Change-Point Analysis for Long-Range Dependent Time Series

Annika Betken

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften
an der Fakultät für Mathematik der Ruhr-Universität Bochum

November 2017

Preface

Many issues in statistics are related to the identification of structural changes in time series. In this context, major challenges consist in determining the point in time when a change occurred and in discriminating between random effects and actual changes in the structure of data-generating stochastic processes. These problems serve as the central motivation for all results presented in the following chapters.
The most common techniques in change-point analysis are based on CUSUM statistics which, as the name indicates, result from a consideration of cumulative sums, and thus are non-robust in the sense that outliers in the data have a significant impact on their values. The focus of this thesis is on outlier robust methods. In particular, most of the considered statistical tools are derived from Wilcoxon rank-sum statistics.
Change-point problems have been widely studied in the case of independent observations. However, for many practical purposes in statistics, the assumption that a given set of observations has been generated by mutually independent random variables does not serve as an adequate model of reality. For a realistic model, one has to allow for dependence between observations. Naturally, this dependence declines as the time lag between observations grows. According to the rate of decay, one differentiates between short- and long-range dependent time series. A relatively fast decay of the autocovariances characterizes short-range dependent time series, while a slow decay defines long-range dependence. At first sight, this classification seems to be quantitative. Yet, in fact, many theoretical results that can be derived under the assumption of long-range dependence differ qualitatively from those obtained under short-range dependence: due to a higher variability in the former case, statistics usually require a stronger scaling in order to converge. In addition, the asymptotic distributions of estimators and test statistics are different from those attained under short-range dependence: statistics computed with respect to independent or weakly dependent observations are asymptotically normal distributed, whereas statistics computed with respect to long-range dependent observations may converge to non-Gaussian limits. As a result, long-range dependence affects data analysis in practice and the proofs of mathematical statements in theory. The different aspects of change-point analysis covered by this thesis and the major accomplishments that underlie the results of the following chapters are all linked to specific features of long-range dependent time series.
The content of this thesis is based on four journal articles: Betken (2016), Betken (2017), Betken and Wendler (2015), and Betken and Kulik (2017). The first-mentioned manuscript partially reproduces results that had been developed in my master thesis (Betken (2013)), prior to the research projects associated with this dissertation. A short remark at the end of Chapter 1 specifies to what extent the results established in Betken (2016) go beyond the scope of Betken (2013). Moreover, the relevant findings of this

article are referred to in the introductory chapter only. Apart from these, Chapter 1 introduces definitions and results that are taken as a basis for the subsequent chapters. Chapter 2, which reproduces the results of Betken (2017), addresses the estimation of change-point locations for mean shifts in long-range dependent time series on the basis of non-self-normalized and self-normalized two-sample Wilcoxon statistics. The resulting estimators are shown to be consistent. Moreover, the rate of convergence and, after suitable standardization, the asymptotic distribution of the Wilcoxon-based estimator are derived. Most notably, long-range dependence considerably affects the asymptotic behavior of change-point estimators.

Betken and Wendler (2015) establishes the validity of a subsampling procedure under the assumption of long-range dependent time series. Although motivated by the need to approximate the distribution of test statistics in change-point analysis, the consistency of subsampling-based estimators is shown to hold for a general class of statistics and under mild assumptions on the data-generating process. Long-range dependence affects the considered subsampling procedure in that the choice of blocklength is restricted by a condition depending on the rate of decay of the time series' autocovariances. In addition to these results, Chapter 3 provides a further consistency result which, on the one hand, requires a more limited choice of the blocklength, but, on the other hand, imposes even less restrictive conditions on the data-generating process.

Betken and Kulik (2017) consider change-point tests for time series that follow the long memory stochastic volatility (LMSV) model. A specific feature of change-point tests for LMSV time series is being observed: while Wilcoxon statistics converge to limits that are typically attained under long-range dependence, the asymptotic distribution of CUSUM statistics does not necessarily correspond to a limit associated with long-range dependence. Chapter 4 is based on the findings in Betken and Kulik (2017). The main theoretical achievement presented in this chapter is the proof of a non-central limit theorem for the two-parameter empirical process of subordinated LMSV time series. In the context of change-point analysis, the corresponding result is needed to derive the asymptotic distribution of test statistics. In general, the theory of empirical processes also applies to many other fields in non-parametric statistics, so that the empirical process limit theorem is of particular and independent interest.

The three main parts of this thesis, i.e. Chapters 2, 3, and 4, are all completed by simulation studies at the end of each chapter. Chapter 5 takes up on the theory developed in the preceding chapters by discussing applications of change-point tests and change-point estimators to real data sets. For this purpose, testing procedures that allow for ties or multiple change-points in the data are needed. Change-point tests based on test statistics resulting from corresponding modifications of Wilcoxon-type statistics are introduced in Appendix B. Appendix A provides a discussion on the topology of path spaces for (two-parameter) empirical processes and on the concept of weak convergence for random variables with values in possibly non-separable spaces.

Acknowledgments

First of all, I would like to express my sincere gratitude to my supervisor Herold Dehling for his various ways of support, many of which cannot be taken for granted; for encouraging independent research while, at the same time, providing all the securities I needed to pursue my studies. I am very grateful to Martin Wendler and Rafał Kulik whose expertise added considerably to my research experience and who gave me the opportunity to visit them in Greifswald and Ottawa. I am thankful to many colleagues and friends, inside and outside university, whose company created the magnificent atmosphere that made the years of study a time that I would never want to miss. Among these, I owe special thanks to Marie-Christine Düker and Tobias Soethe for their thorough proofreading of this thesis. Their comments and suggestions significantly contributed to improving the quality of the manuscript. I highly appreciate the material and non-material funding of the German Academic Foundation, as well as the financial aids of the German Research Foundation through Collaborative Research Center SFB 823 *Statistical modelling of nonlinear dynamic processes* and Research Training Group GRK 2131 *High-dimensional Phenomena in Probability - Fluctuations and Discontinuity.* Last but not least, I would like to thank my siblings, parents, and grandparents for their inspiring serenity and their unconditional support.

Contents

1. Background

This chapter introduces definitions and techniques that are taken as a basis for the research results presented in the subsequent chapters. These include the definition of long-range dependent time series, references to stochastic processes inextricably linked with the concept of long-range dependence, and an introduction of basic model assumptions. What is common to the different chapters of this thesis, is the occurrence of certain statistics designed to identify structural changes in time series. Non-parametric statistics and self-normalization constitute concepts of peculiar interest and therefore are given particular emphasis.

1.1. Long-range dependence

A historical example of a phenomenon that gave rise to the consideration of long-range dependence in statistics and probability theory is the Hurst effect, named after its discoverer, the British hydrologist Harold Edwin Hurst. For his studies of the river Nile's flow in the early 1950s, Hurst analyzed the annual minimum water level of the Nile measured at the Roda gauge near Cairo for the years 622 to 1281; see Figure 1.1. Since Hurst was particularly interested in estimating the storage capacity of water reservoirs, he considered the values of the R/S (rescaled range) statistic in order to assess the variability of the time series. Under the assumption of data generated by a sequence of independent, identically distributed random variables, an increase in the number of observations is expected to entail a growth of the rescaled range statistic of order $\sqrt{n}$, where n denotes the number of observations. Nonetheless, a corresponding prediction concerning the growth rate of the R/S statistic contradicts an observation made by Hurst who found that his empirical analysis of the data suggested a rate of growth that is better approximated by $n^{0.72}$; see Hurst (1956). By dropping the assumption of independent data-generating random variables, the behavior of the R/S statistic can be explained in a statistical model that allows for long-range dependence, i.e. in a model of time series with relatively strong correlation between observations. While a growing distance in time, separating two observations, always goes along with a decay of correlation between these observations, the rate of decay is crucial to the definition of long-range dependent time series. A relatively slow decay of the autocovariances characterizes long-range dependent time series, while a relatively slow decay characterizes short-range dependent processes. Figure 1.1 contrasts the empirical correlation of time series data consisting of the yearly minimal water levels of the Nile river for a time lapse of more than 600 years with the expected behavior of the empirical correlation under the assumption of uncorrelated observations.

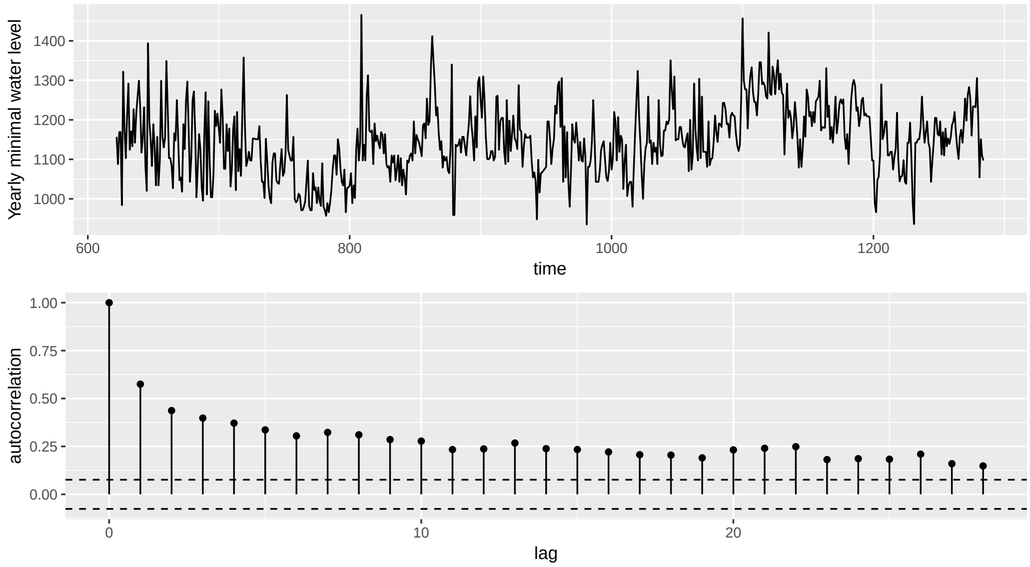

Figure 1.1.: *Yearly minimal water levels of the Nile river for the years 622 to 1284, measured at the Roda gauge near Cairo; see Toussoun (1925). The data has been taken from the* `longmemo` *package in* `R`*. The two dashed horizontal lines in the plot of the autocovariances mark the bounds for the 95% confidence interval under the assumption of data generated by white noise.*

For a description of the covariance structure of long-range dependent processes, the notions of *asymptotic equivalence* and *slowly varying functions* are essential.

