{0}^{\infty} \frac{y_1}{\sqrt{y_1^2+y_2^2}} (f_1(y_1) - f_1(-y_1))dy_1,
\end{aligned}$$

first switching the limits of the first integral and then substituting $-y_1 = y_1$. At this point, the symmetry of f_1 becomes important. The integral equals zero if and only if $f_1(y_1) = f_1(-y_1)$ for all $y_1 > 0$. And this holds true if and only if f_1 is symmetric in zero $\Leftrightarrow \mathbb{E}[Y_{1,1}] = 0$. Similarly one gets

$$\mathbb{E}[Y_{1,2}/\sqrt{Y_{1,1}^2 + Y_{1,2}^2}] = 0 \Leftrightarrow \mathbb{E}[Y_{1,2}] = 0.$$

Concluding that

$$\|\mathbb{E}[h(X_1, Z_2)]\| = \|\mathbb{E}[Y_1/\|Y_1\|]\| = 0 \Leftrightarrow \mathbb{E}[Y_1] = 0,$$

which equivalently means that the change is detectable if $\mathbb{E}[Y_1] \neq 0 \Leftrightarrow \mathbb{E}[X_1] \neq \mathbb{E}[Z_1]$.

From here on, we need one further assumption on the kernel function h to ensure that we can handle "mixed" expectations with one sample from before and one after the change.

Definition 2.9 (Mixed uniform moments)**.** *Let $(X_n, Z_n)_{n\in\mathbb{Z}}$ be a $\mathbb{P}$-NED functional of $(\zeta_n)_{n\in\mathbb{Z}}$. If there exists $M > 0$ such that for all $k, n \in \mathbb{N}$*

$$\begin{aligned}
\mathbb{E}[\|h(X_{0,k}, Z_{n,k})\|^m] &\leq M \\
\mathbb{E}[\|h(X_0, Z_{n,k})\|^m] &\leq M \\
\mathbb{E}[\|h(X_0, Z_n)\|^m] &\leq M,
\end{aligned}$$

we say that the kernel h has mixed uniform m-th moments under approximation.

We will also call this uniform moments, since it is a simple supplement of Definition 2.6.

Theorem 2.2. *Let $(X_n, Z_n)_{n\in\mathbb{Z}}$ be $\mathbb{P}$-NED on an absolutely regular sequence $(\zeta_n)_{n\in\mathbb{Z}}$ such that $a_k\Phi(k^{-8\frac{\delta+3}{\delta}}) = \mathcal{O}(k^{-8\frac{(\delta+3)(\delta+2)}{\delta^2}})$ and $\sum_{k=1}^{\infty} k^2 \beta_k^{\frac{\delta}{4+\delta}} < \infty$ for some $\delta > 0$. Assume that $h : H^2 \to H$ is an antisymmetric kernel that fulfils the variation condition and is either bounded or has uniform $(4+\delta)$-moments under approximation* *for processes $(X_n)_{n\in\mathbb{Z}}$, $(Z_n)_{n\in\mathbb{Z}}$ and $(X_n, Z_n)_{n\in\mathbb{Z}}$ and that $\mathbb{E}[h(X_1, \tilde{Z}_1)] \neq 0$, where $\tilde{Z}_1$ is an independent copy of Z_1. Then*

$$\max_{1\leq k<n} \frac{1}{n^{3/2}} \|U_{n,k}(Y)\| = \max_{1\leq k<n} \frac{1}{n^{3/2}} \Big\| \sum_{i=1}^{k} \sum_{j=k+1}^{n} h(Y_i, Y_j) \Big\| \xrightarrow{\mathbb{P}} \infty.$$

The idea to proof this theorem is to lower bound $\max_{1\leq k<n} \frac{1}{n^{3/2}} \|U_{n,k}(Y)\|$ by $\frac{1}{n^{3/2}} \|U_{n,k^\star}(Y)\|$. Handling $U_{n,k^\star}(Y)$ is easier, since we know that $k^\star$ is the change-point. We can again split $U_{n,k^\star}(Y)$ into a linear and degenerate part, but to use previous results for convergence, some centralisation is needed. This gives an additional term of expectations $\mathbb{E}[h(X_0, Z_0)]$ that dominates the expression and ensures divergence.
Rewriting $U_{n,k^\star}(Y) = \sum_{i=1}^{k^\star} \sum_{j=k^\star+1}^{n} h(Y_i, Y_j)$ with the help of Hoeffding's decomposition reads:

$$\begin{aligned}
U_{n,k^\star}(Y) &= \sum_{i=1}^{k^\star} \sum_{j=k^\star+1}^{n} h(X_i, Z_j) \\
&= \sum_{i=1}^{k^\star} \sum_{j=k^\star+1}^{n} \big(h_1^{(Z)}(X_i) - h_1^{(X)}(Z_j) + h_2^{(X,Z)}(X_i, Z_j)\big) \\
&= (n-k^\star) \sum_{i=1}^{k^\star} h_1^{(Z)}(X_i) - k^\star \sum_{j=k^\star+1}^{n} h_1^{(X)}(Z_j) + \sum_{i=1}^{k^\star} \sum_{j=k^\star+1}^{n} h_2^{(X,Z)}(X_i, Z_j).
\end{aligned}$$

For the behaviour of the three parts, we will give the two-sample counterparts of Propositions 2.2 and 2.3 and their preparatory lemmas. We need to centralise by the expectation to preserve all desired properties.

Proposition 2.4. *Under the assumption of Theorem 2.2 it holds that*

$$\Big(\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor n\lambda \rfloor} \big(h_1^{(Z)}(X_i) - \mathbb{E}[h_1^{(Z)}(X_i)]\big)\Big)_{\lambda\in[0,1]} \xrightarrow{\mathcal{D}} (W_1(\lambda))_{\lambda\in[0,1]}$$

and

$$\Big(\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor n\lambda \rfloor} \big(h_1^{(X)}(Z_i) - \mathbb{E}[h_1^{(X)}(Z_i)]\big)\Big)_{\lambda\in[0,1]} \xrightarrow{\mathcal{D}} (W_2(\lambda))_{\lambda\in[0,1]},$$

where $(W_1(\lambda))_{\lambda\in[0,1]}$, $(W_2(\lambda))_{\lambda\in[0,1]}$ are Brownian motions with covariance operator as definded in Theorem 2.1.

Proof. The proof follows the steps of Proposition 2.2. So, we have to check the assumptions of Theorem 1 Sharipov et al. [2016a]. We will do this for $h_1^{(Z)}(X_i)$, for $h_1^{(X)}(Z_i)$ everything holds similarly. First recall that $\mathbb{E}[h_1^{(Z)}(X_0)] = \mathbb{E}[h(X_0, \tilde{Z}_0)]$.
Assumption 1: $(h_1^{(Z)}(X_n) - \mathbb{E}[h_1^{(Z)}(X_n)])_{n\in\mathbb{Z}}$ is L_1-NED.
Along the lines of the proof of Lemma 2.2 we can show that $(h_1^{(Z)}(X_n))_{n\in\mathbb{Z}}$ is L_2-NED

with approximating constants $a_{k,2} = \mathcal{O}(k^{-4\frac{3+\delta}{\delta}})$, since the variation condition still holds and we assumed mixed uniform moments. By Jensen's inequality it follows that $(h_1^{(Z)}(X_n))_{n\in\mathbb{Z}}$ is L_1-NED with approximating constants $a_{k,1} = a_{k,2}$.
Now observe

$$
\begin{aligned}
&\mathbb{E}\Big[\big\|h_1^{(Z)}(X_0) - \mathbb{E}[h_1^{(Z)}(X_0)] - \mathbb{E}\big[h_1^{(Z)}(X_0) - \mathbb{E}[h_1^{(Z)}(X_0)]|\mathcal{F}_{-k}^k\big]\big\|\Big] \\
&= \mathbb{E}\Big[\big\|h_1^{(Z)}(X_0) - \mathbb{E}[h_1^{(Z)}(X_0)] - \mathbb{E}[h_1^{(Z)}(X_0)|\mathcal{F}_{-k}^k] + \mathbb{E}\big[\mathbb{E}[h_1^{(Z)}(X_0)]|\mathcal{F}_{-k}^k\big]\big\|\Big] \\
&= \mathbb{E}\Big[\big\|h_1^{(Z)}(X_0) - \mathbb{E}[h_1^{(Z)}(X_0)] - \mathbb{E}[h_1^{(Z)}(X_0)|\mathcal{F}_{-k}^k] + \mathbb{E}[h_1^{(Z)}(X_0)]\big\|\Big] \\
&= \mathbb{E}\Big[\big\|h_1^{(Z)}(X_0) - \mathbb{E}[h_1^{(Z)}(X_0)|\mathcal{F}_{-k}^k]\big\|\Big] \\
&\leq Ck^{-4\frac{3+\delta}{\delta}} = a_{k,1}
\end{aligned}
$$

since $(h_1^{(Z)}(X_n))_{n\in\mathbb{Z}}$ is L_1-NED. So, $(h_1^{(Z)}(X_n) - \mathbb{E}[h_1^{(Z)}(X_n)])_{n\in\mathbb{Z}}$ is L_1-NED as well with approximating constants $a_{k,1}$.
Assumption 2: Existing $(4+\delta)$-moments.
Using the triangle inequality, Minkowski's and Jensen's inequality it holds that

$$
\begin{aligned}
&\mathbb{E}\big[\|h_1^{(Z)}(X_0) - \mathbb{E}[h_1^{(Z)}(X_0)]\|^{4+\delta}\big]^{\frac{1}{4+\delta}} \\
&\leq \mathbb{E}\big[\|h_1^{(Z)}(X_0)\|^{4+\delta}\big]^{\frac{1}{4+\delta}} + \mathbb{E}\big[\|\mathbb{E}[h_1^{(Z)}(X_0)]\|^{4+\delta}\big]^{\frac{1}{4+\delta}} \\
&\leq 2\mathbb{E}\big[\|h_1^{(Z)}(X_0)\|^{4+\delta}\big]^{\frac{1}{4+\delta}} \\
&\leq 2M^{\frac{1}{4+\delta}}
\end{aligned}
$$

by uniform $(4+\delta)$ moments of $h_1^{(Z)}$.
Assumption 3: $\sum_{m=1}^{\infty} m^2 a_{m,1}^{\frac{\delta}{3+\delta}} \leq \infty$ follows similar as in the proof of Theorem 2.2.
Assumption 4: $\sum_{m=1}^{\infty} m^2 \beta_m^{\frac{\delta}{4+\delta}} < \infty$ is assumed in Theorem 2.2
Lastly, we observe that

$$
\mathbb{E}\big[h_1^{(Z)}(X_0) - \mathbb{E}[h_1^{(Z)}(X_0)]\big] = \mathbb{E}[h_1^{(Z)}(X_0)] - \mathbb{E}[h_1^{(Z)}(X_0)] = 0
$$

and thus all requirements for Theorem 1 Sharipov et al. [2016a] are given and the statement of the lemma follows. □

Later, in the proof of Theorem 2.2, the following corollary will be of use:

Corollary 2.1. *Under assumptions of Theorem 2.2, it holds that*

$$
\frac{1}{n^{3/2}} \sum_{i=1}^{k^\star} \sum_{j=k^\star+1}^{n} \big(h_1^{(Z)}(X_i) - h_1^{(X)}(Z_j) - 2\mathbb{E}[h(\tilde{X}_0, \tilde{Z}_0)]\big)
$$

is stochastically bounded, where $\tilde{X}_0$ and $\tilde{Z}_0$ are independent copies of X_0 resp. Z_0.

Proof. This follows from Proposition 2.4 above:

$$\Big|\frac{1}{n^{3/2}}\sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n}\big(h_1^{(Z)}(X_i)-h_1^{(X)}(Z_j)-2\mathbb{E}[h(\tilde{X}_0,\tilde{Z}_0)]\big)\Big|$$

$$\leq \Big|\frac{1}{n^{1/2}}\sum_{i=1}^{k^\star}h_1^{(Z)}(X_i)-\mathbb{E}[h(\tilde{X}_0,\tilde{Z}_0)]\Big|+\Big|\frac{1}{n^{1/2}}\sum_{j=k^\star+1}^{n}h_1^{(X)}(Z_j)-\mathbb{E}[h(\tilde{X}_0,\tilde{Z}_0)]\Big|$$

$$\leq \Big|\frac{1}{n^{1/2}}\sum_{i=1}^{k^\star}(h_1^{(Z)}(X_i)-\mathbb{E}[h_1^{(Z)}(X_i)])\Big|+\Big|\frac{1}{n^{1/2}}\sum_{j=k^\star+1}^{n}(h_1^{(X)}(Z_j)-\mathbb{E}[h_1^{(X)}(Z_j)])\Big|.$$

Both summands converge weakly to a Gaussian limit and are stochastically bounded. □

The following lemmas hold as two-sample analogues of Lemmas 2.3 to 2.6. We will only point out the crucial points of the proofs.

Lemma 2.7. *Under the assumptions of Theorem 2.2, there exists a universal constant $C>0$ such that for every $i,k,l\in\mathbb{N}$, $\epsilon>0$ it holds that*

$$\mathbb{E}[\|h_2^{(X,Z)}(X_i,Z_{i+k+2l})-h_2^{(X,Z)}(X_{i,l},Z_{i+k+2l,l})\|^2]^{\frac{1}{2}}\leq C(\sqrt{\epsilon}+\beta_k^{\frac{\delta}{2(s+\delta)}}+(a_l\Phi(\epsilon))^{\frac{\delta}{2(s+\delta)}}).$$

Proof. For the $\mathbb{P}$-NED series $(X_n,Z_n)_{n\in\mathbb{Z}}$, we write $(X_{i,l},Z_{i,l})=f_l(\zeta_{i-l},...,\zeta_{i+l})$. By Lemma D1 Dehling et al. [2017] there exist independent copies $(\zeta_n')_{n\in\mathbb{Z}}$, $(\zeta_n'')_{n\in\mathbb{Z}}$ satisfying

$$\mathbb{P}((\zeta_n')_{n\geq i+k+l}=(\zeta_n)_{n\geq i+k+l})=1-\beta_k \text{ and } \mathbb{P}((\zeta_n'')_{n\leq i+l}=(\zeta_n)_{n\leq i+l})=1-\beta_k.$$

We define

$$(X_i',Z_i')=f((\zeta_{i+n}')_{n\in\mathbb{Z}}),\quad (X_i'',Z_i'')=f((\zeta_{i+n}'')_{n\in\mathbb{Z}})$$
$$(X_{i,l}',Z_{i,l}')=f_l(\zeta_{i-l}',...,\zeta_{i+l}'),\quad (X_{i,l}'',Z_{i,l}'')=f_l(\zeta_{i-l}'',...,\zeta_{i+l}'')$$

and write

$$\begin{aligned}&\mathbb{E}[\|h_2^{(X,Z)}(X_i,Z_{i+k+2l})-h_2^{(X,Z)}(X_{i,l},Z_{i+k+2l,l})\|^2]^{\frac{1}{2}}\\ \leq\ &\mathbb{E}[\|h_2^{(X,Z)}(X_i,Z_{i+k+2l})-h_2^{(X,Z)}(X_i'',Z_{i+k+2l}')\|^2]^{\frac{1}{2}}\\ &+\mathbb{E}[\|h_2^{(X,Z)}(X_i''Z_{i+k+2l}')-h_2^{(X,Z)}(X_{i,l}'',Z_{i+k+2l,l}')\|]^{\frac{1}{2}}\\ &+\mathbb{E}[\|h_2^{(X,Z)}(X_{i,l}'',Z_{i+k+2l,l}')-h_2^{(X,Z)}(X_{i,l},Z_{i+k+2l,l})\|^2]^{\frac{1}{2}}.\end{aligned}$$

The three parts can be handled completely analogous to the ones in Lemma 2.3, thus we will leave it at this. □

Lemma 2.8. *Under the assumptions of Theorem 2.2 it holds for any $n_1<n_2<n_3<n_4\leq n$ and $l=\left\lfloor n_4^{\frac{3}{16}}\right\rfloor$:*

$$\mathbb{E}\Big[\Big(\sum_{n_1\leq i\leq n_2}\sum_{n_3\leq j\leq n_4}\|h_2^{(X,Z)}(X_i,Z_j)-h_2^{(X,Z)}(X_{i,l},Z_{j,l})\|\Big)^2\Big]^{\frac{1}{2}}\leq C(n_4-n_3)n^{\frac{1}{4}}.$$

Proof. This follows from Lemma 2.7 analogous to the proof of Lemma 2.4. $\square$

Lemma 2.9. *Under the assumptions of Theorem 2.2, it holds for any $n_1 < n_2 < n_3 < n_4 \leq n$ and $l = \left\lfloor n_4^{\frac{3}{16}} \right\rfloor$:*

$$\mathbb{E}\Big[\Big(\sum_{n_1\leq i\leq n_2}\sum_{n_3\leq j\leq n_4}\|h_{2,l}^{(X,Z)}(X_{i,l},Z_{j,l}) - h_2^{(X,Z)}(X_{i,l},Z_{j,l})\|\Big)^2\Big]^{\frac{1}{2}} \leq C(n_4-n_3)n^{\frac{1}{4}},$$

where $h_{2,l}^{(X,Z)}(x,y) = h(x,y) - \mathbb{E}[h(x,\tilde{Z}_{0,l})] - \mathbb{E}[h(\tilde{X}_{0,l},y)]\ \forall\, x,y \in H$ and $(\tilde{X}_{i,l},\tilde{Z}_{i,l}) = f_l(\tilde{\zeta}_{i-l},\ldots,\tilde{\zeta}_{i+l})$ for $(\tilde{\zeta}_n)_{n\in\mathbb{Z}}$ an independent copy of $(\zeta_n)_{n\in\mathbb{Z}}$.

We will not give the proof here, since it is a complete analogue of the proof of Lemma 2.5.

For the following two-sample version of Lemma 2.6, we have to add some expectation to $h_{2,l}^{(X,Z)}$, since $h_{2,l}^{(X,Z)}$ itself is not degenerate. As this is an important property used in the proof, so we will consider $h_{2,l}^{(X,Z)}(X_{0,l},Z_{0,l}) - \mathbb{E}[h_{2,l}^{(X,Z)}(\tilde{X}_{0,l},\tilde{Z}_{0,l})]$ instead. A short calculation shows that this is indeed degenerated:

$$\begin{aligned}
&\mathbb{E}_X\big[h_{2,l}^{(X,Z)}(X_{0,l},y) - \mathbb{E}[h_{2,l}^{(X,Z)}(\tilde{X}_{0,l},\tilde{Z}_{0,l})]\big]\\
&= \mathbb{E}_X\Big[h(X_{0,l},y) - \mathbb{E}_Z[h(X_{0,l},\tilde{Z}_{0,l})] - \mathbb{E}_X[h(\tilde{X}_{0,l},y)]\\
&\qquad - \mathbb{E}\Big[h(\tilde{X}_{0,l},\tilde{Z}_{0,l}) - \mathbb{E}_Z[h(\tilde{X}_{0,l},\tilde{Z}_{0,l})] - \mathbb{E}_X[h(\tilde{X}_{0,l},\tilde{Z}_{0,l})]\Big]\Big]\\
&= \mathbb{E}_X[h(X_{0,l},y)] - \mathbb{E}[h(X_{0,l},\tilde{Z}_{0,l})] - \mathbb{E}_X[h(\tilde{X}_{0,l},y)]\\
&\qquad - \mathbb{E}[h(\tilde{X}_{0,l},\tilde{Z}_{0,l})] + \mathbb{E}[h(\tilde{X}_{0,l},\tilde{Z}_{0,l})] + \mathbb{E}[h(\tilde{X}_{0,l},\tilde{Z}_{0,l})]\\
&= 0
\end{aligned}$$

and similarly $\mathbb{E}_Z\big[h_{2,l}^{(X,Z)}(x,Z_{0,l}) - \mathbb{E}[h_{2,l}^{(X,Z)}(\tilde{X}_{0,l},\tilde{Z}_{0,l})]\big] = 0$.

Lemma 2.10. *Under the assumptions of Theorem 2.2, it holds for any $n_1,n_2,n_3,n_4 \leq n$ with $n_1 < n_2$, $n_3 < n_4$ and $l = \lfloor n \rfloor^{\frac{3}{16}}$:*

$$\begin{aligned}
\mathbb{E}\Big[\Big(\|\sum_{n_1\leq i\leq n_2}\sum_{n_3\leq j\leq n_4} h_{2,l}^{(X,Z)}(X_{i,l},Z_{i,l}) - \mathbb{E}[h_{2,l}^{(X,Z)}(\tilde{X}_{i,l},\tilde{Z}_{j,l})]\|\Big)^2\Big]&\\
\leq C(n_4-n_3)(n_2-n_1)n^{\frac{3}{8}}&.
\end{aligned}$$

Since we have a degenerated function, we can prove this completely analogous to Lemma 2.6.

Proposition 2.5. *Under the assumptions of Theorem 2.2 it holds that*

a)

$$\mathbb{E}\Big[\Big(\max_{1\le n_1<n}\Big\|\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n} h_2^{(X,Z)}(X_i,Z_j)-\mathbb{E}[h_2^{(X,Z)}(\tilde{X}_i,\tilde{Z}_j)]\Big\|\Big)^2\Big]^{\frac{1}{2}}\le Cs^2 2^{\frac{5s}{4}}$$

for s large enough that $n\le 2^s$.

b)

$$\max_{1\le n_1<n}\frac{1}{n^{3/2}}\Big\|\sum_{i=1}^{n_1}\sum_{j=n_+1}^{n} h_2^{(X,Z)}(X_i,Z_j)-\mathbb{E}[h_2^{(X,Z)}(\tilde{X}_i,\tilde{Z}_j)]\Big\|\overset{a.s}{\to}0 \text{ for } n\to\infty.$$

Proof. The proof is again similar to the one-sample case stated in Proposition 2.3.
Part a)
We add some zeros and split the expectation with the help of the triangular inequality:

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\max_{1\le n_1<n}\Big\|\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n} h_2^{(X,Z)}(X_i,Z_j)-\mathbb{E}[h_2^{(X,Z)}(\tilde{X}_i,\tilde{Z}_j)]\Big\|\Big)^2\Big]^{\frac{1}{2}}\\
\le\ &\mathbb{E}\Big[\Big(\max_{1\le n_1<n}\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n}\|h_2^{(X,Z)}(X_i,Z_j)-h_2^{(X,Z)}(X_{i,l},Z_{i,l})\|\Big)^2\Big]^{\frac{1}{2}}\\
&+\mathbb{E}\Big[\Big(\max_{1\le n_1<n}\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n}\|\mathbb{E}[h_2^{(X,Z)}(\tilde{X}_i,\tilde{Z}_j)]-\mathbb{E}[h_2^{(X,Z)}(\tilde{X}_{i,l},\tilde{Z}_{i,l})]\|\Big)^2\Big]^{\frac{1}{2}}\\
&+\mathbb{E}\Big[\Big(\max_{1\le n_1<n}\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n}\|h_{2,l}^{(X,Z)}(X_{i,l},Z_{j,l})-h_2^{(X,Z)}(X_{i,l},Z_{i,l})\|\Big)^2\Big]^{\frac{1}{2}}\\
&+\mathbb{E}\Big[\Big(\max_{1\le n_1<n}\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n}\|\mathbb{E}[h_{2,l}^{(X,Z)}(\tilde{X}_{i,l},\tilde{Z}_{j,l})]-\mathbb{E}[h_2^{(X,Z)}(\tilde{X}_{i,l},\tilde{Z}_{i,l})]\|\Big)^2\Big]^{\frac{1}{2}}\\
&+\mathbb{E}\Big[\Big(\max_{1\le n_1<n}\Big\|\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n} h_{2,l}^{(X,Z)}(X_{i,l},Z_{i,l})-\mathbb{E}[h_{2,l}^{(X,Z)}(\tilde{X}_{i,l},\tilde{Z}_{i,l})]\Big\|\Big)^2\Big]^{\frac{1}{2}}\\
\le\ &2\mathbb{E}\Big[\Big(\max_{1\le n_1<n}\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n}\|h_2^{(X,Z)}(X_i,Z_j)-h_2^{(X,Z)}(X_{i,l},Z_{i,l})\|\Big)^2\Big]^{\frac{1}{2}} &(18)\\
&+2\mathbb{E}\Big[\Big(\max_{1\le n_1<n}\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n}\|h_{2,l}^{(X,Z)}(X_{i,l},Z_{j,l})-h_2^{(X,Z)}(X_{i,l},Z_{i,l})\|\Big)^2\Big]^{\frac{1}{2}} &(19)\\
&+\mathbb{E}\Big[\Big(\max_{1\le n_1<n}\Big\|\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n} h_{2,l}^{(X,Z)}(X_{i,l},Z_{i,l})-\mathbb{E}[h_{2,l}^{(X,Z)}(\tilde{X}_{i,l},\tilde{Z}_{i,l})]\Big\|\Big)^2\Big]^{\frac{1}{2}} &(20)
\end{aligned}$$

The statement of Part a) follows by using a suitable partition for all three parts as in the proof of Proposition 2.3. For (18) use Lemma 2.8, for (19) use Lemma 2.9 and for (20) we can use Lemma 2.10.
Part b)
Follows from Part a) as in Proposition 2.3 by using Markov's inequality and the lemma of Borel-Cantelli. □

Proof of Theorem 2.2. We can bound the maximum from below using the reverse triangle inequality and then make use of previous results, recalling that $k^\star = \lfloor n\lambda^\star \rfloor$:

$$\begin{aligned}
&\max_{1\le k\le n} \|\frac{1}{n^{3/2}} U_{n,k}(Y)\| \ge \|\frac{1}{n^{3/2}} U_{n,k^\star}(Y)\| \\
&= \|\frac{1}{n^{3/2}}\big(U_{n,k^\star}(Y) - k^\star(n-k^\star)\mathbb{E}[h(\tilde{X}_0,\tilde{Z}_0)]\big) + \frac{k^\star(n-k^\star)}{n^{3/2}}\mathbb{E}[h(\tilde{X}_0,\tilde{Z}_0)]\| \\
&\ge \Big|\|\frac{1}{n^{3/2}}\big(U_{n,k^\star}(Y) - k^\star(n-k^\star)\mathbb{E}[h(\tilde{X}_0,\tilde{Z}_0)]\big)\| - \|\frac{k^\star(n-k^\star)}{n^{3/2}}\mathbb{E}[h(\tilde{X}_0,\tilde{Z}_0)]\|\Big| \\
&= \Big|\|\frac{1}{n^{3/2}}\sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n}\Big(h_1^{(Z)}(X_i) - h_1^{(X)}(Z_j) + h_2^{(X,Z)}(X_i,Z_j) \\
&\qquad - \mathbb{E}[h_1^{(Z)}(\tilde{X}_0) - h_1^{(X)}(\tilde{Z}_0) + h_2^{(X,Z)}(\tilde{X}_0,\tilde{Z}_0)]\Big)\| - \|\frac{k^\star(n-k^\star)}{n^{3/2}}\mathbb{E}[h(\tilde{X}_0,\tilde{Z}_0)]\|\Big| \\
&\ge \Bigg|\|\frac{1}{n^{3/2}}\sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n}\Big(h_1^{(Z)}(X_i) - h_1^{(X)}(Z_j) - 2\mathbb{E}[h(\tilde{X}_0,\tilde{Z}_0)]\Big)\| \\
&\qquad - \|\frac{1}{n^{3/2}}\sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n}\Big(h_2^{(X,Z)}(X_i,Z_j) - \mathbb{E}[h_2^{(X,Z)}(\tilde{X}_i,\tilde{Z}_j)]\Big)\| \\
&\qquad - \|\frac{k^\star(n-k^\star)}{n^{3/2}}\mathbb{E}[h(\tilde{X}_0,\tilde{Z}_0)]\|\Bigg|.
\end{aligned}$$

By Corollary 2.1 we know that the first part

$$\|\frac{1}{n^{3/2}}\sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n}\big(h_1^{(Z)}(X_i) - h_1^{(X)}(Z_j) - 2\mathbb{E}[h(\tilde{X}_0,\tilde{Z}_0)]\big)\|$$

is stochastically bounded. And by Proposition 2.5 it holds that

$$\begin{aligned}
&\|\frac{1}{n^{3/2}}\sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n} h_2^{(X,Z)}(X_i,Z_j) - \mathbb{E}[h_2^{(X,Z)}(\tilde{X}_i,\tilde{Z}_j)]\| \\
&\le \max_{1\le n_1<n}\frac{1}{n^{3/2}}\|\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n} h_2^{(X,Z)}(X_i,Z_j) - \mathbb{E}[h_2^{(X,Z)}(\tilde{X}_i,\tilde{Z}_j)]\| \xrightarrow{n\to\infty} 0.
\end{aligned}$$

But since $\mathbb{E}[h(\tilde{X}_0, \tilde{Z}_0)] \neq 0$ the last part diverges to infinity:

$$\|\frac{1}{n^{3/2}}k^\star(n-k^\star)\mathbb{E}[h(\tilde{X}_0, \tilde{Z}_0)]\| \approx \|\sqrt{n}\lambda^\star(1-\lambda^\star)\mathbb{E}[h(\tilde{X}_0, \tilde{Z}_0)]\| \xrightarrow{n\to\infty} \infty.$$

And thus $\max\limits_{1\leq k\leq n} \|\frac{1}{n^{3/2}}U_{n,k}(Y)\| \xrightarrow{n\to\infty} \infty.$ □

2.5 Bootstrap

Since theoretical values of the limit distribution of our test-statistic cannot be easily calculated, we perform a bootstrap to find critical values for a test-decision. A simple way to bootstrap U-statistics is to resample the observations $X_1, ..., X_n$ and plug-in the derived bootstrap-observations $X_1^\star, ..., X_n^\star$ into the U-statistic to get the bootstrap variant $\sum_{i=1}^{k}\sum_{j=k+1}^{n} h(X_i^\star, X_j^\star)$ of the U-statistic. In the case of i.i.d. observations one can produce a bootstrap sample by drawing with replacement from the observations. In Janssen [1997] can be found an overview of various consistency results for non-degenerate U-statistics in the independent case. For dependent observations, drawing single observations with replacement does not preserve the dependence structure of the random variables. To transfer the dependence structure of the sample $X_1, ..., X_n$ to the bootstrap sample at least section-wise, block bootstrap methods were established. The book of Lahiri [2003] gives a broad and fundamental overview of block bootstrap methods for dependent data. For a block bootstrap, blocks of $l \geq 1$ consecutive observations $X_i, ..., X_{i+l-1}$, which are drawn and placed one after the other, form the bootstrap sample. Inside each block, the dependency of the original sample is kept. l is naturally called *blocklength.*

Variants of the block bootstrap are for example the moving block bootstrap introduced by Künsch [1989], where blocks are drawn from the set $\{(X_1, ..., X_l), (X_2, ..., X_{l+1}), ..., (X_{n-l+1}, ..., X_n)\}$, or the non-overlapping block bootstrap of Carlstein [1986], where blocks are drawn from the set $\{(X_1, ..., X_l), (X_{l+1}, ..., X_{2l}), ..., (X_{(b-1)l+1}, ..., X_{bl})\}$, with b the largest integer such that $lb \leq n$.

Dehling and Wendler [2010] proved consistency for the bootstrap version of U-statistics for weakly dependent data, using a circular bootstrap. This is essentially an extension of the moving block bootstrap, where blocks are drawn from $\{(X_1, ..., X_l), ..., (X_{n-l+1}, ..., X_n), ..., (X_n, ..., X_{n+l})\}$ with $X_{n+i} = X_i$. Consistency results for U-statistics using the non-overlapping block bootstrap for near-epoch dependent random variables are given in Sharipov et al. [2016b].

An alternative method to transfer the dependence structure of $X_1, ..., X_n$ to $X_1^\star, ..., X_n^\star$, is to use the idea of the dependent wild bootstrap, which was first introduced by Shao [2010], for U-statistics. The original dependent wild bootstrap sample is defined as $X_t^\star = \bar{X}_n + (X_t - \bar{X}_n)\varepsilon_t$, $t = 1, ..., n$, where $\bar{X}_n = n^{-1}\sum_{t=1}^{n} X_t$ and $\varepsilon_1, ..., \varepsilon_n$ are random variables chosen with a certain covariance structure but independent of the original sample $X_1, ..., X_n$. Leucht and Neumann [2013] based their dependent wild bootstrap for degenerate U-statistics on this idea, but created bootstrap samples $(h(X_i, X_j)^\star)_{1\leq i,j\leq n}$ directly instead of plugging in $X_1^\star, ..., X_n^\star$ into the kernel function.

This is incorporated in the definition of our bootstrap version of the non-degenerate U-statistic $U_{n,k}$. For this, let $(\varepsilon_{i,n})_{i\leq n, n\in\mathbb{N}}$ be a row-wise stationary triangular scheme of variables with $\mathcal{N}(0,1)$-marginal distribution (we often drop the second index for notational convenience: $\varepsilon_i = \varepsilon_{i,n}$). The bootstrap version of $U_{n,k}(X)$ is then

$$U^\star_{n,k}(X) := \sum_{i=1}^{k}\sum_{j=k+1}^{n} h(X_i, X_j)(\varepsilon_i + \varepsilon_j).$$

Consistency results for weakly dependent data are given by Leucht and Neumann [2013] for the degenerate case, but since our U-statistic is non-degenerate, we will provide new consistency results for $U^\star_{n,k}$ in the following, validating the usage of the bootstrap. Subsequently, we state the practical bootstrap procedure and construction of the multiplier variables $(\varepsilon_i)_{i\le n}$.

Theorem 2.3. *Let the assumptions of Theorem 2.1 hold for $(X_n)_{n\in\mathbb{Z}}$ and $h : H^2 \to H$. Assume that $(\varepsilon_{i,n})_{i\le n, n\in\mathbb{N}}$ is independent of $(X_n)_{n\in\mathbb{Z}}$, has standard normal marginal distribution and $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = w(|i-j|/q_n)$, where w is symmetric and continuous with $w(0) = 1$ and $\int_{-\infty}^{\infty} |w(t)|dt < \infty$ and $q_n \in \mathbb{R}$ is the so-called bandwidth. Assume that $q_n \xrightarrow{n\to\infty} \infty$ and $q_n/n \xrightarrow{n\to\infty} 0$ Then it holds that*

$$\left(\max_{1\le k<n} \frac{1}{n^{3/2}}\|U_{n,k}(X)\|, \max_{1\le k<n} \frac{1}{n^{3/2}}\|U^\star_{n,k}(X)\|\right) \xrightarrow{\mathcal{D}} \left(\sup_{\lambda\in[0,1]} \|W(\lambda) - \lambda W(1)\|, \sup_{\lambda\in[0,1]} \|W^\star(\lambda) - \lambda W^\star(1)\|\right),$$

where W and $W^\star$ are two independent, H-valued Brownian motions with covariance operator as in Theorem 2.1.

From this statement, it follows that the bootstrap is consistent and it can be evaluated using the Monte Carlo method: Several copies of the bootstrapped test statistic independent conditional on $X_1, .., X_n$ are generated, the empirical quantiles of the bootstrapped test statistics can be used as critical values for the test. For a deeper discussion on bootstrap validity, see Bücher and Kojadinovic [2019]. Of course, in practical applications, the function w and the bandwidth q_n have to be chosen.
The idea of the proof is similar to that of Theorem 2.1. We will show that the degenerate part with multiplier still vanishes and the linear part converge in distribution to the desired process. Additionally, we need some variance result for the linear part.

Proposition 2.6. *Let $(\varepsilon_i)_{i\le n, n\in\mathbb{N}}$ be a triangular scheme of random multiplier independent from $(X_i)_{i\in\mathbb{Z}}$, such that the moment condition $\mathbb{E}[|\varepsilon_i|^2] < \infty$ holds. Then under the Assumptions of Theorem 2.1, it holds that*

$$\max_{1\le k<n} \frac{1}{n^{3/2}} \Big\| \sum_{i=1}^{k}\sum_{j=k+1}^{n} h_2(X_i, X_j)(\varepsilon_i + \varepsilon_j) \Big\| \xrightarrow{a.s.} 0 \quad \text{for } n\to\infty.$$

Proof. The statement follows along the lines of the proofs of Lemmas 2.4 to 2.6 and Proposition 2.3. For this note that by the independence of $(\varepsilon_i)_{i\le n, n\in\mathbb{N}}$ and $(X_i)_{i\in\mathbb{Z}}$ and by Lemma 2.3

$$\mathbb{E}[\|h_2(X_i, X_{i+k+2l})(\varepsilon_i + \varepsilon_{i+k+2l}) - h_2(X_{i,l}, X_{i+k+2l,l})(\varepsilon_i + \varepsilon_{i+k+2l})\|^2]^{\frac{1}{2}}$$

$$
\begin{aligned}
&= \mathbb{E}[\|h_2(X_i, X_{i+k+2l}) - h_2(X_{i,l}, X_{i+k+2l,l})\|^2]^{\frac{1}{2}} \cdot \mathbb{E}[(\varepsilon_i + \varepsilon_{i+k+2l})^2]^{\frac{1}{2}} \\
&\le C\big(\sqrt{\varepsilon} + \beta_k^{\frac{\delta}{2(2+\delta)}} + (a_l\Phi(\varepsilon))^{\frac{\delta}{2(2+\delta)}}\big).
\end{aligned}
$$

From this, we can conclude that for any $n_1 < n_2 < n_3 < n_4 \le n$ and $l = \left\lfloor n^{\frac{3}{16}} \right\rfloor$:

$$
\mathbb{E}\Big[\Big(\sum_{n_1\le i\le n_2}\sum_{n_3\le j\le n_4}\big\|\big(h_2(X_i, X_j) - h_2(X_{i,l}, X_{j,l})\big)(\varepsilon_i + \varepsilon_j)\big\|\Big)^2\Big]^{\frac{1}{2}} \le C(n_4 - n_3)n^{\frac{1}{4}}
$$

as in Lemma 2.4. Similary, we obtain (making use of the independence of $(\varepsilon_i)_{i\le n, n\in\mathbb{N}}$ and $(X_i)_{i\in\mathbb{Z}}$ again)

$$
\begin{aligned}
&\mathbb{E}\big[\big\|\big(h_{2,l}(X_{i,l}, X_{j,l}) - h_2(X_{i,l}, X_{j,l})\big)(\varepsilon_i + \varepsilon_j)\big\|^2\big] \\
&= \mathbb{E}[\|h_{2,l}(X_{i,l}, X_{j,l}) - h_2(X_{i,l}, X_{j,l})\|^2]\mathbb{E}[(\varepsilon_i + \varepsilon_j)^2] \\
&\le C\big(\sqrt{\epsilon} + (a_l\Phi(\epsilon))^{\frac{\delta}{2(2+\delta)}}\big).
\end{aligned}
$$

And along the lines of the proof of Lemma 2.5 for any $n_1 < n_2 < n_3 < n_4 \le n$ and $l = \left\lfloor n^{\frac{3}{16}} \right\rfloor$:

$$
\mathbb{E}\Big[\Big(\sum_{n_1\le i\le n_2}\sum_{n_3\le j\le n_4}\big\|\big(h_{2,l}(X_{i,l}, X_{j,l}) - h_2(X_{i,l}, X_{j,l})\big)(\varepsilon_i + \varepsilon_j)\big\|\Big)^2\Big]^{\frac{1}{2}} \le C(n_4 - n_3)n^{\frac{1}{4}}.
$$

With the same type of argument, we also obtain the analogous result to Lemma 2.6: For any $n_1, n_2, n_3, n_4 \le n$ with $n_1 < n_2$, $n_3 < n_4$ and $l = \left\lfloor n^{\frac{3}{16}} \right\rfloor$ it holds that

$$
\mathbb{E}\Big[\Big(\Big\|\sum_{n_1\le i\le n_2}\sum_{n_3\le j\le n_4} h_{2,l}(X_{i,l}, X_{j,l})(\varepsilon_i + \varepsilon_j)\Big\|\Big)^2\Big] \le C(n_4 - n_3)(n_2 - n_1)n^{\frac{3}{8}}
$$

and then we can proceed as in the proof of Proposition 2.3. □

The following Lemma 2.11 and Proposition 2.7 and their proofs are taken from Wegner and Wendler [2022].

Lemma 2.11. *Under the assumptions of Theorem 2.3, for any $t_0 = 0 < t_1 < t_2, ..., t_k = 1$ and any $a_1, ..., a_k \in H$*

$$
\operatorname{Var}\Bigg[\frac{1}{\sqrt{n}}\sum_{j=1}^{k}\sum_{i=\lfloor nt_{j-1}\rfloor+1}^{\lfloor nt_j\rfloor}\langle a_j, h_1(X_i)\varepsilon_i\rangle\Big| X_1, ..., X_n\Bigg] \xrightarrow{\mathbb{P}} \operatorname{Var}\Bigg[\sum_{j=1}^{k}\langle a_j, W(t_j) - W(t_{j-1})\rangle\Bigg]
$$

Proof. To simplify the notation, we introduce a triangular scheme $V_{i,n} = \langle a_j, h_1(X_i)\rangle$ for $i = \lfloor nt_{j-1}\rfloor + 1, ..., \lfloor nt_j\rfloor$. By our assumptions, $\operatorname{Cov}(\varepsilon_i, \varepsilon_j) = w(|i-j|/q_n)$, so we obtain for the variance condition on $X_1, ..., X_n$:

$$
\operatorname{Var}\Bigg[\frac{1}{\sqrt{n}}\sum_{j=1}^{k}\sum_{i=\lfloor nt_{j-1}\rfloor+1}^{\lfloor nt_j\rfloor}\langle a_j, h_1(X_i)\varepsilon_i\rangle\Big| X_1, ..., X_n\Bigg]
$$

$$= \sum_{i=1}^{n} \sum_{l=1}^{n} V_{i,n} V_{l,n} \operatorname{Cov}(\varepsilon_i, \varepsilon_l) = \sum_{i=1}^{n} \sum_{l=1}^{n} V_{i,n} V_{l,n} w(|i-l|/q_n).$$

This is the kernel estimator for the variance, which is consistent even for heteroscedastic time series under the assumptions of De Jong and Davidson [2000]. The L_2-NED follows by Lemma 2.2. Note that the mixing coefficients for absolute regularity are larger than the strong mixing coefficients used by De Jong and Davidson [2000], so their mixing assumption follows directly from ours. □

Proposition 2.7. *Under the assumptions of Theorem 2.3, we have the weak convergence (in the space $D_{H^2}[0,1]$)*

$$\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} (h_1(X_i), h_1(X_i)\varepsilon_i) \right)_{t \in [0,1]} \xrightarrow{\mathcal{D}} (W(t), W^\star(t))_{t \in [0,1]}$$

in probability.

Proof. We have to proof finite dimensional convergence and tightness. As the tightness for the first component was already established in the proof of Theorem 1 of Sharipov et al. [2016a], we only have to deal with the second component. The tightness of the partial sum process of $(h_1(X_i)\varepsilon_i)_{i \in \mathbb{N}}$, can be shown along the lines of the proof of the same theorem: For this note that by the independence of $(\varepsilon_i)_{i \le n}$ and $X_1, ..., X_n$

$$\begin{aligned}
&|\mathbb{E}\left[\langle h_1(X_i)\varepsilon_i, h_1(X_j)\varepsilon_j \rangle \langle h_1(X_k)\varepsilon_k, h_1(X_l)\varepsilon_l \rangle\right]| \\
=\ &|\mathbb{E}\left[\langle h_1(X_i), h_1(X_j) \rangle \langle h_1(X_k), h_1(X_l) \rangle\right] E[\varepsilon_i \varepsilon_j \varepsilon_k \varepsilon_l]| \\
\le\ &3 \left|\mathbb{E}\left[\langle h_1(X_i), h_1(X_j) \rangle \langle h_1(X_k), h_1(X_l) \rangle\right]\right|.
\end{aligned}$$

The rest follows as in Lemma 2.24 of Borovkova et al. [2001] and in proof of Theorem 1 of Sharipov et al. [2016a].

For the finite dimensional convergence, we will show the weak convergence of the second component conditional on $(h_1(X_i)\varepsilon_i)_{i \in \mathbb{N}}$, because the weak convergence of the first component is already established in Proposition 2.2. By the continuity of the limit process, it is sufficient to study the distribution for $t_1, .., t_k \in \mathbb{Q} \cap [0,1]$ and by the Cramér-Wold-device and the separability of H, it is enough to show the convergence of the condition distribution of $\frac{1}{\sqrt{n}} \sum_{j=1}^{k} \sum_{i=\lfloor nt_{j-1} \rfloor+1}^{\lfloor nt_j \rfloor} \langle a_j, h_1(X_i)\varepsilon_i \rangle$ for $a_1, ..., a_k$ from a countable subset of H. Conditional on $X_1, ..., X_n$, the distribution of $\frac{1}{\sqrt{n}} \sum_{j=1}^{k} \sum_{i=\lfloor nt_{j-1} \rfloor+1}^{\lfloor nt_j \rfloor} \langle a_j, h_1(X_i)\varepsilon_i \rangle$ is Gaussian with expectation 0 and variance converging to the right limit in probability by Lemma 2.11.

Using a well-known characterisation of convergence in probability, for every subseries there is another subseries such that this convergence holds almost surely. So, we can construct a subseries that the almost sure convergence holds for all k, $t_1, .., t_k \in \mathbb{Q} \cap [0,1]$ and all $a_1, ..., a_k$ from the countable subset of H, so we can find a subseries such that the convergence of the finite-dimensional distributions holds almost surely. Thus, the finite-dimensional convergence of the conditional distribution holds in probability and the statement of the proposition is proved. □

Proof of Theorem 2.3. Because the convergence in distribution of $\max_{1\leq k<n} \frac{1}{n^{3/2}}||U_{n,k}||$ has already been established in Theorem 2.1, it is enough to proof the convergence in distribution of $\max_{1\leq k<n} \frac{1}{n^{3/2}}||U^\star_{n,k}||$ conditional on $X_1, ..., X_n$. For this, we apply Hoeffding's decomposition:

$$\frac{1}{n^{3/2}}U^\star_{n,k} = \frac{1}{n^{3/2}}\sum_{i=1}^{k}\sum_{j=k+1}^{n} h(X_i, X_j)(\varepsilon_i + \varepsilon_j)$$

$$= \frac{1}{n^{3/2}}\sum_{i=1}^{k}\sum_{j=k+1}^{n} (h_1(X_i) - h_1(X_j)(\varepsilon_i + \varepsilon_j) + \frac{1}{n^{3/2}}\sum_{i=1}^{k}\sum_{j=k+1}^{n} h_2(X_i, X_j)(\varepsilon_i + \varepsilon_j).$$

The second summand converges to 0 by Proposition 2.6. The first summand can be split into three parts by a short calculation:

$$\begin{aligned}
&\frac{1}{n^{3/2}}\sum_{i=1}^{k}\sum_{j=k+1}^{n} \big(h_1(X_i) - h_1(X_j)\big)(\varepsilon_i + \varepsilon_j)\\
&= \frac{1}{\sqrt{n}}\left(\sum_{i=1}^{k} h_1(X_i)\varepsilon_i + \frac{k}{n}\sum_{i=1}^{n} h_1(X_i)\varepsilon_i\right)\\
&\quad + \frac{1}{n^{3/2}}\sum_{i=1}^{k} h_1(X_i)\sum_{j=1}^{n}\varepsilon_j + \frac{1}{n^{3/2}}\sum_{i=1}^{n} h_1(X_i)\sum_{j=1}^{k}\varepsilon_j.
\end{aligned}$$

By Proposition 2.7 and the continuous mapping theorem, we have the weak convergence

$$\max_{1\leq k<n}\frac{1}{n^{3/2}}\left\|\frac{1}{\sqrt{n}}\Big(\sum_{i=1}^{k} h_1(X_i)\varepsilon_i + \frac{k}{n}\sum_{i=1}^{n} h_1(X_i)\varepsilon_i\Big)\right\| \xrightarrow{\mathcal{D}} \sup_{\lambda\in[0,1]}\|W^\star(\lambda) - \lambda W^\star(1)\|$$

conditional on $X_1, ..., X_n$. For the second part, note that

$$\mathrm{Var}\Big(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\Big) = \frac{1}{n^2}\sum_{i,j=1}^{n} w(|i-j|/q_n) \leq \frac{1}{n}\sum_{i=-n}^{n}|w(i/q_n)| \approx \frac{q_n}{n}\int_{-\infty}^{\infty}|w(x)|dx \to 0$$

for $n\to\infty$ by our assumptions on q_n. So $\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i \to 0$ in probability and

$$\max_{1\leq k\leq n}\Big|\frac{1}{n^{3/2}}\sum_{i=1}^{k} h_1(X_i)\sum_{j=1}^{n}\varepsilon_j\Big| = \max_{1\leq k\leq n}\Big|\frac{1}{n^{1/2}}\sum_{i=1}^{k} h_1(X_i)\Big|\frac{1}{n}\sum_{j=1}^{n}\varepsilon_j \xrightarrow{\mathbb{P}} 0$$

for $n\to\infty$ using the fact that $\frac{1}{n^{1/2}}\sum_{i=1}^{k} h_1(X_i)$ is stochastically bounded, see Proposition 2.2. For the third part, we consider increments of the partial sum and bound the variance of increments similarly as above by

$$\mathrm{Var}\Big(\sum_{i=l+1}^{k}\varepsilon_i\Big) \leq Ckq_n.$$

Because the ε_i are Gaussian, it follows that

$$\mathbb{E}\Big[\Big(\sum_{i=l+1}^{k}\varepsilon_i\Big)^4\Big] \leq C(kq_n)^2.$$

By Theorem 1 of Móricz [1976], we have

$$\mathbb{E}\Big[\max_{1\leq k\leq n}\Big(\sum_{i=1}^{k}\varepsilon_i\Big)^4\Big] \leq C(nq_n)^2$$

and $\frac{1}{n}\max_{1\leq k\leq n}|\sum_{i=1}^{k}\varepsilon_i| \to 0$ in probability because $q_n/n \to 0$. So,

$$\max_{1\leq k\leq n}\Big|\frac{1}{n^{3/2}}\sum_{i=1}^{n}h_1(X_i)\sum_{j=1}^{k}\varepsilon_j\Big| = \Big|\frac{1}{n^{1/2}}\sum_{i=1}^{n}h_1(X_i)\Big|\max_{1\leq k\leq n}\Big|\frac{1}{n}\sum_{j=1}^{k}\varepsilon_j\Big| \xrightarrow{n\to\infty} 0$$

which completes the proof. □

2.5.1 Bootstrap Algorithm

The practical procedure to find the critical value for significance level $\alpha \in (0,1)$ is the following:

- Calculate $h(X_i, X_j)$ for all $i < j$.
- Fix the number of bootstrap iterations m.
- For each of the bootstrap iterations $t = 1, ..., m$:
  - Calculate $h(X_i, X_j)(\varepsilon_i^{(t)} + \varepsilon_j^{(t)})$, where $(\varepsilon_i^{(t)})_{i<n}$ are random multiplier.
  - Calculate $U_{n,k}^{(t)} = \sum_{i=1}^{k}\sum_{j=k+1}^{n} h(X_i, X_j)(\varepsilon_i^{(t)} + \varepsilon_j^{(t)})$ for all $k < n$.
  - Find $\max\limits_{1\leq k<n} \frac{1}{n^{3/2}}\|U_{n,k}^{(t)}\|$.
- Identify the empirical α-quantile U_α of all $\max\limits_{1\leq k<n} \frac{1}{n^{3/2}}\|U_{n,k}^{(1)}\|, ..., \max\limits_{1\leq k<n} \frac{1}{n^{3/2}}\|U_{n,k}^{(m)}\|$.
- Calculate $U_{n,k} = \sum_{i=1}^{k}\sum_{j=k+1}^{n} h(X_i, X_j)$ for all $1 \leq k < n$.
- Test decision: If $\max\limits_{1\leq k<n} \frac{1}{n^{3/2}}\|U_{n,k}\| > U_\alpha$, reject the null hypothesis H_0.

To ensure a certain covariance structure within the multiplier (that fulfils the assumptions of Theorem 2.3), we calculate them as

$$(\varepsilon_i^{(t)})_{i\leq n} = A(\eta_i)_{i\leq n},$$

where $\eta_1, ..., \eta_i$ are i.i.d. $N(0,1)$-distributed and A is the square root of the quadratic spectral covariance matrix constructed with bandwidth-parameter q. That means $AA^t = B$, where B has the entries

$$B_{i,j} = v_{|i-j|} \qquad \forall\, 1 \leq i, j \leq n$$

with

$$
\begin{aligned}
v_0 &= 1\\
v_i &= \frac{25}{12\pi^2(i-1)^2/q^2}\left(\frac{\sin(\frac{6\pi(i-1)/q}{5})}{\frac{6\pi(i-1)/q}{5}} - \cos(\frac{6\pi(i-1)/q}{5})\right) \qquad \forall\, 1 \le i \le n-1.
\end{aligned}
$$

Using the quadratic spectral kernel for the construction of the covariance matrix is based on the recommendation of Rice and Shang [2017]. It turned out to be a good overall-candidate from weakly to highly correlated data in their simulation study of the performance of the long-run variance estimator for $(X_n)_{n\in\mathbb{Z}}$ based on their data-driven bandwidth. As the multiplier should mimic the dependence structure of the original series, the same kernel is used for their construction. In the next subsection, some more details about the connection of long-run variance and bandwidth is given.

2.5.2 Bandwidth and Long-run Variance

The choice of the bandwidth in the bootstrap is not trivial, as it can influence the result of the hypothesis test similarly to the blocklength for a block bootstrap. To illustrate this, we take a look at the close connections between the bandwidth (resp. blocklength) and the estimation of the long-run variance of a time series.
Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary time series in $\mathbb{R}$. Suppose for this excursus without loss of generality $\mathbb{E}[X_t] = 0$ and let $\gamma_k = \mathrm{Cov}(X_0, X_k)$. The long-run variance is defined as

$$\sigma_\infty^2 = \lim_{n\to\infty} \mathrm{Var}(\sqrt{n}\bar{X}_n) = \sum_{k=-\infty}^{\infty} \gamma_k.$$

The long-run variance can also be defined via the spectral density function

$$f(\lambda) = \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} \gamma_k \exp(-ik\lambda)$$

of the process $(X_t)_{t\in\mathbb{Z}}$ by

$$\sigma_\infty^2 = 2\pi f(0).$$

So, estimating σ_∞^2 is equivalent to estimating $f(0)$. The sample autocovariance at lag k, $|k| \le n-1$,

$$\hat{\gamma}_k = \frac{1}{n}\sum_{t=|k|+1}^{n} (X_t - \bar{X}_n)(X_{t-|k|} - \bar{X}_n),$$

where $\bar{X}_n = n^{-1}\sum_{t=1}^{n} X_t$, can be used to construct

$$\hat{f}_n(\lambda) = \frac{1}{2\pi}\sum_{k=1-n}^{n-1} a\left(\frac{k}{l}\right)\hat{\gamma}_k \cos(k\lambda),$$

where $a(\cdot)$ is a lag-window function and l the bandwidth. A function $a(\cdot)$ is called lag-window function (also called kernel but not to be confused with our function h)

if it is symmetric and continuous with $a(0) = 1$ and $\int_{-\infty}^{\infty} |a(t)|dt < \infty$. The long-run variance can then be estimated by

$$\hat{\sigma}_\infty^2 = 2\pi \hat{f}_n(0).$$

If $l \xrightarrow{n\to\infty} \infty$ and $l/n \xrightarrow{n\to\infty} 0$, $\hat{\sigma}_\infty^2$ is a consistent estimator of σ_∞^2, as for example explained by Lahiri [2003].
Alternatively, one can estimate σ_∞^2 via bootstrapping. Let $\bar{X}_n^\star = \frac{1}{n}\sum_{t=1}^n X_t^\star$ be the bootstrap sample mean. Then,

$$\hat{\sigma}_B^2 = \mathrm{Var}^\star(\sqrt{n}\bar{X}_n^\star)$$

is an estimator for σ_∞^2. $\mathrm{Var}^\star$ is the variance of the bootstrap random variable, given the original sample. For a moving block bootstrap, $\hat{\sigma}_B^2 = 2\pi \hat{f}_n(0)$ if $a(\cdot)$ is the Bartlett kernel $a(x) = (1 - |x|)\mathbf{1}_{\{|x|\leq 1\}}$ (see Politis [2003] for more details). This means, choosing the optimal bandwidth l is equivalent to choosing the optimal blocklength in the moving block bootstrap, since it is

$$\mathrm{Var}^\star(\sqrt{n}\bar{X}_n^\star) = l\Big(\frac{1}{N}\sum_{i=1}^N U_{i,MBB}^2 - (\frac{1}{N}\sum_{i=1}^N U_{i,MBB})^2\Big),$$

where $U_{i,MBB} = \frac{1}{l}(X_i + ... + X_{i+l-1})$, $i \geq 1$ are the averages of the moving blocks, $N = n - l + 1$ the number of blocks.
Using the dependent wild bootstrap instead of a block bootstrap, a similar connection can be established. Shao [2010] defines the bootstrap sample of the dependent wild bootstrap as $X_i^\star = \bar{X}_n + (X_i - \bar{X}_n)\varepsilon_i$, $i = 1, ..., n$ with $(\varepsilon_i)_{i\leq n}$ independent of $(X_i)_{i\leq n}$ and $\mathbb{E}[\varepsilon_i] = 0$, $\mathrm{Var}(\varepsilon_i) = 1$ for $i = 1, ..., n$. Now, let $(\varepsilon_i)_{i\leq n}$ be a stationary process with $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = a(\frac{i-j}{l})$, where $a(\cdot)$ is a kernel function and l the bandwidth parameter. Then, it holds again that $\hat{\sigma}_B^2 = 2\pi\hat{f}_n(0)$ for the dependent wild bootstrap, since

$$\begin{aligned}
\mathrm{Var}^\star(\sqrt{n}\bar{X}_n^\star) &= \frac{1}{n}\sum_{i,j=1}^n (X_i - \bar{X}_n)(X_j - \bar{X}_n)\,\mathrm{Cov}^\star(\varepsilon_i, \varepsilon_j) \\
&= \frac{1}{n}\sum_{h=1-n}^{n-1}\sum_{t=\max(1,1-h)}^{\min(n,n-h)} (X_t - \bar{X}_n)(X_{t+h} - \bar{X}_n) a\left(\frac{h}{l}\right) \\
&= \frac{1}{n}\sum_{h=1-n}^{n-1}\sum_{t=|h|+1}^{n} (X_t - \bar{X}_n)(X_{t-|h|} - \bar{X}_n) a\left(\frac{h}{l}\right) \\
&= 2\pi\hat{f}_n(0).
\end{aligned}$$

So, for the moving block bootstrap and the dependent wild bootstrap, it holds that $\hat{\sigma}_B^2 = \hat{\sigma}_\infty^2$. That is, finding an optimal estimator for the long-run variance is the dual problem to finding an optimal l.
For the sake of completeness, note that the equality $\hat{\sigma}_B^2 = \hat{\sigma}_\infty^2$ may not hold if we

use the dependent wild bootstrap for the non-degenerate $U_{n,k}$ as described in the beginning of Chapter 2.5, since it creates $(h(X_i, X_j)^\star)_{1\le i,j\le n}$ instead of a bootstrap sample $X_1^\star, ..., X_n^\star$. Similarly, $\hat{\sigma}_B^2 = \hat{\sigma}_\infty^2$ cannot be assured if a non-overlapping block bootstrap is used instead of a moving block bootstrap because it is

$$\mathrm{Var}(\sqrt{n}\bar{X}_n^\star) = l\Big(\frac{1}{b}\sum_{i=1}^{b} U_{i,NBB}^2 - \big(\frac{1}{b}\sum_{i=1}^{b} U_{i,NBB}\big)^2\Big),$$

where $U_{i,NBB} = \frac{1}{l}(X_{(i-1)l+1} + ... + X_{il})$, $i \ge 1$, are the averages of the non-overlapping blocks and b the number of blocks. In comparison to the variance from the moving block bootstrap, some blocks are "left out" by construction of the bootstrap. As a consequence, $\hat{\sigma}_B^2$ does not necessarily equal $2\pi\hat{f}_n(0)$ if the non-overlapping block bootstrap is used. Nevertheless, $\hat{\sigma}_B^2$ is still a consistent estimator for σ_∞^2 if $l \xrightarrow{n\to\infty} \infty$ and $l/n \xrightarrow{n\to\infty} 0$ (see for example Lahiri [2003] for more details on the consistency of the different estimators).

2.5.3 Choice of the Bandwidth - Algorithm

The practical procedure to evaluate $l = q_{adpt}$ for the multiplier variables used in the bootstrap version $U_{n,k}^\star$ given the sample $X_1, ..., X_n$ is derived from the procedure of Rice and Shang [2017]. Their data-driven proposal ensures consistency of the long-run variance estimator and minimises the asymptotic mean-squared normed error $\mathbb{E}[\|\hat{\sigma}_\infty^2 - \sigma_\infty^2\|^2]$. In Section 2.5.4, we see simulations that indicate this method is a reasonable choice.

- Calculate $\tilde{X}_1, ..., \tilde{X}_n$ where $\tilde{X}_i = \frac{1}{n-1}\sum_{j=1, j\neq i}^{n} h(X_i, X_j)$.
- Determine a starting value $q_0 = n^{1/5}$.
- Calculate matrices $V_k = \frac{1}{n}\sum_{i=1}^{n-(k-1)} \tilde{X}_i \otimes \tilde{X}_k$ for $k = 1, ..., q_0$, where $\otimes$ is the outer product.
- Compute $CP_0 = V_1 + 2\sum_{k=1}^{q_0-1} w(k, q_0)V_{k+1}$
  and $CP_1 = 2\sum_{k=1}^{q_0-1} k\, w(k, q_0)V_{k+1}$,
  where w is a kernel function, we use the quadratic spectral kernel
  $w(k,q) = \frac{25}{12\pi^2 k^2/q^2}\left(\frac{\sin(\frac{6\pi k/q}{5})}{\frac{6\pi k/q}{5}} - \cos(\frac{6\pi k/q}{5})\right)$.
- Receive the data adapted bandwidth

$$q_{adpt} = \left\lceil \left(\frac{3n\sum_{i=1}^{d}\sum_{j=1}^{d} CP_{1i,j}}{\sum_{i=1}^{d}\sum_{j=1}^{d} CP_{0i,j} + \sum_{j=1}^{d} CP_{0j,j}^2} \right)^{1/5} \right\rceil.$$

The bandwidth q_{adpt} resulting from this method reflects the underlying dependence structure of the functional time series: Under weak dependence, one gets a small bandwidth, while under stronger dependence, the bandwidth gets larger.
Note that the first step is only necessary since we bootstrap $h(X_i, X_j)$ instead of X_i. If X_i is bootstrapped directly, then the procedure above can be used with $\tilde{X}_i = X_i$.

2.5.4 Influence of the Bandwidth - A Short Simulation Study

In the following, we see the influence of l on the results of a hypothesis test. For this, we revise a simulation study of Choi and Shin [2020], who compare the performance of different block bootstrap mean break tests for panel data. We will show that the choice of l has an impact on the results and that in particular, the data-adapted method of Rice and Shang [2017] is reasonable to use.
The data model used by Choi and Shin [2020] is a panel data model

$$X_{i,t} = \mu_i + \delta_i \mathbf{1}_{t \geq t_0} + \varepsilon_{i,t},\ 1 \leq i \leq N,\ 1 \leq t \leq T,\ 1 < t_0 < T,$$

where $(\varepsilon_{i,t})_{1 \leq t \leq T}$ is a stationary process for each $1 \leq i \leq N$ with $\mathbb{E}[\varepsilon_{i,t}] = 0$. In this model, the mean of $X_{i,t}$ changes from μ_i to $\mu_i + \delta_i$ at an unknown time t_0. The considered hypothesis of no change in mean is tested against the alternative of a mean in change at some point t_0:

$$H_0 :\ \delta_i = 0 \text{ for all } i \in \{1, ..., N\} \quad \text{vs.} \quad H_1 :\ \delta_i \neq 0 \text{ for some } i \in 1, ..., N.$$

The test statistic we want to focus on was introduced by Sharipov et al. [2016a]:

$$J_{N,T} = \max_{1 \leq k \leq T} \frac{1}{\sqrt{T}} \Big\| \sum_{t=1}^{k} X_t - \frac{k}{T} \sum_{t=1}^{T} X_t \Big\|,$$

where $X_t = (X_{1,t}, ..., X_{N,t})^T$. For theoretical properties and convergence results we refer as well to Sharipov et al. [2016a]. Just note that $J_{N,T}$ resembles a CUSUM-statistic for panel data and recall that we can construct CUSUM-statistics as a special case of $U_{n,k}$ by using the kernel function $h(x, y) = x - y$.
For a given $\alpha \in (0, 1)$, the hypothesis of no break is rejected when $J_{N,T}$ is greater than the right empirical α-quantile of m bootstrap counterparts $J^{\star(j)}_{N,T}$, $k = 1, ..., m$, using a non-overlapping block bootstrap, such that

$$J^{\star(j)}_{N,T} = \max_{1 \leq k < bl} \frac{1}{\sqrt{bl}} \Big\| \sum_{t=1}^{k} X^{\star(j)}_t - \frac{k}{bl} \sum_{t=1}^{bl} X^{\star(j)}_t \Big\|,$$

where $X^{\star(j)}_1, ..., X^{\star(j)}_{bl}$ is the j-th bootstrap sample. We simulate data as described in Choi and Shin [2020] and evaluate the empirical size and power of the test described above, implemented with the blocklength $l = q_{adpt}$ proposed by Rice and Shang [2017]. Simulation results are taken from Wegner and Wendler [2021]. We compare the performance with the results of Choi and Shin [2020] for the same test statistic, while they used the blocklength $l = \hat{l}_{opt}$ recommended by Politis and White [2004]. It is constructed through the following idea: Let $X_1, ..., X_n$ be observations from a stationary real-valued sequence. Using a circular bootstrap, $l_{opt} = \left(\frac{6n}{4}\right)^{1/3} \left(\frac{G}{\sigma^2_\infty}\right)^{2/3}$ is the blocklength such that the mean squared error of $\hat{\sigma}^2_B$ is minimised, where $G = \sum_{k \in \mathbb{Z}} |k| \gamma(k)$. Since l_{opt} contains unknown parameters, especially σ^2_∞ itself, these are estimated. Instead of G it is used $\hat{G} = \sum_{k=-M}^{M} a(\frac{k}{M})|k|\hat{\gamma}(k)$ and σ_{∞}^2 is estimated by $\hat{\sigma}_{\infty,M}^2 = \sum_{k=-M}^{M} a(\frac{k}{M})\hat{\gamma}(k)$, where

$$a(x) = \begin{cases} 1 & \text{if } 0 \leq |x| \leq \frac{1}{2} \\ 2(1-|x|) & \text{if } \frac{1}{2} \leq |x| \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$

Now, $\hat{l}_{opt} = \left(\frac{6n}{4}\right)^{1/3}\left(\frac{\hat{G}}{\hat{\sigma}_{\infty,M}^2}\right)^{2/3}$. To make this usable for the panel data, Choi and Shin [2020] used $\hat{\gamma}(k)$ as the sample autocovariance at lag k of the series $R_1, ..., R_T$. For $t = 1, ..., T$ it is defined as $R_t = \sum_{i=1}^{N}(E_i - \bar{E})$, where $E_i = \left(\frac{X_{i,t}-\bar{X}_i}{sd_i}\right)^2$ and $\bar{E} = \frac{1}{N}\sum_{i=1}^{N} E_i$. Here, sd_i denotes the sample variance of $X_{i,1}, ..., X_{i,T}$ and $\bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{i,t}$ its sample mean. M is chosen to be the nearest integer to $log(T)$.
The essential difference between the two methods of choosing l is that $l = \hat{l}_{opt}$ minimise the MSE of $\hat{\sigma}_B^2$ (given a circular bootstrap is used) and uses the estimate $\hat{\sigma}_{\infty}^2$ for the calculation, while $l = q_{adpt}$ minimises the asymptotic mean-squared normed error of $\hat{\sigma}_{\infty}^2$ itself, without giving a statement about $\hat{\sigma}_B^2$.
Now, coming to the simulation, let

$$X_{i,t} = \delta_i \mathbf{1}_{t>t_0} + \varepsilon_{i,t},\ \varepsilon_{i,t} = \rho_i \varepsilon_{i,t-1} + \eta_{i,t},\ \eta_{i,t} = a_{i,t} + \beta_i f_t \quad i = 1, ..., N,\ t = 1, ..., T.$$

To evaluate empirical size and power for serially and/or cross-sectionally correlated panels, different combinations are simulated. For size analysis, the independent mean zero error terms $a_{i,t}$ and f_t are either standard normal or t_5-distributed. For the serial correlation parameter $\rho_i \in \{0.3, 0.5\}$ is considered. The cross-sectional correlation parameter is chosen as $\beta_i \in \{0.5, 2\}$. Note here, that f_t adds cross-sectional correlation as well.
For power study, $a_{i,t}$, f_t are considered to be standard normal distributed and the correlation parameters are chosen as $\rho_i = \beta_i = 0$. For δ_i two uniform distributions are considered: $\mathcal{U}(-\frac{1}{2}, \frac{1}{2})$ (cancelling break) and $\mathcal{U}(\frac{1}{10}, \frac{1}{2})$ (non-cancelling break). For the time of change, we study $t_0 = 0.3T$ and $t_0 = 0.5T$.
The following Tables 2.1 and 2.2 summarise the results of the simulations. $J^{\star CS}$ denotes the variant of Choi and Shin [2020] and $J^{\star RS}$ the variant with the blocklength of Rice and Shang [2017]. It should be briefly noted that both bootstrap equivalents, $J^{\star CS}$ and $J^{\star RS}$, are produced by using a non-overlapping block bootstrap. This may influence the optimality of $\hat{l}_{opt}$, as it is based on a circular bootstrap, and as we have seen in Section 2.5.2, $\hat{\sigma}_B^2$ is dependent on the choice of the bootstrap procedure.

Empirical size of level $\alpha = 0.05$ tests							
				$\mathcal{N}(0,1)$ error terms		t_5 error terms	
ρ	β	N	T	$J^{\star CS}$	$J^{\star RS}$	$J^{\star CS}$	$J^{\star RS}$
0.3	0	50	50	0.216	0.045	0.237	0.075
		50	100	0.252	0.033	0.230	0.032
		100	50	0.257	0.009	0.252	0.007
		100	100	0.254	0.006	0.242	0.007
		200	100	0.234	0	0.249	0
		100	1000	0.378	0.059	0.389	0.054
0.5	0	50	50	0.447	0.354	0.438	0.329
		50	100	0.487	0.234	0.451	0.175
		100	50	0.454	0.228	0.465	0.192
		100	100	0.481	0.122	0.455	0.108
		200	100	0.473	0.070	0.473	0.045
		100	1000	0.659	0.272	0.646	0.279
0	0.5	50	50	0.024	0.048	0.018	0.038
		50	100	0.032	0.047	0.032	0.029
		100	50	0.016	0.038	0.014	0.030
		100	100	0.026	0.038	0.023	0.023
		200	100	0.034	0.023	0.023	0.015
		100	1000	0.054	0.037	0.056	0.028
0	2	50	50	0.041	0.113	0.043	0.068
		50	100	0.035	0.068	0.033	0.059
		100	50	0.039	0.063	0.032	0.063
		100	100	0.054	0.046	0.043	0.052
		200	100	0.045	0.037	0.045	0.053
		100	1000	0.067	0.043	0.065	0.050
0.3	0.5	50	50	0.123	0.139	0.157	0.111
		50	100	0.156	0.100	0.154	0.083
		100	50	0.157	0.091	0.175	0.072
		100	100	0.149	0.058	0.148	0.068
		200	100	0.154	0.061	0.155	0.065
		100	1000	0.174	0.086	0.154	0.069

Table 2.1: Empirical rejection frequencies for 1000 simulation runs of the tests based on $J^{\star CS}$ and $J^{\star RS}$ for different values of the correlation parameter (ρ_i, β_i) and sample sizes (N, T). Values for $J^{\star CS}$ are taken from Choi and Shin [2020].

Empirical power of level $\alpha = 0.05$ tests						
			$\delta_i \sim \mathcal{U}(-0.5, 0.5)$		$\delta_i \sim \mathcal{U}(0.1, 0.5)$	
t_0	N	T	$J^{\star CS}$	$J^{\star RS}$	$J^{\star CS}$	$J^{\star RS}$
$0.5T$	50	50	0.727	0.506	0.743	0.738
	50	100	0.998	0.987	0.996	1
	100	50	0.819	0.631	0.807	0.868
	100	100	1	1	0.999	1
	200	100	1	1	1	1
	100	1000	1	1	1	1
$0.3T$	50	50	0.310	0.173	0.274	0.340
	50	100	0.958	0.745	0.913	0.915
	100	50	0.308	0.090	0.323	0.281
	100	100	0.966	0.933	0.992	0.993
	200	100	0.957	0.995	0.994	1
	100	1000	1	1	1	1

Table 2.2: Empirical rejection frequencies of the tests based on $J^{\star CS}$ and $J^{\star RS}$ for different times of change t_0 and sample sizes (N, T). Values for $J^{\star CS}$ are taken from Choi and Shin [2020].

It can be observed (Table 2.1) that $J^{\star RS}$ is not oversized at least for mild serial dependence ($\rho = 0.3$) and the size distortion for stronger serial dependence ($\rho = 0.5$) is much less severe compared to $J^{\star CS}$. Under the alternative (Table 2.2), the power does not seem to be much influenced by the choice of the block length. Especially for larger N and T, both variants provide high empirical power. In this setting, the choice of l clearly makes a difference at least for the empirical size of the test, and is thus an important step.

2.6 Simulation Study

In this section we report the results of our simulation study. We compare size and power performance of our Wilcoxon-type test statistic with the well established CUSUM. To do so, we construct different data examples which are described below. Note that the bootstrap and the adapted bandwidth procedure described above works for the Spatial Sign kernel funtion $h(x,y) = (x-y)/\|x-y\|$ as well as for $h(x,y) = x - y$ used for CUSUM. To compare both statistics, we execute the test procedure for both simultaneously on each generated sample.

2.6.1 Generating Sample

We use a functional AR(1)-process on $[0,1]$, where the innovations are standard Brownian motions. We use an approximation on a finite grid with d grid points, if not indicated otherwise. To be more precise, we simulate data as follows:

$$
\begin{aligned}
&X_{-BI} = (\xi_1, \xi_1 + \xi_2, ..., \sum_{i=1}^{d} \xi_i)/\sqrt{d}, \quad \xi_i \text{ i.i.d. } \mathcal{N}(0,1)\text{-distributed.}\\
&X_t = a\,\Phi X_{t-1}^{\mathrm{T}} + W_t \quad \forall -BI < t \leq n,\\
&\text{where } \Phi \in \mathbb{R}^{d\times d} \text{ with entries } \Phi_{i,j} = \begin{cases} i/d^2 & i \leq j \\ j/d^2 & i > j \end{cases} = \min(i,j)/d^2\\
&\text{and } W_t = (\xi_1^{(t)}, \xi_1^{(t)} + \xi_2^{(t)}, ..., \sum_{i=1}^{d} \xi_i^{(t)})/\sqrt{d}, \quad \xi_i^{(t)} \text{ i.i.d. } \mathcal{N}(0,1)\text{-distributed.}
\end{aligned}
$$

The scalar $a \in \mathbb{R}$ is an AR-parameter, we use $a = 1$. The first $(BI + 1)$ simulations are not used, these are so-called burn-in iterations. Through this simulation structure we achieve dependence within n and d. We consider $n = 200$ and $d = 100$.

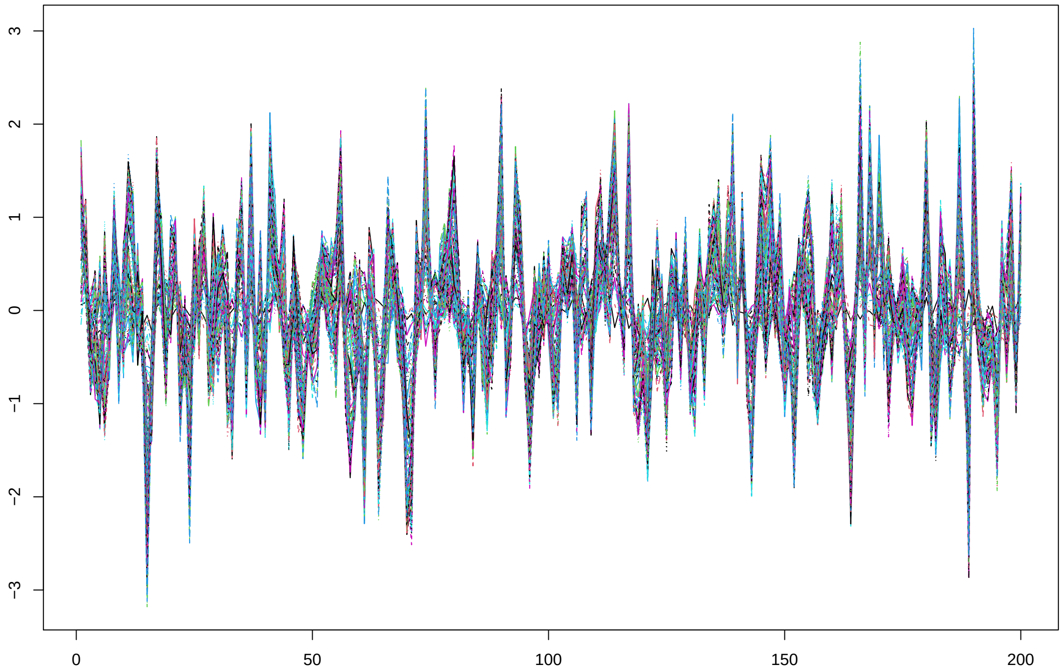


Figure 2.2: One sample of observations under the hypothesis simulated by the procedure described above. Each line corresponds to one $D = 1, ..., d$.