Definition 1 (Bingham et al. (1987))**.** Two real-valued functions f, g are called *asymptotically equivalent*, if

$$\lim_{x\to\infty}\frac{f(x)}{g(x)} = 1.$$

We write $f \sim g$ to indicate asymptotic equivalence.

Definition 2 (Beran et al. (2013))**.** For some $c \geq 0$, let $L : (c, \infty) \longrightarrow \mathbb{R}$ be a positive function satisfying

$$\lim_{x\to\infty}\frac{L(\lambda x)}{L(x)} = 1 \quad \text{for all } \lambda > 0.$$

Then L is said to be *slowly varying at* ∞ (in Karamata's sense). A function L is said to be *slowly varying at the origin* (in Karamata's sense) if the function $\tilde{L}$ defined by $\tilde{L}(x) := L\left(x^{-1}\right)$ is slowly varying at ∞.

A comprehensive treatment of slowly varying functions can be found in Bingham et al. (1987).

Definition 3 (Pipiras and Taqqu (2017))**.** A (second-order) stationary, real-valued time series X_k, $k \in \mathbb{Z}$, is called *long-range dependent* if its autocovariance function γ satisfies

$$\gamma(k) := \mathrm{Cov}(X_1, X_{k+1}) \sim k^{-D} L_\gamma(k), \quad \text{as } k \to \infty,$$

with $D \in (0,1)$ for some slowly varying function L_γ. We refer to D as *long-range dependence (LRD) parameter.*

Apart from Definition 3, various notions of long-range dependence can be found in the literature. Most of these are related, but, in general, not equivalent. In defining long-range dependence with respect to the behavior of the autocovariances, Definition 3 is sometimes referred to as being expressed in the time domain. However, in some cases, it is useful to take a spectral domain perspective by relating long-range dependence in time series to the behavior of its spectral density at the origin.

Definition 4 (Brockwell and Davis (2002))**.** Given a stationary time series X_k, $k \in \mathbb{Z}$, with autocovariance function γ, a function f is called the *spectral density* of this time series if

a) $f(\lambda) \geq 0$ for all $\lambda \in (0, \pi]$,

b) $\gamma(k) = \int_{-\pi}^{\pi} \mathrm{e}^{ik\lambda} f(\lambda)\, d\lambda$ for all $k \in \mathbb{Z}$.

A definition of long-range dependence in the spectral domain is given by the following condition on the spectral density f of the considered time series:

$$f(\lambda) \sim |\lambda|^{D-1} L_f(\lambda), \quad \text{as } \lambda \to 0,$$

for some at the origin slowly varying function L_f.
The above characterization of long-range dependence in the spectral domain is not equivalent to the notion of long-range dependence in the time domain. Nonetheless, it can be shown that the defining conditions on the spectral density and the autocovariance function are equivalent under certain assumptions concerning the slowly varying functions L_γ and L_f; see Beran et al. (2013).

1.1.1. Fractional Brownian motion and fractional Gaussian noise

A process which is closely connected to the concept of long-range dependence is the so-called fractional Brownian motion. As it is a Gaussian process, a definition only requires the specification of mean and covariances.

Definition 5 (Beran et al. (2013))**.** A Gaussian process $B_H(t)$, $t \in \mathbb{R}$, with mean 0 and covariance function

$$\mathrm{Cov}\left(B_H(s), B_H(t)\right) = \frac{\sigma^2}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right),$$

where $\sigma^2 = \mathrm{Var}\, B_H(1)$ and $0 < H < 1$, is called *fractional Brownian motion.* It is called *standard fractional Brownian motion* if $\sigma^2 = 1$.

Remark 1. When $H = \frac{1}{2}$, the process $B := B_{\frac{1}{2}}$ is a usual Brownian motion.

Fractional Brownian motions with parameter $H \in \left(\frac{1}{2}, 1\right)$ characterize asymptotic distributions in limit theorems for time series with long-range dependence and are in this respect related to the concept of long-range dependence. In fact, partial sums of long-range dependent random variables with LRD parameter D converge to a fractional Brownian motion with parameter $H = 1 - \frac{D}{2}$, when appropriately standardized. In contrast to the classical central limit theorem for partial sums of independent (or short-range dependent) random variables which require a normalization of order $\sqrt{n}$, where n denotes the number of observations, the stronger normalization n^H is needed in the long-range dependent case. This observation corresponds to the previously mentioned Hurst effect. For this reason, the parameter H in the definition of fractional Brownian motion processes is also called *Hurst parameter* or *Hurst index*.

The sample paths of a fractional Brownian motion B_H are almost surely Hölder continuous of any order strictly less than H, i.e. for every $\beta \in (0, H)$ and every compact interval I, there exists a constant $C > 0$ such that

$$|B_H(t) - B_H(s)| \leq C\,|t - s|^{\beta} \tag{1.1}$$

for all $s, t \in I$. Moreover, it can be shown that the trajectories of B_H are almost surely not β-Hölder continuous for $\beta \geq H$, and, in particular, nowhere differentiable. Heuristically, the Hölder exponent β in (1.1) characterizes the roughness of the sample paths of B_H. As a consequence, a Hurst parameter H which is close to 1 points towards a smooth and regular behavior of the sample paths of B_H, while an H which is close to 0 indicates a rougher behavior of sample paths characterized by relatively high local variability.
A precise description of the path behavior of fractional Brownian motion processes is given by the corresponding law of the iterated logarithm; see Arcones (1995).

Theorem 1 (Law of the iterated logarithm)**.** *Let $B_H(t)$, $t \in \mathbb{R}$, be a standard fractional Brownian motion. Then, for all $t \in \mathbb{R}$,*

$$\limsup_{\varepsilon \searrow 0} \frac{B_H(t+\varepsilon) - B_H(t)}{\sqrt{2\varepsilon^{2H} \log\log(1/\varepsilon)}} = 1 \quad a.s.$$

Strictly speaking, the definition of long-range dependence does not apply to fractional Brownian motions, since these are not stationary. However, a fractional Brownian motion has stationary increments and therefore gives rise to a construction of stationary, long-range dependent Gaussian processes.

Definition 6 (Taqqu (2003))**.** Let $B_H(t)$, $t \in \mathbb{R}$, be a fractional Brownian motion. The process $\xi_H(k)$, $k \in \mathbb{Z}$, defined by

$$\xi_H(k) := B_H(k+1) - B_H(k)$$

is called *fractional Gaussian noise* with Hurst parameter H.

The autocovariance function of the process $\xi_H(k)$, $k \in \mathbb{Z}$, is given by

$$\gamma(k) = \frac{1}{2}\left(|k+1|^{2H} - 2\,|k|^{2H} + |k-1|^{2H}\right).$$

It can be shown that

$$\gamma(k) \sim H(2H-1)|k|^{2H-2}, \quad \text{as } k \to \infty,$$

if $H \neq \frac{1}{2}$. Thus, the process ξ_H exhibits long-range dependence when $H \in \left(\frac{1}{2}, 1\right)$. If $H = \frac{1}{2}$, the variables $\xi_H(k)$, $k \in \mathbb{Z}$, are uncorrelated and ξ_H is a Gaussian white noise process. Observations generated by fractional Gaussian noise ξ_H with $H \in \left(0, \frac{1}{2}\right)$ tend to have opposite algebraic signs due to negative correlation of the variables $\xi_H(k)$, $k \in \mathbb{Z}$, resulting in realizations which are distinctively zigzagging. In contrast, the behavior of observations generated by fractional Gaussian noise processes ξ_H with $H \in \left(\frac{1}{2}, 1\right)$ is characterized by less pronounced zigzagging. As a result, the behavior of ξ_H is often referred to as being *antipersistent* for $H \in \left(0, \frac{1}{2}\right)$, *chaotic* for $H = \frac{1}{2}$ and *persistent* for $H \in \left(\frac{1}{2}, 1\right)$; see Taqqu (2003).

1.1.2. Subordinated Gaussian processes

Chapter 2 and 3 of this thesis focus on the consideration of long-range dependent time series generated by transformations of Gaussian processes. We will refer to this model as *Gaussian subordination.*

Definition 7. Let ξ_t, $t \in T$, be a Gaussian process with index set T. A process Y_t, $t \in T$, satisfying $Y_t = G(\xi_t)$ for some measurable function $G : \mathbb{R} \longrightarrow \mathbb{R}$ is called *subordinated Gaussian process.*

Remark 2. For any particular distribution function F, an appropriate choice of the transformation G in Definition 7 yields subordinated Gaussian processes with marginal distribution F. Moreover, there exist algorithms for generating Gaussian processes that, after suitable transformation, yield subordinated Gaussian processes with marginal distribution F and a predefined covariance structure; see Pipiras and Taqqu (2017).