2.6.2 Size

To calculate the empirical size, the data simulation and test procedure via bootstrap is repeated $S = 3000$ times with $m = 1000$ bootstrap repetitions. We count the number of times the hypothesis was rejected both for the CUSUM-type and the Wilcoxon-type statistic. To analyse how good the test statistics perform if outliers are present or if gaussianity is not given, we study two additional simulations:

- Data simulated as above, but with presence of outliers:

$$Y_i = \begin{cases} X_i & i \notin \{0.2n, 0.4n, 0.6n, 0.8n\} \\ 10X_i & i \in \{0.2n, 0.4n, 0.6n, 0.8n\} \end{cases}$$

- Data simulated similar to the above, but with $\xi_i, \xi_i^{(t)} \sim t_1 \, \forall i \leq d$, $-BI \leq t \leq n$, i.e. heavy tailed data.

As we can see in Table 2.3, Spatial Sign and CUSUM perform almost similarly under normality. In the presence of outliers and for heavy tailed data, CUSUM shows a slightly smaller size, whereas Spatial Sign is still not oversized. In summary, we note that Spatial Sign is neither oversized in all observed simulations.

Empirical size						
	Gaussian		outliers		heavy tails	
α	CUSUM	Spatial Sign	CUSUM	Spatial Sign	CUSUM	Spatial Sign
0.1	0.07967	0.07767	0.051	0.086	0.018	0.077
0.05	0.0333	0.032	0.0153	0.035	0.00267	0.02967
0.025	0.008	0.00083	0.0043	0.0123	0.0003	0.01
0.01	0.002	0.0023	0.001	0.00267	0	0.00167

Table 2.3: Empirical size of CUSUM and Spatial Sign for different significance level α.

2.6.3 Power

To evaluate the performance of the test statistics in presence of a change in mean, we construct four scenarios.

Scenario 1: Uniform jump of +0.3 after $n/2$ observations:

$$Y_i = \begin{cases} X_i & i < n/2 \\ X_i + 0.3u & i \geq n/2 \end{cases}$$

where $u = (1, ..., 1)^T$.

Scenario 2: Sinus-jump after $n/2$ of observations:

$$Y_i = \begin{cases} X_i & i < n/2 \\ X_i + \frac{1}{2\sqrt{2}}(\sin(\pi D/d))_{D \leq d} & i \geq n/2 \end{cases}.$$

Scenario 3: Uniform jump of +0.3 after $n/2$ observations in presence of outliers at $0.2n, 0.4n, 0.6n, 0.8n$:

$$Y_i = \begin{cases} X_i & i < n/2, i \notin \{0.2n, 0.4n\} \\ 10X_i & i \in \{0.2n, 0.4n\} \\ X_i + 0.3u & i \geq n/2, i \notin \{0.6n, 0.8n\} \\ 10X_i + 0.3u & i \in \{0.6n, 0.8n\} \end{cases}$$

where $u = (1, ..., 1)^T$.

Scenario 4: Heavy tails: In the simulation of $(X_i)_{i \leq n}$ we use $\xi_i, \xi_i^{(t)} \sim t_1$ (Cauchy distributed) $\forall i \leq d, -BI < t \leq n$ and a uniform jump of +5 after $n/2$ observations.

As we can also see in Figure 2.3, the four scenarios resemble different difficulties that can occur in a sample. Scenarios 1 and 2 only have rather small changes in mean that may be difficult to detect. Scenario 3 has some additional outliers, while the change in Scenario 4 is rather large but the heavy tails may overshadow this.

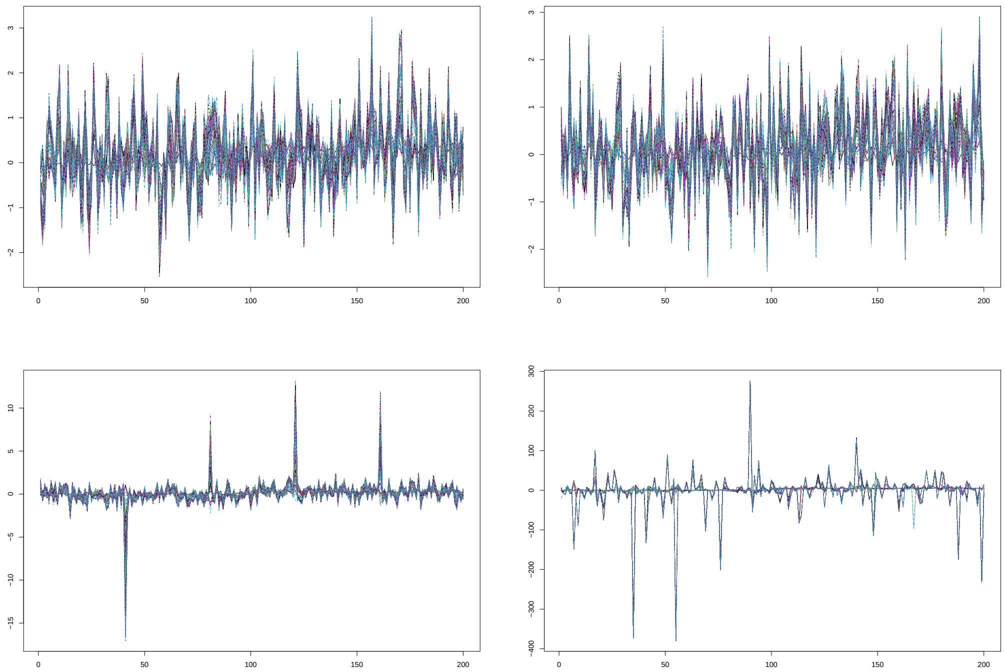


Figure 2.3: One sample of observations for each of the Scenarios 1 (top left), 2 (top right), 3 (bottom left) and 4 (bottom right) simulated by the procedure described above.

As in the analysis under the hypothesis, we chose $m = 1000$ bootstrap repetitions. The data simulation and test procedure via bootstrap is repeated $S = 3000$ times for each scenario and the number of times the hypothesis was rejected is counted to calculate the empirical power. To compare our test-statistic with CUSUM, we calculate the Spatial Sign and CUSUM simultaneously in each simulation run.
Comparing the size-power-plots for both test statistics (Figure 2.4), we see that Spatial Sign outperforms CUSUM clearly in Scenarios 1 and 2. For these two scenarios with a jump after $n/2$ of the observations, Spatial Sign provides lower empirical size and at the same time higher empirical power. In Scenario 1, we see that Spatial Sign provides empirical power larger than 0.9 for $\alpha \in \{0.1, 0.05\}$. For smaller α, the empirical power declines but not as drastically as for CUSUM, which provides for $\alpha = 0.01$ only empirical power of about 0.4, while Spatial Sign still shows empirical power greater than 0.6. In the second scenario we see a smaller empirical power for both statistics compared to the first scenario. Nevertheless, the empirical power of Spatial Sign is for all observed α greater than the one of CUSUM.

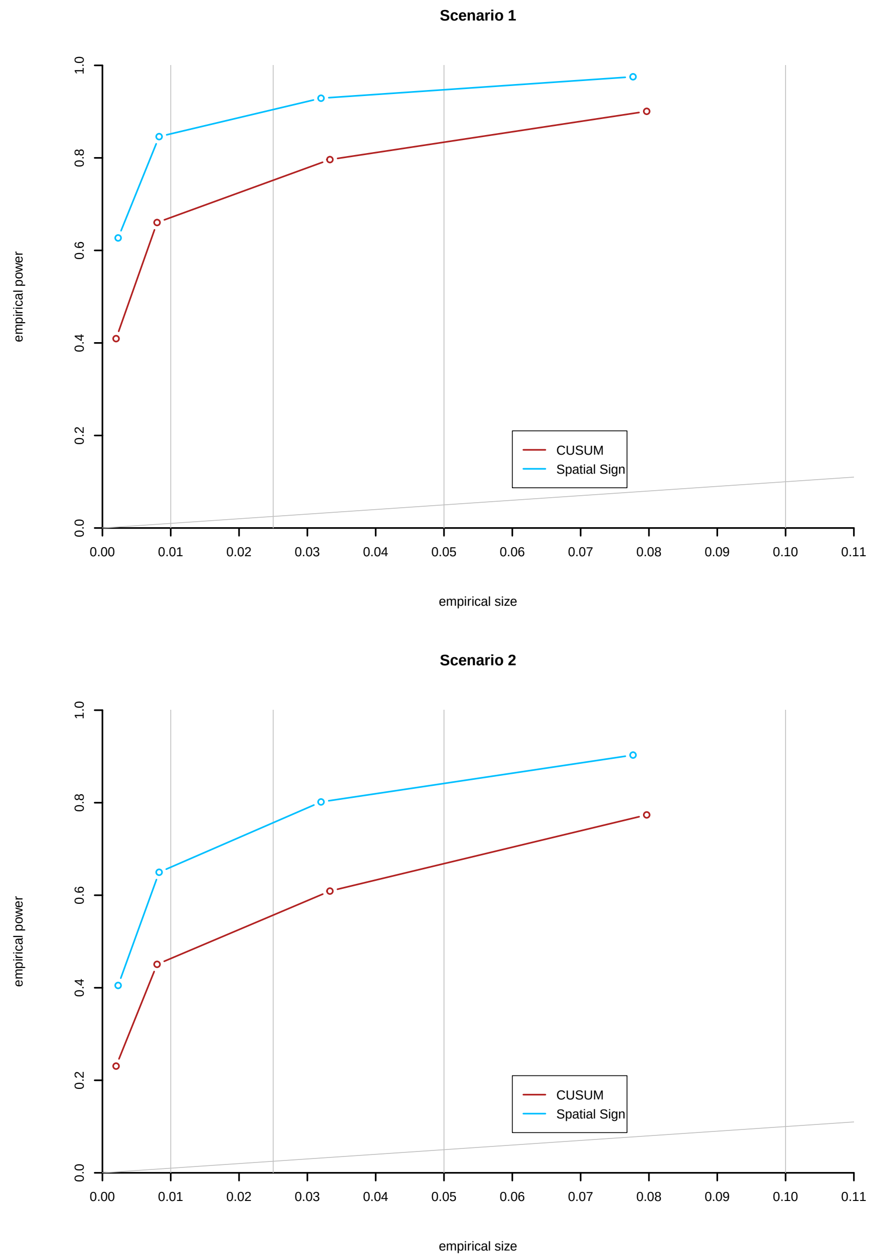


Figure 2.4: Size-Power-Plot for CUSUM and Spatial Sign statistic, Scenario 1 and 2.

In the third scenario, the jump with outliers in the data, we see that CUSUM shows a lower empirical size than Spatial Sign (Figure 2.5). On the other hand, Spatial Sign shows clearly more empirical power. For larger $\alpha \in \{0.1, 0.05\}$, the empirical power of Spatial Sign is over 0.9 while CUSUM just provides an empirical power of about 0.6 for $\alpha = 0.1$.

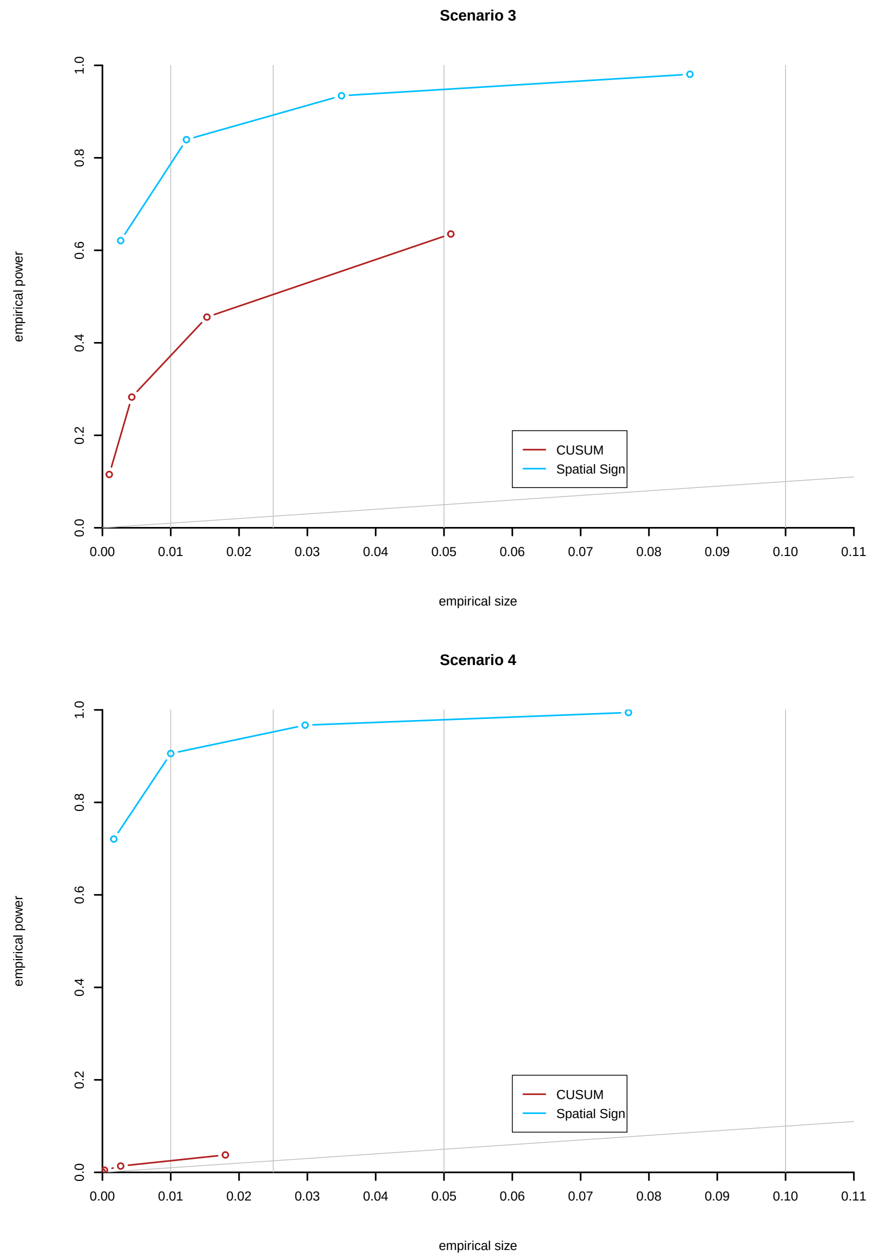


Figure 2.5: Size-Power-Plot for CUSUM and Spatial Sign statistic, Scenario 3 and 4.

In Scenario 4, we see that CUSUM barely provides any empirical power at all (Figure 2.5). Even for $\alpha = 0.1$ CUSUM shows an empirical power < 0.04. In heavy contrast, Spatial Sign shows relatively large empirical power, being greater than 0.9 for $\alpha \geq 0.025$.
For exact values of the empirical power in each scenario, see Table 2.4.
Typically, change-point tests lose power, if the change-point lies closer to the beginning (or symmetrically to the end) of the observations. To compare the effect of such

Empirical power								
	Scenario 1		Scenario 2		Scenario 3		Scenario 4	
α	CUSUM	Spatial Sign	CUSUM	Spatial Sign	CUSUM	Spatial Sign	CUSUM	Spatial Sign
0.1	0.90667	0.97533	0.77367	0.903	0.63533	0.98067	0.03767	0.9943
0.05	0.79633	0.929	0.609	0.80167	0.45567	0.93433	0.01367	0.967
0.025	0.66033	0.846	0.45067	0.64967	0.28267	0.83933	0.00467	0.90567
0.01	0.40933	0.627	0.23067	0.405	0.11533	0.621	0.00067	0.72067

Table 2.4: Empirical power of CUSUM and Spatial Sign for different significance level α, Scenario 1-4.

change-point locations on the Wilcoxon-type statistic with the effect on CUSUM, we simulated data with the change-point after 30% resp. 15% of the observations. Another interesting question is how size and power of the test changes, if $d >> n$. This might happen in practice if we only have few observations given. We call these two additional problem sets Scenario 5 and 6.

Scenario 5: Uniform Jump of +0.3 after γn observations:

$$Y_i = \begin{cases} X_i & i < \gamma n \\ X_i + 0.3u & i \geq \gamma n \end{cases} \quad \text{with } \gamma = 0.3 \text{ and } \gamma = 0.15 \text{ resp.}$$

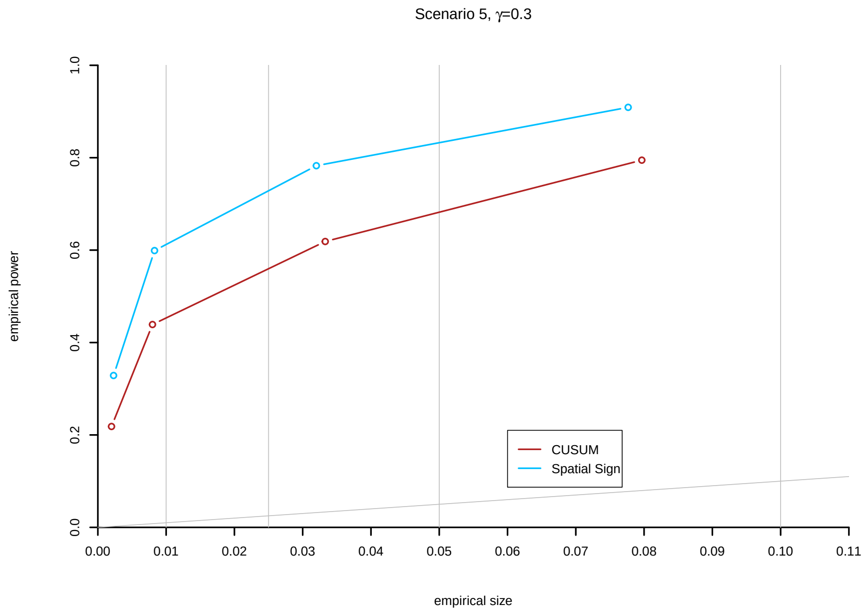
where $u = (1, ..., 1)^T$.

Scenario 6: As Scenario 1 but with $n = 150$, $d = 350$.

Empirical power						
	Scenario 5, $\gamma = 0.3$		Scenario 5, $\gamma = 0.15$		Scenario 6	
α	CUSUM	Spatial Sign	CUSUM	Spatial Sign	CUSUM	Spatial Sign
0.1	0.79467	0.909	0.3373	0.3723	0.75867	0.903
0.05	0.61867	0.78267	0.173	0.1933	0.58633	0.757
0.025	0.439	0.599	0.076	0.08433	0.391	0.552
0.01	0.21833	0.32867	0.024	0.027	0.14533	0.239

Table 2.5: Empirical power of CUSUM and Spatial Sign for different significance level α, Scenario 5 and 6.

The size-power-plots of Scenarios 5 and 6 (Figures 2.6 and 2.7) show that Spatial Sign suffers less loss in power than CUSUM if the change-point lies closer to the beginning of the observations or if d becomes larger than n.
In particular we see that in Scenario 5 with $\gamma = 0.3$ (Table 2.5), the power of both statistics is smaller than in Scenario 1 where the change-point is in the middle of the observations. Nevertheless, the empirical power of Spatial Sign is still larger than the empirical power of CUSUM and for $\alpha = 0.1$ Spatial Sign still provides empirical power of about 0.9. For $\gamma = 0.15$ we see a drastic decline in power for both statistics, with empirical power smaller than 0.4 even for $\alpha = 0.1$. Spatial Sign, nevertheless, keeps a small advantage over CUSUM in this scenario.



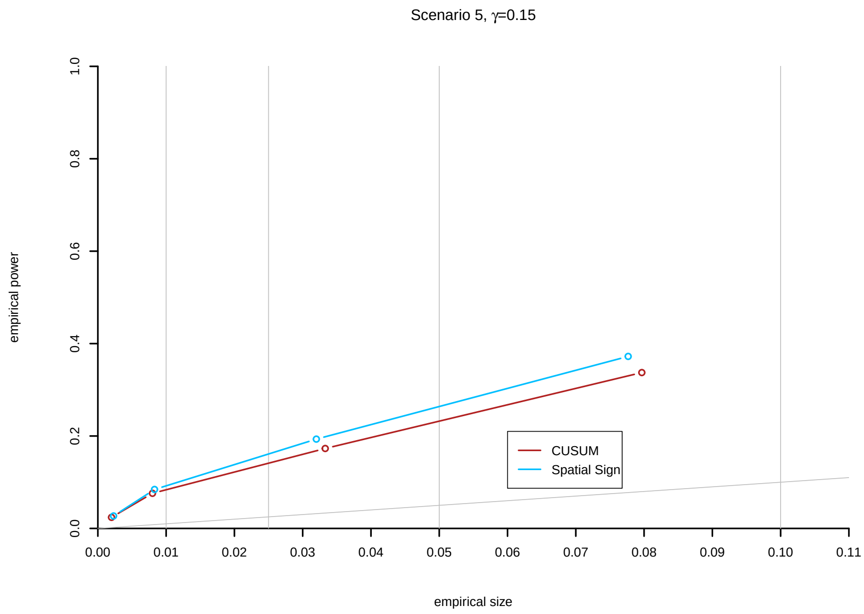


Figure 2.6: Size-Power-Plot for CUSUM and Spatial Sign statistic, Scenario 5.

Empirical Size - Scenario 6		
α	CUSUM	Spatial Sign
0.1	0.06733	0.064
0.05	0.025	0.01967
0.025	0.00467	0.0067
0.01	0.00067	0.00167

Table 2.6: Empirical size of CUSUM and Spatial Sign for different significance level α, Scenario 6.

In the last scenario we observe the situation of $d >> n$. For empirical size, we generated data as described in Chapter 2.6.1, but with $n = 150$ and $d = 350$ and received the values presented in Table 2.6. We see that the size of both statistics is even smaller than under Scenario 1. However, looking at the empirical power (Table 2.5), we see a reduction of power for both statistics compared to Scenario 1. Nevertheless, we can still observe that Spatial Sign provides a greater empirical power than CUSUM. Particularly for $\alpha = 0.1$, Spatial Sign still shows a power of about 0.9.

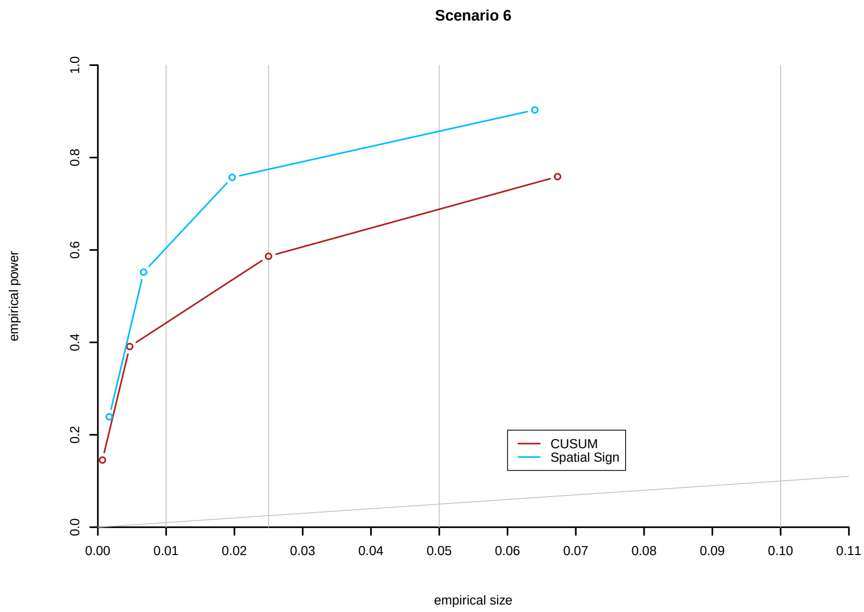


Figure 2.7: Size-Power-Plot for CUSUM and Spatial Sign statistic, Scenario 6.

2.7 Data Example

We look at data of 344 measuring stations of the Umweltbundesamt [2020] for air pollutants located all over Germany. The data of interest is the daily average of particulate matter with particles smaller than 10 μm (PM_{10}) measured in $\mu g/m^3$ from January 1, 2020 to May 31, 2020. This means we have $n = 152$ observations and treat the measurement of all stations on one day as data from $\mathbb{R}^{344}$.
Since the official restrictions of the German Government in course of the COVID-19 pandemic came into force on March 22, 2020, an often asked question was whether these restrictions (e.g. social distancing, closed gastronomy, closed/reduced work or work from home) had an effect on the air quality in Germany. This question comes from the assumption that the restrictions may lead to reduced traffic, resulting in reduced amount of particulate matter.
There are several publications from various countries studying the effects of lockdown measures on air pollution parameters like nitrogen oxides (NO, NO_2), ozone (O_3) and particulate matter (PM_{10}, $PM_{2.5}$). For example, Lian et al. [2020] investigated data from the city of Wuhan, or Zangari et al. [2020] for New York City.
Data for Berlin, as for 19 other cities around the world, are investigated by Fu et al. [2020]. They observed a decline in particular matter (PM_{10} and $PM_{2.5}$, significant for $PM_{2.5}$) in the period of lockdown. Although, the observed time period of one month is rather short (Mar. 17 to Apr. 19, 2020) and the findings for a densely populated city may not simply be transferred to the whole of Germany.
In contrast to that, we use data from measuring stations located across the whole country, to investigate an overall and long-lasting effect on particulate matter.

Looking at the empirical p-values of CUSUM and Spatial Sign test resulting from $m = 3000$ bootstrap iterations in Table 2.7, we see that with CUSUM H_0 is never rejected for any significance level $\alpha < 0.2$. But the Spatial Sign test rejects H_0 for significance level α larger than 0.03.

p-values	
CUSUM	Spatial Sign
0.226	0.027

Table 2.7: Empirical p-values for CUSUM and Spatial Sign test with data adapted bandwidth. $m = 3000$ bootstrap iterations were used.

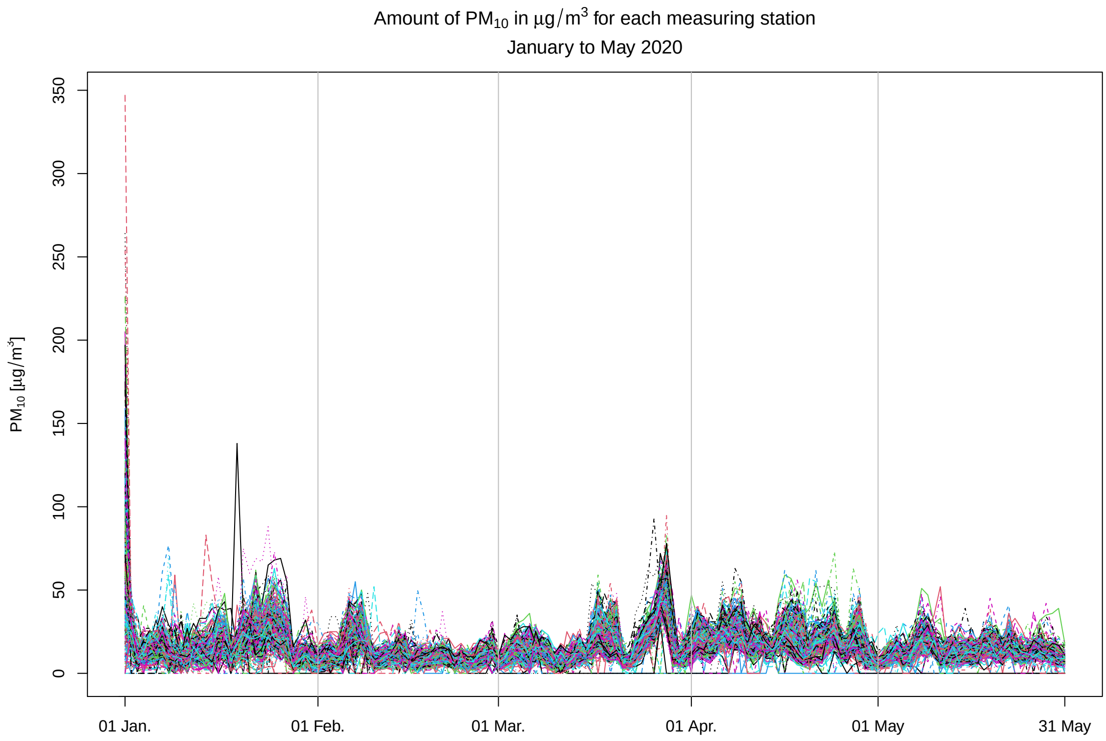


Figure 2.8: Daily average of PM_{10} in $\mu g/m^3$ for 344 measure stations from January 1, 2020 to May 31, 2020. Each line corresponds to one measuring station. The massive outlier on January 1 should result from New Year's fireworks.

Since the data exhibits a massive outlier located on January 1 (likely due to New Year's firework), we repeated the test procedure without the data of this day. We observed that the resulting p-value for Spatial Sign changed just slightly. Whereas the p-value for CUSUM decreased notably to around 0.08 (Table 2.8).
This example shows again that CUSUM is clearly more influenced by the outlier in the data than Spatial Sign. Evaluation showed that the data adapted bandwidth was set to $q_{adpt} = 3$ for both CUSUM and Spatial Sign for both scenarios.

p-values (data excluding Jan. 1)	
CUSUM	Spatial Sign
0.078	0.03

Table 2.8: Empirical p-values for CUSUM and Spatial Sign test with data adapted bandwidth for data excluding January 1, 2020. $m = 3000$ bootstrap iterations were used.

3 Time and Direction of Change

After receiving a significant result from the hypothesis test from Chapter 2, it is a comprehensible next step to estimate the time of change and its direction. In the course of this chapter, we will present consistent estimators for both of these problems.

3.1 Estimating the Time of Change

For real-valued i.i.d. settings, first works on the estimation of the time of change in some distribution parameter were already published in the 1970s and 1980s. Hinkley [1970] and Cobb [1978] addressed the problem via a maximum-likelihood approach. Based on the likelihood ratio test, Worsley [1986] and Siegmund [1988] constructed confidence intervals for the time of change for sequences from exponential families. In a different approach, Antoch et al. [1995] proposed to find the value $1 < k < n$ that maximises some CUSUM-like partial sum process and proved consistency for this estimator.
Moving away from the i.i.d set-up, there have been several developments for weakly dependent functional data in recent time. Aston and Kirch [2012] estimated the time of change in the framework of principal component analysis, while Aue et al. [2018] proposed a procedure based on the full functional information without dimension reduction.
For a change in the mean, a CUSUM-type test for dependent data was introduced by Kokoszka and Leipus [1998] and rates of convergence are given. Since CUSUM-type estimators encounter some robustness problems, Gerstenberger [2018] developed a Wilcoxon-type estimator, based on the test by Dehling et al. [2015], for near-epoch dependent processes.
Expanding this idea to $\mathbb{P}$-NED functional data in Hilbert spaces, we will show that the $1 \leq k < n$ that maximises the Wilcoxon-type statistic introduced in Chapter 2, is a robust and consistent estimator of the true time of change.

We recall the following model: We have a stationary, $H \otimes H$-valued sequence $(X_n, Z_n)_{n \in \mathbb{Z}}$ and we observe $Y_1, ..., Y_n$ with

$$Y_i = \begin{cases} X_i & \text{for } i \leq \lfloor n\lambda^\star \rfloor = k^\star \\ Z_i & \text{for } i > \lfloor n\lambda^\star \rfloor = k^\star \end{cases},$$

so $\lambda^\star \in (0,1)$ is the proportion of observations after which the change happens. The function $h : H^2 \to H$ denotes a kernel function. Assume that it is of such form that $\mathbb{E}[h(X_0, \tilde{Z}_0)] = \Delta \neq 0$.

Notation: We will often assume that n is eventually large enough that we can omit the floor function and write $k \approx \lambda n$ for some $\lambda \in (0,1)$.

We know by the previous chapters that the Wilcoxon-type test statistic $\max_{1\leq k<n} \|\frac{1}{n^{3/2}}U_{n,k}(Y)\|$ is suitable to detect change-points under the concepts and assumptions given in Chapter 2. The natural choice for an estimator for the time of change is given by

$$\hat{k} = \min\{k : \|\frac{1}{n^{3/2}}U_{n,k}(Y)\| = \max_{1\leq j<n} \|\frac{1}{n^{3/2}}U_{n,j}(Y)\|\},$$

since $\|\mathbb{E}[\frac{1}{n^{3/2}}U_{n,k}(Y)]\|$ achieves its maximum at $k = k^\star$, which is the true change point. This can be seen by a short calculation:
If $k \leq k^\star$,

$$\mathbb{E}[\frac{1}{n^{3/2}}U_{n,k}(Y)] = \frac{1}{n^{3/2}}\sum_{i=1}^{k}\sum_{j=k^\star+1}^{n} \mathbb{E}[h(X_i, Z_j)] + \frac{1}{n^{3/2}}\sum_{i=1}^{k}\sum_{j=k+1}^{k^\star} \mathbb{E}[h(X_i, X_j)]$$
$$= \frac{1}{n^{3/2}}k(n-k^\star)\Delta \approx \frac{1}{n^{3/2}}n\lambda \cdot n(1-\lambda^\star)\Delta = n^{1/2}\lambda(1-\lambda^\star)\Delta.$$

If $k \geq k^\star$,

$$\mathbb{E}[\frac{1}{n^{3/2}}U_{n,k}(Y)] = \frac{1}{n^{3/2}}\sum_{i=1}^{k^\star}\sum_{j=k+1}^{n} \mathbb{E}[h(X_i, Z_j)] + \frac{1}{n^{3/2}}\sum_{i=k^\star+1}^{k}\sum_{j=k+1}^{n} \mathbb{E}[h(Z_i, Z_j)]$$
$$= \frac{1}{n^{3/2}}k^\star(n-k)\Delta \approx \frac{1}{n^{3/2}}n\lambda^\star \cdot n(1-\lambda)\Delta = n^{1/2}\lambda^\star(1-\lambda)\Delta.$$

Here, it was used that $\mathbb{E}[h(X_i, Z_j)] = \Delta$, but $\mathbb{E}[h(X_i, X_j)] = \mathbb{E}[h(Z_i, Z_j)] = 0$. Both cases combined, read

$$\mathbb{E}[\frac{1}{n^{3/2}}U_{n,k}(Y)] \approx \begin{cases} \sqrt{n}\Delta(1-\lambda^\star)\lambda & \text{if } k \leq k^\star \Leftrightarrow \lambda \leq \lambda^\star \\ \sqrt{n}\Delta\lambda^\star(1-\lambda) & \text{if } k \geq k^\star \Leftrightarrow \lambda \geq \lambda^\star \end{cases},$$

where we can see that in both cases the maximum is attained at $\lambda^\star$, so

$$\max_{1\leq k\leq n} \|\mathbb{E}[\frac{1}{n^{3/2}}U_{n,k}(Y)]\| = \|\mathbb{E}[\frac{1}{n^{3/2}}U_{n,k^\star}(Y)]\| \approx \|\sqrt{n}\Delta(1-\lambda^\star)\lambda^\star\|$$

Taking up the notation of $\lambda^\star$, let $\hat{\lambda} \in (0,1)$ be the estimated proportion of observations after which the change happens, meaning that we can also write $\hat{k} = \lfloor n\hat{\lambda} \rfloor$. Note that while we assume that $\lambda^\star$ is fixed, due to the estimation, $\hat{\lambda} = \hat{\lambda}(n)$ might change with n.

First, we will see that a similar result as Theorem 2.1 holds for $\max\limits_{1\leq k\leq n} \|\frac{1}{n^{3/2}}\big(U_{n,k}(Y) - \mathbb{E}[U_{n,k}(Y)]\big)\|$, i.e. it converges in distribution:

Theorem 3.1. *Let $(X_n, Z_n)_{n\in\mathbb{Z}}$ be $\mathbb{P}$-NED on an absolutely regular sequence $(\zeta_n)_{n\in\mathbb{Z}}$ such that $a_k\Phi(k^{-8\frac{\delta+3}{\delta}}) = \mathcal{O}(k^{-8\frac{(\delta+3)(\delta+2)}{\delta^2}})$ and $\sum_{k=1}^\infty k^2\beta_k^{\frac{\delta}{4+\delta}} < \infty$ for some $\delta > 0$.*

Assume that $h : H^2 \to H$ is an antisymmetric kernel that fulfils the variation condition and is either bounded or has uniform $(4+\delta)$-moments under approximation for processes $(X_n)_{n\in\mathbb{Z}}$, $(Z_n)_{n\in\mathbb{Z}}$ and $(X_n, Z_n)_{n\in\mathbb{Z}}$, and that $\mathbb{E}[h(X_1, \tilde{Z}_1)] \neq 0$, were $\tilde{Z}_1$ is an independent copy of Z_1. Then it holds that

$$\max_{1\leq k<n} \| \frac{1}{n^{3/2}} \big(U_{n,k}(Y) - \mathbb{E}[U_{n,k}(Y)]\big) \|$$

is bounded in probability.

To prove this theorem, Hoeffding's decomposition is useful again. With the help of the decomposition for h, we can as well decompose $U_{n,k}(Y)$ into separate parts of $h_1^{(X)}, h_1^{(Z)}, h_2^{(X,X)}, h_2^{(Z,Z)}$ and $h_2^{(X,Z)}$. We have to consider two cases, depending on whether k is larger or smaller than $k^\star$.

If $k \leq k^\star$,

$$\begin{aligned}
U_{n,k}(Y) &= \sum_{i=1}^{k} \sum_{j=k+1}^{n} h(Y_i, Y_j) \\
&= \sum_{i=1}^{k} \sum_{j=k^\star+1}^{n} h(X_i, Z_j) + \sum_{i=1}^{k} \sum_{j=k+1}^{k^\star} h(X_i, X_j) \\
&= \sum_{i=1}^{k} \sum_{j=k^\star+1}^{n} \big(h_1^{(Z)}(X_i) - h_1^{(X)}(Z_j) + h_2^{(X,Z)}(X_i, Z_j)\big) \\
&\quad + \sum_{i=1}^{k} \sum_{j=k+1}^{k^\star} \big(h_1^{(X)}(X_i) - h_1^{(X)}(X_j) + h_2^{(X,X)}(X_i, X_j)\big) \\
&= (n-k^\star) \sum_{i=1}^{k} h_1^{(Z)}(X_i) - k \sum_{j=k+1}^{n} h_1^{(X)}(Z_j) + \sum_{i=1}^{k} \sum_{j=k^\star+1}^{n} h_2^{(X,Z)}(X_i, Z_j) \\
&\quad + (k^\star - k) \sum_{i=1}^{k} h_1^{(X)}(X_i) - k \sum_{j=k+1}^{k^\star} h_1^{(X)}(X_j) + \sum_{i=1}^{k} \sum_{j=k+1}^{k^\star} h_2^{(X,X)}(X_i, X_j).
\end{aligned}$$

If $k \geq k^\star$,

$$\begin{aligned}
U_{n,k}(Y) &= \sum_{i=1}^{k} \sum_{j=k+1}^{n} h(Y_i, Y_j) \\
&= \sum_{i=1}^{k^\star} \sum_{j=k+1}^{n} h(X_i, Z_j) + \sum_{i=k^\star+1}^{k} \sum_{j=k+1}^{n} h(Z_i, Z_j) \\
&= \sum_{i=1}^{k^\star} \sum_{j=k+1}^{n} \big(h_1^{(Z)}(X_i) - h_1^{(X)}(Z_j) + h_2^{(X,Z)}(X_i, Z_j)\big) \\
&\quad + \sum_{i=k^\star+1}^{k} \sum_{j=k+1}^{n} \big(h_1^{(Z)}(Z_i) - h_1^{(Z)}(Z_j) + h_2^{(Z,Z)}(Z_i, Z_j)\big)
\end{aligned}$$

$$= (n-k)\sum_{i=1}^{k^\star} h_1^{(Z)}(X_i) - k^\star \sum_{j=k+1}^{n} h_1^{(X)}(Z_j) + \sum_{i=1}^{k^\star}\sum_{j=k+1}^{n} h_2^{(X,Z)}(X_i, Z_j)$$
$$+ (n-k)\sum_{i=k^\star+1}^{k} h_1^{(Z)}(Z_i) - (k-k^\star)\sum_{j=k+1}^{n} h_1^{(Z)}(Z_j) + \sum_{i=k^\star+1}^{k}\sum_{j=k+1}^{n} h_2^{(Z,Z)}(Z_i, Z_j).$$

In the following, we will mostly focus on the case $k \geq k^\star$. This choice is arbitrary and the other case works completely analogously.