Example 1. For $k, \alpha > 0$, the cumulative distribution function

$$F_{\alpha,k}(x) := \begin{cases} 1 - \left(\frac{x}{k}\right)^{-\alpha} & \text{if } x \geq k, \\ 0 & \text{else,} \end{cases}$$

characterizes the Pareto(α, k) distribution with scale parameter k and shape parameter α, known as *tail index.* A Pareto(α, k)-distributed random variable Y has finite expectation when $\alpha > 1$ and finite variance when $\alpha > 2$. In these cases, the expected value and the variance are given by

$$\mathrm{E}\,Y = \frac{\alpha k}{\alpha - 1}, \ \alpha > 1,$$

$$\operatorname{Var} Y = \frac{\alpha k^2}{(\alpha-1)^2(\alpha-2)}, \ \alpha > 2.$$

To generate a Pareto-distributed time series Y_t, $t \in T$, which has a representation as a subordinated Gaussian process, i.e. which satisfies $Y_t = G(\xi_t)$ for some measurable function G and Gaussian random variables ξ_t, $t \in T$, we consider the quantile transformation $G(x) := k(\Phi(x))^{-\frac{1}{\alpha}}$ with Φ denoting the standard normal distribution function. In order to obtain standardized observations, i.e. random variables Y_t, $t \in T$, with mean 0 and variance 1, we have to choose

$$G(x) = \left(\frac{\alpha k^2}{(\alpha-1)^2(\alpha-2)}\right)^{-\frac{1}{2}} \left(k(\Phi(x))^{-\frac{1}{\alpha}} - \frac{\alpha k}{\alpha-1}\right).$$

The subordinated random variables $Y_t = G(\xi_t)$, $t \in T$, can be considered as elements of the Hilbert space $L^2(\mathbb{R}, \varphi(x)dx)$, i.e. the space of all measurable, real-valued functions which are square-integrable with respect to the measure $\varphi(x)dx$ associated with the standard normal density function φ. For two functions $G_1, G_2 \in L^2(\mathbb{R}, \varphi(x)dx)$ the corresponding inner product is defined by

$$\langle G_1, G_2 \rangle_{L^2} := \int_{-\infty}^{\infty} G_1(x)G_2(x)\varphi(x)dx = \operatorname{E} G_1(X)G_2(X)$$

with X denoting a standard normally distributed random variable.
A collection of orthogonal elements in $L^2(\mathbb{R}, \varphi(x)dx)$ is given by the sequence of Hermite polynomials.

Definition 8 (Pipiras and Taqqu (2017))**.** For $n \geq 0$, the *Hermite polynomial* of order n is defined by

$$H_n(x) = (-1)^n \mathrm{e}^{\frac{1}{2}x^2} \frac{d^n}{dx^n} \mathrm{e}^{-\frac{1}{2}x^2}, \ x \in \mathbb{R}.$$

Orthogonality of the sequence H_n, $n \geq 0$, in $L^2(\mathbb{R}, \varphi(x)dx)$ follows from

$$\langle H_n, H_m \rangle_{L^2} = \begin{cases} n! & \text{if } n = m, \\ 0 & \text{if } n \neq m. \end{cases}$$

Moreover, it can be shown that the Hermite polynomials form an orthogonal basis of $L^2(\mathbb{R}, \varphi(x)dx)$. As a result, every $G \in L^2(\mathbb{R}, \varphi(x)dx)$ has an expansion in Hermite polynomials, i.e. for $G \in L^2(\mathbb{R}, \varphi(x)dx)$ and X standard normally distributed, we have

$$G(X) = \sum_{r=0}^{\infty} \frac{J_r(G)}{r!} H_r(X), \tag{1.2}$$

where the so-called *Hermite coefficient* $J_r(G)$ is given by

$$J_r(G) := \langle G, H_r \rangle_{L^2} = \operatorname{E} G(X)H_r(X).$$

Equation (1.2) holds in an L^2-sense, meaning

$$\lim_{n\to\infty} \left\| G(X) - \sum_{r=0}^{n} \frac{J_r(G)}{r!} H_r(X) \right\|_{L^2} = 0,$$

where $\|\cdot\|_{L^2}$ denotes the norm induced by the inner product $\langle\cdot,\cdot\rangle_{L^2}$.

Given the Hermite expansion (1.2), it is possible to characterize the dependence structure of subordinated Gaussian time series $G(\xi_n)$, $n \in \mathbb{N}$. Under the assumption that the Gaussian sequence ξ_n, $n \in \mathbb{N}$, is stationary and that G is a one-to-one function, the behavior of the autocorrelations of the transformed process is completely determined by the dependence structure of the underlying process. However, this is not the case in general. In fact, it holds that

$$\operatorname{Cov}(G(\xi_1), G(\xi_{k+1})) = \sum_{r=1}^{\infty} \frac{J_r^2(G)}{r!} (\gamma(k))^r, \tag{1.3}$$

where γ denotes the autocovariance function of ξ_n, $n \in \mathbb{N}$. Under the assumption that, as k tends to ∞, $\gamma(k)$ converges to 0 with a certain rate, the asymptotically dominating term in the series (1.3) is the summand corresponding to the smallest integer r for which the Hermite coefficient $J_r(G)$ is non-zero. This index, which decisively depends on G, is called *Hermite rank*.

Definition 9 (Pipiras and Taqqu (2017))**.** Let $G \in L^2(\mathbb{R}, \varphi(x)dx)$, $\operatorname{E} G(X) = 0$ for standard normally distributed X and let $J_r(G)$, $r \geq 0$, be the Hermite coefficients in the Hermite expansion of G. The smallest index $k \geq 1$ for which $J_k(G) \neq 0$ is called the *Hermite rank* of G, i.e.

$$r := \min\{k \geq 1 : J_k(G) \neq 0\}.$$

It follows from (1.3) that subordination of long-range dependent Gaussian time series potentially generates time series whose autocovariances decay faster than the autocovariances of the underlying Gaussian process. In some cases, the subordinated time series is long-range dependent as well, in other cases subordination may even yield short-range dependence.

Theorem 2 (Pipiras and Taqqu (2017))**.** *Let ξ_n, $n \in \mathbb{N}$, be a stationary, long-range dependent Gaussian series with mean 0, variance 1 and $\operatorname{Cov}(\xi_1, \xi_{k+1}) \sim k^{-D}L(k)$, as $k \to \infty$, and let $G \in L^2(\mathbb{R}, \varphi(x)dx)$ be a function with Hermite rank r. Then*

$$\operatorname{Cov}(G(\xi_1), G(\xi_{k+1})) \sim J_r^2(G) r! k^{-Dr} L^r(k), \quad \text{as } k \to \infty.$$

It immediately follows from the above theorem that subordinated Gaussian time series $G(\xi_n)$, $n \in \mathbb{N}$, are long-range dependent with LRD parameter $D_G := Dr$ and slowly-varying function $L_G(k) = J_r^2(G) r! L^r(k)$ whenever $Dr < 1$.

Given the previous definitions, it is possible to specify model assumptions that are taken as a basis for the results presented in the following chapters.

Model 1. Let $Y_n = G(\xi_n)$, where ξ_n, $n \in \mathbb{N}$, is a stationary, long-range dependent Gaussian time series with mean 0, variance 1 and LRD parameter D, and let F denote the marginal distribution function of Y_n, $n \in \mathbb{N}$. Moreover, let $J_r(G; x)$ denote the r-th Hermite coefficient in the Hermite expansion of $1_{\{G(\xi_i) \leq x\}} - F(x)$, i.e.

$$J_r(G; x) := \mathrm{E}\left(1_{\{G(\xi_i) \leq x\}} H_r(\xi_i)\right).$$

We assume that $Dr < 1$, where r denotes the Hermite rank of the class of functions $1_{\{G(\xi_1) \leq x\}} - F(x)$, $x \in \mathbb{R}$, defined by

$$r := \min_{x \in \mathbb{R}} r(x), \quad r(x) := \min\left\{q \geq 1 : J_q(G; x) \neq 0\right\}.$$

Moreover, we assume that $G : \mathbb{R} \longrightarrow \mathbb{R}$ is a measurable function and that F is continuous.

1.2. Change-point identification

The following subsections address aspects of change-point analysis that are relevant to this thesis. In particular, hypothesis tests that differentiate between data generated by stationary time series and data generated by time series containing a structural change are considered. The presented results focus on non-parametric change-point tests, i.e. on testing procedures that do not necessarily rest upon the assumption that the considered data stems from a parametric family of probability distributions. The emphasis is laid on change-point tests based on Wilcoxon-type statistics. These are robust in the sense that outliers in the data do not have a significant impact on test decisions. Moreover, particular emphasis is given to self-normalized statistics, i.e. statistics that, due to standardization by data-dependent quantities, can be considered as parameter-free.

1.2.1. Change-point problems

The most frequently considered change-point problems relate to the identification of shifts in the mean value of time series; see Figure 1.2 for an illustration. Formally, we refer to the following assumption when considering time series with a change in the mean:

Assumption 1. We assume that a given set of observations $X_1, \ldots, X_n$ is generated by a sequence of random variables X_n, $n \in \mathbb{N}$, where

$$X_n = \mu_n + Y_n$$

for a sequence of unknown constants μ_n, $n \in \mathbb{N}$, and a mean-zero stochastic process Y_n, $n \in \mathbb{N}$. A *change-point in the mean* of the time series X_n, $n \in \mathbb{N}$, is characterized by a sequence μ_n, $n \in \mathbb{N}$, satisfying

$$\mu_k = \begin{cases} \mu & \text{for } k = 1, \ldots, k_0, \\ \mu + h_n & \text{for } k = k_0 + 1, \ldots, n \end{cases}$$

for some $k_0 = \lfloor n\tau \rfloor$, $0 < \tau < 1$, denoting the change-point location, and a deterministic sequence of shift heights h_n, $n \in \mathbb{N}$, with $h_n \neq 0$ for all $n \in \mathbb{N}$.
We differentiate between fixed and local changes:

- $h_n = h$ for some $h \neq 0$ *(fixed changes)*;
- $\lim_{n\to\infty} h_n = 0$ *(local changes)*.