Proof of Theorem 3.1. Let $k \geq k^\star$. Using Hoeffding's decomposition, we get that

$$\begin{aligned}
&\max_{1\leq k<n} \Big\| \frac{1}{n^{3/2}} \big(U_{n,k}(Y) - \mathbb{E}[U_{n,k}(Y)]\big) \Big\| \\
= &\max_{1\leq k<n} \Big\| \frac{1}{n^{3/2}} (n-k) \sum_{i=1}^{k^\star} \big(h_1^{(Z)}(X_i) - \mathbb{E}[h_1^{(Z)}(X_i)]\big) \\
&\quad - \frac{1}{n^{3/2}} k^\star \sum_{j=k+1}^{n} \big(h_1^{(X)}(Z_j) - \mathbb{E}[h_1^{(X)}(Z_j)]\big) \\
&\quad + \frac{1}{n^{3/2}} \sum_{i=1}^{k^\star}\sum_{j=k+1}^{n} \big(h_2^{(X,Z)}(X_i, Z_j) - \mathbb{E}[h_2^{(X,Z)}(X_i, Z_j)]\big) \\
&\quad + \frac{1}{n^{3/2}} (n-k) \sum_{i=k^\star+1}^{k} \big(h_1^{(Z)}(Z_i) - \underbrace{\mathbb{E}[h_1^{(Z)}(Z_i)]}_{=0}\big) \\
&\quad - \frac{1}{n^{3/2}} (k-k^\star) \sum_{j=k+1}^{n} \big(h_1^{(Z)}(Z_j) - \underbrace{\mathbb{E}[h_1^{(Z)}(Z_j)]}_{=0}\big) \\
&\quad + \frac{1}{n^{3/2}} \sum_{i=k^\star+1}^{k}\sum_{j=k+1}^{n} \big(h_2^{(Z,Z)}(Z_i, Z_j) - \underbrace{\mathbb{E}[h_2^{(Z,Z)}(Z_i, Z_j)]}_{=0}\big) \Big\| \\
\leq &\max_{1\leq k<n} \Big\| \frac{n-k}{n^{3/2}} \sum_{i=1}^{k^\star} \big(h_1^{(Z)}(X_i) - \mathbb{E}[h_1^{(Z)}(X_i)]\big) - \frac{k^\star}{n^{3/2}} \sum_{j=k+1}^{n} \big(h_1^{(X)}(Z_j) - \mathbb{E}[h_1^{(X)}(Z_j)]\big) \\
&\qquad + \frac{n-k}{n^{3/2}} \sum_{i=k^\star+1}^{k} h_1^{(Z)}(Z_i) - \frac{k-k^\star}{n^{3/2}} \sum_{j=k+1}^{n} h_1^{(Z)}(Z_j) \Big\| \\
&+ \max_{1\leq k<n} \Big\| \frac{1}{n^{3/2}} \sum_{i=1}^{k^\star}\sum_{j=k+1}^{n} \big(h_2^{(X,Z)}(X_i, Z_j) - \mathbb{E}[h_2^{(X,Z)}(X_i, Z_j)]\big) \Big\| \\
&+ \max_{1\leq k<n} \Big\| \frac{1}{n^{3/2}} \sum_{i=k^\star+1}^{k}\sum_{j=k+1}^{n} h_2^{(Z,Z)}(Z_i, Z_j) \Big\|.
\end{aligned}$$

By using an analogue of Proposition 2.2 and Proposition 2.4 we get that for n large enough

$$\max_{1\leq k<n} \Big\| \frac{n-k}{n^{3/2}} \sum_{i=1}^{k^\star} \big(h_1^{(Z)}(X_i) - \mathbb{E}[h_1^{(Z)}(X_i)]\big) - \frac{k^\star}{n^{3/2}} \sum_{j=k+1}^{n} \big(h_1^{(X)}(Z_j) - \mathbb{E}[h_1^{(X)}(Z_j)]\big)$$

$$
\begin{aligned}
&\quad + \frac{n-k}{n^{3/2}} \sum_{i=k^\star+1}^{k} h_1^{(Z)}(Z_i) - \frac{k-k^\star}{n^{3/2}} \sum_{j=k+1}^{n} h_1^{(Z)}(Z_j) \big\| \\
&\le \max_{1\le k<n} \Big(\big\| \frac{n-k}{n^{3/2}} \sum_{i=1}^{k^\star} \big(h_1^{(Z)}(X_i) - \mathbb{E}[h_1^{(Z)}(X_i)]\big) \big\| \\
&\quad + \big\| \frac{k^\star}{n^{3/2}} \sum_{j=k+1}^{n} \big(h_1^{(X)}(Z_j) - \mathbb{E}[h_1^{(X)}(Z_j)]\big) \big\| \\
&\quad + \big\| \frac{n-k}{n^{3/2}} \sum_{i=k^\star+1}^{k} h_1^{(Z)}(Z_i) \big\| + \big\| \frac{k-k^\star}{n^{3/2}} \sum_{j=k+1}^{n} h_1^{(Z)}(Z_j) \big\| \Big) \\
&\approx \sup_{\lambda\in[0,1]} \Big(\big\| \frac{1-\lambda}{n^{1/2}} \sum_{i=1}^{\lfloor \lambda^\star n\rfloor} \big(h_1^{(Z)}(X_i) - \mathbb{E}[h_1^{(Z)}(X_i)]\big) \big\| \\
&\quad + \big\| \frac{\lambda^\star}{n^{1/2}} \sum_{j=\lfloor \lambda n\rfloor+1}^{n} \big(h_1^{(X)}(Z_j) - \mathbb{E}[h_1^{(X)}(Z_j)]\big) \big\| \\
&\quad + \big\| \frac{1-\lambda}{n^{1/2}} \sum_{i=\lfloor \lambda^\star n\rfloor+1}^{\lfloor \lambda n\rfloor} h_1^{(Z)}(Z_i) \big\| + \big\| \frac{\lambda-\lambda^\star}{n^{1/2}} \sum_{j=\lfloor \lambda n\rfloor+1}^{n} h_1^{(Z)}(Z_j) \big\| \Big) \\
&\xrightarrow{\mathcal{D}} \sup_{\lambda\in[0,1]} \Big(\|(1-\lambda)W_1(\lambda^\star)\| + \|\lambda^\star(W_2(1)-W_2(\lambda))\| + \|(1-\lambda)(\tilde{W}(\lambda)-\tilde{W}(\lambda^\star))\| \\
&\quad + \|(\lambda-\lambda^\star)(\tilde{W}(1)-\tilde{W}(\lambda))\| \Big).
\end{aligned}
$$

Where $(\tilde{W}(\lambda))_{\lambda\in[0,1]}$ is a Brownian motion with covariance operator $\langle \tilde{S}(x), y\rangle = \sum_{i=-\infty}^{\infty} \operatorname{Cov}(\langle h_1^{(Z)}(Z_0), x\rangle, \langle h_1^{(Z)}(Z_i), y\rangle)$ similar to the Brownian motion given in Proposition 2.2, but based on $(h_1^{(Z)}(Z_i))_{i\le n}$ instead of $(h_1^{(X)}(X_i))_{i\le n}$.
This implies that

$$
\begin{aligned}
&\max_{1\le k<n} \big\| \frac{n-k}{n^{3/2}} \sum_{i=1}^{k^\star} \big(h_1^{(Z)}(X_i) - \mathbb{E}[h_1^{(Z)}(X_i)]\big) - \frac{k^\star}{n^{3/2}} \sum_{j=k+1}^{n} \big(h_1^{(X)}(Z_j) - \mathbb{E}[h_1^{(X)}(Z_j)]\big) \\
&\qquad + \frac{n-k}{n^{3/2}} \sum_{i=k^\star+1}^{k} h_1^{(Z)}(Z_i) - \frac{k-k^\star}{n^{3/2}} \sum_{j=k+1}^{n} h_1^{(Z)}(Z_j) \big\| = \mathcal{O}_P(1).
\end{aligned}
$$

Furthermore, using Propositions 2.3 and 2.5 gives

$$
\begin{aligned}
&\max_{1\le k<n} \big\| \frac{1}{n^{3/2}} \sum_{i=1}^{k^\star} \sum_{j=k+1}^{n} \big(h_2^{(X,Z)}(X_i, Z_j) - \mathbb{E}[h_2^{(X,Z)}(X_i, Z_j)]\big) \big\| \\
&+ \max_{1\le k<n} \big\| \frac{1}{n^{3/2}} \sum_{i=k^*+1}^{k} \sum_{j=k+1}^{n} h_2^{(Z,Z)}(Z_i, Z_j) \big\| \xrightarrow{\text{a.s.}} 0 ,
\end{aligned}
$$

which proves the theorem.
Note that we need to do one additional approximation step in the proofs of Propo-

sitions 2.3 and 2.5, since the double sum looks a little bit different. For example, if Proposition 2.3 is used for $\max_{1\leq k<n}\|\frac{1}{n^{3/2}}\sum_{i=k^\star+1}^{k}\sum_{j=k+1}^{n}h_2^{(Z,Z)}(Z_i,Z_j)\|$, note that

$$
\begin{aligned}
&\mathbb{E}\Big[\Big(\max_{1\leq k<n}\|\sum_{i=k^\star+1}^{k}\sum_{j=k+1}^{n}h_2^{(Z,Z)}(Z_i,Z_j)\|\Big)^2\Big]^{1/2}\\
\leq{}&\mathbb{E}\Big[\Big(\max_{1\leq k<n}\sum_{i=k^\star+1}^{k}\sum_{j=k+1}^{n}\|h_2^{(Z,Z)}(Z_i,Z_j)-h_2^{(Z,Z)}(Z_{i,l},Z_{j,l})\|\Big)^2\Big]^{1/2}\\
&+\mathbb{E}\Big[\Big(\max_{1\leq k<n}\sum_{i=k^\star+1}^{k}\sum_{j=k+1}^{n}\|h_{2,l}^{(Z,Z)}(Z_{i,l},Z_{j,l})-h_2^{(Z,Z)}(Z_{i,l},Z_{j,l})\|\Big)^2\Big]^{1/2}\\
&+\mathbb{E}\Big[\Big(\max_{1<k<n}\|\sum_{i=k^\star+1}^{k}\sum_{j=k+1}^{n}h_{2,l}^{(Z,Z)}(Z_{i,l},Z_{j,l})\|\Big)^2\Big]^{1/2}\\
\leq{}&\mathbb{E}\Big[\Big(\max_{1\leq k<n}\sum_{i=1}^{k}\sum_{j=k+1}^{n}\|h_2^{(Z,Z)}(Z_i,Z_j)-h_2^{(Z,Z)}(Z_{i,l},Z_{j,l})\|\Big)^2\Big]^{1/2}\\
&+\mathbb{E}\Big[\Big(\max_{1\leq k<n}\sum_{i=1}^{k}\sum_{j=k+1}^{n}\|h_{2,l}^{(Z,Z)}(Z_{i,l},Z_{j,l})-h_2^{(Z,Z)}(Z_{i,l},Z_{j,l})\|\Big)^2\Big]^{1/2}\\
&+\mathbb{E}\Big[\Big(\max_{1\leq k<n}\|\sum_{i=1}^{k}\sum_{j=k+1}^{n}h_{2,l}^{(Z,Z)}(Z_{i,l},Z_{j,l})-\sum_{i=1}^{k^\star}\sum_{j=k+1}^{n}h_{2,l}^{(Z,Z)}(Z_{i,l},Z_{j,l})\|\Big)^2\Big]^{1/2}\\
\leq{}&\mathbb{E}\Big[\Big(\max_{1\leq k<n}\sum_{i=1}^{k}\sum_{j=k+1}^{n}\|h_2^{(Z,Z)}(Z_i,Z_j)-h_2^{(Z,Z)}(Z_{i,l},Z_{j,l})\|\Big)^2\Big]^{1/2}\\
&+\mathbb{E}\Big[\Big(\max_{1\leq k<n}\sum_{i=1}^{k}\sum_{j=k+1}^{n}\|h_{2,l}^{(Z,Z)}(Z_{i,l},Z_{j,l})-h_2^{(Z,Z)}(Z_{i,l},Z_{j,l})\|\Big)^2\Big]^{1/2}\\
&+2\mathbb{E}\Big[\Big(\max_{1\leq n_1<n}\max_{1\leq n_2<n}\sum_{i=1}^{n_1}\sum_{j=n_2}^{n}\|h_2^{(Z,Z)}(Z_i,Z_j)-h_2^{(Z,Z)}(Z_{i,l},Z_{j,l})\|\Big)^2\Big]^{1/2}
\end{aligned}
$$

and then we can proceed as in the proof of Proposition 2.3.
For $\max_{1\leq k<n}\|\frac{1}{n^{3/2}}\sum_{i=1}^{k^\star}\sum_{j=k+1}^{n}\big(h_2^{(X,Z)}(X_i,Z_j)-\mathbb{E}[h_2^{(X,Z)}(X_i,Z_j)]\big)\|$ we can do similar adjustments. $\square$

We will now formulate the main theorem of this section, stating that $\hat{\lambda}$ is an consistent estimator for $\lambda^\star$.

Theorem 3.2. *Let $(X_n,Z_n)_{n\in\mathbb{Z}}$ be P-NED on an absolutely regular sequence $(\zeta_n)_{n\in\mathbb{Z}}$ such that $a_k\Phi(k^{-8\frac{\delta+3}{\delta}})=\mathcal{O}(k^{-8\frac{(\delta+3)(\delta+2)}{\delta^2}})$ and $\sum_{k=1}^{\infty}k^2\beta_k^{\frac{\delta}{4+\delta}}<\infty$ for some $\delta>0$. Assume that $h:H^2\to H$ is an antisymmetric kernel that fulfils the variation condition and is either bounded or has uniform $(4+\delta)$-moments under approximation for processes $(X_n)_{n\in\mathbb{Z}}$, $(Z_n)_{n\in\mathbb{Z}}$ and $(X_n,Z_n)_{n\in\mathbb{Z}}$, and that $\mathbb{E}[h(X_1,\tilde{Z}_1)]\neq 0$, were $\tilde{Z}_1$ is an independent copy of Z_1.*
Then it holds that $|\lambda^\star-\hat{\lambda}|\xrightarrow{a.s.}0$ as $n\to\infty$.

For the proof of this, one preparatory lemma is needed:

Lemma 3.1. *Under the assumptions of Theorem 3.2 it holds that*

$$\max_{1\le k<n} \Big\| \frac{1}{n^2}\sum_{i=1}^{k}\sum_{j=k+1}^{n} (h(Y_i,Y_j) - \mathbb{E}[h(Y_i,Y_j)]) \Big\| \xrightarrow{a.s.} 0.$$

Proof. Let $k \ge k^\star$. Using Hoeffding's decomposition once again, we can write

$$\begin{aligned}
&\max_{1\le k<n} \Big\| \frac{1}{n^2}\sum_{i=1}^{k}\sum_{j=k+1}^{n} \big(h(Y_i,Y_j) - \mathbb{E}[h(Y_i,Y_j)]\big) \Big\| \\
\le\ &\max_{1\le k<n} \Big\| \frac{1}{n^2}(n-k)\sum_{i=1}^{k^\star} \big(h_1^{(Z)}(X_i) - \mathbb{E}[h_1^{(Z)}(X_i)]\big) \Big\| \\
&+ \max_{1\le k<n} \Big\| \frac{1}{n^2}k^\star \sum_{j=k+1}^{n} \big(h_1^{(X)}(Z_j) - \mathbb{E}[h_1^{(X)}(Z_j)]\big) \Big\| \\
&+ \max_{1\le k<n} \Big\| \frac{1}{n^2}(n-k) \sum_{i=k^\star+1}^{k} \big(h_1^{(Z)}(Z_i) - \mathbb{E}[h_1^{(Z)}(Z_i)]\big) \Big\| \\
&+ \max_{1\le k<n} \Big\| \frac{1}{n^2}(k-k^\star) \sum_{j=k+1}^{n} \big(h_1^{(Z)}(Z_j) - \mathbb{E}[h_1^{(Z)}(Z_j)]\big) \Big\| \\
&+ \max_{1\le k<n} \Big\| \frac{1}{n^2}\sum_{i=1}^{k^\star}\sum_{j=k+1}^{n} \big(h_2^{(X,Z)}(X_i,Z_j) - \mathbb{E}[h_2^{(X,Z)}(X_i,Z_j)]\big) \Big\| \\
&+ \max_{1\le k<n} \Big\| \frac{1}{n^2}\sum_{i=k^\star+1}^{k}\sum_{j=k+1}^{n} \big(h_2^{(Z,Z)}(Z_i,Z_j) - \mathbb{E}[h_2^{(Z,Z)}(Z_i,Z_j)]\big) \Big\|.
\end{aligned}$$

For the convergence of the linear parts, we can rewrite each term and use Birkhoff's ergodic theorem for stationary ergodic processes: Simply note that the sequence $(X_n, Z_n)_{n\in\mathbb{Z}}$ is a functional of a stationary absolutely regular sequence $(\zeta_n)_{n\in\mathbb{Z}}$, which means that it is stationary itself and furthermore ergodicity follows from the mixing condition of $(\zeta_n)_{n\in\mathbb{Z}}$.
For example, for the first term, write

$$\begin{aligned}
&\Big\| \frac{1}{n^2}(n-k)\sum_{i=1}^{k^\star} \big(h_1^{(Z)}(X_i) - \mathbb{E}[h_1^{(Z)}(X_i)]\big) \Big\| \\
\approx\ &\Big\| \frac{1}{n^2}n(1-\lambda)\sum_{i=1}^{\lfloor\lambda^\star n\rfloor} \big(h_1^{(Z)}(X_i) - \mathbb{E}[h_1^{(Z)}(X_i)]\big) \Big\| \\
\le\ &\Big\| (1-\lambda)\frac{1}{\lfloor\lambda^\star n\rfloor}\sum_{i=1}^{\lfloor\lambda^\star n\rfloor} \big(h_1^{(Z)}(X_i) - \mathbb{E}[h_1^{(Z)}(X_i)]\big) \Big\| \xrightarrow{n\to\infty} 0
\end{aligned}$$

Since this convergence holds a.s. uniformly, it holds for the maximum as well. The other linear parts converge a.s. to zero by similar considerations.
For the degenerate parts, convergence holds by Proposition 2.3 resp. 2.5:

$$\begin{aligned}
&\max_{1\le k<n} \Big\| \frac{1}{n^2} \sum_{i=k^\star+1}^{k} \sum_{j=k+1}^{n} \big(h_2^{(Z,Z)}(Z_i,Z_j) - \mathbb{E}[h_2^{(Z,Z)}(Z_i,Z_j)]\big) \Big\| \\
\le &\max_{k^\star<k<n} \Big\| \frac{1}{n^{3/2}} \sum_{i=k^\star+1}^{k} \sum_{j=k+1}^{n} h_2^{(Z,Z)}(Z_i,Z_j) \Big\|
\end{aligned}$$

since $h_2^{(Z,Z)}$ is degenerated. Here, we use Proposition 2.3 with a suitable adjustment in the proof, as explained in the proof of Theorem 3.1 and get that

$$\max_{k^\star<k<n} \Big\| \frac{1}{n^{3/2}} \sum_{i=k^\star+1}^{k} \sum_{j=k+1}^{n} h_2^{(Z,Z)}(Z_i,Z_j) \Big\| \xrightarrow{n\to\infty} 0 \text{ almost surely.}$$

For the second degenerate part, we use Proposition 2.5. □

We have now everything we need to prove Theorem 3.2.

Proof of Theorem 3.2. For n large enough, calculate the two expectations as done prior to the statement of Theorem 3.2:

$$\begin{aligned}
\Big\|\mathbb{E}\big[\frac{1}{n^{3/2}} U_{n,k^\star}(Y)\big]\Big\| &= \sqrt{n}\|\Delta\|\lambda^\star(1-\lambda^\star) \text{ and} \\
\Big\|\mathbb{E}\big[\frac{1}{n^{3/2}} U_{n,\hat{k}}(Y)\big]\Big\| &\approx \begin{cases} \sqrt{n}\|\Delta\|(1-\lambda^\star)\hat{\lambda} & \hat{\lambda}\le\lambda^\star \\ \sqrt{n}\|\Delta\|\lambda^\star(1-\hat{\lambda}) & \hat{\lambda}\ge\lambda^\star \end{cases}
\end{aligned}$$

So,

$$\begin{aligned}
&\Big\|\mathbb{E}\big[\frac{1}{n^{3/2}} U_{n,k^\star}(Y)\big]\Big\| - \Big\|\mathbb{E}\big[\frac{1}{n^{3/2}} U_{n,\hat{k}}(Y)\big]\Big\| \approx \begin{cases} \sqrt{n}\|\Delta\|(1-\lambda^\star)(\lambda^\star-\hat{\lambda}) & \hat{\lambda}\le\lambda^\star \\ \sqrt{n}\|\Delta\|\lambda^\star(\hat{\lambda}-\lambda^\star) & \hat{\lambda}\ge\lambda^\star \end{cases} \\
&= \sqrt{n}\|\Delta\|\bar{\lambda}|\lambda^\star-\hat{\lambda}| \text{ , where } \bar{\lambda} = \min\{\lambda^\star, 1-\lambda^\star\} \qquad (21)
\end{aligned}$$

Now,

$$\begin{aligned}
&\Big\|\frac{1}{n^{3/2}} U_{n,\hat{k}}(Y)\Big\| - \Big\|\frac{1}{n^{3/2}} U_{n,k^\star}(Y)\Big\| \\
\le &\Big\|\frac{1}{n^{3/2}} U_{n,\hat{k}}(Y) - \mathbb{E}\big[\frac{1}{n^{3/2}} U_{n,\hat{k}}(Y)\big]\Big\| + \Big\|\mathbb{E}\big[\frac{1}{n^{3/2}} U_{n,\hat{k}}(Y)\big]\Big\| \\
&+ \Big\|\frac{1}{n^{3/2}} U_{n,k^\star}(Y) - \mathbb{E}\big[\frac{1}{n^{3/2}} U_{n,k^\star}(Y)\big]\Big\| - \Big\|\mathbb{E}\big[\frac{1}{n^{3/2}} U_{n,k^\star}(Y)\big]\Big\| \\
\le &\, 2\max_{1\le k<n} \Big\|\frac{1}{n^{3/2}} U_{n,k}(Y) - \mathbb{E}\big[\frac{1}{n^{3/2}} U_{n,k}(Y)\big]\Big\| \\
&+ \Big\|\mathbb{E}\big[\frac{1}{n^{3/2}} U_{n,\hat{k}}(Y)\big]\Big\| - \Big\|\mathbb{E}\big[\frac{1}{n^{3/2}} U_{n,k^\star}(Y)\big]\Big\|
\end{aligned}$$

and rearranging give us

$$
\begin{aligned}
&\big\|\mathbb{E}\big[\frac{1}{n^{3/2}}U_{n,k^\star}(Y)\big]\big\| - \big\|\mathbb{E}\big[\frac{1}{n^{3/2}}U_{n,\hat{k}}(Y)\big]\big\| \\
&\le 2\max_{1\le k<n}\big\|\frac{1}{n^{3/2}}U_{n,k}(Y) - \mathbb{E}\big[\frac{1}{n^{3/2}}U_{n,k}(Y)\big]\big\| + \underbrace{\big\|\frac{1}{n^{3/2}}U_{n,k^\star}(Y)\big\|}_{\le\|\frac{1}{n^{3/2}}U_{n,\hat{k}}(Y)\|} - \big\|\frac{1}{n^{3/2}}U_{n,\hat{k}}(Y)\big\| \\
&\le 2\max_{1\le k<n}\big\|\frac{1}{n^{3/2}}U_{n,k}(Y) - \mathbb{E}\big[\frac{1}{n^{3/2}}U_{n,k}(Y)\big]\big\| \\
&\overset{(21)}{\Leftrightarrow} \sqrt{n}\|\Delta\|\bar{\lambda}|\lambda^\star - \hat{\lambda}| \le 2\max_{1\le k<n}\big\|\frac{1}{n^{3/2}}U_{n,k}(Y) - \mathbb{E}\big[\frac{1}{n^{3/2}}U_{n,k}(Y)\big]\big\| \\
&\Leftrightarrow |\lambda^\star - \hat{\lambda}| \le \frac{2}{\|\Delta\|\bar{\lambda}} 2\max_{1\le k<n}\big\|\frac{1}{n^{2}}U_{n,k}(Y) - \mathbb{E}\big[\frac{1}{n^{2}}U_{n,k}(Y)\big]\big\| \xrightarrow{a.s.} 0 \text{ for } n\to\infty
\end{aligned}
$$

by Lemma 3.1. And thus, $|\lambda^\star - \hat{\lambda}| \xrightarrow{a.s.} 0$ for $n\to\infty$. □

3.1.1 Shrinking Magnitude of Change

So far, we have dealt with a fixed difference in location, in other words, with a fixed magnitude of change. For some applications it is reasonable to allow the difference in location to shrink with larger sample size n

More formally, let $(X_i^{(n)}, Z_i^{(n)})_{i\le n,\, n\in\mathbb{N}}$ be a triangular scheme of $H\otimes H$-valued row-wise stationary random variables. For each $n\in\mathbb{N}$, a series of observations $Y_1^{(n)},...,Y_n^{(n)}$ is given by

$$
Y_i^{(n)} = \begin{cases} X_i^{(n)} & \text{for } i\le\lfloor n\lambda^\star\rfloor = k^\star \\ Z_i^{(n)} & \text{for } i>\lfloor n\lambda^\star\rfloor = k^\star \end{cases},
$$

assuming that $\mathbb{E}[h(X_0^{(n)}, \tilde{Z}_0^{(n)})] = \Delta_n \neq 0$ is dependent on n with $\|\Delta_n\| \xrightarrow{n\to\infty} 0$ but $\sqrt{n}\|\Delta_n\| \xrightarrow{n\to\infty} \infty$, which characterises the speed of $\|\Delta_n\|$ going to zero when n grows to infinity. Note that $k^\star = \lfloor n\lambda^\star\rfloor$ is proportional on n. We will proof that the estimator of the time of change

$$
\hat{k}(n) = \hat{k} = \min\big\{k : \|\frac{1}{n^{3/2}}U_{n,k}(Y^{(n)})\| = \max_{1\le j<n}\|\frac{1}{n^{3/2}}U_{n,j}(Y^{(n)})\|\big\}
$$

is still a consistent estimator of $k^\star$ with a rate of consistency dependent on Δ_n. Gerstenberger [2018] used a Wilcoxon-type estimator to formulate a result for real L_1-NED processes. We will extend and generalise the results for $\mathbb{P}$-NED functional processes in H.

Theorem 3.3. *Let $(X_i^{(n)}, Z_i^{(n)})_{i\le n, n\in\mathbb{Z}}$ be $\mathbb{P}$-NED, meaning it holds for every $n\in\mathbb{Z}$ that $(X_i^{(n)}, Z_i^{(n)})_{i\le n}$ is $\mathbb{P}$-NED on an absolutely regular sequence $(\zeta_i^{(n)})_{i\in\mathbb{Z}}$ such that $a_k^{(n)}\Phi^{(n)}(k^{-8\frac{\delta+3}{\delta}}) = \mathcal{O}(k^{-8\frac{(\delta+3)(\delta+2)}{\delta^2}})$ and $\sum_{k=1}^\infty k^2 \beta_k^{(n)\frac{\delta}{4+\delta}} < \infty$ for some $\delta>0$ uniformly in n. Assume that $h: H^2\to H$ is an antisymmetric kernel that fulfils the variation condition and is either bounded or has uniform $(4+\delta)$-moments under*

approximation for processes $(X_i^{(n)})_{i\leq n}$*,* $(Z_i^{(n)})_{i\leq n}$ *and* $(X_i^{(n)}, Z_i^{(n)})_{i\leq n}$*. Furthermore assume that* $\mathbb{E}[h(X_0^{(n)}, \tilde{Z}_0^{(n)})] = \Delta_n$*, where* $\tilde{Z}_0^{(n)}$ *is an independent copy of* $Z_0^{(n)}$ *and* $\|\Delta_n\| \xrightarrow{n\to\infty} 0$*,* $\sqrt{n}\|\Delta_n\| \xrightarrow{n\to\infty} \infty$*.*
Then it holds that $|k^\star - \hat{k}| = \mathcal{O}_P(\frac{1}{\|\Delta_n\|^2})$*.*

In order to prove this theorem, we will start by proving the following (slightly simpler) lemma:

Lemma 3.2. *Under the assumptions of Theorem 3.3, it holds that* $|k^\star - \hat{k}| = o_P(k^\star)$*.*

Proof. First of all, note that for $\epsilon > 0$ equivalently

$$|k^\star - k| \geq \epsilon k^\star \Leftrightarrow k \notin [k^\star(1-\epsilon), k^\star(1+\epsilon)]$$
$$|k^\star - k| \leq \epsilon k^\star \Leftrightarrow k \in [k^\star(1-\epsilon), k^\star(1+\epsilon)].$$

Using this, we have the following equivalent statements

$$\begin{aligned} &|k^\star - \hat{k}| = o_P(k^\star) \\ \Leftrightarrow &\forall \epsilon > 0 \ \mathbb{P}(|k^\star - \hat{k}| \leq \epsilon k^\star) \overset{n\to\infty}{\to} 1 \\ \Leftrightarrow &\forall \epsilon > 0 \ \mathbb{P}(\hat{k} \in [k^\star(1-\epsilon), k^\star(1+\epsilon)]) \overset{n\to\infty}{\to} 1. \end{aligned} \tag{22}$$

So, we want to show that with probability growing to 1, $\hat{k}$ lies in an arbitrary small neighbourhood of the true change-point $k^\star$. Or, in other words, the maximum of $\|U_{n,k}(Y^{(n)})\|$ is attained in the neighbourhood of $k^\star$. We will prove this by showing that *outside* this neighbourhood, $\|U_{n,k}(Y^{(n)})\|$ does not exceed $\|U_{n,k^\star}(Y^{(n)})\|$ (in probability) for $n \to \infty$:

$$\mathbb{P}\Big(\max_{k:|k-k^\star|\geq\epsilon k^*} \|U_{n,k}(Y^{(n)})\| < \|U_{n,k^\star}(Y^{(n)})\|\Big) \xrightarrow{n\to\infty} 1. \tag{23}$$

Then, (22) is a consequence of (23) by the following arguments:
Let $\epsilon > 0$ and assume (23) holds. By the definition, it is

$$\hat{k} = \hat{k}(n) = \underset{1\leq k\leq n}{\operatorname{argmax}} \|U_{n,k}(Y^{(n)})\|.$$

For the maximum over all k *inside* the neighbourhood of $k^\star$ it is

$$\max_{k:|k-k^\star|\leq\epsilon k^\star} \|U_{n,k}(Y^{(n)})\| \geq \|U_{n,k^\star}(Y^{(n)})\|.$$

On the other hand, for the maximum over all k *outside* the neighbourhood of $k^\star$ it is

$$\max_{k:|k-k^\star|\geq\epsilon k^\star} \|U_{n,k}(Y^{(n)})\| \leq \|U_{n,k^\star}(Y^{(n)})\| \text{ (in probability for } n\to\infty) \text{ by (23).}$$

This means, the maximum is indeed attained inside the neighbourhood of $k^\star$, i.e. it must be

$$
\begin{aligned}
&\|U_{n,\hat{k}}(Y^{(n)})\| = \max_{k:|k-k^\star|\leq \epsilon k^\star} \|U_{n,k}(Y^{(n)})\| \text{ (in probability for } n\to\infty)\\
&\Leftrightarrow \hat{k} \in [k^\star(1-\epsilon), k^\star(1+\epsilon)] \text{ (in probability for } n\to\infty)\\
&\Leftrightarrow \mathbb{P}\big(\hat{k} \in [k^\star(1-\epsilon), k^\star(1+\epsilon)]\big) \overset{n\to\infty}{\to} 1.
\end{aligned}
$$

It is left to show that (23) holds. We will do this by bounding the difference $\|U_{n,k^\star}(Y^{(n)})\| - \max_{k:|k-k^\star|>\epsilon k^\star} \|U_{n,k}(Y^{(n)})\|$ from below.
Starting with $U_{n,k^\star}$, adding expectations and using the reverse triangle inequality, yields

$$
\begin{aligned}
&\|U_{n,k^\star}(Y^{(n)})\|\\
&= \Big\|\sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n} h(Y_i^{(n)}, Y_j^{(n)}) - \mathbb{E}[h(Y_i^{(n)}, Y_j^{(n)})] + \sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n} \mathbb{E}[h(Y_i^{(n)}, Y_j^{(n)})]\Big\|\\
&\geq \Big|\Big\|\sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n} \mathbb{E}[h(X_i^{(n)}, Z_j^{(n)})]\Big\| - \Big\|\sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n} h(X_i^{(n)}, Z_j^{(n)}) - \mathbb{E}[h(X_i^{(n)}, Z_j^{(n)})]\Big\|\Big|\\
&\geq k^\star(n-k^\star)\|\Delta_n\| - \mathcal{O}_P(n^{3/2}).
\end{aligned}
$$

Here, we used Hoeffding's decomposition and Corollary 2.1 and Proposition 2.5 for the first part. For $\max_{k:|k-k^\star|>\epsilon k^\star} \|U_{n,k}(Y^{(n)})\|$ we have to consider two cases, $k < k^\star$ and $k > k^\star$. We will only consider the first case here, since the other case works very similarly.

$$
\begin{aligned}
&\max_{k:|k-k^\star|>\epsilon k^\star} \|U_{n,k}(Y^{(n)})\| = \max_{1\leq k<k^\star(1-\epsilon)} \|U_{n,k}(Y^{(n)})\|\\
&\leq \max_{1\leq k<k^\star(1-\epsilon)} \Big\|\sum_{i=1}^{k}\sum_{j=k+1}^{n} h(Y_i^{(n)}, Y_j^{(n)}) - \mathbb{E}[h(Y_i^{(n)}, Y_j^{(n)})]\Big\|\\
&\quad + \max_{1\leq k<k^\star(1-\epsilon)} \sum_{i=1}^{k}\sum_{j=k^\star+1}^{n} \|\mathbb{E}[h(X_i^{(n)}, Z_j^{(n)})]\|\\
&\quad + \max_{1\leq k<k^\star(1-\epsilon)} \sum_{i=1}^{k}\sum_{j=k+1}^{k^\star} \underbrace{\|\mathbb{E}[h(X_i^{(n)}, X_j^{(n)})]\|}_{=0}\\
&= \mathcal{O}_P(n^{3/2}) + k^\star(1-\epsilon)(n-k^\star)\|\Delta_n\|
\end{aligned}
$$

by using Theorem 3.1 for the first part. Combining this, we get for the difference

$$
\begin{aligned}
&\|U_{n,k^\star}(Y^{(n)})\| - \max_{k:|k-k^\star|>\epsilon k^\star} \|U_{n,k}(Y^{(n)})\|\\
&\geq \epsilon k^\star(n-k^\star)\|\Delta_n\| - \mathcal{O}_P(n^{3/2}) = \epsilon\delta_n - \mathcal{O}_P(n^{3/2}),
\end{aligned}
$$

where we define $\delta_n := k^\star(n-k^\star)\|\Delta_n\|$.
Now, it holds that $\delta_n^{-1} = o(n^{3/2})$, since for n large enough it is

$$n^{3/2}\delta_n^{-1} \approx n^{3/2}\frac{1}{\lambda^\star n}\frac{1}{n-\lambda^\star n}\frac{1}{\|\Delta_n\|} = \frac{1}{\lambda^\star(1-\lambda^\star)}\frac{1}{\sqrt{n}\|\Delta_n\|} \xrightarrow{n\to\infty} 0$$

by the assumption that $\sqrt{n}\|\Delta_n\| \xrightarrow{n\to\infty} \infty$.
Thus, it holds that

$$\begin{aligned}\epsilon\delta_n - \mathcal{O}_P(n^{3/2}) &= \epsilon\delta_n - \epsilon\delta_n\delta_n^{-1}\mathcal{O}_P(n^{3/2})\\ &= \epsilon\delta_n(1-o_P(1))\end{aligned}$$

and it follows, using $\delta_n \xrightarrow{n\to\infty} \infty$,

$$\|U_{n,k^\star}(Y^{(n)})\| - \max_{k:|k-k^\star|>\epsilon k^\star}\|U_{n,k}(Y^{(n)})\| \geq \epsilon\delta_n(1-o_P(1)) \xrightarrow{n\to\infty} 0$$

in probability. And it follows that

$$\mathbb{P}\big(\max_{k:|k-k^\star|>\epsilon k^\star}\|U_{n,k}(Y^{(n)})\| < \|U_{n,k^\star}(Y^{(n)})\|\big) \xrightarrow{n\to\infty} 1.$$

□

For the proof of Theorem 3.3, we will need two further preparing lemmas.

Lemma 3.3. *Under the assumptions of Theorem 3.3, it holds for any $n_1, n_2, n_3, n_4 \leq n$ with $n_1 < n_2$, $n_3 < n_4$ that*

$$\mathbb{E}\Big[\|\sum_{i=n_1}^{n_2}\sum_{j=n_3}^{n_4} h_2^{(X,X)}(X_i^{(n)}, X_j^{(n)})\|^2\Big] \leq C(n_2-n_1)(n_4-n_3). \tag{a1}$$

$$\mathbb{E}\Big[\|\sum_{i=n_1}^{n_2}\sum_{j=n_3}^{n_4} h_2^{(X,Z)}(X_i^{(n)}, Z_j^{(n)})\|^2\Big] \leq C(n_2-n_1)(n_4-n_3). \tag{a2}$$

$$\mathbb{E}\Big[\|\sum_{i=n_1}^{n_2} h_2^{(X,X)}(X_i^{(n)}, X_r^{(n)})\|^2\Big] \leq C(n_2-n_1) \quad r \notin \{n_1,...,n_2\}. \tag{b1}$$

$$\mathbb{E}\Big[\|\sum_{i=n_1}^{n_2} h_2^{(X,Z)}(X_i^{(n)}, Z_r^{(n)})\|^2\Big] \leq C(n_2-n_1) \quad r \notin \{n_1,...,n_2\}. \tag{b2}$$

Proof. The statements of this lemma are variants of Lemmas 2.6 and 2.10 with the difference that we do not work with the "truncated" versions $h_{2,l}$ and $X_{i,l}$, $Z_{i,l}$. The proof follows largely the proof of the cited lemmas. To make the difference clear, we will shortly sketch the proof for (a1).
Rewrite the double sum as in the proof of Lemma 2.6:

$$\mathbb{E}\Big[\|\sum_{n_1\leq i\leq n_2}\sum_{n_3\leq j\leq n_4} h_2^{(X,X)}(X_i^{(n)}, X_j^{(n)})\|^2\Big]$$

$$
\begin{aligned}
&= \mathbb{E}\big[\langle \sum_{n_1\le i\le n_2}\sum_{n_3\le j\le n_4} h_2^{(X,X)}(X_i^{(n)},X_j^{(n)}), \sum_{n_1\le i\le n_2,}\sum_{n_3\le j\le n_4} h_2^{(X,X)}(X_i^{(n)},X_j^{(n)})\rangle\big]\\
&= \sum_{n_1\le i_1\le n_2}\sum_{\substack{n_3\le j_1\le n_4\\ (i_1\neq i_2)\text{ or }(j_1\neq j_2)\text{ or both}}}\sum_{n_1\le i_2\le n_2}\sum_{n_3\le j_2\le n_4} \mathbb{E}[\langle h_2^{(X,X)}(X_{i_1}^{(n)},X_{j_1}^{(n)}), h_2^{(X,X)}(X_{i_2}^{(n)},X_{j_2}^{(n)})\rangle] \quad (24)\\
&\quad + \sum_{n_1\le i\le n_2}\sum_{n_3\le j\le n_4} \mathbb{E}[\langle h_2^{(X,X)}(X_i^{(n)},X_j^{(n)}), h_2^{(X,X)}(X_i^{(n)},X_j^{(n)})\rangle]. \quad (25)
\end{aligned}
$$

We know by the uniform moments under approximation that (25) is bounded by the following:

$$
(25) = \sum_{n_1\le i\le n_2}\sum_{n_3\le j\le n_4} \mathbb{E}[\|h_2^{(X,X)}(X_i^{(n)},X_j^{(n)})\|^2] \le (n_2-n_1)(n_4-n_3)M.
$$

For (24), we use the lemma of Yoshihara [1976] just as in the proof of Lemma 2.6 and get

$$
(24) = \sum_{n_1\le i_1\le n_2}\sum_{\substack{n_3\le j_1\le n_4\\ (i_1\neq i_2)\text{ or }(j_1\neq j_2)\text{ or both}}}\sum_{n_1\le i_2\le n_2}\sum_{n_3\le j_2\le n_4} C{\beta_m^{(n)}}^{\tilde\delta/(1+\tilde\delta)}.
$$

Note that since we do not use the "truncated" versions, we get $\beta_m^{(n)}$ here instead of $\beta_{m-2l}^{(n)}$. With the combinatorical arguments given in the proof of Lemma 2.6, we get

$$
(24) \le C(n_4-n_3)(n_2-n_1)\sum_{m=1}^{n} m{\beta_m^{(n)}}^{\tilde\delta/(1+\tilde\delta)} \le C(n_4-n_3)(n_2-n_1)\underbrace{\sum_{m=1}^{\infty} m^2{\beta_m^{(n)}}^{\tilde\delta/(1+\tilde\delta)}}_{<\infty}
$$

and thus, combining (25) and (24) gives (a1).
(a2) can be similarly proven as Lemma 2.10 with the same difference for the mixing coefficient as explained above. Furthermore, (b1) and (b2) are special cases of the first two statements. □

The next lemma is a Hájek-Rényi-type inequality for Hilbert space-valued random variables which adapts the idea of Theorem 4.1 in Kokoszka and Leipus [2000].

Lemma 3.4. *Let $(R_n)_{n\ge1}$ be a series of $H-$valued random variables with $\mathbb{E}[\|R_n\|^2] < \infty$ and $(c_n)_{n\ge1}\in\mathbb{R}$ non-negative constants. Then it holds for any $\epsilon>0$*

$$
\begin{aligned}
&\epsilon^2\mathbb{P}\Big(\max_{1\le k\le m} c_k\Big\|\sum_{i=1}^{k} R_i\Big\| > \epsilon\Big)\\
&\le c_1^2\mathbb{E}[\|R_1\|^2] + \sum_{k=1}^{m-1}\Big(|c_{k+1}^2-c_k^2|\mathbb{E}\big[\Big\|\sum_{i=1}^{k} R_i\Big\|^2\big]\\
&\qquad + 2c_{k+1}^2\mathbb{E}\big[\Big\|\sum_{i=1}^{k} R_i\Big\|\,\|R_{k+1}\|\big] + c_{k+1}^2\mathbb{E}[\|R_{k+1}\|^2]\Big).
\end{aligned}
$$

Proof. Lemma 4.1 of Kokoszka and Leipus [2000] says:
For any series of random variables $(X_n)_{n\in\mathbb{Z}}$ *and events* $A = \{\max_{1\le k\le m} X_k > \epsilon\}$, $D_k = \{X_1 \le \epsilon, ..., X_k \le \epsilon\}$ *it holds*

$$\epsilon \boldsymbol{1}_A \le X_1 + \sum_{k=1}^{m-1} (X_{k+1} - X_k) \boldsymbol{1}_{D_k} - X_m \boldsymbol{1}_{D_m}.$$

We will use the above cited lemma for random variables $X_k = c_k^2 \| \sum_{i=1}^k R_i \|^2$, ϵ^2 instead of ϵ and take expectations. This yields

$$\begin{aligned}
&\epsilon^2 \mathbb{P}\Big(\max_{1\le k\le m} c_k^2 \Big\| \sum_{i=1}^k R_i \Big\|^2 > \epsilon^2 \Big) \\
&\le \mathbb{E}[c_1^2 \|R_1\|^2] + \sum_{k=1}^{m-1} \mathbb{E}\Big[\Big(\big(c_{k+1} \Big\| \sum_{i=1}^{k+1} R_i \Big\|\big)^2 - \big(c_k \Big\| \sum_{i=1}^{k} R_i \Big\|\big)^2 \Big) \mathbf{1}_{D'_k} \Big] \\
&\qquad \underbrace{- \mathbb{E}\Big[\big(c_m \Big\| \sum_{i=1}^m R_i \Big\|\big)^2 \mathbf{1}_{D'_m} \Big]}_{\ge 0} \\
&\le \mathbb{E}[c_1^2 \|R_1\|^2] + \sum_{k=1}^{m-1} \mathbb{E}\Big[\Big(\big(c_{k+1} \Big\| \sum_{i=1}^{k+1} R_i \Big\|\big)^2 - \big(c_k \Big\| \sum_{i=1}^{k} R_i \Big\|\big)^2 \Big) \mathbf{1}_{D'_k} \Big],
\end{aligned}$$

where $D'_k = \{X_1 \le \epsilon^2, ..., X_k \le \epsilon^2\}$. For each summand we calculate

$$\begin{aligned}
&\Big(\big(c_{k+1} \Big\| \sum_{i=1}^{k+1} R_i \Big\|\big)^2 - \big(c_k \Big\| \sum_{i=1}^{k} R_i \Big\|\big)^2 \Big) \mathbf{1}_{D'_k} \\
&\le \Big(c_{k+1}^2 \Big\| \sum_{i=1}^k R_i \Big\|^2 + 2c_{k+1}^2 \Big\| \sum_{i=1}^k R_i \Big\| \, \|R_{k+1}\| + c_{k+1}^2 \|R_{k+1}\|^2 - c_k^2 \Big\| \sum_{i=1}^k R_i \Big\|^2 \Big) \mathbf{1}_{D'_k} \\
&\le |c_{k+1}^2 - c_k^2| \Big\| \sum_{i=1}^k R_i \Big\|^2 + 2c_{k+1}^2 \Big\| \sum_{i=1}^k R_i \Big\| \, \|R_{k+1}\| + c_{k+1}^2 \|R_{k+1}\|^2.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\epsilon^2 \mathbb{P}\Big(\max_{1\le k\le m} c_k \Big\| \sum_{i=1}^k R_i \Big\| > \epsilon \Big) = \epsilon^2 \mathbb{P}\Big(\max_{1\le k\le m} c_k^2 \Big\| \sum_{i=1}^k R_i \Big\|^2 > \epsilon^2 \Big) \\
&\le c_1^2 \mathbb{E}[\|R_1\|^2] + \sum_{k=1}^{m-1} |c_{k+1}^2 - c_k^2| \mathbb{E}\Big[\Big\| \sum_{i=1}^k R_i \Big\|^2 \Big] + 2c_{k+1}^2 \mathbb{E}\Big[\Big\| \sum_{i=1}^k R_i \Big\| \, \|R_{k+1}\| \Big] \\
&\quad + c_{k+1}^2 \mathbb{E}[\|R_{k+1}\|^2].
\end{aligned}$$

□

Now, we can prove Theorem 3.3.

Proof of Theorem 3.3. Let $a(n) = \frac{M}{\|\Delta_n\|^2}$, where $M > 0$ and $a(n) \xrightarrow{n\to\infty} \infty$, as $\|\Delta_n\| \xrightarrow{n\to\infty} 0$.
In the manner of Lemma 3.2, we will show

$$\lim_{n\to\infty} \mathbb{P}(|k^* - \hat{k}| \leq a(n)) \xrightarrow{M\to\infty} 1$$

by showing

$$\lim_{n\to\infty} \mathbb{P}\big(\max_{k:|k-k^\star|\geq a(n)} \|U_{n,k}(Y^{(n)})\| \leq \|U_{n,k^\star}(Y^{(n)})\|\big) \xrightarrow{M\to\infty} 1. \tag{26}$$

To do so, define $V_k(Y^{(n)}) := \|U_{n,k}(Y^{(n)})\|^2 - \|U_{n,k^\star}(Y^{(n)})\|^2$ and note that whatever $1 \leq k \leq n$ maximises $\|U_{n,k}(Y^{(n)})\|$ also maximises $V_k(Y^{(n)})$, meaning that

$$\hat{k} = \min\big\{k : \|U_{n,k}(Y^{(n)})\| = \max_{1\leq l\leq n} \|U_{n,l}(Y^{(n)})\|\big\} = \min\big\{k : V_k(Y^{(n)}) = \max_{1\leq l\leq n} V_l(Y^{(n)})\big\}.$$

So instead of (26), show

$$\lim_{n\to\infty} \mathbb{P}\big(\max_{k:|k-k^\star|\geq a(n)} V_k(Y^{(n)}) < 0\big) \xrightarrow{M\to\infty} 1. \tag{27}$$

It suffices to consider $\max\{k : |k - k^\star| \leq \epsilon k^\star, |k - k^\star| \geq a(n)\}$ as a subset of $\max\{k : |k - k^\star| \geq a(n)\}$, because of $\mathbb{P}(\hat{k} \in [(1-\epsilon)k^\star, (1+\epsilon)k^\star]) \xrightarrow{n\to\infty} 1$ (Lemma 3.2). Thus, if we define

$$\tilde{k} = \min\big\{k : |k - k^\star| \leq \epsilon k^\star, V_k(Y^{(n)}) = \max_{(1-\epsilon)k^\star\leq l\leq(1+\epsilon)k^\star} V_l(Y^{(n)})\big\}$$

it holds that $\lim_{n\to\infty} \mathbb{P}(\hat{k} = \tilde{k}) = 1$. Meaning that if $V_k(Y^{(n)})$ attains its maximum, it will be attained inside the interval $[(1-\epsilon)k^\star, (1+\epsilon)k^\star]$ (in probability, for $n \to \infty$). This results into two cases:

$$|k - k^\star| \leq \epsilon k^\star, |k - k^\star| \geq a(n) = \begin{cases} (1-\epsilon)k^\star \leq k \leq k^\star - a(n) & k < k^\star \\ k^\star + a(n) \leq k \leq (1+\epsilon)k^\star & k > k^\star \end{cases}.$$

Again, we will present the case $k < k^\star$, the other case works very similarly.
Since $k^\star - k > 0$, (27) holds if

$$\lim_{n\to\infty} \mathbb{P}\Big(\max_{(1-\epsilon)k^\star\leq k\leq k^\star-a(n)} \frac{V_k(Y^{(n)})}{(n(k^\star - k))^2} < 0\Big) \xrightarrow{M\to\infty} 1. \tag{28}$$

Now, we can write

$$\frac{-V_k(Y^{(n)})}{(n(k^\star - k))^2} = \frac{\|U_{n,k^\star}(Y^{(n)})\|^2 - \|U_{n,k}(Y^{(n)})\|^2}{(n(k^\star - k))^2}$$

$$= -\Big(\frac{\|U_{n,k^\star}(Y^{(n)})\| - \|U_{n,k}(Y^{(n)})\|}{n(k^\star - k)}\Big)^2 \tag{29}$$

$$
\begin{aligned}
&+2\Big(\frac{\|U_{n,k^\star}(Y^{(n)})\|-\|U_{n,k}(Y^{(n)})\|}{n(k^\star-k)}\Big)\frac{\|U_{n,k^\star}(Y^{(n)})\|}{n(k^\star-k)}\\
=&-\Big(\frac{\|U_{n,k^\star}(Y^{(n)})\|-\|U_{n,k}(Y^{(n)})\|}{n(k^\star-k)}-\frac{n-k^\star}{n}\|\Delta_n\|+\frac{n-k^\star}{n}\|\Delta_n\|\Big)^2\\
&+2\Big(\frac{\|U_{n,k^\star}(Y^{(n)})\|-\|U_{n,k}(Y^{(n)})\|}{n(k^\star-k)}-\frac{n-k^\star}{n}\|\Delta_n\|+\frac{n-k^\star}{n}\|\Delta_n\|\Big)\\
&\cdot\Big(\frac{\|U_{n,k^\star}(Y^{(n)})\|-(n-k^\star)k^\star\|\Delta_n\|+(n-k^\star)k^\star\|\Delta_n\|}{n(k^\star-k)}\Big). \qquad (30)
\end{aligned}
$$

After this preparatory work, the main part of the proof is to show that

$$
\max_{k:(k^\star-k)>a(n)}\Big(\frac{\|U_{n,k^\star}(Y^{(n)})\|-\|U_{n,k}(Y^{(n)})\|}{n(k^\star-k)}-\frac{n-k^\star}{n}\|\Delta_n\|\Big)=o_P(\|\Delta_n\|). \qquad (31)
$$

By triangular inequality it holds that

$$
\begin{aligned}
&\frac{\|U_{n,k^\star}(Y^{(n)})\|-\|U_{n,k}(Y^{(n)})\|}{n(k^\star-k)}-\frac{n-k^\star}{n}\|\Delta_n\|\\
\le&\Big\|\frac{U_{n,k^\star}(Y^{(n)})-U_{n,k}(Y^{(n)})}{n(k^\star-k)}-\frac{n-k^\star}{n}\Delta_n\Big\|\\
=&\Big\|\frac{1}{n(k^\star-k)}\Big(-\sum_{i=1}^{k}\sum_{j=k+1}^{k^\star}h(X_i^{(n)},X_j^{(n)})+\sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n}h(X_i^{(n)},Z_j^{(n)})\\
&\quad-\sum_{i=1}^{k}\sum_{j=k^\star+1}^{n}h(X_i^{(n)},Z_j^{(n)})\Big)-\frac{n-k^\star}{n}\Delta_n\Big\|\\
\le&\Big\|\frac{1}{n(k^\star-k)}\sum_{i=1}^{k}\sum_{j=k+1}^{k^\star}h(X_i^{(n)},X_j^{(n)})\Big\|\\
&+\Big\|\frac{1}{n(k^\star-k)}\Big(\sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n}h(X_i^{(n)},Z_j^{(n)})-\sum_{i=1}^{k}\sum_{j=k^\star+1}^{n}h(X_i^{(n)},Z_j^{(n)})\Big)-\frac{n-k^\star}{n}\Delta_n\Big\|.
\end{aligned}
$$

So, if both

$$
\mathbb{P}\Big(\frac{1}{n\|\Delta_n\|}\max_{k:(k^\star-k)>a(n)}\Big\|\frac{1}{k^\star-k}\sum_{i=1}^{k}\sum_{j=k+1}^{k^\star}h(X_i^{(n)},X_j^{(n)})\Big\|>\epsilon\Big)\xrightarrow{n,M\to\infty}0 \qquad (32)
$$

and

$$
\begin{aligned}
\mathbb{P}\Big(\frac{1}{n\|\Delta_n\|}\max_{k:(k^\star-k)>a(n)}\Big\|\frac{1}{k^\star-k}\Big(\sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n}h(X_i^{(n)},Z_j^{(n)})-\sum_{i=1}^{k}\sum_{j=k^\star+1}^{n}h(X_i^{(n)},Z_j^{(n)})\Big)&\\
-(n-k^\star)\Delta_n\Big\|>\epsilon\Big)\xrightarrow{n,M\to\infty}0& \qquad (33)
\end{aligned}
$$

hold, then (31) holds as well.
For (32), use Hoeffding's decomposition to further deconstruct:

$$
\begin{aligned}
&\sum_{i=1}^{k}\sum_{j=k+1}^{k^\star} h(X_i^{(n)}, X_j^{(n)})\\
&= (k^\star - k)\underbrace{\sum_{i=1}^{k} h_1^{(X)}(X_i^{(n)})}_{=:\tilde{S}_k^X(X^{(n)})} - k\sum_{j=k+1}^{k^\star} h_1^{(X)}(X_j^{(n)}) + \sum_{i=1}^{k}\sum_{j=k+1}^{k^\star} h_2^{(X,X)}(X_i^{(n)}, X_j^{(n)})\\
&= (k^\star - k)\tilde{S}_k^X(X^{(n)}) - k(\tilde{S}_{k^\star}^X(X^{(n)}) - \tilde{S}_k^X(X^{(n)})) + \sum_{i=1}^{k}\sum_{j=k+1}^{k^\star} h_2^{(X,X)}(X_i^{(n)}, X_j^{(n)}).
\end{aligned}
$$

Thus, it follows:

$$
\begin{aligned}
&\frac{1}{(k^\star - k)n\|\Delta_n\|}\Big\|\sum_{i=1}^{k}\sum_{j=k+1}^{k^\star} h(X_i^{(n)}, X_j^{(n)})\Big\|\\
&\le \underbrace{\frac{1}{\|\Delta_n\|}\frac{\|\tilde{S}_k^X(X^{(n)})\|}{n}}_{\nu^{(1)}(X^{(n)})} + \underbrace{\frac{1}{\|\Delta_n\|}\frac{\|\tilde{S}_{k^\star}^X(X^{(n)}) - \tilde{S}_k^X(X^{(n)})\|}{k^\star - k}}_{\nu^{(2)}(X^{(n)})}\\
&\quad + \underbrace{\frac{1}{(k^\star - k)n\|\Delta_n\|}\Big\|\sum_{i=1}^{k}\sum_{j=k+1}^{k^\star} h_2^{(X,X)}(X_i^{(n)}, X_j^{(n)})\Big\|}_{\nu^{(3)}(X^{(n)})}.
\end{aligned}
$$

Equation (32) holds, if $\mathbb{P}(\max_{1\le k\le k^\star - a(n)} \nu^{(l)}(X^{(n)}) > \epsilon) \overset{M,n\to\infty}{\to} 0$ for $l = 1, 2, 3$.
We start with $\nu^{(1)}(X^{(n)})$:

$$
\frac{\|\tilde{S}_k^X(X^{(n)})\|}{n} \le \frac{\|\tilde{S}_n^X(X^{(n)}) - \tilde{S}_k^X(X^{(n)})\|}{n} + \frac{\|\tilde{S}_n^X(X^{(n)})\|}{n}.
$$

We will use Lemma A.2 from the online supplementary appendix of Gerstenberger [2018] which states that $\frac{1}{n}\|\tilde{S}_n^X(X^{(n)})\| = \mathcal{O}_P(n^{-1/2})$ and recall that $\frac{1}{n}\|\tilde{S}_n^X(X^{(n)}) - \tilde{S}_k^X(X^{(n)})\| = \frac{1}{n}\|\sum_{i=k+1}^{n} h_1^{(X)}(X_i^{(n)})\|$.
Now, by stationarity, it is

$$
\begin{aligned}
&\Big\{\Big\|\sum_{i=k+1}^{n} h_1^{(X)}(X_i^{(n)})\Big\|,\ \ 1 \le k \le k^\star - a(n)\Big\}\\
&\overset{\mathcal{D}}{=} \Big\{\Big\|\sum_{i=1}^{n-k} h_1^{(X)}(X_i^{(n)})\Big\|,\ \ 1 \le k \le k^\star - a(n)\Big\}.
\end{aligned}
$$

So it follows by Lemma A.2 Gerstenberger [2018] (online appendix)

$$
\max_{1\le k\le k^\star - a(n)} \frac{\|\tilde{S}_n^X(X^{(n)}) - \tilde{S}_k^X(X^{(n)})\|}{n} \overset{\mathcal{D}}{=} \max_{1\le k\le k^\star - a(n)} \frac{\|\tilde{S}_{n-k}^X(X^{(n)})\|}{n}
$$

$$\leq \max_{1\leq k\leq k^\star - a(n)} \frac{\|\tilde{S}^X_{n-k}(X^{(n)})\|}{n-k} \overset{\mathcal{D}}{\leq} \max_{a(n)\leq j\leq n} \frac{\|\tilde{S}^X_j(X^{(n)})\|}{j} = \mathcal{O}_P(\frac{1}{\sqrt{a(n)}}).$$

And thus,

$$\begin{aligned}
&\max_{1\leq k\leq k^\star - a(n)} \nu^{(1)}(X^{(n)}) \\
&\leq \max_{1\leq k\leq k^\star - a(n)} \frac{1}{\|\Delta_n\|} \frac{\|\tilde{S}^X_n(X^{(n)}) - \tilde{S}^X_k(X^{(n)})\|}{n} + \frac{1}{\|\Delta_n\|}\frac{\|\tilde{S}^X_n(X^{(n)})\|}{n} \\
&= \mathcal{O}_P\Big(\frac{1}{\|\Delta_n\|\sqrt{a(n)}}\Big) + \mathcal{O}_P\Big(\frac{1}{\|\Delta_n\|\sqrt{n}}\Big) = \mathcal{O}_P\Big(\frac{1}{\sqrt{M}}\Big) + \mathcal{O}_P\Big(\frac{1}{\|\Delta_n\|\sqrt{n}}\Big) \\
&= o_P(1) \text{ as } M, n \to \infty.
\end{aligned}$$

For $\nu^{(2)}(X^{(n)})$ a similar argument is used:
By stationarity it is

$$\begin{aligned}
&\{\|\tilde{S}^X_{k^\star}(X^{(n)}) - \tilde{S}^X_k(X^{(n)})\|,\ \ 1\leq k\leq k^\star - a(n)\} \\
\overset{\mathcal{D}}{=} &\{\|\tilde{S}^X_{k^\star - k}(X^{(n)})\|,\ \ 1\leq k\leq k^\star - a(n)\}.
\end{aligned}$$

Thus it follows by Lemma A.2 Gerstenberger [2018] (online appendix)

$$\begin{aligned}
&\max_{1\leq k\leq k^\star - a(n)} \frac{\|\tilde{S}^X_{k^\star}(X^{(n)}) - \tilde{S}^X_k(X^{(n)})\|}{k^\star - k} \overset{\mathcal{D}}{=} \max_{1\leq k\leq k^\star - a(n)} \frac{\|\tilde{S}^X_{k^\star - k}(X^{(n)})\|}{k^\star - k} \\
\overset{\mathcal{D}}{=} &\max_{a(n)\leq j\leq n} \frac{\|\tilde{S}^X_j(X^{(n)})\|}{j} = \mathcal{O}_P\Big(\frac{1}{\sqrt{a(n)}}\Big)
\end{aligned}$$

So, it holds that

$$\begin{aligned}
&\max_{1\leq k\leq k^\star - a(n)} \nu^{(2)}(X^{(n)}) = \max_{1\leq k\leq k^\star - a(n)} \frac{1}{\|\Delta_n\|} \frac{\|\tilde{S}^X_{k^\star}(X^{(n)}) - \tilde{S}^X_k(X^{(n)})\|}{k^\star - k} \\
&= \frac{1}{\|\Delta_n\|}\mathcal{O}_P\Big(\frac{1}{\sqrt{a(n)}}\Big) = \mathcal{O}_P\Big(\frac{1}{\|\Delta_n\|\sqrt{a(n)}}\Big) = \mathcal{O}_P\Big(\frac{1}{\sqrt{M}}\Big) \\
&= o_P(1) \text{ as } M \to \infty.
\end{aligned}$$

To handle $\nu^{(3)}(X^{(n)})$, a little more effort is needed:
Define $X^{(2)}(k) := \sum_{i=1}^k \sum_{j=k+1}^{k^\star} h_2^{(X,X)}(X_i^{(n)}, X_j^{(n)})$, $R_k := X^{(2)}(k) - X^{(2)}(k-1)$, with $X^{(2)}(0) = 0$, $R_0 = 0$, and constants $c_k := (k^\star - k)^{-1}$. Then, it is $\sum_{i=1}^k R_i = X^{(2)}(k)$ and by Lemma 3.4

$$\begin{aligned}
&\mathbb{P}\big(\max_{1\leq k\leq k^\star - a(n)} \nu^{(3)}(X^{(2)}(k)) > \epsilon\big) = \mathbb{P}\big(\max_{1\leq k\leq k^\star - a(n)} c_k \big\|\sum_{i=1}^k R_i\big\| > \epsilon n\|\Delta_n\|\big) \\
\leq &\frac{1}{(\epsilon n\|\Delta_n\|)^2}\Big(c_1\mathbb{E}[\|R_1\|^2] + \sum_{k=1}^{k^\star - a(n)} (c_{k+1}^2 - c_k^2)\mathbb{E}\big[\big\|\sum_{i=1}^k R_i\big\|^2\big]
\end{aligned}$$

$$+ 2c_{k+1}^2 \mathbb{E}\big[\big\| \sum_{i=1}^{k} R_i \big\| \, \|R_{k+1}\|\big] + c_{k+1}^2 \mathbb{E}[\|R_{k+1}\|^2]\Big).$$

Upper bound the expectations by Lemma 3.3:

$$\mathbb{E}\big[\big\| \sum_{i=1}^{k} R_i \big\|^2\big] = \mathbb{E}[\|X^{(2)}(k)\|^2] = \mathbb{E}\big[\big\| \sum_{i=1}^{k} \sum_{j=k+1}^{k^\star} h_2^{(X,X)}(X_i^{(n)}, X_j^{(n)}) \big\|^2\big]$$
$$\leq Ck(k^\star - k)$$

and

$$\begin{aligned}
&\mathbb{E}[\|R_{k+1}\|^2] = \mathbb{E}[\|X^{(2)}(k+1) - X^{(2)}(k)\|^2] \\
&= \mathbb{E}\big[\big\| \sum_{i=1}^{k+1} \sum_{j=k+2}^{k^\star} h_2^{(X,X)}(X_i^{(n)}, X_j^{(n)}) - \sum_{i=1}^{k} \sum_{j=k+1}^{k^\star} h_2^{(X,X)}(X_i^{(n)}, X_j^{(n)}) \big\|^2\big] \\
&= \mathbb{E}\big[\big\| \sum_{j=k+2}^{k^\star} h_2^{(X,X)}(X_{k+1}^{(n)}, X_j^{(n)}) - \sum_{i=1}^{k} h_2^{(X,X)}(X_i^{(n)}, X_{k+1}^{(n)}) \big\|^2\big] \\
&= \mathbb{E}\big[\big\| \sum_{j=k+2}^{k^\star} h_2^{(X,X)}(X_{k+1}^{(n)}, X_j^{(n)}) + \sum_{i=1}^{k} h_2^{(X,X)}(X_{k+1}^{(n)}, X_i^{(n)}) \big\|^2\big] \\
&= \mathbb{E}\big[\big\| \sum_{\substack{i=1 \\ i\neq k+1}}^{k^\star} h_2^{(X,X)}(X_{k+1}^{(n)}, X_i^{(n)}) \big\|^2\big] \\
&\leq Ck^\star
\end{aligned}$$

using the anti-symmetry of $h_2^{(X,X)}$.
By Hölder's inequality and the above inequality

$$\mathbb{E}\big[\big\| \sum_{i=1}^{k} R_i \big\| \, \|R_{k+1}\|\big] \leq \mathbb{E}\big[\big\| \sum_{i=1}^{k} R_i \big\|^2\big]^{\frac{1}{2}} \mathbb{E}[\|R_{k+1}\|^2]^{\frac{1}{2}} \leq C\sqrt{(k^\star - k)k}\,\sqrt{k^\star}.$$

Additionally, note that $c_{k+1}^2 - c_k^2 \leq (k^\star - k - 1)^{-2}$. Thus, it follows from the above calculations:

$$\begin{aligned}
&\mathbb{P}(\max_{1\leq k\leq k^\star - a(n)} \nu^{(3)}(X^{(n)}) > \epsilon) \\
&\leq \frac{C}{(\epsilon n \|\Delta_n\|)^2}\Big(\frac{k^\star}{(k^\star-1)^2} + \sum_{k=1}^{k^\star - a(n)} \frac{(k^\star - k)k + \sqrt{(k^\star - k)k}\sqrt{k^\star} + k^\star}{(k^\star - k - 1)^2}\Big) \\
&\lesssim \frac{C}{(\epsilon n \|\Delta_n\|)^2}\Big(\frac{1}{n} + \frac{n^2}{a(n)}\Big) \text{ (for } n \text{ large)} \\
&= \frac{C}{\epsilon}\Big(\frac{1}{n^3\|\Delta_n\|^2} + \frac{1}{M}\Big) \xrightarrow{n,M\to\infty} 0.
\end{aligned}$$

Here, we used the following inequality to bound the sum: For some monotone decreasing function f, which takes non-negative values on the interval $[a, b]$ and its integral $\int_a^b f(x)dx$ exists, it holds

$$\sum_{n=a+1}^{b-1} f(n) \leq \int_a^b f(x)\,dx.$$

Thus, we can bound the sum by:

$$\sum_{k=1}^{k^\star - a(n)} \frac{(k^\star - k)k + \sqrt{(k^\star - k)k}\sqrt{k^\star} + k^\star}{(k^\star - k - 1)^2} \leq n^2 \sum_{k=1}^{k^\star - a(n)} \frac{1}{(k^\star - k - 1)^2}$$
$$\leq n^2 \int_0^{k^\star - a(n)+1} \frac{1}{(k^\star - x - 1)^2}\,dx = n^2\Big(\frac{1}{a(n)} - \frac{1}{k^\star - 1}\Big) \leq \frac{n^2}{a(n)}.$$

With this, we have shown that (32) holds. Next, we turn to equation (33). As we have done for (32), we first use Hoeffding's decomposition again:

$$\begin{aligned}
&\sum_{i=1}^{k} \sum_{j=k^\star+1}^{n} h(X_i^{(n)}, X_j^{(n)}) - (n - k^\star)k\Delta_n \\
&= \sum_{i=1}^{k} \sum_{j=k^\star+1}^{n} \big(h(X_i^{(n)}, Z_j^{(n)}) - \mathbb{E}[h(X_i^{(n)}, Z_j^{(n)})]\big) \\
&= (n - k^\star) \underbrace{\sum_{i=1}^{k} \big(h_1^{(Z)}(X_i^{(n)}) - \mathbb{E}[h_1^{(Z)}(X_i^{(n)})]\big)}_{=:S_k^Z(X^{(n)})} - k \underbrace{\sum_{j=k^\star+1}^{n} \big(h_1^{(X)}(Z_j^{(n)}) - \mathbb{E}[h_1^{(X)}(Z_j^{(n)})]\big)}_{=:S_n^X(Z^{(n)}) - S_{k^\star}^X(Z^{(n)})} \\
&\quad + \sum_{i=1}^{k} \sum_{j=k^\star+1}^{n} \big(h_2^{(X,Z)}(X_i^{(n)}, Z_j^{(n)}) - \mathbb{E}[h_2^{(X,Z)}(X_i^{(n)}, Z_j^{(n)})]\big) \\
&= (n - k^\star) S_k^Z(X^{(n)}) - k\big(S_n^X(Z^{(n)}) - S_{k^\star}^X(Z^{(n)})\big) \\
&\quad + \sum_{i=1}^{k} \sum_{j=k^\star+1}^{n} \big(h_2^{(X,Z)}(X_i^{(n)}, Z_j^{(n)}) - \mathbb{E}[h_2^{(X,Z)}(X_i^{(n)}, Z_j^{(n)})]\big)
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{i=1}^{k^\star} \sum_{j=k^\star+1}^{n} h(X_i^{(n)}, Z_j^{(n)}) - (n - k^\star)k^\star\Delta_n \\
&= (n - k^\star) S_{k^\star}^Z(X^{(n)}) - k^\star\big(S_n^X(Z^{(n)}) - S_{k^\star}^X(Z^{(n)})\big) \\
&\quad + \sum_{i=1}^{k^\star} \sum_{j=k^\star+1}^{n} \big(h_2^{(X,Z)}(X_i^{(n)}, Z_j^{(n)}) - \mathbb{E}[h_2^{(X,Z)}(X_i^{(n)}, Z_j^{(n)})]\big).
\end{aligned}$$

With this, we get

$$
\begin{aligned}
&\frac{1}{n\|\Delta_n\|}\Big\|\frac{1}{k^\star-k}\Big(\sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n} h(X_i^{(n)},Z_j^{(n)})\\
&\qquad -\sum_{i=1}^{k}\sum_{j=k^\star}^{n} h(X_i^{(n)},Z_j^{(n)})-(k^\star-k)(n-k^\star)\Delta_n\Big)\Big\|\\
&\le \underbrace{\frac{1}{n\|\Delta_n\|}\frac{n-k^\star}{k^\star-k}\|S_{k^\star}^Z(X^{(n)})-S_k^Z(X^{(n)})\|}_{=\iota^{(1)}(X^{(n)},Z^{(n)})}+\underbrace{\frac{1}{n\|\Delta_n\|}\|S_n^X(Z^{(n)})-S_{k^\star}^X(Z^{(n)})\|}_{=\iota^{(2)}(X^{(n)},Z^{(n)})}\\
&\quad+\underbrace{\frac{1}{n\|\Delta_n\|(k^\star-k)}\sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n}\big(h_2^{(X,Z)}(X_i^{(n)},Z_j^{(n)})-\mathbb{E}[h_2^{(X,Z)}(X_i^{(n)},Z_j^{(n)})]\big)}_{=\iota^{(3)}(X^{(n)},Z^{(n)})}\\
&\quad+\underbrace{\frac{1}{n\|\Delta_n\|(k^\star-k)}\sum_{i=1}^{k}\sum_{j=k^\star+1}^{n}\big(h_2^{(X,Z)}(X_i^{(n)},Z_j^{(n)})-\mathbb{E}[h_2^{(X,Z)}(X_i^{(n)},Z_j^{(n)})]\big)}_{=\iota^{(4)}(X^{(n)},Z^{(n)})}
\end{aligned}
$$

So, equation (33) holds if $\mathbb{P}(\max_{1\le k\le k^\star-a(n)}\iota^{(l)}(X^{(n)},Z^{(n)})>\epsilon)\xrightarrow{M,n\to\infty}0$ for $l=1,...4$. Considerations similar to those seen before help us to handle $\iota^{(1)}(X^{(n)},Z^{(n)})$: By stationarity it is

$$\{\|S_{k^\star}^Z(X^{(n)})-S_k^Z(X^{(n)})\|,\ 1\le k\le k^\star-a(n)\}\stackrel{\mathcal{D}}{=}\{\|S_{k^\star-k}^Z(X^{(n)})\|,\ 1\le k\le k^\star-a(n)\}.$$

So by Lemma A.2 Gerstenberger [2018] (online appendix),

$$
\begin{aligned}
&\max_{1\le k\le k^\star-a(n)}\frac{\|S_{k^\star}^Z(X^{(n)})-S_k^Z(X^{(n)})\|}{k^\star-k}\stackrel{\mathcal{D}}{=}\max_{1\le k\le k^\star-a(n)}\frac{\|S_{k^\star-k}^Z(X^{(n)})\|}{k^\star-k}\\
&\stackrel{\mathcal{D}}{=}\max_{a(n)\le j\le n}\frac{\|S_j^Z(X^{(n)})\|}{j}=\mathcal{O}_P\Big(\frac{1}{\sqrt{a(n)}}\Big)
\end{aligned}
$$

and thus

$$
\begin{aligned}
&\max_{1\le k\le k^\star-a(n)}\iota^{(1)}(X^{(n)},Z^{(n)})\le\frac{1}{\|\Delta_n\|}\max_{1\le k\le k^\star-a(n)}\frac{\|S_{k^\star}^Z(X^{(n)})-S_k^Z(X^{(n)})\|}{k^\star-k}\\
&=\mathcal{O}_P\Big(\frac{1}{\|\Delta_n\|\sqrt{a(n)}}\Big)=\mathcal{O}_P\Big(\frac{1}{\sqrt{M}}\Big)\\
&=o_P(1)\text{ as }M\to\infty.
\end{aligned}
$$

For $\iota^{(2)}(X^{(n)},Z^{(n)})$, we use the convergence in distribution of the process $(\frac{1}{\sqrt{n}}S_{\lfloor n\lambda\rfloor}^X(Z^{(n)}))_{\lambda\in(0,1)}$. It is

$$\frac{1}{n\|\Delta_n\|}\|S_n^X(Z^{(n)})-S_{k^\star}^X(Z^{(n)})\|\le\frac{1}{\sqrt{n}\|\Delta_n\|}\Big(\frac{\|S_n^X(Z^{(n)})\|}{\sqrt{n}}+\frac{\|S_{k^\star}^X(Z^{(n)})\|}{\sqrt{n}}\Big)$$

and we know that by Proposition 2.4,

$$\Big(\frac{1}{\sqrt{n}}S^X_{\lfloor n\lambda\rfloor}(Z^{(n)})\Big)_{\lambda\in(0,1)} = \Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor n\lambda\rfloor} h_1^{(X)}(Z_i^{(n)}) - \mathbb{E}[h_1^{(X)}(Z_i^{(n)})]\Big)_{\lambda\in(0,1)}$$

converges in distribution to some Brownian motion. Therefore, boundedness in probability follows, i.e. $\frac{\|S^X_n(Z^{(n)})\|}{\sqrt{n}} = \mathcal{O}_P(1)$ and $\frac{\|S^X_{k^\star}(Z^{(n)})\|}{\sqrt{n}} = \mathcal{O}_P(1)$. We can thus conclude:

$$\max_{1\le k\le k^\star - a(n)} \iota^{(2)}(X^{(n)}, Z^{(n)}) = \mathcal{O}_P\Big(\frac{1}{\sqrt{n}\|\Delta_n\|}\Big) = o_P(1) \text{ as } n\to\infty.$$