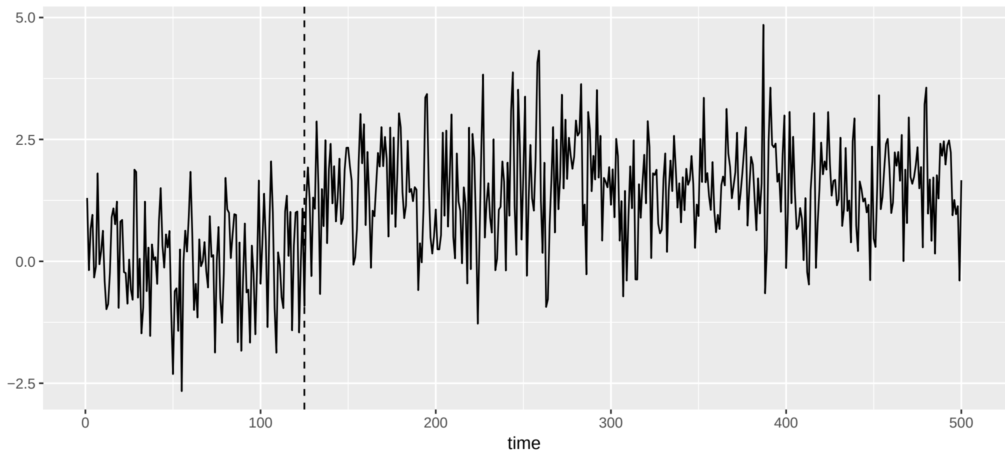

Figure 1.2.: *Time series of length* $n = 500$ *generated by fractional Gaussian noise with Hurst parameter* $H = 0.6$ *and a change in the mean of height* $h = 1.5$ *in* $k_0 = \lfloor n\tau \rfloor$ *with* $\tau = 0.25$.

In fact, various different change-point problems can be reduced to identifying changes in the mean of transformed observations $\psi(X_1), \ldots, \psi(X_n)$, where ψ is a suitably chosen function. Possible choices include:

- $\psi(x) = x$ in order to detect changes in the mean *(change in location)*;
- $\psi(x) = x^2$ in order to detect changes in the variance *(change in volatility)*;
- $\psi(x) = \log(x^2)$ or $\psi(x) = \log(|x|)$ in order to detect changes in the tail parameter α of heavy-tailed observations *(change in the tail index)*.

Given a sequence of observations with a structural change, we are interested in identifying the location of the change, i.e. the point in time when the change occurred. Estimators for the change-point location are considered in Chapter 2.
An important issue in change-point analysis that precedes the estimation of change-point locations can be described by the question whether the data-generating process that underlies a given set of observations changes at all. Since observational data is subject to random fluctuations, it is often difficult to discriminate between structural changes in data-generating processes and random effects that induce ostensible changes.

Given observations $X_1, \dots, X_n$ and a function ψ, chosen according to the specific change-point problem, we consider the testing problem (H, A):

$$H: \ \mathrm{E}\,\psi(X_1) = \cdots = \mathrm{E}\,\psi(X_n)$$

against

$$\begin{aligned} A: \ &\mathrm{E}\,\psi(X_1) = \cdots = \mathrm{E}\,\psi(X_k) \neq \mathrm{E}\,\psi(X_{k+1}) = \cdots = \mathrm{E}\,\psi(X_n) \\ &\text{for some} \quad k \in \{1, \dots, n-1\}\,. \end{aligned}$$

The above choice of the alternative hypothesis implies that under the assumption of a structural change, the location of the change-point is unknown. To motivate the design of test statistics for deciding on the change-point problem, we temporarily assume that the change-point location is known, i.e. for a given $k \in \{1, \dots, n-1\}$ we consider the testing problem (H, A_k):

$$H: \ \mathrm{E}\,\psi(X_1) = \cdots = \mathrm{E}\,\psi(X_n)$$

against

$$A_k: \ \mathrm{E}\,\psi(X_1) = \cdots = \mathrm{E}\,\psi(X_k) \neq \mathrm{E}\,\psi(X_{k+1}) = \cdots = \mathrm{E}\,\psi(X_n).$$

Within this setting, the two-sample CUSUM test is a commonly used non-parametric test. The corresponding test statistic is given by the cumulative sum of differences between single values and the overall average. More precisely, the two-sample statistic is defined by

$$C_{k,n} := \sum_{i=1}^{k} \psi(X_i) - \frac{k}{n} \sum_{j=1}^{n} \psi(X_j).$$

Under the assumption of a change in location, that means if for some point in time k the values of $\psi(X_1), \dots, \psi(X_k)$ tend to be above or below average, the absolute value of $C_{k,n}$ will be exceptionally large. For this reason, the two-sample CUSUM test rejects the hypothesis for values of $|C_{k,n}|$ that exceed a predefined threshold.
Another non-parametric test for the two-sample testing problem (H, A_k) is the Wilcoxon rank test which is based on the test statistic

$$W_{k,n} := \sum_{i=1}^{k} \sum_{j=k+1}^{n} \left(1_{\{\psi(X_i) \le \psi(X_j)\}} - \frac{1}{2} \right).$$

By definition, the value of $W_{k,n}$ corresponds to the number of times one of the observations $\psi(X_1), \dots, \psi(X_k)$ exceeds one of the observations $\psi(X_{k+1}), \dots, \psi(X_n)$. Therefore, the two-sample Wilcoxon test rejects the hypothesis for exceptionally large values of $|W_{k,n}|$.

Under the assumption that the change-point location is unknown under the alternative, it seems natural to consider the statistics $|C_{k,n}|$ and $|W_{k,n}|$ for every possible change-point location k and to decide in favor of the alternative hypothesis A if the maximum exceeds a predefined critical value. As a result, CUSUM and Wilcoxon change-point tests base test decisions on the values of the statistics

$$C_n := \max_{1\leq k\leq n-1} |C_{k,n}| \quad \text{and} \quad W_n := \max_{1\leq k\leq n-1} |W_{k,n}| \,. \tag{1.4}$$

1.2.2. Wilcoxon change-point test

Since the exact distribution of the Wilcoxon statistic is unknown and, in general, hard to obtain, test decisions are based on a comparison of the value of the test statistic with quantiles of its limit distribution. For the determination of the asymptotic distribution of the Wilcoxon statistic W_n, computed with respect to time series data Y_n, $n \in \mathbb{N}$, with continuous marginal distribution function F, it is useful to note that

$$W_{k,n} = \sum_{i=1}^{k} \sum_{j=k+1}^{n} \left(1_{\{Y_i\leq Y_j\}} - \frac{1}{2}\right) = (n-k)\,k \left(\int_{\mathbb{R}} F_k(x)dF_{k+1,n}(x) - \int_{\mathbb{R}} F(x)dF(x)\right),$$

where F_k and $F_{k+1,n}$ denote the empirical distribution functions of the first k and last $n-k$ realizations of $Y_1, \ldots, Y_n$, i.e.

$$F_k(x) := \frac{1}{k}\sum_{i=1}^{k} 1_{\{Y_i\leq x\}},$$

$$F_{k+1,n}(x) := \frac{1}{n-k}\sum_{i=k+1}^{n} 1_{\{Y_i\leq x\}}.$$

As shown in Dehling et al. (2013), due to the above representation, the asymptotic distribution of the Wilcoxon statistic can be derived from the following empirical process limit theorem:

Theorem 3 (Dehling and Taqqu (1989))**.** *Let $Y_n = G(\xi_n)$, $n \in \mathbb{N}$, be a subordinated Gaussian sequence according to Model 1 and define the sequence $d_{n,r}$, $n \in \mathbb{N}$, by*

$$d_{n,r}^2 := \operatorname{Var}\left(\sum_{i=1}^{n} H_r(\xi_i)\right) \tag{1.5}$$

with H_r denoting the r-th order Hermite polynomial and r the Hermite rank of the class of functions $1_{\{G(\xi_1)\leq x\}} - F(x)$, $x \in \mathbb{R}$. Then, as $n \to \infty$,

$$d_{n,r}^{-1}\lfloor nt\rfloor (F_{\lfloor nt\rfloor}(x) - F(x)) \xrightarrow{\mathcal{D}} \frac{1}{r!} J_r(x) Z_{r,H}(t), \;\; x \in [-\infty, \infty]\,, t \in [0,1]\,,$$

where $Z_{r,H}$ is an r-th order Hermite process, $H = 1-\frac{rD}{2}$, and $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution with respect to the σ-field generated by the open balls in $D\left([-\infty,\infty] \times [0,1]\right)$, equipped with the supremum norm.

Remark 3.

1. Since the space $D([-\infty,\infty]\times[0,1])$, equipped with the supremum norm, is non-separable, the standard definition of weak convergence cannot be applied in this context. A detailed discussion of convergence in distribution in non-separable càdlàg spaces can be found in Appendix A.

2. If $r=1$, the Hermite process $Z_{r,H}$ equals a standard fractional Brownian motion with Hurst parameter $H=1-\frac{D}{2}$. We refer to Taqqu (1979) for a general definition of Hermite processes.

We write

$$\begin{aligned}e_n(x,t) &:= d_{n,r}^{-1}\lfloor nt\rfloor(F_{\lfloor nt\rfloor}(x)-F(x)),\\ e(x,t) &:= \frac{1}{r!}J_r(x)Z_{r,H}(t),\end{aligned}$$

so that e_n, $n\in\mathbb{N}$, can be considered as a sequence of random variables with values in $D([-\infty,\infty]\times[0,1])$ converging in distribution to e. Note that J_r is bounded and continuous. Moreover, the Hermite process $Z_{r,H}$ is almost surely continuous; see Mikosch (1998).