For $\iota^{(3)}(X^{(n)}, Z^{(n)})$ observe that by Markov's inequality and a variant of Lemma 3.3 for $h_2^{(X,Z)}(\cdot,\cdot) - \mathbb{E}[h_2^{(X,Z)}(\cdot,\cdot)]$, which is a degenerate kernel it holds that

$$\begin{aligned}
&\mathbb{P}(\max_{1\le k\le k^\star - a(n)} \iota^{(3)}(X^{(n)}, Z^{(n)}) > \epsilon)\\
&= \mathbb{P}\Big(\max_{1\le k\le k^\star - a(n)} \frac{1}{n\|\Delta_n\|}\frac{\|\sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n} h_2^{(X,Z)}(X_i^{(n)}, Z_j^{(n)}) - \mathbb{E}[h_2^{(X,Z)}(X_i^{(n)}, Z_j^{(n)})]\|}{k^\star - k} \ge \epsilon\Big)\\
&\le \mathbb{P}\Big(\frac{1}{a(n)}\|\sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n} h_2^{(X,Z)}(X_i^{(n)}, Z_j^{(n)}) - \mathbb{E}[h_2^{(X,Z)}(X_i^{(n)}, Z_j^{(n)})]\| > \epsilon n\|\Delta_n\|\Big)\\
&\le \frac{1}{(\epsilon n\|\Delta_n\|)^2}\mathbb{E}\Big[\frac{1}{a(n)^2}\Big\|\sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n} h_2^{(X,Z)}(X_i^{(n)}, Z_j^{(n)}) - \mathbb{E}[h_2^{(X,Z)}(X_i^{(n)}, Z_j^{(n)})]\Big\|^2\Big]\\
&\le \frac{1}{(\epsilon n\|\Delta_n\|)^2}\frac{C(n-k^\star)k^\star}{a(n)^2}\\
&\le \frac{C}{\epsilon^2}\frac{1}{\|\Delta_n\|^2 a(n)^2} \le \frac{C}{\epsilon^2}\frac{1}{\|\Delta_n\|^2 a(n)} \text{ for } n \text{ large}\\
&= \frac{C}{\epsilon^2}\frac{1}{M} \xrightarrow{M\to\infty} 0.
\end{aligned}$$

For $\iota^{(4)}(X^{(n)}, Z^{(n)})$ we proceed similarly to $\nu^{(3)}(X^{(n)})$:
Define $Y^{(2)}(k) := \sum_{i=1}^{k}\sum_{j=k^\star+1}^{n} h_2^{(X,Z)}(X_i^{(n)}, Z_j^{(n)}) - \mathbb{E}[h_2^{(X,Z)}(X_i^{(n)}, Z_j^{(n)})]$. Further let $R_k := Y^{(2)}(k) - Y^{(2)}(k-1)$, with $Y^{(2)}(0) = 0$, $R_0 = 0$, and constants $c_k := (k^\star - k)^{-1}$. Then, it is $\sum_{i=1}^{k} R_i = Y^{(2)}(k)$ and by Lemma 3.4

$$\begin{aligned}
\mathbb{P}(\max_{1\le k\le k^\star - a(n)} \iota^{(4)}(X^{(n)}, Z^{(n)}) > \epsilon) &= \mathbb{P}(\max_{1\le k\le k^\star - a(n)} c_k\Big\|\sum_{i=1}^{k} R_i\Big\| > \epsilon n\|\Delta_n\|)\\
\le \frac{1}{(\epsilon n\|\Delta_n\|)^2}\Big(c_1\mathbb{E}[\|R_1\|^2] &+ \sum_{k=1}^{k^\star - a(n)}(c_{k+1}^2 - c_k^2)\mathbb{E}\Big[\Big\|\sum_{i=1}^{k} R_i\Big\|^2\Big]\\
&+ 2c_{k+1}^2\mathbb{E}\Big[\Big\|\sum_{i=1}^{k} R_i\Big\|\,\|R_{k+1}\|\Big] + c_{k+1}^2\mathbb{E}[\|R_{k+1}\|^2]\Big).
\end{aligned}$$

We upper bound the expectations by a variant of Lemma 3.3 for $h_2^{(X,Z)}(\cdot,\cdot) - \mathbb{E}[h_2^{(X,Z)}(\cdot,\cdot)]$, so that we get

$$
\begin{aligned}
&\mathbb{E}\big[\big\|\sum_{i=1}^{k} R_i\big\|^2\big] = \mathbb{E}[\|Y^{(2)}(k)\|^2] \\
&= \mathbb{E}\big[\big\|\sum_{i=1}^{k}\sum_{j=k^\star+1}^{n} h_2^{(X,Z)}(X_i^{(n)}, Z_j^{(n)}) - \mathbb{E}[h_2^{(X,Z)}(X_i^{(n)}, Z_j^{(n)})]\big\|^2\big] \\
&\leq C(n-k^*)k
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}[\|R_{k+1}\|^2] = \mathbb{E}[\|Y^{(2)}(k+1) - Y^{(2)}(k)\|^2] \\
&= \mathbb{E}\big[\big\|\sum_{j=k^\star+1}^{n} h_2^{(X,Z)}(X_{k+1}^{(n)}, Z_j^{(n)}) - \mathbb{E}[h_2^{(X,Z)}(X_{k+1}^{(n)}, Z_j^{(n)})]\big\|^2\big] \\
&\leq C(n-k^\star).
\end{aligned}
$$

By Hölder's inequality and the above inequality it holds:

$$
\mathbb{E}\big[\big\|\sum_{i=1}^{k} R_i\big\|\,\|R_{k+1}\|\big] \leq \mathbb{E}\big[\big\|\sum_{i=1}^{k} R_i\big\|^2\big]^{\frac{1}{2}}\,\mathbb{E}[\|R_{k+1}\|^2]^{\frac{1}{2}} \leq C\sqrt{(n-k^\star)k}\sqrt{(n-k^\star)}.
$$

Noting that $c_{k+1}^2 - c_k^2 \leq (k^\star - k - 1)^{-1}$, it follows from the calculations above:

$$
\begin{aligned}
&\mathbb{P}\big(\max_{1\leq k\leq k^\star - a(n)} \iota^{(4)}(X^{(n)}, Z^{(n)}) > \epsilon\big) \\
&\leq \frac{C}{(\epsilon n\|\Delta_n\|)^2}\Big(\frac{n-k^\star}{(k^\star-1)^2} + \sum_{k=1}^{k^\star - a(n)} \frac{(n-k^\star)k + (n-k^\star)\sqrt{k} + (n-k^\star)}{(k^\star-k-1)^2}\Big) \\
&\lesssim \frac{C}{(\epsilon n\|\Delta_n\|)^2}\Big(\frac{1}{n} + \frac{n^2}{a(n)}\Big) \text{ (for } n \text{ large)} \\
&= \frac{C}{\epsilon}\Big(\frac{1}{n^3\|\Delta_n\|^2} + \frac{1}{M}\Big) \xrightarrow{n,M\to\infty} 0.
\end{aligned}
$$

The sum is bounded by the same inequality as in the calculations for $\nu^{(3)}(X^{(n)})$. This means that (33) holds. So, we have that (31) holds. Therefore, we can conclude that

$$
\begin{aligned}
&\max_{k:(k^\star-k)>a(n)} \Big(\frac{\|U_{n,k^\star}(Y^{(n)})\| - \|U_{n,k}(Y^{(n)})\|}{n(k^\star-k)} - \frac{n-k^\star}{n}\|\Delta_n\| + \frac{n-k^\star}{n}\|\Delta_n\|\Big) \\
&\approx o_P(\|\Delta_n\|) + (1-\lambda^\star)\|\Delta_n\| \\
&= (1-\lambda^\star)\|\Delta_n\|(1+o_P(1)). \qquad (34)
\end{aligned}
$$

We can now draw the conclusion for $V_k(Y^{(n)})$:
We know by Theorem 3.1 that

$$
\frac{1}{n}\big\|\sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n} h(Y_i^{(n)}, Y_j^{(n)}) - \mathbb{E}[h(Y_i^{(n)}, Y_j^{(n)})]\big\| = \mathcal{O}_P(n^{\frac{1}{2}}).
$$

And thus it holds that

$$\begin{aligned}
&\frac{1}{n}\|U_{n,k^\star}(Y^{(n)})\| - (n-k^\star)k^\star\|\Delta_n\| \\
&\leq \frac{1}{n}\Big\|\sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n} h(Y_i^{(n)}, Y_j^{(n)}) - \mathbb{E}[h(Y_i^{(n)}, Y_j^{(n)})]\Big\| = \mathcal{O}_P(n^{\frac{1}{2}}).
\end{aligned}$$

So, it follows that

$$\begin{aligned}
&\max_{(1-\epsilon)k^\star \leq k \leq k^\star - a(n)} \frac{\|U_{n,k^\star}(Y^{(n)})\| - (n-k^\star)k^\star\|\Delta_n\| + (n-k^\star)k^\star\|\Delta_n\|}{n(k^\star - k)} \\
&\geq \frac{1}{\epsilon}\Big(\frac{\|U_{n,k^\star}(Y^{(n)})\| - (n-k^\star)k^\star\|\Delta_n\|}{n} + \frac{(n-k^\star)k^\star\|\Delta_n\|}{n}\Big) \\
&\approx \frac{1}{\epsilon}\big(\mathcal{O}_P(n^{\frac{1}{2}}) + \frac{1}{n}\delta_n\big) = \frac{1}{\epsilon}\frac{1}{n}\delta_n(1+o_P(1)) \geq \frac{1}{\epsilon}\lambda^\star(1-\lambda^\star)\|\Delta_n\|(1+o_P(1)) \\
&\geq \frac{1}{\epsilon}(1-\lambda^\star)\|\Delta_n\|(1+o_P(1)).
\end{aligned} \tag{35}$$

Combining (34) and (35), we receive for $V_k(Y^{(n)})$ (recall (30)):

$$\begin{aligned}
&\max_{(1-\epsilon)k^\star \leq k \leq k^\star - a(n)} \frac{-V_k(Y^{(n)})}{(n(k^\star - k))^2} \\
&\geq \; -\Big((1-\lambda^\star)\|\Delta_n\|(1+o_P(1))\Big)^2 \\
&\qquad + 2\Big((1-\lambda^\star)\|\Delta_n\|(1+o_P(1))\Big)\frac{1}{\epsilon}(1-\lambda^\star)\|\Delta_n\|(1+o_P(1)) \\
&= \; -(1-\lambda^\star)^2\|\Delta_n\|^2(1+o_P(1)) + \frac{2}{\epsilon}(1-\lambda^\star)^2\|\Delta_n\|^2(1+o_P(1)) \\
&= (\frac{2}{\epsilon} - 1)(1-\lambda^\star)^2\|\Delta_n\|^2(1+o_P(1)) \\
&> 0 \quad (\forall \epsilon < 2),
\end{aligned}$$

which means that (28) holds and the statement is proven. □

3.2 Estimating the Direction of Change

In this section, we turn back to the fixed magnitude of change as stated in Section 3.1. We now focus our attention on the next step of our estimations, the direction of change. For that we will utilise the quantile function of our process. In a change-point setting, the median of the differences before and after the change equals the direction of the change. We prove that we can use the estimated empirical median based on the estimated change-point $\hat{k}$ as a suitable estimator for the true median. The idea of using the empirical median to quantify changes in location has a long tradition:

For a sample $x_1, ..., x_n \in \mathbb{R}$, the (sample) median is defined as the value $Q \in \mathbb{R}$ that minimises the distance to all values: $\sum_{i=1}^{n} |x_i - Q|$. One well-known example for the usage of the median in statistics was introduced by Hodges and Lehmann [1963]. They proposed to use the median of all pair-wise averages $\{(x_i + x_j)/2, 1 \leq i, j \leq n, i \neq j\}$ for a robust estimation of location. Since then, there have been several extensions to the median and Hodges-Lehmann-type estimators.

For a random variable X in $\mathbb{R}^d$, define $Q(u), u \in B_{\mathbb{R}^d}(0,1)$ as the value minimising $\mathbb{E}[\|X - Q\| - \|X\|] - \langle u, Q \rangle$ with respect to Q. This type of function is called *spatial quantile function*. *Spatial* originates from the fact that $Q(u)$ is equivalent to the solution Q of $\mathbb{E}[S(Q - X)] = u$, where $S(\cdot)$ is the spatial sign function. Setting $u = 0$ gives the median, thus $Q(0)$ is also called the *spatial median*. For the univariate case, Bahadur [1966] approximated the quantile function by its empirical analogue and provided asymptotic results, also known as Bahadur-Kiefer representation (see also Kiefer [1967]). Chaudhuri [1996] established the version for the multivariate case, as well as asymptotic results for a multivariate Hodges-Lehmann-type estimator with $m-$wise averages. Chakraborty and Chaudhuri [2014] introduced the extension of the concept to infinite dimensional Banach spaces as well as asymptotic Bahadur-Kiefer-type results.

Generalizing the multivariate m-th order Hodges-Lehmann estimator, Zhou and Serfling [2008] present a Bahadur-Kiefer representation for univariate spatial U-quantiles: For i.i.d. random variables $X_1, ..., X_n$ and a kernel function h, the *spatial U-quantile function* is defined as the solution Q of $\mathbb{E}[S(Q - h(X_{i_1}, ..., X_{i_m}))] = u$, where $i_1, ..., i_m$ are distinct indices from $\{1, ..., n\}$. Univariate *empirical U-quantiles* have been studied for example by Dehling et al. [1987] in the i.i.d. case. Wendler [2011] presented a Bahadur-Kiefer representation for functionals of absolutely regular processes.

Transferring the concepts of spatial U-quantiles to a two-sample setting, gives us tools for the estimation of the direction of a detected change. To get the idea, assume we have two samples $x_1, ..., x_{n_1} \in \mathbb{R}$ and $z_1, ..., z_{n_2} \in \mathbb{R}$. The difference in location can be estimated by $Q = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i - \frac{1}{n_2}\sum_{j=1}^{n_2} z_j$. This is equivalent to finding the solution Q of

$$\frac{1}{n_1}\sum_{i=1}^{n_1} x_i - \frac{1}{n_2}\sum_{j=1}^{n_2} z_j - Q = 0$$

$$\Leftrightarrow \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (x_i - z_j - Q) = 0. \tag{36}$$

So, (36) can be seen as some *two-sample empirical quantile function* (for u = 0). Asymptotic results for two-sample empirical U-quantiles and U-statistics for weakly dependent data are given by Dehling and Fried [2012].
For our purposes we have to connect all of the above to define the two-sample quantile function in (infinite) Hilbert spaces and achieve asymptotic results for the empirical version of weakly dependent data. Additionally, we have to handle the problem of the unknown but estimated time of change.

First of all, we define the quantile function:

Definition 3.1. *For two random variables $X, Z \in H$, the quantile function at $u \in B_H(0,1)$ is the solution $Q \in H$ of*

$$\mathbb{E}[S(Q - f(X,Z))] = u,$$

where $f(x,y) = x - y$.

Remark 3.1 (Uniqueness)**.** *Let μ be the law of $f(X,Z)$, assuming μ is non-atomic. Since a Hilbert space is a smooth, strictly convex Banach space, Theorem 3.1 of Chakraborty and Chaudhuri [2014] gives that the map $x \mapsto \mathbb{E}[S(x - f(X,Z))]$ is strictly monotone if μ is not entirely supported on a straight line in H. This means that in this case the solution Q of $\mathbb{E}[S(Q - f(X,Z))] = u$ is unique.*

Example 3.1 (Non-uniqueness)**.** *Uniqueness arguments hold analogously for the empirical quantile function: If all observations lie on a straight line, uniqueness is not given. For a minimal example of non-uniqueness in $\mathbb{R}^2$, look for the solution $(q_1, q_2)^T$ of*

$$\frac{1}{2}\Big(S\big((q_1,q_2)^T - (-1,-1)^T\big) + S\big((q_1,q_2)^T - (1,1)\big)^T\Big) = 0,$$

which is equivalent to solving the equations

$$(q_1 - 1)\sqrt{(q_1+1)^2 + (q_2+1)^2} + (q_1+1)\sqrt{(q_1-1)^2 + (q_2-1)^2} = 0$$
$$(q_2 - 1)\sqrt{(q_1+1)^2 + (q_2+1)^2} + (q_2+1)\sqrt{(q_1-1)^2 + (q_2-1)^2} = 0$$

simultaneously. In fact, any $(q_1, q_2)^T = (q,q)^T$ with $q \in [-1,1]$ solves the equations.

Given a sample of functional data $(Y_i)_{i \le n}$ with an existing but unknown change-point, the quantile function can only be estimated, splitting the sample at the estimated time of change. A more convenient way to formulate the spatial quantile function and its empirical analogue is the following:

- Define $Q^\star(u)$ as the minimiser (with respect to Q) of

  $$\varphi(Q) := \mathbb{E}[\|Q - f(X_1,Z_1)\| - \|f(X_1,Z_1)\|] - \langle u, Q\rangle\,,$$

  where $\langle \cdot, \cdot \rangle$ is the scalar product in H.

- Define $\hat{Q}(u)$ as the minimiser (with respect to Q) of

$$\hat{\varphi}(Q) := \frac{1}{\hat{k}(n-\hat{k})} \sum_{i=1}^{\hat{k}} \sum_{j=\hat{k}+1}^{n} (\|Q - f(Y_i, Y_j)\| - \|f(Y_i, Y_j)\|) - \langle u, Q \rangle ,$$

where $(Y_i)_{i \le n}$ is defined as stated in the beginning of this chapter. Call $\hat{Q}(u)$ the *estimated* empirical version of $Q(u)$.

Lemma 3.5. *Q is solution of $\mathbb{E}[S(Q - f(X_1, Z_1))] = u$ if and only if it minimises $\varphi(Q(u))$.*

Proof. The proof is given by showing both implications of the statement.
First, assume that Q_0 minimises φ. If φ is Gateaux-differentiable, we will utilise the fact that the derivative will equal 0 at Q_0. Let $w \in H$. The norm function $w \mapsto \|w\|$ is Gateaux-differentiable with derivative $\frac{w}{\|w\|}$ and for any $v \in H$, the inner product $w \mapsto \langle w, v \rangle$ is also Gateaux-differentiable with derivative v. Since φ is composed of such functions, it is thus Gateaux-differentiable as well with its derivative

$$\varphi'(Q) = \mathbb{E}\Big[\frac{Q - f(X_1, Z_1)}{\|Q - f(X_1, Z_1)\|}\Big] - u.$$

Inserting $Q = Q_0$, we get that

$$\begin{aligned} &\varphi'(Q_0) = 0 \\ \Leftrightarrow\ &\mathbb{E}\left[\frac{Q_0 - f(X_1, Z_1)}{\|Q_0 - f(X_1, Z_1)\|}\right] - u = 0 \\ \Leftrightarrow\ &\mathbb{E}\left[\frac{Q_0 - f(X_1, Z_1)}{\|Q_0 - f(X_1, Z_1)\|}\right] = u, \end{aligned}$$

which concludes the first part of the lemma.

For the other implication, we will show that if Q_0 solves $\mathbb{E}\Big[\frac{Q_0 - f(X_1,Z_1)}{\|Q_0 - f(X_1,Z_1)\|}\Big] = u$, it is $\varphi(Q_0) \le \varphi(Q)$ for any $Q \in H$.
Note that for $x \in H$ it is

$$\langle u, x \rangle = \left\langle \mathbb{E}\left[\frac{Q_0 - f(X_1, Z_1)}{\|Q_0 - f(X_1, Z_1)\|}\right], x \right\rangle = \mathbb{E}\left[\frac{1}{\|Q_0 - f(X_1, Z_1)\|} \langle Q_0 - f(X_1, Z_1), x \rangle\right]$$

since Q_0 solves $\mathbb{E}\Big[\frac{Q_0 - f(X_1,Z_1)}{\|Q_0 - f(X_1,Z_1)\|}\Big] = u$. Now, upper bound the difference by using the Cauchy-Schwarz-inequality

$$\begin{aligned} &\varphi(Q_0) - \varphi(Q) \\ &= \mathbb{E}[\|Q_0 - f(X_1, Z_1)\| - \|f(X_1, Z_1)\|] - \langle u, Q_0 \rangle \\ &\quad - \mathbb{E}[\|Q - f(X_1, Z_1)\| - \|f(X_1, Z_1)\|] + \langle u, Q \rangle \\ &= \mathbb{E}[\|Q_0 - f(X_1, Z_1)\| - \|Q - f(X_1, Z_1)\|] - \langle u, Q_0 - Q \rangle \end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}[\|Q_0 - f(X_1, Z_1)\| - \|Q - f(X_1, Z_1)\|] \\
&\quad - \mathbb{E}\left[\frac{1}{\|Q_0 - f(X_1, Z_1)\|}\langle Q_0 - f(X_1, Z_1), Q_0 - Q\rangle\right] \\
&= \mathbb{E}[\|Q_0 - f(X_1, Z_1)\| - \|Q - f(X_1, Z_1)\|] \\
&\quad - \mathbb{E}\Bigg[\frac{1}{\|Q_0 - f(X_1, Z_1)\|}\Big(\langle Q_0 - f(X_1, Z_1), Q_0 - f(X_1, Z_1)\rangle \\
&\qquad\quad - \langle Q_0 - f(X_1, Z_1), Q - f(X_1, Z_1)\rangle\Big)\Bigg] \\
&\leq \mathbb{E}[\|Q_0 - f(X_1, Z_1)\| - \|Q - f(X_1, Z_1)\|] \\
&\quad - \mathbb{E}\Bigg[\frac{1}{\|Q_0 - f(X_1, Z_1)\|}\Big(\|Q_0 - f(X_1, Z_1)\|^2 \\
&\qquad\quad - \|Q_0 - f(X_1, Z_1)\|\,\|Q - f(X_1, Z_1)\|\Big)\Bigg] \\
&= \mathbb{E}[\|Q_0 - f(X_1, Z_1)\| - \|Q - f(X_1, Z_1)\|] \\
&\quad - \mathbb{E}[\|Q_0 - f(X_1, Z_1)\| - \|Q - f(X_1, Z_1)\|] \\
&= 0.
\end{aligned}
$$

Thus it is $\varphi(Q_0) - \varphi(Q) \leq 0 \Leftrightarrow \varphi(Q_0) \leq \varphi(Q)$ and the lemma is proven. □

Since $Q^\star(u)$ minimises $\varphi(Q)$ by definition, the lemma above gives us that $Q^\star(u)$ is actually the quantile function given in Definition 3.1. As an analogue, we will call $\hat{Q}(u)$ the estimated quantile function.
Our goal is to show that $\|\hat{Q}(0) - Q^\star(0)\| \xrightarrow{n\to\infty} 0$ in probability, which means that the estimated spatial median is a suitable approximation of the spatial median.
The first step is to show that $\hat{Q}(0)$ "approximately" minimises φ in an almost sure sense:

Theorem 3.4. *Let $(X_n, Z_n)_{n\in\mathbb{Z}}$ be $\mathbb{P}$-NED on an absolutely regular sequence $(\zeta_n)_{n\in\mathbb{Z}}$ such that $a_k\Phi(k^{-\frac{\delta+3}{\delta}}) = \mathcal{O}(k^{-8\frac{(\delta+3)(\delta+2)}{\delta^2}})$ and $\sum_{k=1}^\infty k^2\beta_k^{\frac{\delta}{4+\delta}} < \infty$ for some $\delta > 0$. Denote $f(X_1, Z_1) \sim \mu$ and assume that μ is δ-tight, i.e. for all $\epsilon > 0$ there exists a compact set K and a sequence $\delta_n \xrightarrow{n\to\infty} 0$ such that*

$$\mu(K^{\delta_n}) = \mu(\{x \in H | d(x, K) < \delta_n\}) > 1 - \epsilon \quad \forall n.$$

*Assume that $\|Q - f(x, z)\| - \|f(x, z)\|$ fulfils the variation condition and is either bounded or has uniform $(4+\delta)$-moments under approximation.
Then it holds for $\hat{Q}(0)$ and $Q^\star(0)$ with $\|Q^\star(0)\| \leq M < \infty$, that*

$$\varphi(\hat{Q}(0)) - \varphi(Q^\star(0)) \to 0 \textit{ in probability as } n \to \infty.$$

For the purpose of proving Theorem 3.4, recall that $Q^\star$ minimises φ and extend $\varphi(\hat{Q}(0)) - \varphi(Q^\star(0))$ by $[-\hat{\varphi}(\hat{Q}(0)) + \hat{\varphi}(\hat{Q}(0))]$ to get

$$\varphi(\hat{Q}(0)) - \varphi(Q^\star(0)) = |\varphi(\hat{Q}(0)) - \hat{\varphi}(\hat{Q}(0)) + \hat{\varphi}(\hat{Q}(0)) - \varphi(Q^\star(0))|$$

$$\leq |\varphi(\hat{Q}(0)) - \hat{\varphi}(\hat{Q}(0))| + |\hat{\varphi}(Q^\star(0)) - \varphi(Q^\star(0))|.$$

We will show the convergence of the two parts separately. Starting with the second part, we will introduce a new kernel function and bound the difference $|\varphi(Q)-\hat{\varphi}(Q)|$ for arbitrary $Q \in H$ with $\|Q\| \leq M < \infty$. In order to do this, we will rewrite the double sum of $\hat{\varphi}(Q)$ and make use of asymptotic results from the last chapter.

Define a new kernel $h^Q(x,y) := \|Q - f(x,y)\| - \|f(x,y)\|$. Then, split up $\hat{\varphi}(Q)$ into two parts:

$$\hat{\varphi}(Q) = \frac{1}{\hat{k}(n-\hat{k})} \sum_{i=1}^{\hat{k}} \sum_{j=\hat{k}+1}^{n} h^Q(Y_i, Y_j) - \langle u, Q\rangle$$

$$= \begin{cases} \frac{1}{\hat{k}(n-\hat{k})}\big(\sum_{i=1}^{k^\star}\sum_{j=\hat{k}+1}^{n} h^Q(Y_i,Y_j) + \sum_{i=k^\star+1}^{\hat{k}}\sum_{j=\hat{k}+1}^{n} h^Q(Y_i,Y_j)\big) - \langle u, Q\rangle & \hat{k} \geq k^\star \\ \frac{1}{\hat{k}(n-\hat{k})}\big(\sum_{i=1}^{\hat{k}}\sum_{j=k^\star+1}^{n} h^Q(Y_i,Y_j) + \sum_{i=1}^{\hat{k}}\sum_{j=\hat{k}+1}^{k^\star} h^Q(Y_i,Y_j)\big) - \langle u, Q\rangle & \hat{k} \leq k^\star \end{cases}$$

$$= \begin{cases} \frac{1}{\hat{k}(n-\hat{k})}\big(\sum_{i=1}^{k^\star}\sum_{j=\hat{k}+1}^{n} h^Q(X_i,Z_j) + \sum_{i=k^\star+1}^{\hat{k}}\sum_{j=\hat{k}+1}^{n} h^Q(Z_i,Z_j)\big) - \langle u, Q\rangle & \hat{k} \geq k^\star \\ \frac{1}{\hat{k}(n-\hat{k})}\big(\sum_{i=1}^{\hat{k}}\sum_{j=k^\star+1}^{n} h^Q(X_i,Z_j) + \sum_{i=1}^{\hat{k}}\sum_{j=\hat{k}+1}^{k^\star} h^Q(X_i,X_j)\big) - \langle u, Q\rangle & \hat{k} \leq k^\star \end{cases}.$$

From here on forward, we will handle the case $\hat{k} \geq k^\star$. The other case works very similar.
It follows for the difference

$$|\varphi(Q) - \hat{\varphi}(Q)| = \Big|\frac{1}{\hat{k}(n-\hat{k})} \sum_{i=1}^{\hat{k}} \sum_{j=\hat{k}+1}^{n} h^Q(Y_i, Y_j) - \mathbb{E}[h^Q(X_0, Z_0)]\Big|$$

$$= \Big|\frac{1}{\hat{k}(n-\hat{k})} \sum_{i=1}^{\hat{k}} \sum_{j=\hat{k}+1}^{n} (h^Q(Y_i, Y_j) - \mathbb{E}[h^Q(X_i, Z_j)])\Big|$$

$$= \Big|\frac{1}{\hat{k}(n-\hat{k})}\Big(\sum_{i=1}^{k^\star} \sum_{j=\hat{k}+1}^{n} (h^Q(X_i, Z_j) - \mathbb{E}[h^Q(X_i, Z_j)])$$

$$+ \sum_{i=k^\star+1}^{\hat{k}} \sum_{j=\hat{k}+1}^{n} (h^Q(Z_i, Z_j) - \mathbb{E}[h^Q(X_i, Z_j)])\Big)\Big|$$

$$\leq \Big|\frac{1}{\hat{k}(n-\hat{k})} \sum_{i=1}^{k^\star} \sum_{j=\hat{k}+1}^{n} (h^Q(X_i, Z_j) - \mathbb{E}[h^Q(X_i, Z_j)])\Big|$$

$$+ \Big|\frac{1}{\hat{k}(n-\hat{k})} \sum_{i=k^\star+1}^{\hat{k}} \sum_{j=\hat{k}+1}^{n} (h^Q(Z_i, Z_j) - \mathbb{E}[h^Q(X_i, Z_j)])\Big|. \tag{37}$$

We will now show that both parts converge almost surely to zero.

Lemma 3.6. *Under the assumptions of Theorem 3.4 it holds for any $Q \in H$ that*
$$\Big|\frac{1}{\hat{k}(n-\hat{k})}\sum_{i=1}^{k^*}\sum_{j=\hat{k}+1}^{n}\left(h^Q(X_i,Z_j)-\mathbb{E}[h^Q(X_i,Z_j)]\right)\Big| \xrightarrow{a.s.} 0 \text{ as } n \to \infty.$$

Proof. Hoeffding's decomposition of $h^Q(X_s, Z_t)$ reads
$$h^Q(X_s,Z_t) = \vartheta^{X,Z} + h_1^{Z,Q}(X_s) + h_1^{X,Q}(Z_t) + h_2^{(X,Z),Q}(X_s,Z_t), \quad s,t \in \mathbb{Z},$$
$$\text{where } \vartheta^{X,Z} = \mathbb{E}[h^Q(\tilde{X}_0,\tilde{Z}_0)], \quad h_1^{Z,Q}(x) = \mathbb{E}[h^Q(x,\tilde{Z}_0)] - \vartheta^{X,Z}$$
$$h_1^{X,Q}(x) = \mathbb{E}[h^Q(\tilde{X}_0,x)] - \vartheta^{X,Z},$$
$$h_2^{(X,Z),Q}(x,y) = h^Q(x,y) - h_1^{Z,Q}(x) - h_1^{X,Q}(y) - \vartheta^{X,Z},$$
where $\tilde{X}_0$ and $\tilde{Z}_0$ are independent copies of X_0 and Z_0. Note that Hoeffding's decomposition looks a little bit different than for other kernels used before, since the kernel h^Q is not antisymmetric. Nevertheless, by this decomposition $h_2^{(X,Z),Q}$ is still degenerated.
We split our sum into three parts and use the decomposition:
$$\begin{aligned}
&\Big|\frac{1}{\hat{k}(n-\hat{k})}\sum_{i=1}^{k^\star}\sum_{j=\hat{k}+1}^{n}\left(h^Q(X_i,Z_j)-\mathbb{E}[h^Q(X_i,Z_j)]\right)\Big| \\
\le &\Big|\frac{1}{\hat{k}(n-\hat{k})}\sum_{i=1}^{k^\star}\sum_{j=\hat{k}+1}^{n} h_1^{Z,Q}(X_i)\Big| + \Big|\frac{1}{\hat{k}(n-\hat{k})}\sum_{i=1}^{k^\star}\sum_{j=\hat{k}+1}^{n} h_1^{X,Q}(Z_j)\Big| \\
&+ \Big|\frac{1}{\hat{k}(n-\hat{k})}\sum_{i=1}^{k^\star}\sum_{j=\hat{k}+1}^{n} h_2^{(X,Z),Q}(X_i,Z_j)\Big| \\
= &\underbrace{\Big|\frac{1}{\hat{k}}\sum_{i=1}^{k^\star} h_1^{Z,Q}(X_i)\Big|}_{(a)} + \underbrace{\Big|\frac{k^\star}{\hat{k}}\frac{1}{n-\hat{k}}\sum_{j=\hat{k}+1}^{n} h_1^{X,Q}(Z_j)\Big|}_{(b)} + \underbrace{\Big|\frac{1}{\hat{k}(n-\hat{k})}\sum_{i=1}^{k^\star}\sum_{j=\hat{k}+1}^{n} h_2^{(X,Z),Q}(X_i,Z_j)\Big|}_{(c)}.
\end{aligned}$$
Each of the parts converges almost surely to zero. We show this by arguments similar to those seen before:
$$\begin{aligned}
(a) &\le \Big|\frac{1}{k^\star}\sum_{i=1}^{k^\star} h_1^{Z,Q}(X_i)\Big| = \Big|\frac{1}{k^\star}\sum_{i=1}^{k^\star} h_1^{Z,Q}(X_i) - \underbrace{\mathbb{E}[h_1^{Z,Q}(X_i)]}_{=0}\Big| \\
&= \Big|\frac{1}{\lfloor\lambda^\star n\rfloor}\sum_{i=1}^{\lfloor\lambda^\star n\rfloor} h_1^{Z,Q}(X_i) - \mathbb{E}[h_1^{Z,Q}(X_i)]\Big| \xrightarrow{a.s.} 0
\end{aligned}$$
as $n \to \infty$ by Birkhoff's ergodic theorem.
For (b) we observe that for n large enough
$$\Big|\underbrace{\frac{k^\star}{\hat{k}}}_{\le 1}\frac{1}{n-\hat{k}}\sum_{j=\hat{k}+1}^{n} h_1^{X,Q}(Z_j)\Big| \le \Big|\frac{1}{n-\hat{k}}\sum_{j=\hat{k}+1}^{n} h_1^{X,Q}(Z_j)\Big|$$

$$
\begin{aligned}
&= \Big|\frac{1}{n-\hat{k}} \sum_{j=\hat{k}+1}^{n} h_1^{X,Q}(Z_j) - \underbrace{\mathbb{E}[h_1^{X,Q}(Z_j)]}_{=0}\Big| = \Big|\frac{1}{n-\hat{k}} \sum_{j=\hat{k}+1}^{n} h_1^{X,Q}(Z_j) - \mathbb{E}[h_1^{X,Q}(Z_j)]\Big| \\
&= \Big|\frac{1}{n-\hat{k}} \Big(\sum_{j=1}^{n} h_1^{X,Q}(Z_j) - \mathbb{E}[h_1^{X,Q}(Z_j)] - \sum_{j=1}^{\hat{k}} h_1^{X,Q}(Z_j) - \mathbb{E}[h_1^{X,Q}(Z_j)]\Big)\Big| \\
&\leq \Big|\frac{1}{n-\hat{k}} \sum_{j=1}^{n} h_1^{X,Q}(Z_j) - \mathbb{E}[h_1^{X,Q}(Z_j)]\Big| + \Big|\frac{1}{n-\hat{k}} \sum_{j=1}^{\hat{k}} h_1^{X,Q}(Z_j) - \mathbb{E}[h_1^{X,Q}(Z_j)]\Big| \\
&= \Big|\frac{1}{(n-\lfloor \hat{\lambda} n\rfloor)} \sum_{j=1}^{n} h_1^{X,Q}(Z_j) - \mathbb{E}[h_1^{X,Q}(Z_j)]\Big| \\
&\quad + \Big|\frac{1}{(n-\lfloor \hat{\lambda} n\rfloor)} \sum_{j=1}^{\lfloor \hat{\lambda} n\rfloor} h_1^{X,Q}(Z_j) - \mathbb{E}[h_1^{X,Q}(Z_j)]\Big| \\
&\leq \frac{n}{(n-\lfloor \hat{\lambda} n\rfloor)} \Big|\frac{1}{n} \sum_{j=1}^{n} h_1^{X,Q}(Z_j) - \mathbb{E}[h_1^{X,Q}(Z_j)]\Big| \\
&\quad + \frac{n}{(n-\lfloor \hat{\lambda} n\rfloor)} \max_{0\leq\lambda\leq 1} \Big|\frac{1}{n} \sum_{j=1}^{\lfloor \lambda n\rfloor} h_1^{X,Q}(Z_j) - \mathbb{E}[h_1^{X,Q}(Z_j)]\Big| \\
&\approx \frac{1}{1-\hat{\lambda}} \Big|\frac{1}{n} \sum_{j=1}^{n} h_1^{X,Q}(Z_j) - \mathbb{E}[h_1^{X,Q}(Z_j)]\Big| \\
&\quad + \frac{1}{1-\hat{\lambda}} \max_{0\leq\lambda\leq 1} \Big|\frac{1}{n} \sum_{j=1}^{\lfloor \lambda n\rfloor} h_1^{X,Q}(Z_j) - \mathbb{E}[h_1^{X,Q}(Z_j)]\Big| \\
&\quad \xrightarrow{a.s.} 0,
\end{aligned}
$$

since by continuous mapping theorem, $\frac{1}{1-\hat{\lambda}} \xrightarrow{n\to\infty} \frac{1}{1-\lambda^*}$ and the sums converge (uniformly) a.s. to zero as $n \to \infty$ by Birkhoff's ergodic theorem.
Lastly, (c) converges to zero as well, since for n large enough

$$
\begin{aligned}
&\Big|\frac{1}{\hat{k}(n-\hat{k})} \sum_{i=1}^{k^\star} \sum_{j=\hat{k}+1}^{n} h_2^{(X,Z),Q}(X_i, Z_j)\Big| \approx \frac{1}{\hat{\lambda}(1-\hat{\lambda})} \Big|\frac{1}{n^2} \sum_{i=1}^{k^\star} \sum_{j=\hat{k}+1}^{n} h_2^{(X,Z),Q}(X_i, Z_j)\Big| \\
&\leq \frac{1}{\hat{\lambda}(1-\hat{\lambda})} \max_{k^\star\leq k<n} \Big|\frac{1}{n^2} \sum_{i=1}^{k^\star} \sum_{j=k+1}^{n} h_2^{(X,Z),Q}(X_i, Z_j)\Big| \xrightarrow{a.s.} 0
\end{aligned}
$$

because $\frac{1}{\hat{\lambda}(1-\hat{\lambda})} \xrightarrow{n\to\infty} \frac{1}{\lambda^\star(1-\lambda^\star)}$ (a.s.) and the maximum converges a.s. to zero by using an analogue of Proposition 2.5, where $h_2^{(X,Z)}(X_i, Z_j) - \mathbb{E}[h_2^{(X,Z)}(\tilde{X}_i, \tilde{Z}_j)]$ is replaced by $h_2^{(X,Z),Q}(X_i, Z_j)$. Since this is degenerate, the proof can be adapted. Just note that minor adjustments of the proof have to be made, since $h_2^{(X,Z),Q}$ is not

antisymmetric. See also Remark 2.6 for this.
Combining all three parts, the statement of the lemma is proven. □

We will now take care of the second summand of (37):

Lemma 3.7. *Under the assumptions of Theorem 3.4 it holds for any $Q \in H$ with $\|Q\| \leq M < \infty$ that*

$$\Big|\frac{1}{\hat{k}(n-\hat{k})}\sum_{i=k^\star+1}^{\hat{k}}\sum_{j=\hat{k}+1}^{n}(h^Q(Z_i,Z_j)-\mathbb{E}[h^Q(X_i,Z_j)])\Big| \xrightarrow{a.s.} 0 \text{ as } n\to\infty.$$

Proof. We will make use of the fact that $\|Q\|$ is bounded and the estimated time of change converges to the real time of change:

$$\begin{aligned}
&\Big|\frac{1}{\hat{k}(n-\hat{k})}\sum_{i=k^\star+1}^{\hat{k}}\sum_{j=\hat{k}+1}^{n}(h^Q(Z_i,Z_j)-\mathbb{E}[h^Q(X_i,Z_j)])\Big| \\
&\leq \frac{1}{\hat{k}(n-\hat{k})}\sum_{i=k^\star+1}^{\hat{k}}\sum_{j=\hat{k}+1}^{n}\big|\|Q-f(Z_i,Z_j)\|-\|f(Z_i,Z_j)\|\big| \\
&\qquad\qquad + \big|\mathbb{E}[\|Q-f(X_i,Z_j)\|-\|f(X_i,Z_j)\|]\big| \\
&\leq \frac{1}{\hat{k}(n-\hat{k})}\sum_{i=k^\star+1}^{\hat{k}}\sum_{j=\hat{k}+1}^{n}2\|Q\| \leq 2M\frac{(\hat{k}-k^\star)(n-\hat{k})}{\hat{k}(n-\hat{k})} \\
&\leq 2M\frac{\hat{\lambda}-\lambda^\star}{\lambda^\star} \xrightarrow{a.s.} 0 \text{ as } n\to\infty.
\end{aligned}$$

□

Combining Lemmas 3.6 and 3.7, we can already conclude that

$$|\varphi(Q^\star(0)) - \hat{\varphi}(Q^\star(0))| \xrightarrow{n\to\infty} 0,$$

since $\|Q^\star(0)\| < M$ under the assumptions of Theorem 3.4.
The next step is to show that $|\varphi(\hat{Q}(0)) - \hat{\varphi}(\hat{Q}(0))| \xrightarrow{n\to\infty} 0$. We follow an idea of Cadre [2001], where we need a metric of measures to show this.

Definition 3.2. *Let L be the class of measurable functions $g : H \to \mathbb{R}$ that are 1-Lipschitz and absolutely bounded by 1, i.e. $|g(x)-g(y)| \leq \|x-y\|$ for all $x, y \in H$ and $\sup_{x\in H}|g(x)| \leq 1$. Then, for any laws P and $\tilde{P}$, the metric $\beta(\cdot,\cdot)$ is defined as*

$$\beta(P,\tilde{P}) := \sup_{g\in L}\Big|\int gd(P-\tilde{P})\Big|.$$

Furthermore, denote

$$\mu_n = \frac{1}{k^\star(n-k^\star)}\sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n}\delta_{f(X_i,Z_j)}$$

as the empirical measure based on the (unknown) true change-point $k^\star$ and

$$\hat{\mu}_n = \frac{1}{\hat{k}(n-\hat{k})} \sum_{i=1}^{\hat{k}} \sum_{j=\hat{k}+1}^{n} \delta_{f(Y_i,Y_j)}$$

as the *estimated* empirical measure resulting from the estimated time of change $\hat{k}$. Now, Lemma 1 Cadre [2001] gives us for any $Q \in H$ that

$$\sup_{n\geq 1} \left| \frac{\hat{\varphi}(Q)}{\|Q\|} - 1 \right| \to 0 \text{ as } \|Q\| \to \infty.$$

Additionally, in the proof of Lemma 2 i) of Cadre [2001] it reads: *Let* $r_1 = \sup_{n\geq 1} \|\hat{Q}(0)\|$. *Since* $\hat{Q} = \operatorname{argmin}_{Q\in H} \hat{\varphi}$ *for all* $n \geq 1$, *Lemma 1 Cadre [2001] shows that* $r_1 < \infty$. *Then, for all* $n \geq 1$*:*

$$\begin{aligned}
&|\hat{\varphi}(\hat{Q}(0)) - \varphi(\hat{Q}(0))| = \Big| \int (\|\hat{Q}(0) - x\| - \|x\|)(\hat{\mu}_n - \mu)dx \Big| \\
&= 3\max(1, \|\hat{Q}(0)\|) \Big| \int \frac{\|\hat{Q}(0) - x\| - \|x\|}{3\max(1, \|\hat{Q}(0)\|)} (\hat{\mu}_n - \mu)dx \Big| \\
&\leq 3\max(1, r_1)\beta(\hat{\mu}_n, \mu).
\end{aligned}$$

This means, $|\varphi(\hat{Q}(0)) - \hat{\varphi}(\hat{Q}(0))| \xrightarrow{n\to\infty} 0$ will follow if the estimated empirical measure $\hat{\mu}_n$ converges to μ in metric $\beta(\cdot,\cdot)$.

Remark 3.2. *It is actually given by Theorem 11.3.3 Dudley [2018] that on separable metric spaces* S*, the equivalence*

$$\beta(P_n, P) \to 0 \Leftrightarrow P_n \to P$$

holds for laws P_n *and* P*. Where* $P_n \to P$ *means that* $\int f dP_n \xrightarrow{n\to\infty} \int f dP$ *for all bounded, continuous and real-valued functions* f *on* S*.*

We will use the fact that $\beta(\hat{\mu}_n, \mu) \leq \beta(\hat{\mu}_n, \mu_n) + \beta(\mu_n, \mu)$ and show convergence for both parts.

Lemma 3.8. *Under the assumptions of Theorem 3.4,*

$$\beta(\hat{\mu}_n, \mu_n) \xrightarrow{n\to\infty} 0 \textit{ alomst surely.}$$

Proof. The convergence of the measures will follow from the convergence of the estimated change-point to the true change-point. Let $g \in L$ be arbitrary. Use the fact that $|g|$ is bounded by 1 and calculate for the case where $\hat{k} \geq k^\star$:

$$
\begin{aligned}
&\Big|\int g d(\mu_n - \hat{\mu}_n)\Big| \\
&= \Big|\frac{1}{k^\star(n-k^\star)} \sum_{i=1}^{k^\star} \sum_{j=k^\star+1}^{n} g(f(X_i, Z_j)) - \frac{1}{\hat{k}(n-\hat{k})} \sum_{i=1}^{\hat{k}} \sum_{j=\hat{k}+1}^{n} g(f(Y_i, Y_j))\Big| \\
&= \Big|\frac{1}{k^\star(n-k^\star)} \sum_{i=1}^{k^\star} \sum_{j=k^\star+1}^{n} g(f(X_i, Z_j)) - \frac{1}{\hat{k}(n-\hat{k})} \sum_{i=1}^{k^\star} \sum_{j=k^\star+1}^{n} g(f(X_i, Z_j)) \\
&\quad + \frac{1}{\hat{k}(n-\hat{k})} \sum_{i=1}^{k^\star} \sum_{j=k^\star+1}^{\hat{k}} g(f(X_i, Z_j)) - \frac{1}{\hat{k}(n-\hat{k})} \sum_{i=k^\star+1}^{\hat{k}} \sum_{j=\hat{k}+1}^{n} g(f(Z_i, Z_j))\Big| \\
&\le \Big(\frac{1}{k^\star(n-k^\star)} - \frac{1}{\hat{k}(n-\hat{k})}\Big) k^\star(n-k^\star) + \frac{k^\star(\hat{k}-k^\star)}{\hat{k}(n-\hat{k})} + \frac{(\hat{k}-k^\star)(n-\hat{k})}{\hat{k}(n-\hat{k})} \\
&\approx \Big(\frac{1}{\lambda^\star n(n-\lambda^\star n)} - \frac{1}{\hat{\lambda} n(n-\hat{\lambda} n)}\Big) \lambda^\star n(n-\lambda^\star n) \\
&\quad + \frac{\lambda^\star n(\hat{\lambda} n - \lambda^\star n)}{\hat{\lambda} n(n-\hat{\lambda} n)} + \frac{(\hat{\lambda} n - \lambda^\star n)(n-\hat{\lambda} n)}{\hat{\lambda} n(n-\hat{\lambda} n)} \\
&= \Big(\frac{1}{\lambda^\star(1-\lambda^\star)} - \frac{1}{\hat{\lambda}(1-\hat{\lambda})}\Big) \lambda^\star(1-\lambda^\star) + \frac{\lambda^\star(\hat{\lambda}-\lambda^\star)}{\hat{\lambda}(1-\hat{\lambda})} + \frac{(\hat{\lambda}-\lambda^\star)(1-\hat{\lambda})}{\hat{\lambda}(1-\hat{\lambda})} \\
&\xrightarrow{a.s.} 0 \text{ when } n \to \infty, \text{ since } |\hat{\tau} - \tau^\star| \xrightarrow{n\to\infty} 0 \text{ almost surely.}
\end{aligned}
$$

This almost sure convergence does not depend on the choice of g and thus uniform convergence over L follows. □

The convergence of μ_n to μ is based on Theorem 2.1 Politis et al. [1999] however it has to be adapted to the change-point setting.

Lemma 3.9. *Under the assumptions of Theorem 3.4,*

$$\beta(\mu_n, \mu) \xrightarrow{n\to\infty} 0 \textit{ in probability.}$$

Proof. Let L be as in Definition 3.2. For any function $g \in L$, it holds that $\mu_n g - \mu g = \int g d(\mu_n - \mu)$ has mean zero and variance converging to zero, which can be seen by calculating:

$$
\begin{aligned}
\mathbb{E}[\mu_n g] &= \mathbb{E}\Big[\frac{1}{k^\star(n-k^\star)} \sum_{i=1}^{k^\star} \sum_{j=k^\star+1}^{n} g(f(X_i, Z_j))\Big] \\
&= \frac{1}{k^\star(n-k^\star)} \sum_{i=1}^{k^\star} \sum_{j=k^\star+1}^{n} \mathbb{E}\big[g(f(X_i, Z_j))\big] = \mathbb{E}\big[g(f(X_1, Z_1))\big] = \mu g.
\end{aligned}
$$

And thus $\mathbb{E}[\mu_n g - \mu g] = 0$. Use this for the variance and get:

$$\mathrm{Var}(\mu_n g - \mu g) = \mathrm{Var}(\mu_n g) = \mathbb{E}\big[(\mu_n g - \mathbb{E}[\mu_n g])^2\big] = \mathbb{E}\big[(\mu_n g - \mu g)^2\big]$$

$$= \mathbb{E}\Big[\Big(\frac{1}{k^\star(n-k^\star)}\sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n} g(f(X_i,Z_j)) - \mathbb{E}[g(f(X_i,Z_j))]\Big)^2\Big]$$

$$= \Big(\frac{1}{k^\star(n-k^\star)}\Big)^2\mathbb{E}\Big[\Big(\sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n} g(f(X_i,Z_j)) - \mathbb{E}[g(f(X_i,Z_j))]\Big)^2\Big]. \tag{38}$$

For the expectation, a variant of Lemma 2.6 can be applied with the difference that here not the "truncated" versions of the process and kernel are used. This causes the factor $n^{3/8}$ to disappear. For more details on this, also see Lemma 3.3.
If the function $\hat{g}(x,y) := g(f(x,y)) - \mathbb{E}[g(f(x,y))]$ is degenerate, fulfils the variation condition and is bounded, the arguments of Lemma 2.6 can be used. These properties can be verified by some short calculations, using the fact that $g \in L = \{g : H \to \mathbb{R} \mid \ |g(x)-g(y)| \le \|x-y\| \text{ and } \sup_{x\in H}|g(x)| \le 1\}$ and recalling that $f(x,y) = x - y$. Thus, it indeed holds that

- $\hat{g}$ is degenerate: $\mathbb{E}[\hat{g}(X,y)] = \mathbb{E}\big[g(f(X,y))\big] - \mathbb{E}\big[\mathbb{E}[g(f(X,y))]\big] = 0$ and $\mathbb{E}[\hat{g}(x,Y)] = 0$ as well.

- $\hat{g}$ fulfils the variation condition:

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\sup_{\substack{\|x-X\|\le\epsilon\\ \|y-Z\|\le\epsilon}} |\hat{g}(x,y)-\hat{g}(X,Z)|\Big)^2\Big]\\
&\le \mathbb{E}\Big[\Big(\sup_{\substack{\|x-X\|\le\epsilon\\ \|y-Z\|\le\epsilon}} (|g(f(x,y))-g(f(X,Z))| + \mathbb{E}\big[|g(f(x,y))-g(f(X,Z))|\big])\Big)^2\Big]\\
&\le \mathbb{E}\Big[\Big(\sup_{\substack{\|x-X\|\le\epsilon\\ \|y-Z\|\le\epsilon}} (\|f(x,y)-f(X,Z)\| + \mathbb{E}[\|f(x,y)-f(X,Z)\|])\Big)^2\Big]\\
&\le \mathbb{E}\Big[\Big(\sup_{\substack{\|x-X\|\le\epsilon\\ \|y-Z\|\le\epsilon}} (\|x-X\|+\|y-Z\| + \mathbb{E}[\|x-X\|+\|y-Z\|])\Big)^2\Big]\\
&\le L\epsilon.
\end{aligned}$$

- $\hat{g}$ is bounded:

$$\begin{aligned}
|\hat{g}(x,y)| &\le \big|g(f(x,y))\big| + \big|\mathbb{E}\big[g(f(x,y))\big]\big|\\
&\le \big|g(f(x,y))\big| + \mathbb{E}\big[\big|g(f(x,y))\big|\big]\\
&\le 2.
\end{aligned}$$

Having checked the properties of $\hat{g}$, the before mentioned variation of Lemma 2.6 gives

$$(38) = \Big(\frac{1}{k^\star(n-k^\star)}\Big)^2\mathbb{E}\Big[\Big(\sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n} g(f(X_i,Z_j)) - \mathbb{E}[g(f(X_i,Z_j))]\Big)^2\Big]$$

$$= \Big(\frac{1}{k^\star(n-k^\star)}\Big)^2 \mathbb{E}\Big[\Big(\sum_{i=1}^{k^\star}\sum_{j=k^\star+1}^{n} \hat{g}(X_i, Z_j)\Big)^2\Big]$$
$$\leq \Big(\frac{1}{k^\star(n-k^\star)}\Big)^2 Ck^\star(n-k^\star) = C\frac{1}{k^\star(n-k^\star)} \xrightarrow{n\to\infty} 0.$$

The rest of the proof follows from Theorem 2.1 Politis et al. [1999]. For the sake of completeness, the arguments are included:
Let L_n be functions of the form $g(x)\mathbf{1}_{\{x\in K^{\delta_n}\}}$, $g \in L$. It suffices to show $\sup_{g\in L_n} |\int gd(\mu_n - \mu)| \to 0$, since two properties hold:

A)
$$\mu(K^{\delta_n}) \geq 1 - \epsilon$$
since μ is δ-tight.

B) By Chebychev's inequality and the above variance calculation for $\mathbf{1}_{K^{\delta_n}} \in L$:
$$\begin{aligned}&\mathbb{P}\big(1-\mu_n(K^{\delta_n}) \geq 2\epsilon\big) = \mathbb{P}\big(\mu_n(K^{\delta_n}) - \mu(K^{\delta_n}) \leq 1 - \mu(K^{\delta_n}) - 2\epsilon\big)\\ &\leq \mathbb{P}\big(\mu_n(K^{\delta_n}) - \mu(K^{\delta_n}) \leq -\epsilon\big) \text{ by A)}\\ &\leq \mathbb{P}\big(|\mu_n(K^{\delta_n}) - \mu(K^{\delta_n})| \geq \epsilon\big)\\ &\leq \frac{1}{\epsilon^2}\operatorname{Var}\big(\mu_n(K^{\delta_n})\big)\\ &\leq \frac{1}{\epsilon^2}\frac{1}{k^\star(n-k^\star)} \xrightarrow{n\to\infty} 0.\end{aligned}$$

Then,
$$\sup_{g\in L}\Big|\int gd(\mu_n-\mu)\Big| \leq \sup_{g\in L_n}\Big|\int gd(\mu_n-\mu)\Big| + \sup_{g\in L_n'}\Big|\int gd(\mu_n-\mu)\Big|,$$
where L_n' are functions of the form $g(x)(1-\mathbf{1}_{\{x\in K^{\delta_n}\}})$, $g\in L$. However,
$$\begin{aligned}&\sup_{g\in L_n'}\Big|\int gd(\mu_n-\mu)\Big| \leq \sup_{g\in L_n'}\Big|\int gd\mu_n\Big| + \sup_{g\in L_n'}\Big|\int gd\mu\Big|\\ &\leq \big(1-\mu_n(K^{\delta_n})\big) + \epsilon \leq 2\epsilon + \epsilon\end{aligned}$$
with probability 1, by A) and B). Thus, it is left to show that
$$\sup_{g\in L_n}\Big|\int gd(\mu_n-\mu)\Big| \to 0 \text{ in probability.}$$

Fix some arbitrary $\epsilon > 0$ and let L_K be the functions of the form $g(x)\mathbf{1}_{x\in K}$, $g \in L$. Let $\{g_1, ..., g_{m_\epsilon}\}$ be and ϵ-net for L_K, i.e. if $g \in L$, there exists (at least) one function g_i from the collection $\{g_1, ..., g_{m_\epsilon}\}$ such that $\sup_{x\in K}|g(x) - g_i(x)| < \epsilon$. Note that since L_K is a family of uniformly bounded and equicontinuous (since 1-Lipschitz)

functions and K is compact, L_K is relatively compact (also called precompact) by the Arzela-Ascoli Theorem. Since the space of bounded functions $g : H \to \mathbb{R}$ is a complete metric space, L_K is equivalently totally bounded. Which means that the number m_ϵ of functions in the ϵ-net is finite. Furthermore, it can be assumed that $g_i \in L$ for $i = 1, ..., m_\epsilon$ (see Theorems 6.1.1 and 11.2.3 Dudley [2018]).
Now, by the definition of K^{δ_n}, if $x \in K^{\delta_n}$, there exists $\tilde{x} \in K$ such that $\|x-\tilde{x}\| < \delta_n$. Then, for any $g \in L$ it holds

$$|g(x) - g(\tilde{x})| \leq \|x - \tilde{x}\| < \delta_n \text{ by Lipschitz property of } g$$

Recall that $g_i \in L$ as well, and thus

$$\begin{aligned} g(x) &\leq g(\tilde{x}) + \delta_n = [g(\tilde{x}) - g_i(\tilde{x})] + [g_i(\tilde{x}) - g_i(x)] + g_i(x) + \delta_n \\ &\leq \epsilon + \delta_n + g_i(x) + \delta_n = g_i(x) + \epsilon + 2\delta_n. \end{aligned}$$

This means for any $g \in L_n$, there exists $i \leq m_\epsilon$ satisfying

$$\sup_{x \in K^{\delta_n}} |g(x) - g_i(x)| \leq \epsilon + 2\delta_n. \tag{39}$$

Given g, let $\tilde{g}$ be the $g_i \in \{g_1, ..., g_{m_\epsilon}\}$ satisfying (39). Then:

$$\begin{aligned} &\sup_{g \in L_n} \Big| \int g d(\mu_n - \mu) \Big| \\ &\leq \sup_{g \in L_n} \Big| \int (g - \tilde{g}) d(\mu_n - \mu) \Big| + \max_{1 \leq i \leq m_\epsilon} \Big| \int g_i d(\mu_n - \mu) \Big| \\ &\leq 2\epsilon + 4\delta_n + \max_{1 \leq i \leq m_\epsilon} \Big| \int g_i d(\mu_n - \mu) \Big|. \end{aligned}$$

Fix $\eta > 0$ and let $\epsilon = \eta/8$. Using the above calculations yields

$$\begin{aligned} &\mathbb{P}\Big(\sup_{g \in L_n} \Big| \int g d(\mu_n - \mu) \Big| > \eta \Big) \\ &\leq \mathbb{P}\Big(\sup_{g \in L_n} \Big| \int (g - \tilde{g}) d(\mu_n - \mu) \Big| > \eta/2 \Big) + \mathbb{P}\Big(\max_{1 \leq i \leq m_\epsilon} \Big| \int g_i d(\mu_n - \mu) \Big| > \eta/2 \Big) \\ &\leq \mathbb{P}\Big(\max_{1 \leq i \leq m_\epsilon} \Big| \int g_i d(\mu_n - \mu) \Big| > \eta/2 \Big) \end{aligned}$$

as soon as $2\epsilon + 4\delta_n = \eta/4 + 4\delta_n < \eta/2$ holds true. Equivalently, this is the case when $\eta > 16\delta_n$.
Finally, the last term can be bounded by

$$\frac{4m_\epsilon}{\eta^2} \Big(\frac{1}{k^\star(n - k^\star)} \Big)^2,$$

which converges in probability to zero, as $n \to \infty$. $\square$

Proof of Theorem 3.4. The statement of the theorem follows directly by combining Lemmas 3.6 to 3.9 as described before. $\square$

The desired convergence of the estimated spatial median to the spatial median finally follows from the above theorem and Proposition 1 i) of Asplund [1968] and is formulated in the following corollary:

Corollary 3.1. *Under the assumptions of Theorem 3.4, it holds for $Q^\star(0)$ and $\hat{Q}(0)$ with $\|Q^\star(0)\| \leq M < \infty$ that*

$$\|\hat{Q}(0) - Q^\star(0)\| \xrightarrow{n\to\infty} 0 \textit{ in probability.}$$

Proof of Corollary 3.1. Let Γ_U be the class of convex and left-continuous functions $\gamma : [0, \infty) \to [0, \infty]$ with $\gamma(0) = 0$ and $\gamma(t) > 0$ if $t > 0$. Note that this characterization of Γ_U means that any $\gamma \in \Gamma_U$ is in fact continuous (especially in zero), strictly convex and thus strictly monotone. It follows that γ is bijective and its inverse function γ^{-1} exists and is continuous itself.

Proposition 1 i) Asplund [1968] applied to our setting gives that for some $\gamma \in \Gamma_U$

$$\begin{aligned}&\varphi(Q(u)) + \langle u, Q(u)\rangle \\ &\geq \varphi(Q^\star(u)) + \langle u, Q^\star(u)\rangle + \langle Q(u) - Q^\star(u), u\rangle + \gamma(\|Q(u) - Q^\star(u)\|)\ \forall\ Q \in H.\end{aligned}$$

We can use this especially for $u = 0$ and $Q = \hat{Q}$ to achieve

$$\begin{aligned}&\varphi(\hat{Q}(0)) \geq \varphi(Q^\star(0)) + \gamma(\|\hat{Q}(0) - Q^\star(0)\|)\\ \Rightarrow\ &\varphi(\hat{Q}(0)) - \varphi(Q^\star(0)) \geq \gamma(\|\hat{Q}(0) - Q^\star(0)\|)\\ \Rightarrow\ &\gamma^{-1}\big(\varphi(\hat{Q}(0)) - \varphi(Q^\star(0))\big) \geq \|\hat{Q}(0) - Q^\star(0)\|.\end{aligned}$$

Thus, it can be concluded that for any $\epsilon > 0$

$$\mathbb{P}(\|\hat{Q}(0) - Q^\star(0)\| > \epsilon) \leq \mathbb{P}\Big(\gamma^{-1}\big(\varphi(\hat{Q}(0)) - \varphi(Q^\star(0))\big) > \epsilon\Big) \overset{n\to\infty}{\to} 0$$

by the continuous mapping theorem and since $\varphi(\hat{Q}(0)) - \varphi(Q^\star(0)) \to 0$ in probability by Theorem 3.4.

□

3.3 Simulation Study

In this section, we present the results of the estimator for time and direction in a simulation study.

3.3.1 Estimating the Time

Recall that we estimate the location of a change by the smallest $1 \leq k < n$ for which the test statistic attains its maximum:

$$\hat{k} = \min\left\{k : \left\|\frac{1}{n^{3/2}}U_{n,k}\right\| = \max_{1\leq j<n}\left\|\frac{1}{n^{3/2}}U_{n,j}\right\|\right\}.$$

We compare it to the change-point estimator given by the $1 \leq k < n$ that maximises the CUSUM statistic for the scenarios introduced in Section 2.6.3, to evaluate the performance of the test:

Scenario 1: Uniform jump of $+0.3$ after $n/2$ observations:

$$Y_i = \begin{cases} X_i & i < n/2 \\ X_i + 0.3u & i \geq n/2 \end{cases}$$

where $u = (1, ..., 1)^T$.

Scenario 2: Sinus-jump after $n/2$ of observations:

$$Y_i = \begin{cases} X_i & i < n/2 \\ X_i + \frac{1}{2\sqrt{2}}(\sin(\pi D/d))_{D\leq d} & i \geq n/2 \end{cases}.$$

Scenario 3: Uniform jump of $+0.3$ after $n/2$ observations in presence of outliers at $0.2n, 0.4n, 0.6n, 0.8n$:

$$Y_i = \begin{cases} X_i & i < n/2, i \notin \{0.2n, 0.4n\} \\ 10X_i & i \in \{0.2n, 0.4n\} \\ X_i + 0.3u & i \geq n/2, i \notin \{0.6n, 0.8n\} \\ 10X_i + 0.3u & i \in \{0.6n, 0.8n\} \end{cases}$$

where $u = (1, ..., 1)^T$.

Scenario 4: Heavy tails: In the simulation of $(X_i)_{i\leq n}$ we use $\xi_i, \xi_i^{(t)} \sim t_1$ (Cauchy distributed) $\forall i \leq d, -BI < t \leq n$ and a uniform jump of $+5$ after $n/2$ observations.

Scenario 5: Uniform Jump of $+0.3$ after γn observations:

$$Y_i = \begin{cases} X_i & i < \gamma n \\ X_i + 0.3u & i \geq \gamma n \end{cases} \quad \text{with } \gamma = 0.3 \text{ and } \gamma = 0.15 \text{ resp.}$$

where $u = (1, ..., 1)^T$.

Scenario 6: As Scenario 1 but with $n = 150$, $d = 350$.

If not stated otherwise, we simulated 3000 samples with $n = 200$ and $d = 100$ for each scenario and calculated $\hat{k}$ for CUSUM and our Wilcoxon-type statistic for each sample simultaneously. The 3000 evaluated $\hat{k}$ are presented in some histograms (Figures 3.1 to 3.3) to compare both procedures. Furthermore, Table 3.1 gives the proportion of $\hat{k}$ in the 20 %-interval around $k^\star$ given by $[k^\star - 0.1n, k^\star + 0.1n]$. We shortly remark that we did no pre-testing before calculating $\hat{k}$, meaning $\hat{k}$ was calculated no matter how the hypothesis-test for existence of a change-point would have turned out.

Scenario	CUSUM	Spatial Sign
1	0.867	0.937
$5, \gamma = 0.3n$	0.776	0.835
$5, \gamma = 0.15n$	0.459	0.495
2	0.442	0.89
3	0.652	0.928
4	0.293	0.956
6	0.792	0.893

Table 3.1: Proportion of $\hat{k}$ in $[k^\star - 0.1n, k^\star + 0.1n]$ for 3000 simulations of Scenarios 1-6.

The results for the simple jump after $0.5n$, resp. $0.3n$ and $0.15n$ of the observations (Scenarios 1 and 5) are grouped together in Figure 3.1. It is apparent that Spatial Sign estimates the real time of change correctly more often than CUSUM does. The advantage is the biggest in the case where $k^\star = 0.5n$, however it still holds when the change moves more to the beginning of the series. Especially for $k^\star = 0.3n$, the estimation of Spatial Sign is still superior.

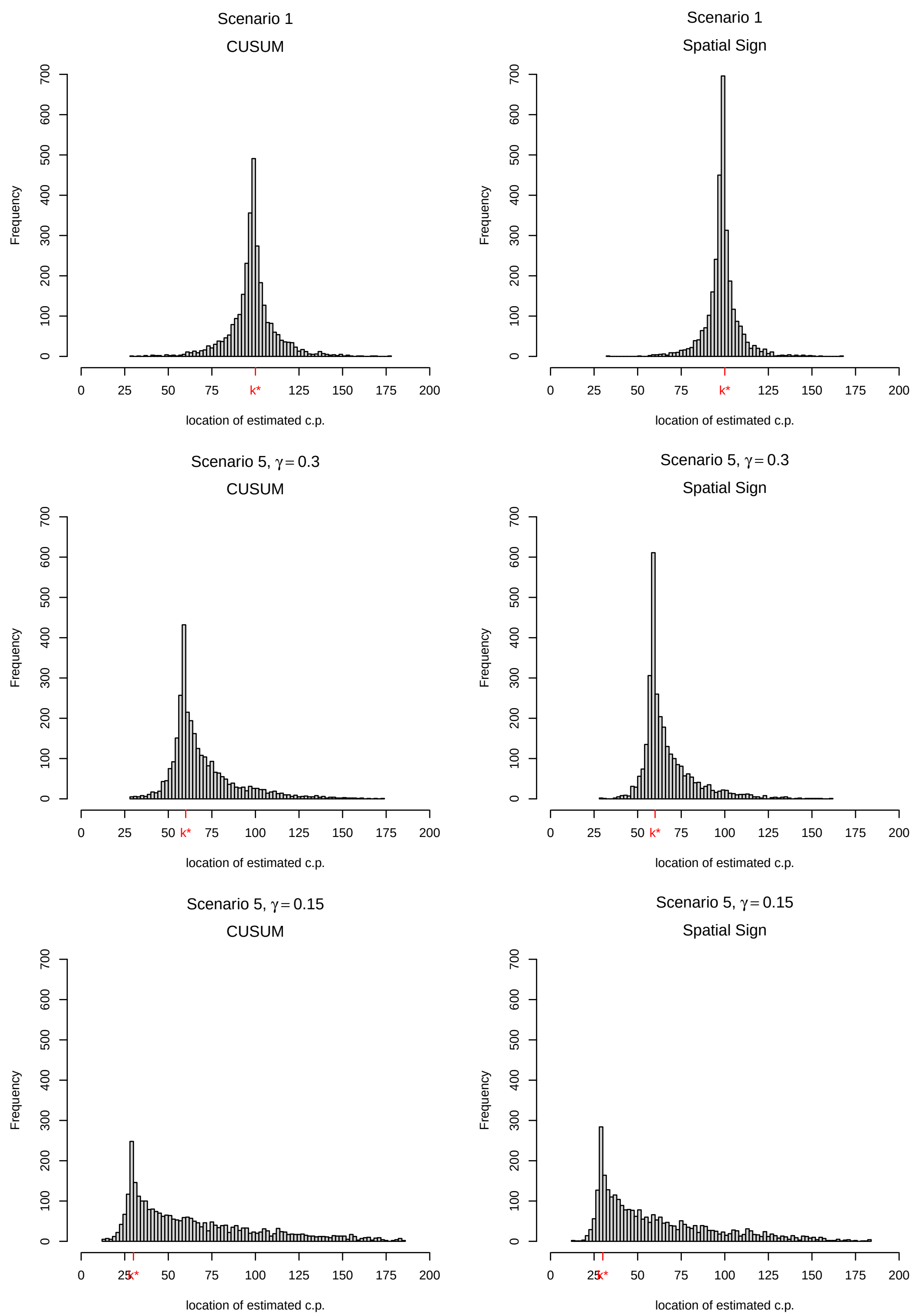


Figure 3.1: Histograms of $\hat{k}$ from CUSUM and Spatial Sign for 3000 samples in each of the Scenarios 1 and 5. $k^{\star}$ (red) marks the true change-point (c.p.) of the simulated series.

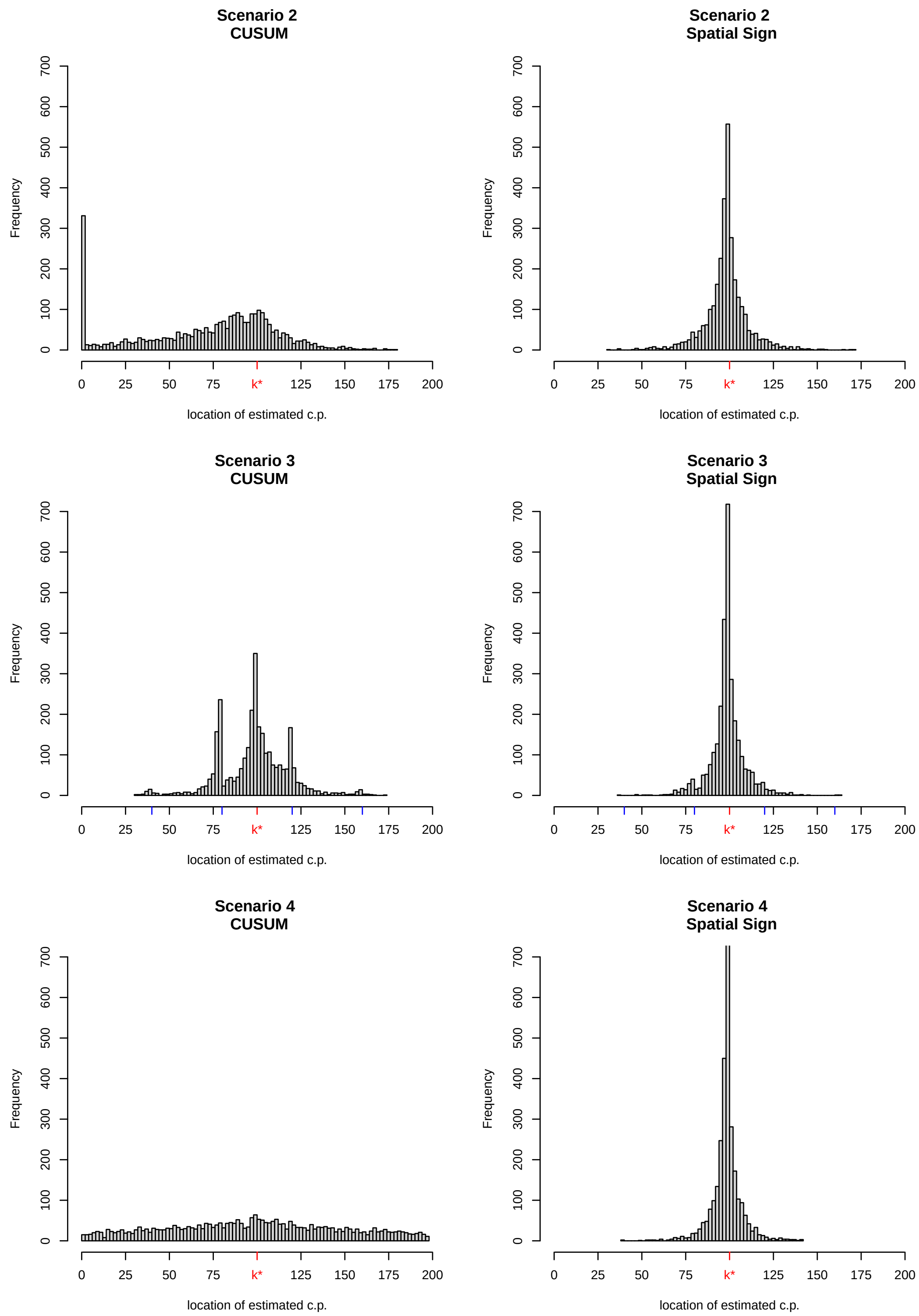


Figure 3.2: Histograms of $\hat{k}$ from CUSUM and Spatial Sign for 3000 samples in each of the Scenarios 2 to 4. $k^{\star}$ (red) marks the true change-point (c.p.) of the simulated series. For Scenario 3, blue ticks mark the time of the simulated outliers.

In Scenarios 2 to 4, the advantage of the Wilcoxon-type test is very prominent, as we can see in Figure 3.2. For Scenario 2, CUSUM has some big distortion at 0 and additionally no clear cumulation of $\hat{k}$ around $k^\star$, while the Wilcoxon-type test clearly does. In fact, for the Wilcoxon-type test, nearly 90 % of $\hat{k}$ lie in the interval $[k^\star - 0.1n, k^\star + 0.1n]$, while for CUSUM it is just half of it (see Table 3.1).
For Scenario 3 we can see that CUSUM is clearly influenced by the outliers, as we see high frequencies in the histogram around these times. For Spatial Sign, we do not see such high frequencies around the time of the outliers, instead there is a clear maximum at $k^\star$.
The difference for the two procedures is strongly pronounced for Scenario 4. The histogram of $\hat{k}$ from CUSUM looks almost like the result of a uniform distribution. Even though the frequency around $k^\star$ is slightly increased, the 20 %-interval around $k^\star$ only contains around 30 % of the $\hat{k}$. In contrast to this, Spatial Sign is not really influenced by the heavy tails. The histogram looks similar to that of Scenario 1, showing a clear maximum at $k^\star$ and also around the same proportion of $\hat{k}$ inside the interval $[k^\star - 0.1n, k^\star + 0.1n]$.

If $d >> n$ (Scenario 6, see Figure 3.3), we still have a similar picture as in Scenario 1. As we can see in Table 3.1, the proportion of $\hat{k}$ in the 20 %-interval around $k^\star$ is smaller for both, the CUSUM and the Wilcoxon-type test, nevertheless the Wilcoxon-type test is still superior in this scenario. The histogram shows high frequencies around $k^\star$ and clearly a maximum at $k^\star$.

In summary, the Wilcoxon-type test statistic is superior to CUSUM in all of the investigated scenarios. Especially in the cases of the sinus-jump, outliers or heavy tails, Wilcoxon is not influenced much by this, while simultaneously we observe a strong effect on CUSUM, partly resulting in drastically reduced proportions of estimations around the true change-point.

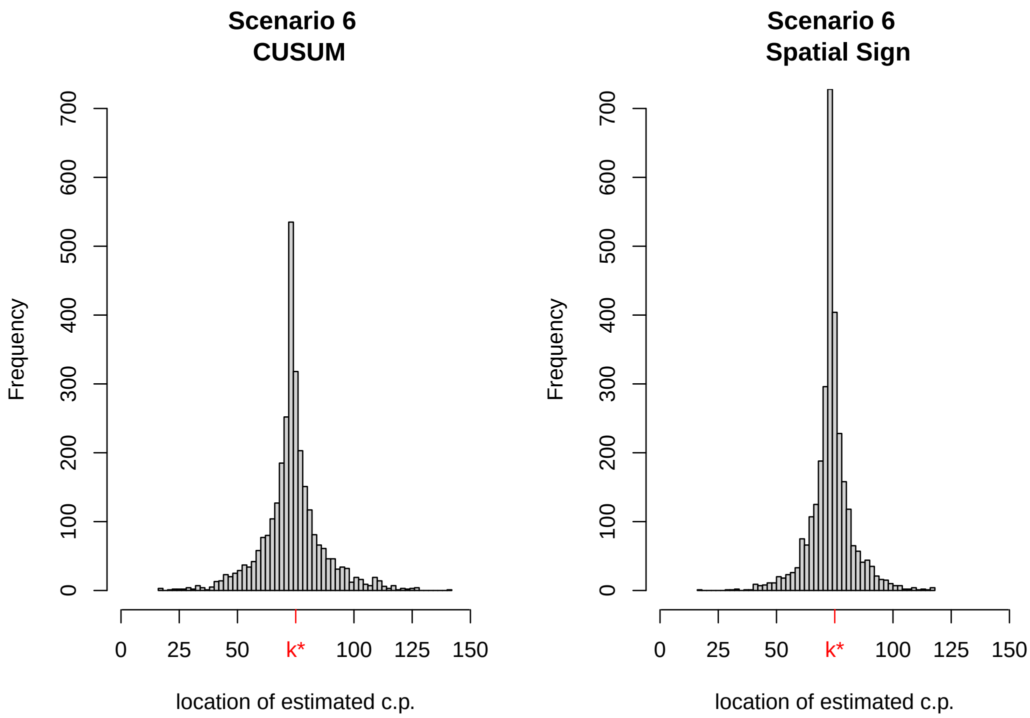


Figure 3.3: Histograms of $\hat{k}$ from CUSUM and Spatial Sign for 3000 samples in Scenario 6. $k^\star$ (red) marks the true change-point (c.p.) of the simulated series.