With $C([-\infty,\infty]\times[0,1])$ denoting the set of all continuous, real-valued functions with domain $[-\infty,\infty]\times[0,1]$, it follows that $e\in C([-\infty,\infty]\times[0,1])$ almost surely. Since $C([-\infty,\infty]\times[0,1])$ is a separable subset of $D([-\infty,\infty]\times[0,1])$, the Dudley-Wichura version of Skorohod's representation theorem implies that there exists another probability space $(\Omega^\star,\mathcal{F}^\star,P^\star)$ and random variables $e_n^\star$, $n\in\mathbb{N}$, and $e^\star$ defined on that space with

$$e_n^\star\overset{\mathcal{D}}{=}e_n \quad\text{and}\quad e^\star\overset{\mathcal{D}}{=}e \quad\text{such that}\quad \sup_{x\in[-\infty,\infty],t\in[0,1]}|e_n^\star(x,t)-e^\star(x,t)|\xrightarrow{\text{a.s.}}0$$

see Shorack and Wellner (1986), Theorem 2.3.4. For notational convenience, we write

$$\sup_{x\in[-\infty,\infty],t\in[0,1]}\left|d_{n,r}^{-1}\lfloor nt\rfloor\left(F_{\lfloor nt\rfloor}(x)-F(x)\right)-\frac{1}{r!}J_r(x)Z_{r,H}(t)\right|\xrightarrow{\text{a.s.}}0,\ \text{as } n\to\infty, \tag{1.6}$$

to indicate the existence of random variables $e_n^\star$ and $e^\star$ with these properties, although, generally speaking, it is not possible to infer that $\sup_{x\in[-\infty,\infty],t\in[0,1]}|e_n(x,t)-e(x,t)|$ converges to 0 almost surely. Since, whenever the argument in the proofs is based on the almost sure convergence in (1.6), we are only interested in distributional properties, this notation is justified.

With the objective of calculating the asymptotic distribution of the Wilcoxon test statistic, we consider the stochastic process

$$W_{n,r}(t) := \frac{1}{nd_{n,r}}W_{\lfloor nt\rfloor,n}=\frac{1}{nd_{n,r}}\sum_{i=1}^{\lfloor nt\rfloor}\sum_{j=\lfloor nt\rfloor+1}^{n}\left(1_{\{X_i\le X_j\}}-\frac{1}{2}\right),\ t\in[0,1]. \tag{1.7}$$

Given the asymptotic distribution of this process, the limit distribution of the Wilcoxon test statistic can be derived directly by an application of the continuous mapping theorem.
To fully characterize the asymptotic behavior of the process, we differentiate the following cases:

1. $h_n = o\left(\frac{d_{n,r}}{n}\right)$;

2. $h_n \sim c\frac{d_{n,r}}{n}$ for some constant c;

3. $h_n^{-1} = o\left(\frac{n}{d_{n,r}}\right)$.

If there is a change-point in the mean, the behavior of the process $W_{n,r}(t)$, $t \in [0,1]$, is influenced by a deterministic component depending on the height of the level shift as well as the location of the change-point. For a detailed description of the asymptotics, define the function $\delta_\tau : [0,1] \longrightarrow \mathbb{R}$ by

$$\delta_\tau(t) = \begin{cases} t(1-\tau) & \text{for } t \leq \tau, \\ (1-t)\tau & \text{for } t \geq \tau. \end{cases}$$

In the presence of local changes, the asymptotic behavior of the process is characterized by the following theorem:

Theorem 4 (Dehling et al. (2017a))**.** *Let X_n, $n \in \mathbb{N}$, with $X_n = \mu_n + Y_n$ denote a time series with a local change in the mean with shift height h_n, $n \in \mathbb{N}$, as defined in Assumption 1 and let Y_n, $n \in \mathbb{N}$, be a subordinated Gaussian sequence according to Model 1. Then*

$$W_{n,r}(t) - \frac{n}{d_{n,r}}\delta_\tau(t)\int_{\mathbb{R}} (F(x+h_n) - F(x))\, dF(x), \ t \in [0,1],$$

converges in distribution to

$$\frac{1}{r!}\left(Z_{r,H}(t) - tZ_{r,H}(1)\right)\int_{\mathbb{R}} J_r(x) dF(x), \ t \in [0,1],$$

in $D[0,1]$.

Remark 4. Choosing $h_n \equiv 0$, Theorem 4 provides the asymptotic distribution of the Wilcoxon process under the hypothesis of stationarity.

Dependent on the convergence rate of the shift height, it is possible to determine the asymptotic behavior of $W_{n,r}(t)$, $t \in [0,1]$, under local changes. To see this, note that, under the assumption that F has a bounded density f,

$$\frac{n}{d_{n,r}}\delta_\tau(t)\int_{\mathbb{R}} (F(x+h_n) - F(x))\, dF(x), \ t \in [0,1],$$

converges to

$$\begin{cases} 0 & \text{if } h_n = o\left(\frac{d_{n,r}}{n}\right); \\ \delta_\tau(t) \int_{\mathbb{R}} f^2(x)dx & \text{if } h_n \sim \frac{d_{n,r}}{n}; \\ \infty & \text{if } h_n^{-1} = o\left(\frac{n}{d_{n,r}}\right); \end{cases}$$

uniformly in $t \in [0,1]$. As a result, $W_{n,r}(t)$, $t \in [0,1]$, converges in distribution to

$$\begin{cases} \frac{1}{r!}\left(Z_{r,H}(t) - tZ_{r,H}(1)\right) \int_{\mathbb{R}} J_r(x)dF(x) & \text{if } h_n = o\left(\frac{d_{n,r}}{n}\right); \\ \frac{1}{r!}\left(Z_{r,H}(t) - tZ_{r,H}(1)\right) \int_{\mathbb{R}} J_r(x)dF(x) + \delta_\tau(t) \int_{\mathbb{R}} f^2(x)dx & \text{if } h_n \sim \frac{d_{n,r}}{n}; \\ \infty & \text{if } h_n^{-1} = o\left(\frac{n}{d_{n,r}}\right); \end{cases}$$

in $D[0,1]$. In the latter case, i.e. when $W_{n,r}(t)$, $t \in [0,1]$, diverges, a stronger normalization of $W_{n,r}(t)$, $t \in [0,1]$, ensures convergence to a proper deterministic function; see Lemma 1 in Chapter 2.

1.2.3. Self-normalized change-point tests

An application of the Wilcoxon change-point test to a given data set presupposes determination of the scaling factor $d_{n,r}$. With r denoting the Hermite rank of the class of functions $1_{\{G(\xi_1)\leq x\}} - F(x)$, $x \in \mathbb{R}$, we have $d_{n,r}^2 \sim c_r n^{2-rD} L^r(n)$, where c_r is a constant depending on r and D; see Dehling et al. (2013). In statistical practice, the parameters D, r and the function L are usually unknown. While in many cases $r = 1$, and while there are methods to estimate D, estimating L seems to be hardly possible. In order to avoid a normalization depending on these unknown quantities, it seems reasonable to replace the deterministic normalization by a data-driven one, i.e. by a normalizing sequence that depends on the given realizations only and which is therefore referred to as *self-normalization*.

The concept of self-normalization has recently been applied to several testing procedures in change-point analysis. Originally established by Lobato (2001) in another testing context, it has been adapted to the change-point problem in Shao and Zhang (2010) by definition of a self-normalized Kolmogorov-Smirnov test statistic. In these papers, short-range dependent processes are considered. An extension to possibly long-range dependent processes was introduced by Shao, who established a self-normalized change-point test based on the CUSUM statistic; see Shao (2011). A self-normalized version of the Wilcoxon change-point test is considered in Betken (2016).

The definition of the self-normalized Wilcoxon statistic is best motivated with reference to an application of the self-normalization procedure to the CUSUM statistic. Given a stochastic process X_n, $n \in \mathbb{N}$, the asymptotic distribution of the CUSUM statistic is usually derived from a central limit theorem for the sequential partial sum process

$$\frac{1}{\sigma_n} \sum_{i=1}^{\lfloor nt \rfloor} \left(X_i - \operatorname{E} X_1\right), \ t \in [0,1],$$

where $\sigma_n^2 := \operatorname{Var} S_n$ with $S_n := \sum_{i=1}^n X_i$. Under the assumption of independent, identically distributed random variables, $\sigma_n^2 = n\sigma^2$ with σ^2 denoting the variance of the marginal distribution. In this case, a natural estimate for σ_n^2 is the empirical variance

$$\hat{\sigma}_n^2 := \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}_n\right)^2, \quad \text{where} \quad \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

For stationary time series generated by a sequence of dependent random variables X_n, $n \in \mathbb{N}$, the variance of partial sums cannot be derived from the variance of the marginal distribution. In fact, we have

$$\lim_{n\to\infty} \frac{1}{n}\sigma_n^2 = \sum_{j=-\infty}^{\infty} \gamma(j),$$

so that the normalization of the partial sum process depends on the covariance structure of the considered time series. For this reason, the empirical variance $\hat{\sigma}_n$ can no longer be considered as a suitable normalization. Instead, we consider the variance of S_k, $k = 1, \dots, n$, as an estimate for σ_n^2. Under the assumption of stationary time series data, the empirical variance of the partial sums corresponds to

$$\hat{\sigma}_n^2 := \frac{1}{n} \sum_{k=1}^{n} (S_k - \operatorname{E} S_k)^2 = \frac{1}{n} \sum_{k=1}^{n} \left(\sum_{i=1}^{k} X_i - k \operatorname{E} X_1 \right)^2.$$

By replacing the expected value of X_1 in the previous formula with its natural estimate $\bar{X}_n$, a parameter-free normalizing sequence for the partial sum process is defined by

$$V_n^2 := \frac{1}{n} \sum_{k=1}^{n} \left(S_k - k\bar{X}_n\right)^2 = \frac{1}{n} \sum_{k=1}^{n} \left(\sum_{i=1}^{k} \left(X_i - \bar{X}_n\right) \right)^2.$$

Remark 5. Given a sequence of dependent random variables X_n, $n \in \mathbb{N}$, a commonly used estimator for the long-run variance $\sum_{j=-\infty}^{\infty} \gamma(j)$ is the lag-window estimate

$$\sum_{j=-(n-1)}^{n-1} K\left(\frac{j}{b_n}\right) \hat{\gamma}_n(j), \ \hat{\gamma}_n(j) := \frac{1}{n} \sum_{i=1}^{n-|j|} \left(X_i - \bar{X}_n\right) \left(X_{i+|j|} - \bar{X}_n\right),$$

where K denotes a kernel function and b_n is a bandwidth parameter. Choosing K as the Bartlett kernel, i.e.

$$K(x) := (1 - |x|)\, 1_{\{|x| \leq 1\}},$$

and $b_n = n$, it follows that

$$\sum_{j=-(n-1)}^{n-1} K\left(\frac{j}{b_n}\right) \hat{\gamma}_n(j) = \frac{2}{n} V_n^2.$$

Therefore, self-normalization can be regarded as a special case of kernel-based estimation for the long-run variance; see Kiefer and Vogelsang (2002) and Shao (2010).

Taking the possibility of a structural change at time k into consideration, a normalization for the two-sample CUSUM statistic is obtained by combining the values of V_n computed with respect to the separate samples $X_1, \ldots, X_k$ and $X_{k+1}, \ldots, X_n$. Accordingly, define

$$V_{k,n}^2 := \frac{1}{n} \sum_{t=1}^{k} S_t^2(1,k) + \frac{1}{n} \sum_{t=k+1}^{n} S_t^2(k+1,n)$$

with

$$S_t(j,k) := \sum_{h=j}^{t} \left(X_h - \bar{X}_{j,k} \right), \quad \bar{X}_{j,k} := \frac{1}{k-j+1} \sum_{t=j}^{k} X_t,$$

as normalizing sequence and define the self-normalized CUSUM statistic by

$$SC_n(\tau_1, \tau_2) := \max_{\lfloor n\tau_1 \rfloor \le k \le \lfloor n\tau_2 \rfloor} |SC_{k,n}|, \quad SC_{k,n} := \frac{C_{k,n}}{V_{k,n}},$$

where $0 < \tau_1 < \tau_2 < 1$. The corresponding change-point test for the test problem (H, A), considered in Section 1.2.1, rejects the hypothesis for large values of $SC_n(\tau_1, \tau_2)$.
Note that the proportion of the two-sample statistics that are included in the calculation of $SC_n(\tau_1, \tau_2)$ is restricted by the choice of τ_1 and τ_2. Structural breaks at the beginning or the end of a sample are hard to detect since there is a lack of information concerning the behavior of the time series before or after a potential break point. Hence, the interval $[\tau_1, \tau_2]$ must be small enough for the critical values not to get too large on the one hand, yet large enough to include potential break points on the other hand. A common choice is $\tau_1 = 1 - \tau_2 = 0.15$; see Andrews (1993).
Replacing the original observations $X_1, \ldots, X_n$ in the CUSUM test statistic by their ranks $R_1, \ldots, R_n$ results in the following identity:

$$\left| \sum_{i=1}^{k} R_i - \frac{k}{n} \sum_{i=1}^{n} R_i \right| = \left| \sum_{i=1}^{k} \sum_{j=k+1}^{n} \left(1_{\{X_i \le X_j\}} - \frac{1}{2} \right) \right|, \quad R_i := \operatorname{rank}(X_i) = \sum_{j=1}^{n} 1_{\{X_j \le X_i\}},$$

i.e. the Wilcoxon statistic arises from an application of the CUSUM statistic to the ranks. Therefore, it seems natural to choose a data-driven normalization for the Wilcoxon statistic by evaluation of the self-normalized CUSUM statistic in $R_1, \ldots, R_n$. For this reason, we define the self-normalized two-sample Wilcoxon statistic by

$$SW_{k,n} := \frac{\sum_{i=1}^{k} R_i - \frac{k}{n} \sum_{i=1}^{n} R_i}{\left\{ \frac{1}{n} \sum_{t=1}^{k} S_t^2(1,k) + \frac{1}{n} \sum_{t=k+1}^{n} S_t^2(k+1,n) \right\}^{1/2}}, \tag{1.8}$$

where

$$S_t(j,k) := \sum_{h=j}^{t} \left(R_h - \bar{R}_{j,k} \right) \quad \text{with} \quad \bar{R}_{j,k} := \frac{1}{k-j+1} \sum_{t=j}^{k} R_t.$$

The self-normalized Wilcoxon change-point test for the test problem (H, A), considered in Section 1.2.1, rejects the hypothesis for large values of

$$SW_n(\tau_1, \tau_2) := \max_{\lfloor n\tau_1 \rfloor \leq k \leq \lfloor n\tau_2 \rfloor} |SW_{k,n}|, \tag{1.9}$$

where $0 < \tau_1 < \tau_2 < 1$.
Note that

$$SW_n(t) := SW_{\lfloor nt \rfloor, n} = G_{W_n}(t) + \mathcal{O}_P(1), \quad \tau_1 \leq t \leq \tau_2, \tag{1.10}$$

where for $f \in D\,[0,1]$ the function $G_f \in D\,[0,1]$ is defined by

$$G_f(t) := \frac{f(t)}{V_f(t)},\ V_f(t) := \left\{\int_0^t \left(f(s) - \frac{s}{t} f(t)\right)^2 ds + \int_t^1 \left(f(s) - \frac{1-s}{1-t} f(t)\right)^2 ds\right\}^{\frac{1}{2}},$$

so that the limit of the self-normalized process $SW_n(t)$, $\tau_1 \leq t \leq \tau_2$, can be derived from the asymptotic behavior of the Wilcoxon process $W_n(t)$, $t \in [0,1]$, described in Section 1.2.2. In particular, it follows that under local changes the process $SW_n(t)$, $\tau_1 \leq t \leq \tau_2$, converges to

- $G_{W_{r,H}}(t)$, $\tau_1 \leq t \leq \tau_2$, with

  $$W_{r,H}(t) := \frac{1}{r!}\left(Z_{r,H}(t) - tZ_{r,H}(1)\right) \int_{\mathbb{R}} J_r(x) dF(x)$$

  if $h_n = o\left(\frac{d_{n,r}}{n}\right)$;

- $G_{W_{r,H,\tau}}(t)$, $\tau_1 \leq t \leq \tau_2$, with

  $$W_{r,H,\tau}(t) := \frac{1}{r!}\left(Z_{r,H}(t) - tZ_{r,H}(1)\right) \int_{\mathbb{R}} J_r(x) dF(x) + \delta_\tau(t) \int f^2(x) dx$$

  if $h_n \sim \frac{d_{n,r}}{n}$.

Remark 6. For $r = 1$, the asymptotic distribution of the self-normalized Wilcoxon statistic is derived in Betken (2013), while formal proofs for corresponding limit theorems allowing for $r > 1$ can only be found in Betken (2016). Beside generalizations of the results established in Betken (2013), Betken (2016) provides a proof for the consistency of the self-normalized Wilcoxon change-point test under the assumption of changes in the mean with fixed change-point height.

2. Wilcoxon-type change-point estimators

Recall that, given a sample of observations $X_1, \ldots, X_n$, the two-sample Wilcoxon statistic is defined by

$$W_{k,n} = \sum_{i=1}^{k} \sum_{j=k+1}^{n} \left(1_{\{X_i \leq X_j\}} - \frac{1}{2}\right)$$

for $k \in \{1, \ldots, n-1\}$. Since the value of $W_{k,n}$ is determined by the number of times one of the observations $X_{k+1}, \ldots, X_n$ exceeds one of the observations $X_1, \ldots, X_k$, we expect the absolute value of $W_{k_0,n}$ to exceed the absolute value of $W_{l,n}$ for any $l \neq k_0$ if k_0 denotes the location of a change-point in the mean. This observation is illustrated by Figure 2.1 which depicts the values of the two-sample statistic $W_{k,n}$ as a function of k for a sample of long-range dependent observations with a change-point located in k_0.

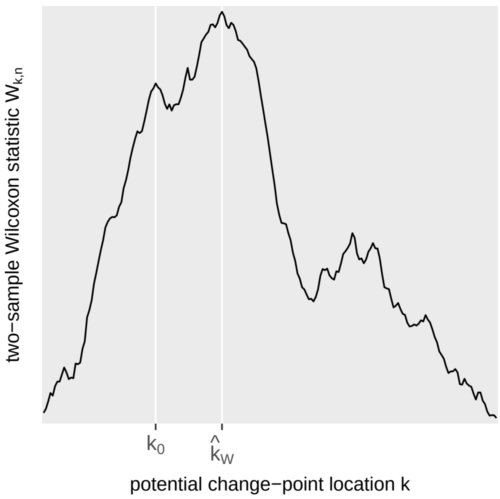

Figure 2.1.: *Values of the two-sample Wilcoxon statistic $W_{k,n}$ for a time series of length $n = 200$ generated by fractional Gaussian noise with Hurst parameter $H = 0.9$ and a change in the mean of height $h = 1$ in $k_0 = \lfloor n\tau \rfloor$ with $\tau = 0.25$.*

Apparently, the value of the two-sample statistic increase when k approaches the change-point location k_0. Therefore, it seems natural to define an estimator of k_0 by

$$\hat{k}_W = \hat{k}_W(n) := \min\left\{k : |W_{k,n}| = \max_{1 \leq i \leq n-1} |W_{i,n}|\right\}. \tag{2.1}$$

In the same way, the change-point estimator

$$\hat{k}_{SW} = \hat{k}_{SW}(n) := \min\left\{k : |SW_{k,n}| = \max_{\lfloor n\tau_1\rfloor \leq i \leq \lfloor n\tau_2\rfloor} |SW_{i,n}|\right\}$$

arises from the self-normalized Wilcoxon two-sample statistic $SW_{k,n}$ defined by (1.8) in Section 1.2.3.