3.3.2 Estimating the Direction

As explained in the introduction of Section 3.2, the spatial median $Q^\star(0)$ would be a natural estimator of the direction of the change, if we knew the true change-point $k^\star$. In application, we only know $\hat{k}$. However, we know by Corollary 3.1, that the estimated spatial median is an adequate substitute for the spatial median. Hence, to estimate the direction of the change, we calculate $\hat{Q}(0)$ which minimises $\hat{\varphi}(Q(0))$. In order to evaluate the results of the estimation, we repeatedly simulated samples similar to the scenarios described in Section 2.6.1 with some changes in the height of the change, since the results are better to observe and compare.

Scenario 1: Uniform jump of +1 after $k^\star = \gamma n$ observations, $\gamma \in \{0.25, 0.5\}$:

$$Y_i = \begin{cases} X_i & i < \gamma n \\ X_i + u & i \geq \gamma n \end{cases}$$

where $u = (1, ..., 1)^T$.

Scenario 2: Sinus-jump after $k^\star = n/2$ of observations:

$$Y_i = \begin{cases} X_i & i < n/2 \\ X_i + (\sin(\pi D/d))_{D \leq d} & i \geq n/2 \end{cases}.$$

Scenario 3: Uniform jump of $+1$ after $k^\star = n/2$ observations in presence of outliers at $0.2n, 0.4n, 0.6n, 0.8n$:

$$Y_i = \begin{cases} X_i & i < n/2, i \notin \{0.2n, 0.4n\} \\ 10X_i & i \in \{0.2n, 0.4n\} \\ X_i + u & i \geq n/2, i \notin \{0.6n, 0.8n\} \\ 10X_i + u & i \in \{0.6n, 0.8n\} \end{cases}$$

where $u = (1, ..., 1)^T$.

Scenario 4: Heavy tails: In the simulation of $(X_i)_{i \leq n}$ we use $\xi_i, \xi_i^{(t)} \sim t_1$ (Cauchy distributed) $\forall i \leq d, -BI \leq t \leq n$ and a uniform jump of $+10$ after $k^\star = n/2$ observations.

First, $\hat{k}$ is estimated and then $\hat{Q}(0)$ evaluated, using a steep gradient descent approach. Since the minimisation is time-consuming, we simulated 500 samples for each scenario with $n = 100$ and $d = 30$. Again, we compare the two performances in the sense that $\hat{k}$ is calculated based on CUSUM and the Wilcoxon-type statistic simultaneously for each sample and $\hat{Q}(0)$ is evaluated for both $\hat{k}$. The results are presented as box plots (Figures 3.4 to 3.6), illustrating the deviation from the true direction of change for each coordinate $D = 1, ..., d$.

Starting with Scenario 1, we make multiple observations (see Figure 3.4): While the boxes of the Wilcoxon-type procedure are a little bit smaller, meaning that the first and third quantile of the estimations are closer together, the Wilcoxon-type procedure produces in turn more outliers farer away from the true direction then the CUSUM procedure. This effect enhances, if the time of change $k^\star$ moves closer to the beginning of the series. Over all, we can see that both procedures get less precise if $k^\star$ moves away from $n/2$.

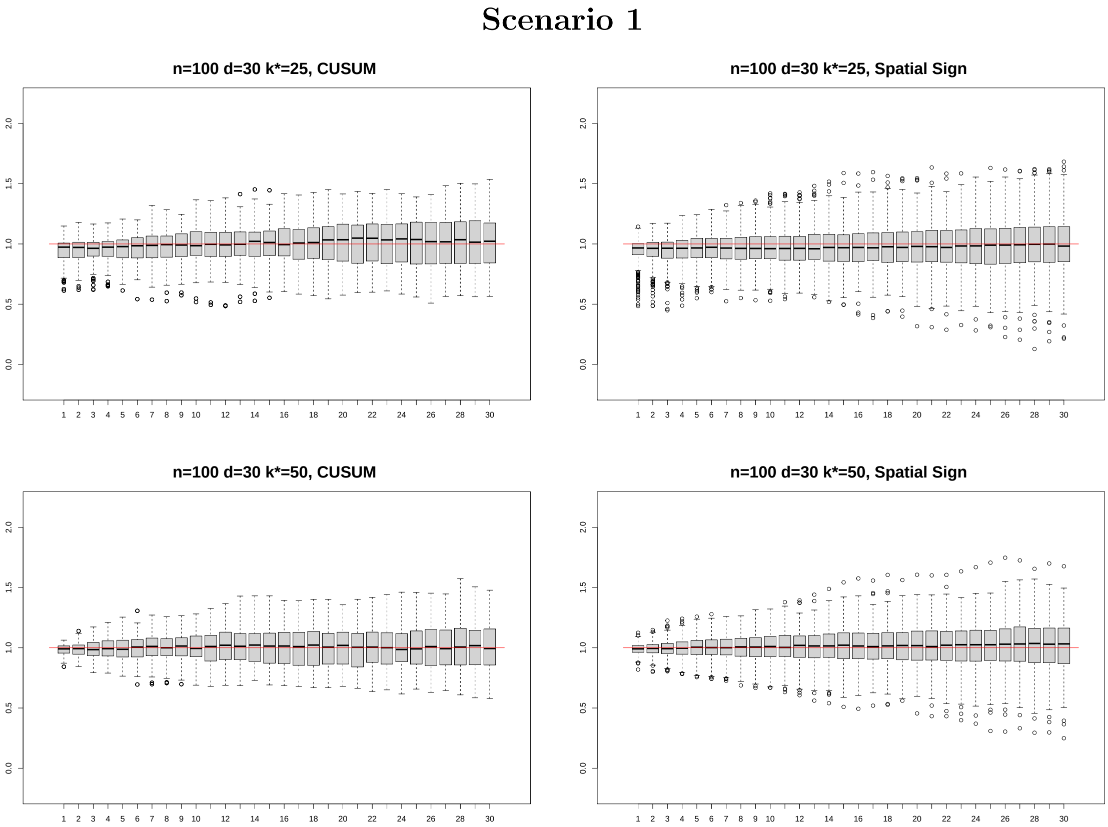


Figure 3.4: Box plots of $\hat{Q}(0)$ based on $\hat{k}$ received from CUSUM (left) resp. Wilcoxon-type statistic (right) for $k^\star = 25$ (top row) and $k^\star = 50$ (bottom row) in Scenario 1. The red lines mark the true direction of the change for each coordinate $D = 1, ..., d$ presented on the x-axis.

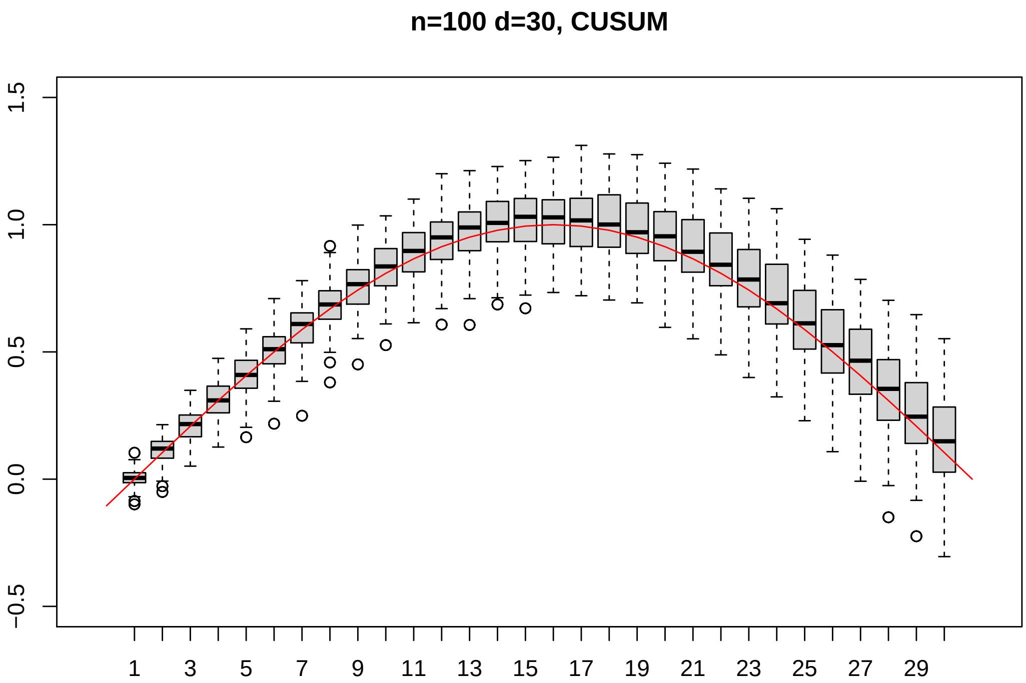
For both procedures we observe the phenomenon that the estimations gain more variability for growing coordinate $D = 1, ..., d$. This can be attributed to the variance structure of our simulated data. Recall that $X_{-BI} = (\xi_1, \xi_1 + \xi_2, ..., \sum_{i=1}^d \xi_i)/\sqrt{d}$ with ξ_i i.i.d. $\mathcal{N}(0,1)$-distributed (and W_t, $-BI < t \leq n$, has the same structure). So

$$\begin{aligned}\mathrm{Var}(X_{-BI}) &= \Big(\mathrm{Var}\left(\xi_1/\sqrt{d}\right), \mathrm{Var}\left((\xi_1+\xi_2)/\sqrt{d}\right), ..., \mathrm{Var}\left((\sum_{i=1}^{d}\xi_i)/\sqrt{d}\right)\Big)^T \\ &= \Big(\frac{1}{d}, \frac{2}{d}, ..., 1\Big)^T.\end{aligned}$$

The same holds for W_t. Meaning that a change in the direction of $+1$ is in relation much larger for lower coordinates than for higher ones. Thus, the direction is detected with higher precision for smaller D.

In Scenario 2 (Figure 3.5), we make similar observations as for Scenario 1. The median in the Wilcoxon-type-estimation of $\hat{Q}(0)$ lies closer to the true direction of change. In turn, the estimations show more outliers than the ones from CUSUM. Nevertheless, overall the difference between the two procedures is not very big.

Scenario 2



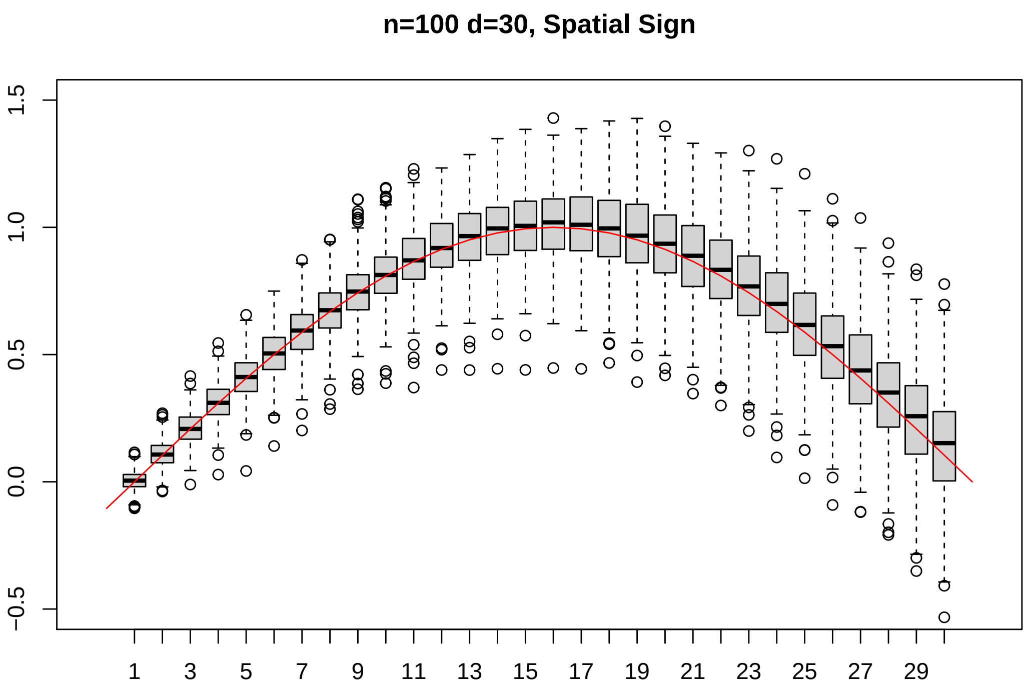


Figure 3.5: Box plots of $\hat{Q}(0)$ based on $\hat{k}$ received from CUSUM (top) resp. Wilcoxon-type statistic (bottom) for $k^\star = 50$ in Scenario 2. The red lines mark the true direction of the change for each coordinate $D = 1, ..., d$ presented on the x-axis.

The results for Scenario 3 and 4 are more interesting (see Figure 3.6). Compared to the results of Scenario 1, the Wilcoxon-type procedure loses precision. However, compared to the estimations based on CUSUM it can handle outliers in the data or heavy tails quiet well. In Scenario 3, CUSUM over-estimates the direction of change by about 0.34 ($D = 1$) to 0.56 ($D = 30$) measured from the median of all 500 repetitions. The Wilcoxon-type on the other hand only over-estimates the direction by a maximum of about 0.05.
For heavy tails (Scenario 4), the problems of the CUSUM-based procedure become even more apparent. The misestimation of the Spatial Sign procedure grows as well, now as a maximum we have +0.29. Whereas the CUSUM procedure overestimates the maximum by around 4.2, which represents a deviation of nearly 50 % of the size of change. Furthermore, the CUSUM procedure leads to very large outliers in the estimation of the direction. We get results in the range of $[-50, 100]$. Thus, we can conclude that the CUSUM procedure is hardly usable for heavy tailed data or if outliers are present in the sample, which is very common for real-world examples.

Scenario 3 and 4

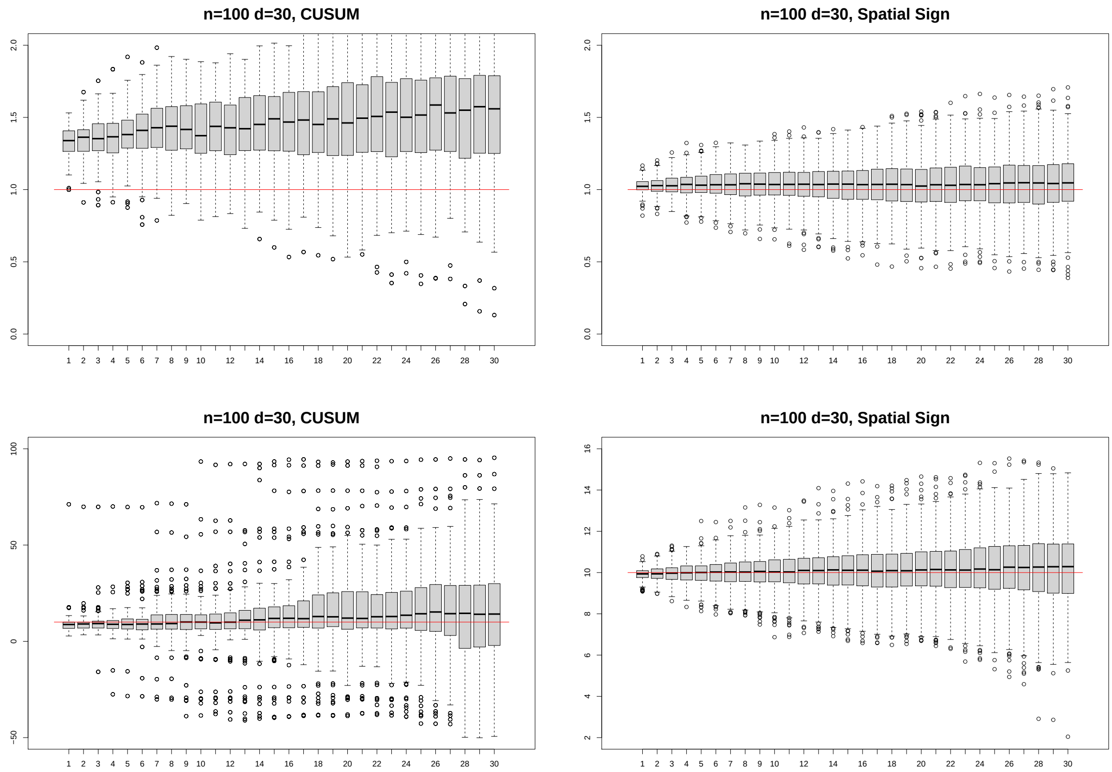


Figure 3.6: Box plots of $\hat{Q}(0)$ based on $\hat{k}$ received from CUSUM (left) resp. Wilcoxon-type statistic (right) for Scenario 3 (top row) and Scenario 4 (bottom row). The red lines mark the true direction of the change for each coordinate $D = 1, ..., d$ presented on the x-axis.

3.4 Data Example

We continue our example from Section 2.7 of particular matter in Germany in the first five months of 2020. We have seen that the Wilcoxon-type test statistic gives a significant result for a change-point. Thus, we want to estimate the location of that change. The maximum of the Spatial Sign test statistic, which marks the estimated change point, is received at March 15, 2020. (The maximum of the CUSUM statistic is indeed located at the same point although it gave no significant result for the existence of a change-point.)

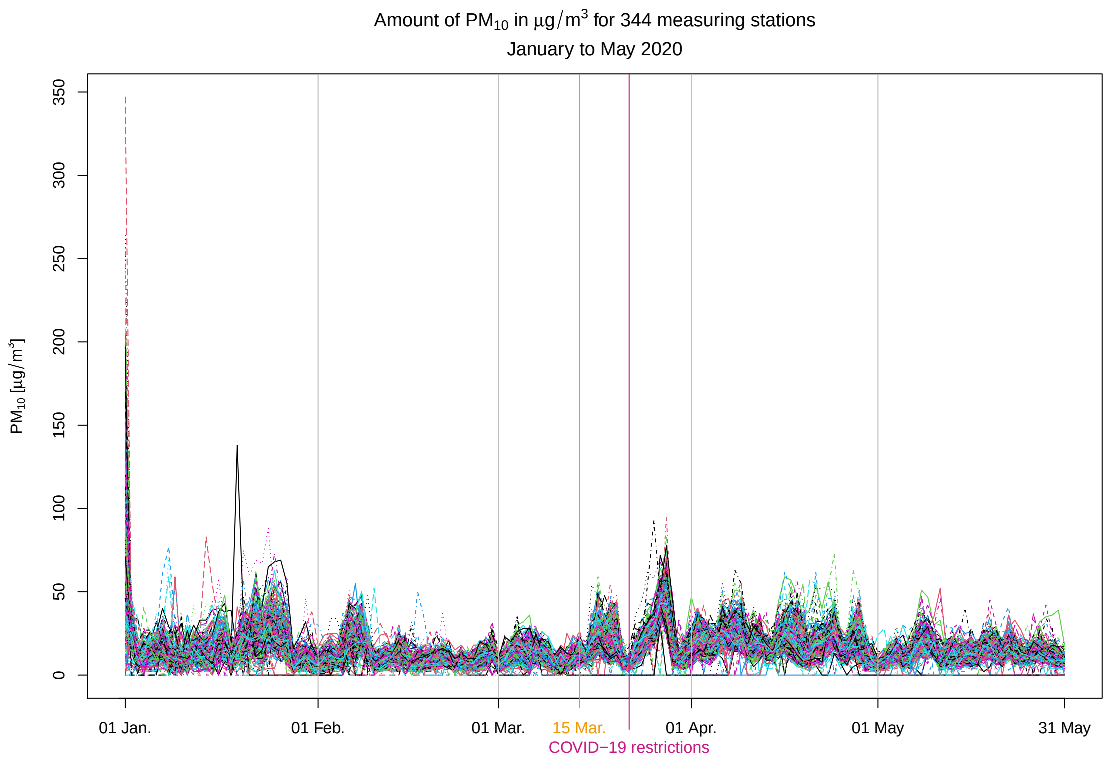


Figure 3.7: Daily average of PM_{10} in $\mu g/m^3$ for 344 measure stations from January 1, 2020 to May 31, 2020. The vertical orange line denotes the estimated change, the vertical magenta line marks the beginning of the official COVID-19 restrictions in Germany.

The estimated change-point in our example lies one week before the official restrictions regarding COVID-19 were imposed (see Figure 3.7). It is possible that the citizen, being aware of the situation, changed their behaviour beforehand, without strict official restrictions. Data projects using mobile phone data (e.g., Covid-19 Mobility Project [2022] and Destatis - Deutsches Statistisches Bundesamt [2022]) indeed show a decline in mobility preceding the official restrictions on March 22,

2020 by around a week.
However, if we look at the data (Figure 2.8), it seems like a change in mean would rather be upwards than downwards, meaning that the daily average pollution increased after March 15, 2020 compared to the beginning of the year.
We estimate the direction of the change by calculating $\hat{Q}(0)$ and see that the change is indeed overall upwards. In Table 3.2 a summary of the estimation is given. We see that the median change is of around +3.7, the lower quantile is greater than +2, the upper quantile greater than +5. Even though the changes are not so pronounced, they are consistent over most stations. The vast majority of stations recorded an upwards change after March 15, with the maximum upwards change estimated of slightly over +15. Only for 15 stations the estimated direction of change is downwards, and only by −4.44 in the most extreme case.
This means the findings do not support the theory of reduced PM_{10} after (or even slightly ahead of) the COVID-19 restrictions.

lower quantile	median	upper quantile
2.359	3.696	5.146

max. negative change	number of stations with negative change	max. positive change	number of stations with positive change
−4.44	15	15.055	329

Table 3.2: Summary of the estimation of direction of change after March 15, 2020 for PM_{10} in $\mu g/m^3$ for 344 measure stations.

Similar findings were made by Ropkins and Tate [2021]. They studied the impact of the COVID-19 lockdown on air quality across the UK. While using long-term data (Jan. 2015 to Jun. 2020) from rural background, urban background and urban traffic stations, they observed an increase of PM_{10} and $PM_{2.5}$ during active restrictions. Noting that this trend is ‘highly inconsistent with an air quality response to the lockdown’, they discussed the possibility that the lockdown did not greatly limit the largest impacts on particulate matter. We assume that the findings are to some extend comparable to Germany due to the similar geographic and demographic characteristics of the countries.
Furthermore, the German Umweltbundesamt [2020] states that traffic is not the main contributor to PM_{10} in Germany any more and that other sources of particular matter (e.g., fertilization, Saharan dust, soil erosion, fires) can overlay effects of reduced traffic. It is known that one mayor meteorological effect on particulate matter is precipitation, since it washes the dust out of the air (scavenging). Comparing the data with the meteorological recordings (Fig. 3.8) another explanation for the change-point gets visible:
While January was relatively warm and dry, February and first half of March showed increased precipitation. Beginning in the middle of March, a relatively drought period started and lasted throughout April and May (Data extracted from DWD Climate Data Center (CDC)).

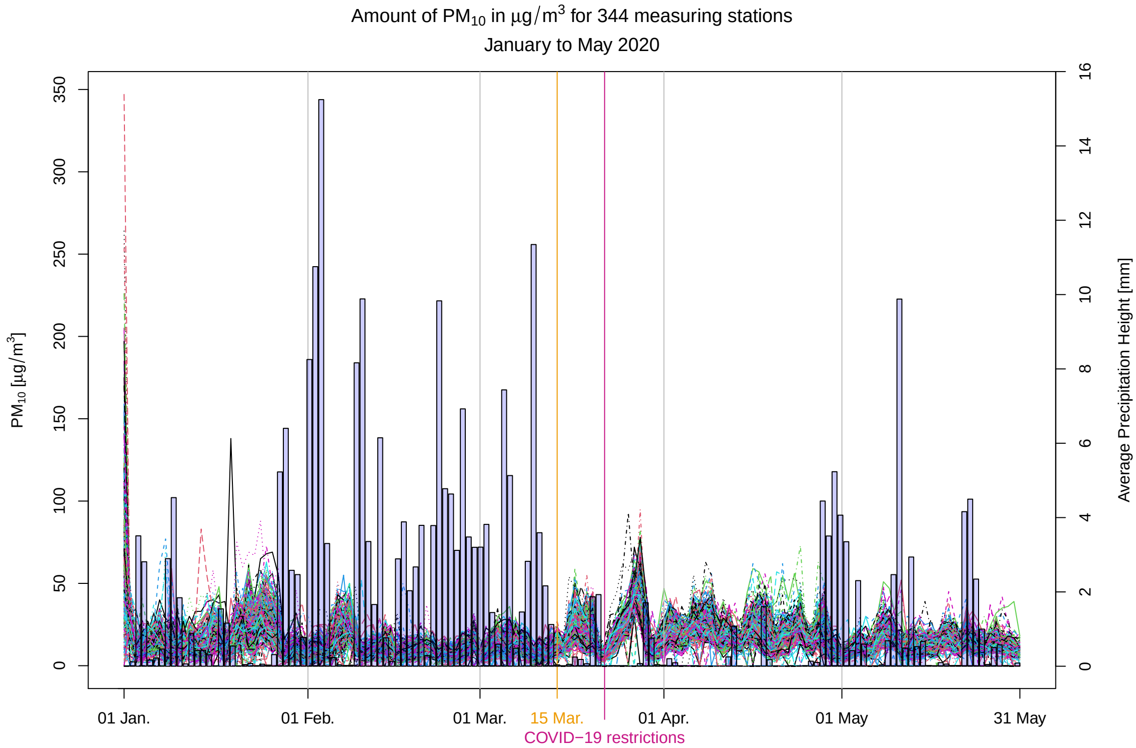


Figure 3.8: Daily average of PM_{10} in $\mu g/m^3$ for 344 measure stations from January 1, 2020 to May 31, 2020 and daily precipitation in mm in Germany averaged over 1637 weather stations.

We see that this fits the PM_{10}-data quite well. Especially in February and the first half of March, where it was very wet, we have relatively low quantity of PM_{10}. Beginning with the drought weather, the concentration of PM_{10} increases and especially the lows are now higher than before, meaning that days with a concentration of PM_{10} as low as in the beginning of the year, are more rare.
We would like to note that these findings do not contradict the satellite data published by the European Space Agency [2020] which shows a reduced air pollution over Europe in 2020 compared to 2019. While the satellites measure atmospheric pollution, the data of the Umweltbundesamt is collected at stations at ground level. It is known that there is a difference between these two sorts of pollution.

References

J. Antoch, M. Hušková, and N. Veraverbeke. Change-point problem and bootstrap. *Journaltitle of Nonparametric Statistics*, 5(2):123–144, 1995.

E. Asplund. Fréchet differentiability of convex functions. *Acta Mathematica*, 121: 31–47, 1968.

J.A.D. Aston and C. Kirch. Detecting and estimating changes in dependent functional data. *Journal of Multivariate Analysis*, 109:204–220, 2012.

A. Aue, G. Rice, and O. Sönmez. Detecting and dating structural breaks in functional data without dimension reduction. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(3):509–529, 2018.

R.R. Bahadur. A note on quantiles in large samples. *The Annals of Mathematical Statistics*, 37(3):577–580, 1966.

I. Berkes, R. Gabrys, L. Horváth, and P. Kokoszka. Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(5):927–946, 2009.

S. Borovkova, R. Burton, and H. Dehling. Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation. *Transactions of the American Mathematical Society*, 353(11):4261–4318, 2001.

D. Bosq. *Nonparametric statistics for stochastic processes: Estimation and prediction*. Springer, 1996.

R.C. Bradley. Basic properties of strong mixing conditions. A survey and some open questions. *Probability surveys*, 2:107–144, 2005.

B. Bucchia and M. Wendler. Change-point detection and bootstrap for Hilbert space valued random fields. *Journal of Multivariate Analysis*, 155:344–368, 2017.

A. Bücher and I. Kojadinovic. A note on conditional versus joint unconditional weak convergence in bootstrap consistency results. *Journal of Theoretical Probability*, 32(3):1145–1165, 2019.

B. Cadre. Convergent estimators for the L_1-median of banach valued random variable. *Statistics: A Journal of Theoretical and Applied Statistics*, 35(4):509–521, 2001.

E. Carlstein. The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The annals of statistics*, pages 1171–1179, 1986.

A. Chakraborty and P. Chaudhuri. The spatial distribution in infinite dimensional spaces and related quantiles and depths. *The Annals of Statistics*, 42(3):1203–1231, 2014.

A. Chakraborty and P. Chaudhuri. Tests for high-dimensional data based on means, spatial signs and spatial ranks. *The Annals of Statistics*, 45(2):771–799, 2017.

P. Chaudhuri. On a geometric notion of quantiles for multivariate data. *Journal of the American statistical association*, 91(434):862–872, 1996.

J. Choi and D.W. Shin. Block bootstrapping for a panel mean break test. *Journal of the Korean Statistical Society*, 49(3):802–821, 2020.

G.W. Cobb. The problem of the Nile: Conditional solution to a changepoint problem. *Biometrika*, 65(2):243–251, 1978.

Covid-19 Mobility Project. Mobilitätsdaten, 2022. URL `https://www.covid-19-mobility.org/de/data-info/`. Last accessed on 19.12.2022.

M. Csörgő and L. Horváth. Invariance principles for changepoint problems. In *Multivariate statistics and probability*, pages 151–168. Elsevier, 1989.

M. Csörgö and L. Horváth. Limit theorems in change-point analysis. 1997.

B.S. Darkhovsky. A non-parametric method for a posteriori detection of the disorder time for a sequence of independent random variables. *Teoriya Veroyatnostei i ee Primeneniya*, 21(1):180–184, 1976.

R.M. De Jong and J. Davidson. Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices. *Econometrica*, 68(2):407–423, 2000.

H. Dehling and R. Fried. Asymptotic distribution of two-sample empirical U-quantiles with applications to robust tests for shifts in location. *Journal of Multivariate Analysis*, 105(1):124–140, 2012.

H. Dehling and T. Mikosch. Random quadratic forms and the bootstrap for U-statistics. *Journal of Multivariate Analysis*, 51(2):392–413, 1994.

H. Dehling and M. Wendler. Central limit theorem and the bootstrap for U-statistics of strongly mixing data. *Journal of Multivariate Analysis*, 101(1):126–137, 2010.

H. Dehling, M. Denker, and W. Philipp. The almost sure invariance principle for the empirical process of U-statistic structure. In *Annales de l'IHP Probabilités et statistiques*, volume 23, pages 121–134, 1987.

H. Dehling, R. Fried, I. Garcia, and M. Wendler. Change-point detection under dependence based on two-sample U-statistics. In *Asymptotic laws and methods in stochastics*, pages 195–220. Springer, 2015.

H. Dehling, D. Vogel, M. Wendler, and D. Wied. Testing for changes in Kendall's tau. *Econometric Theory*, 33(6):1352–1386, 2017.

H. Dehling, K. Vuk, and M. Wendler. Change-point detection based on weighted two-sample U-statistics. *Electronic Journal of Statistics*, 16(1):862–891, 2022.

M. Denker and G. Keller. Rigorous statistical procedures for data from dynamical systems. *Journal of Statistical Physics*, 44(1):67–93, 1986.

Destatis - Deutsches Statistisches Bundesamt. Mobilitätsindikatoren auf Basis von Mobilfunkdaten, 2022. URL `https://www.destatis.de/DE/Service/EXDAT/Datensaetze/mobilitaetsindikatoren-mobilfunkdaten.html`. Last accessed on 19.12.2022.

H. Dette, K. Kokot, and A. Aue. Functional data analysis in the Banach space of continuous functions. *The Annals of Statistics*, 48(2):1168–1192, 2020.

P. Doukhan. Mixing. In *Mixing*, pages 15–23. Springer, 1994.

R.M. Dudley. *Real analysis and probability.* CRC Press, 2018.

DWD Climate Data Center (CDC). Daily station observations precipitation height in mm for germany, version v19.3. Last accessed on 02.09.2020.

European Space Agency. Air pollution remains low as Europeans stay at home, 2020. URL `https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-5P/Air_pollution_remains_low_as_Europeans_stay_at_home`. Last accessed on 19.12.2022.

F. Fu, K.L. Purvis-Roberts, and B. Williams. Impact of the COVID-19 pandemic lockdown on air pollution in 20 major cities around the world. *Atmosphere*, 11 (11):1189, 2020.

C. Gerstenberger. Robust Wilcoxon-type estimation of change-point location under short-range dependence. *Journal of Time Series Analysis*, 39(1):90–104, 2018.

D.V. Hinkley. Inference about the change-point in a sequence of random variables. *Biometrica*, 57(1):1–17, 1970.

J.L. Hodges and E.L. Lehmann. Estimates of location based on rank tests. *The Annals of Mathematical Statistics*, 34(2):598–611, 1963.

W. Hoeffding. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19(3):293–325, 1948.

S. Hörmann and P. Kokoszka. Functional time series. In *Handbook of statistics*, volume 30, pages 157–186. Elsevier, 2012.

L. Horváth and M. Hušková. Change-point detection in panel data. *Journal of Time Series Analysis*, 33(4):631–648, 2012.

L. Horváth, P. Kokoszka, and G. Rice. Testing stationarity of functional time series. *Journal of Econometrics*, 179(1):66–82, 2014.

I.A. Ibragimov. Some limit theorems for stationary processes. *Theory of Probability & Its Applications*, 7(4):349–382, 1962.

P. Janssen. Bootstrapping U-statistics. *South African Statistical Journal*, 31(2): 185–216, 1997.

J. Kiefer. On Bahadur's representation of sample quantiles. *The Annals of Mathematical Statistics*, 38(5):1323–1342, 1967.

P. Kokoszka and R. Leipus. Change-point in the mean of dependent observations. *Statistics & probability letters*, 40(4):385–393, 1998.

P. Kokoszka and R. Leipus. Change-point estimation in ARCH models. *Bernoulli*, pages 513–539, 2000.

J.P. Kreiß and G. Neuhaus. *Einführung in die Zeitreihenanalyse*. Springer-Verlag, 2006.

H. R. Künsch. The jackknife and the bootstrap for general stationary observations. *The annals of Statistics*, pages 1217–1241, 1989.

S.N. Lahiri. *Resampling methods for dependent data*. Springer Science & Business Media, 2003.

A. Leucht and M.H. Neumann. Dependent wild bootstrap for degenerate U-and V-statistics. *Journal of Multivariate Analysis*, 117:257–280, 2013.

X. Lian, J. Huang, R. Huang, C. Liu, L. Wang, and T. Zhang. Impact of city lockdown on the air quality of COVID-19-hit of Wuhan city. *Science of the Total Environment*, 742:140556, 2020.

H.B. Mann and D.R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pages 50–60, 1947.

F. Móricz. Moment inequalities and the strong laws of large numbers. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 35(4):299–314, 1976.

E.S. Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954.

A.N. Pettitt. A non-parametric approach to the change-point problem. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 28(2):126–135, 1979.

D. Politis, J.P. Romano, and M. Wolf. Weak convergence of dependent empirical measures with application to subsampling in function spaces. *Journal of statistical planning and inference*, 79(2):179–190, 1999.

D.N. Politis. The impact of bootstrap methods on time series analysis. *Statistical Science*, pages 219–230, 2003.

D.N. Politis and H. White. Automatic block-length selection for the dependent bootstrap. *Econometric reviews*, 23(1):53–70, 2004.

J. Qiu and Z. Lin. The functional central limit theorem for linear processes with strong near-epoch dependent innovations. *Journal of mathematical analysis and applications*, 376(1):373–382, 2011.

J.O. Ramsay and B.W. Silverman. *Functional Data Analysis*. Springer Series in Statistics, 2005.

G. Rice and H.L. Shang. A plug-in bandwidth selection procedure for long-run covariance estimation with stationary functional time series. *Journal of Time Series Analysis*, 38(4):591–609, 2017.

K. Ropkins and J.E. Tate. Early observations on the impact of the COVID-19 lockdown on air quality trends across the UK. *Science of the Total Environment*, 754:142374, 2021.

X. Shao. The dependent wild bootstrap. *Journal of the American Statistical Association*, 105(489):218–235, 2010.

O. Sharipov, J. Tewes, and M. Wendler. Sequential block bootstrap in a Hilbert space with application to change point analysis. *Canadian Journal of Statistics*, 44(3):300–322, 2016a.

S.O. Sharipov, J. Tewes, and M. Wendler. Bootstrap for U-statistics: A new approach. *Journal of Nonparametric Statistics*, 28(3):576–594, 2016b.

D. Siegmund. Confidence sets in change-point problems. *International Statistical Review/Revue Internationale de Statistique*, pages 31–48, 1988.

Umweltbundesamt. Luftdaten, 2020. URL `https://www.umweltbundesamt.de/daten/luft/luftdaten/stationen`. Last accessed on 06.08.2020.

Umweltbundesamt. FAQ: Auswirkungen der Corona-Krise auf die Luftqualität, 2020. URL `https://www.umweltbundesamt.de/faq-auswirkungen-der-corona-krise-auf-die#welche-auswirkungen-hat-die-corona-krise-auf-die-feinstaub-pm10-belastung`. Last accessed on 19.12.2022.

V.A. Volkonskii and Y.A. Rozanov. Some limit theorems for random functions. i. *Theory of Probability & Its Applications*, 4(2):178–197, 1959.

L. Wegner and M. Wendler. Block length choice for the bootstrap of dependent panel data–a comment on Choi and Shin (2020). *arXiv preprint arXiv:2103.11805*, 2021.

L. Wegner and M. Wendler. Robust change-point detection for functional time series based on U-statistics and dependent wild bootstrap. *arXiv preprint arXiv:2206.01458*, 2022.

M. Wendler. Bahadur representation for U-quantiles of dependent data. *Journal of Multivariate Analysis*, 102(6):1064–1079, 2011.

F. Wilcoxon. Individual comparisons by ranking methods. In *Breakthroughs in statistics*, pages 196–202. Springer, 1992.

K.J. Worsley. Confidence regions and tests for a change-point in a sequence of exponential family random variables. *Biometrika*, 73(1):91–104, 1986.

Yahoo!Finance. URL `https://finance.yahoo.com`. Last accessed on 08.02.2022.

K. Yoshihara. Limiting behavior of U-statistics for stationary, absolutely regular processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 35 (3):237–252, 1976.

M. Yu and X. Chen. A robust bootstrap change point test for high-dimensional location parameter. *Electronic Journal of Statistics*, 16(1):1096–1152, 2022.

S. Zangari, D.T. Hill, A.T. Charette, and J.E. Mirowsky. Air quality changes in New York City during the COVID-19 pandemic. *Science of the Total Environment*, 742:140496, 2020.

W. Zhou and R. Serfling. Multivariate spatial U-quantiles: A Bahadur–Kiefer representation, a Theil–Sen estimator for multiple regression, and a robust dispersion estimator. *Journal of Statistical Planning and Inference*, 138(6):1660–1678, 2008.