In order to characterize the asymptotic behavior of the change-point estimators $\hat{k}_W$ and $\hat{k}_{SW}$, we impose the following assumption:

Assumption 2. Let X_n, $n \in \mathbb{N}$, denote a time series with a change in the mean in $k_0 = \lfloor n\tau \rfloor$ with shift height h_n, $n \in \mathbb{N}$. More precisely, assume that

$$X_k = \begin{cases} Y_k & \text{for } k \leq k_0 \\ Y_k + h_n & \text{for } k > k_0 \end{cases}$$

for a deterministic sequence of unknown constants h_n, $n \in \mathbb{N}$, and a mean-zero subordinated Gaussian sequence $Y_n = G(\xi_n)$, $n \in \mathbb{N}$, according to Model 1. In particular, assume that

$$\gamma(k) := \operatorname{Cov}(\xi_1, \xi_{k+1}) \sim k^{-D} L_\gamma(k), \quad \text{as } k \to \infty,$$

for some $D \in (0,1)$ and some slowly varying function L_γ. With r denoting the Hermite rank of the class of functions $1_{\{G(\xi_1) \leq x\}} - F(x)$, $x \in \mathbb{R}$, define

$$g_{D,r}(t) := t^{\frac{rD}{2}} L_\gamma^{-\frac{r}{2}}(t).$$

Remark 7. The function $g_{D,r}$ relates to the normalizing sequence $d_{n,r}$, defined by (1.5) in Section 1.2.2, as follows:

$$d_{n,r} \sim \frac{n}{g_{D,r}(n)} c_r, \text{ as } n \to \infty, \text{ where } c_r := \sqrt{\frac{2r!}{(1-Dr)(2-Dr)}}.$$

Since $g_{D,r}$ is a regularly varying function, there exists a function $g^{-}_{D,r}$ such that

$$\left(g_{D,r} \circ g^{-}_{D,r}\right)(t) \sim \left(g^{-}_{D,r} \circ g_{D,r}\right)(t) \sim t, \quad \text{as } t \to \infty;$$

see Theorem 1.5.12 in Bingham et al. (1987). We refer to $g^{-}_{D,r}$ as the *asymptotic inverse* of $g_{D,r}$.

The majority of articles that address the problem of estimating the change-point location refers to a family of estimators that can be derived from the two-sample CUSUM test statistics, defined by

$$C_{k,n}(\beta) := \left(\frac{k(n-k)}{n}\right)^{1-\beta} \left(\frac{1}{k}\sum_{i=1}^{k} X_i - \frac{1}{n-k}\sum_{i=k+1}^{n} X_i\right)$$

for $k \in \{1, \ldots, n-1\}$ and a parameter $0 \leq \beta < 1$. For $\beta = 0$, the corresponding statistic has already been considered in Section 1.2.1. Based on the values of $C_{k,n}(\beta)$, the location of the change-point is approximated by

$$\hat{k}_{C,\beta} = \hat{k}_{C,\beta}(n) := \min\left\{k : |C_{k,n}(\beta)| = \max_{1 \leq i \leq n-1} |C_{i,n}(\beta)|\right\}. \tag{2.2}$$

Under non-restrictive constraints on the dependence structure of the data-generating process (including long-range dependent time series), Kokoszka and Leipus (1998) prove consisteny of $\hat{k}_{C,\beta}$ for both, fixed and certain local changes. For a constant jump height, these authors establish convergence rates that depend on the intensity of dependence in the data-generating random variables. Under the additional assumption that the considered data is generated by Gaussian LRD processes, Horváth and Kokoszka (1997) derive the asymptotic distribution of the estimator $\hat{k}_{C,\beta}$.
Bai (1994) establishes an estimator for the location of a shift in the mean by the method of least squares. He proves consistency, determines the rate of convergence of the change-point estimator and derives its asymptotic distribution. These results are shown to hold for weakly dependent observations that satisfy a linear model and cover, for example, ARMA(p, q)-processes. Moreover, Bai extended these results to the estimation of the location of a parameter change in multiple regression models that also allow for lagged dependent variables and trending regressors; see Bai (1997). A generalization of these results to possibly long-range dependent data-generating processes (including fractionally integrated processes) is given in Kuan and Hsu (1998) and Lavielle and Moulines (2000).
Under the assumption of independent data, Darkhovskh (1976) establishes an estimator for the location of a change in distribution based on the two-sample Mann-Whitney test statistic. He obtains a convergence rate that has order $\frac{1}{n}$, where n is the number of observations. Allowing for strong dependence in the data, Giraitis et al. (1996) consider Kolmogorov-Smirnov and Cramér-von-Mises-type test statistics for the detection of a change in the marginal distribution of the random variables that underlie the observed data. Consistency of the corresponding change-point estimators is proved under the assumption that the jump height approaches 0.

In the following sections, we prove consistency of the estimators $\hat{k}_W$ and $\hat{k}_{SW}$, establish an optimal convergence rate for $\hat{k}_W$ and finally derive its asymptotic distribution. For $\hat{k}_{SW}$, neither the asymptotic distribution nor an optimal convergence rate is obtained. Although self-normalized statistics have advantages over non-self normalized statistics when testing for structural changes in time series, these benefits do not necessarily transfer to the corresponding change-point estimators. A change-point estimator based on a self-normalized CUSUM statistic has been applied in Shao (2011) to real data sets. Eventhough Shao assumes validity of using the estimator, the article does not cover a formal proof of consistency. Moreover, it has been noted by Shao and Zhang (2010) that, even under the assumption of short-range dependence, it seems difficult to derive the asymptotic distribution of the estimate.

2.1. Consistency

It has been noted in Section 1.2.2 that for local changes with shift height h_n, $n \in \mathbb{N}$, both Wilcoxon test statistics converge to the limits that are obtained under the assumption of stationarity if h_n decreases relatively fast, while they tend to ∞ if h_n converges to 0 sufficiently slow. For this reason, we cannot expect the change-point estimators $\hat{k}_W$ and $\hat{k}_{SW}$ to approach the true change-point location in the former case, but we may conclude that consistency holds in the latter case. The following proposition verifies this conjecture for subordinated Gaussian time series:

Proposition 1 (Betken (2017))**.** *Let X_n, $n \in \mathbb{N}$, denote a time series with a change in the mean in $k_0 = \lfloor n\tau \rfloor$, $0 < \tau < 1$, with shift height h_n. Suppose that X_n, $n \in \mathbb{N}$, satisfies Assumption 2. Then, as $n \to \infty$,*

$$\hat{\tau}_W := \frac{\hat{k}_W}{n} \xrightarrow{P} \tau, \qquad \hat{\tau}_{SW} := \frac{\hat{k}_{SW}}{n} \xrightarrow{P} \tau$$

if either

- $h_n = h$ *with* $h \neq 0$

or

- $\lim_{n\to\infty} h_n = 0$, $h_n^{-1} = o(g_{D,r}(n))$ *with D and r as in Assumption 2, and F has a bounded density f.*

In both cases, the test statistics

$$W_n := \max_{1\leq k\leq n-1} |W_{k,n}| \quad \text{and} \quad SW_n(\tau_1, \tau_2) := \max_{\lfloor n\tau_1\rfloor \leq k \leq \lfloor n\tau_2 \rfloor} |SW_{k,n}|$$

tend to ∞ in probability, implying consistency of the corresponding change-point tests.

Given the assumptions of Proposition 1, the results of Kokoszka and Leipus (1998) imply consistency of the CUSUM-based estimator $\hat{k}_{C,0}$ under fixed changes and under local changes with shift height h_n satisfying $h_n^{-1} = o(n^{rD/2})$. Hence, $\hat{k}_W$ and $\hat{k}_{C,0}$ are asymptotically unbiased estimators under essentially the same constraints on the convergence rate of the change-point height.

The proof of Proposition 1 is based on an application of the following lemma:

Lemma 1 (Betken (2017))**.** *Let X_n, $n \in \mathbb{N}$, denote a time series with a change in the mean in $k_0 = \lfloor n\tau \rfloor$, $0 < \tau < 1$, with shift height h_n. Suppose that X_n, $n \in \mathbb{N}$, satisfies Assumption 2. If $h_n^{-1} = o(g_{D,r}(n))$ with D and r as in Assumption 2,*

$$\frac{1}{n^2 h_n} \sum_{i=1}^{\lfloor nt \rfloor} \sum_{j=\lfloor nt \rfloor+1}^{n} \left(1_{\{X_i \leq X_j\}} - \frac{1}{2}\right) \xrightarrow{P} C\delta_\tau(t), \;\; 0 \leq t \leq 1,$$

where

$$C := \begin{cases} \frac{1}{h}\int_{\mathbb{R}} (F(x+h) - F(x))\, dF(x) & \text{if } h_n = h,\ h \neq 0, \\ \int_{\mathbb{R}} f^2(x)dx & \text{if } \lim_{n\to\infty} h_n = 0 \text{ and } F \text{ has a bounded density } f. \end{cases}$$

To simplify notation, we write $\int$ instead of $\int_{\mathbb{R}}$ for the proof of Lemma1 and all other proofs in this chapter.

Proof. First, consider the case $h_n = h$ with $h \neq 0$. For $\lfloor nt \rfloor \leq \lfloor n\tau \rfloor$, we have

$$\begin{aligned} &\frac{1}{n^2}\sum_{i=1}^{\lfloor nt \rfloor}\sum_{j=\lfloor nt \rfloor+1}^{n}\left(1_{\{X_i \leq X_j\}} - \frac{1}{2}\right) \\ =&\frac{1}{n^2}\sum_{i=1}^{\lfloor nt \rfloor}\sum_{j=\lfloor n\tau \rfloor+1}^{n}\left(1_{\{Y_i \leq Y_j + h\}} - \frac{1}{2}\right) + \frac{1}{n^2}\sum_{i=1}^{\lfloor nt \rfloor}\sum_{j=\lfloor nt \rfloor+1}^{\lfloor n\tau \rfloor}\left(1_{\{Y_i \leq Y_j\}} - \frac{1}{2}\right). \end{aligned}$$

By Lemma 1 in Betken (2016), the first summand on the right-hand side of the equation converges in probability to $t(1-\tau)\int (F(x+h) - F(x))\, dF(x)$, uniformly in $t \in [0, \tau]$. The second summand vanishes as n tends to ∞.
Whenever $\lfloor nt \rfloor > \lfloor n\tau \rfloor$,

$$\begin{aligned} &\frac{1}{n^2}\sum_{i=1}^{\lfloor nt \rfloor}\sum_{j=\lfloor nt \rfloor+1}^{n}\left(1_{\{X_i \leq X_j\}} - \frac{1}{2}\right) \\ =&\frac{1}{n^2}\sum_{i=1}^{\lfloor n\tau \rfloor}\sum_{j=\lfloor nt \rfloor+1}^{n}\left(1_{\{Y_i \leq Y_j + h\}} - \frac{1}{2}\right) + \frac{1}{n^2}\sum_{i=\lfloor n\tau \rfloor+1}^{\lfloor nt \rfloor}\sum_{j=\lfloor nt \rfloor+1}^{n}\left(1_{\{Y_i \leq Y_j\}} - \frac{1}{2}\right). \end{aligned}$$

In this case, the first summand on the right-hand side of the equation converges in probability to $(1-t)\tau\int (F(x+h) - F(x))\, dF(x)$, uniformly in $t \in [\tau, 1]$, while the second summand converges to 0 in probability. All in all, it follows that

$$\frac{1}{n^2}\sum_{i=1}^{\lfloor nt \rfloor}\sum_{j=\lfloor nt \rfloor+1}^{n}\left(1_{\{X_i \leq X_j\}} - \frac{1}{2}\right) \xrightarrow{P} \delta_\tau(t)\int (F(x+h) - F(x))\, dF(x)$$

uniformly in $t \in [0, 1]$.
If $\lim_{n\to\infty} h_n = 0$, the process

$$W_{n,r}(t) - \frac{n}{d_{n,r}}\delta_\tau(t)\int (F(x+h_n) - F(x))\, dF(x),\ 0 \leq t \leq 1,$$

converges in distribution to

$$\frac{1}{r!}\left(Z_{r,H}(t) - tZ_{r,H}(1)\right)\int J_r(x)dF(x),\ 0 \leq t \leq 1,$$

due to Theorem 4 in Chapter 1. Since $h_n^{-1} = o(g_{D,r}(n))$ by assumption, it follows that

$$\frac{1}{n^2 h_n} \sum_{i=1}^{\lfloor nt \rfloor} \sum_{j=\lfloor nt \rfloor+1}^{n} \left(1_{\{X_i \leq X_j\}} - \frac{1}{2} \right) \xrightarrow{P} \delta_\tau(t) \int f^2(x) dx, \ 0 \leq t \leq 1,$$

as $n \to \infty$. □

Proof of Proposition 1. According to Lemma 1, it holds that, under the assumptions of Proposition 1,

$$\frac{1}{n^2 h_n} \sum_{i=1}^{\lfloor nt \rfloor} \sum_{j=\lfloor nt \rfloor+1}^{n} \left(1_{\{X_i \leq X_j\}} - \frac{1}{2} \right) \xrightarrow{P} C\delta_\tau(t), \ 0 \leq t \leq 1, \tag{2.3}$$

as $n \to \infty$, where $\delta_\tau : [0,1] \longrightarrow \mathbb{R}$ is defined by

$$\delta_\tau(t) := \begin{cases} t(1-\tau) & \text{for } t \leq \tau \\ (1-t)\tau & \text{for } t \geq \tau \end{cases}$$

and C denotes some non-zero constant. Since $h_n^{-1} = o(g_{D,r}(n))$ by assumption, it directly follows that $\frac{1}{nd_{n,r}} \max_{1 \leq k \leq n-1} |W_{k,n}|$ tends to ∞ in probability.
To show consistency of the change-point estimator $\hat{\tau}_W$, define

$$Z_{n,\varepsilon} := \frac{1}{n^2 h_n} \max_{1 \leq k \leq \lfloor n\tau \rfloor} |W_{k,n}| - \frac{1}{n^2 h_n} \max_{1 \leq k \leq \lfloor n(\tau-\varepsilon) \rfloor} |W_{k,n}|$$

for $\varepsilon > 0$. Due to (2.3), $Z_{n,\varepsilon}$ converges in probability to $C(1-\tau)\varepsilon$. It follows that

$$\lim_{n\to\infty} P(\hat{k}_W < \lfloor n(\tau-\varepsilon) \rfloor) = \lim_{n\to\infty} P(Z_{n,\varepsilon} = 0) = 0.$$

An analogous argument yields $\lim_{n\to\infty} P(\hat{k}_W > \lfloor n(\tau+\varepsilon) \rfloor) = 0$.
All in all, it follows that for any $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left| \frac{\hat{k}_W}{n} - \tau \right| > \varepsilon \right) = 0.$$

In order to show that $\hat{\tau}_{SW}$ is a consistent estimator, we consider the process $SW_n(t)$, $0 \leq t \leq 1$, defined by (1.10). According to Betken (2016), the limit of the self-normalized Wilcoxon test statistic can be obtained by an application of the continuous mapping theorem to the (suitably standardized) process $W_n(t)$, $0 \leq t \leq 1$. Therefore, it follows by the corresponding argument in Betken (2016) that $SW_n(t)$ converges in probability to

$$\left\{ \int_0^t \left(\delta_\tau(s) - \frac{s}{t}\delta_\tau(t) \right)^2 ds + \int_t^1 \left(\delta_\tau(s) - \frac{1-s}{1-t}\delta_\tau(t) \right)^2 ds \right\}^{-\frac{1}{2}} |\delta_\tau(t)|$$

uniformly in $t \in [0,1]$. As a result, elementary calculations yield

$$\max_{\lfloor n\tau_1 \rfloor \le k \le k_0 - n\varepsilon} SW_{k,n} \xrightarrow{P} \sup_{t \in [\tau_1, \tau - \varepsilon]} \frac{\sqrt{3}t\sqrt{1-t}}{(\tau - t)},$$

$$\max_{k_0 + n\varepsilon \le k \le \lfloor n\tau_2 \rfloor} SW_{k,n} \xrightarrow{P} \sup_{t \in [\tau + \varepsilon, \tau_2]} \frac{\sqrt{3}\sqrt{t}(1-t)}{(\tau - t)}.$$

As $SW_{k_0,n}$ tends to ∞ in probability due to Theorem 2 in Betken (2016), it is possible to conclude that $P(\hat{k}_{SW} > k_0 + n\varepsilon)$ and $P(\hat{k}_{SW} < k_0 - n\varepsilon)$ converge to 0 in probability. This proves consistency of the change-point estimator $\hat{\tau}_{SW}$. □

The boxplots in Figure 2.2 illustrate the consistency of the change-point estimators $\hat{\tau}_W$ and $\hat{\tau}_{SW}$ for two different values of H: the median of the estimated values approaches the true value of the change-point location while the interquartile range and the length of boxplot whiskers diminish as the sample size increases. Obviously, a higher correlation within a sample of observations, characterized by the larger value of H, goes along with a higher variability of the estimated values. Moreover, it can be observed that the change-point estimator $\hat{\tau}_{SW}$ yields slightly better results than the estimator $\hat{\tau}_W$.

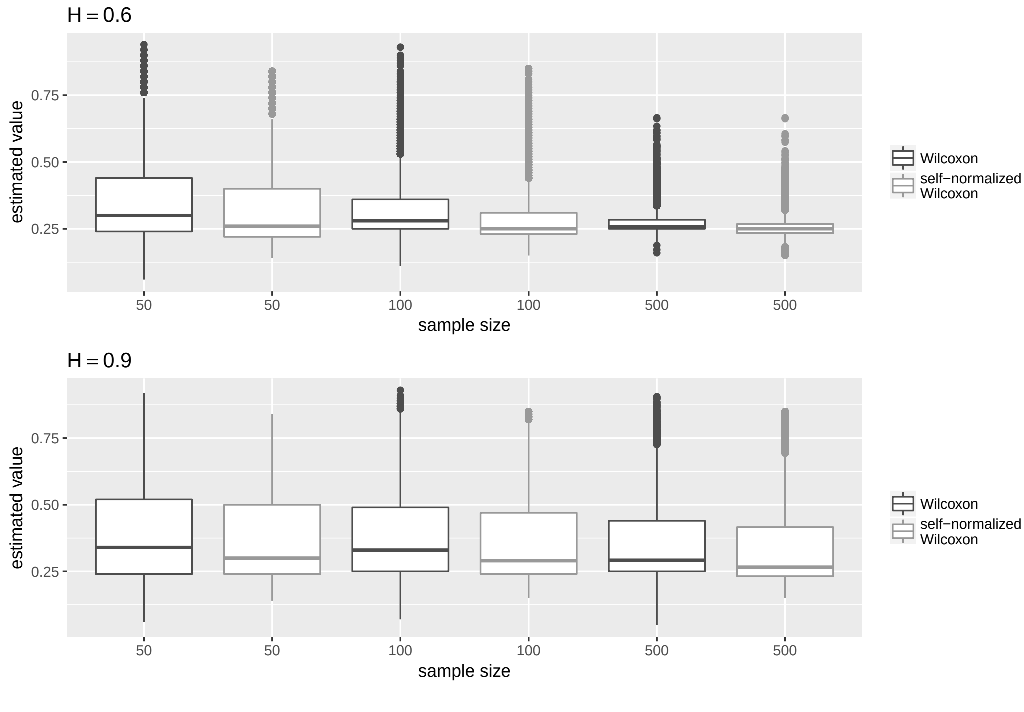

Figure 2.2.: *Boxplots of the estimators* $\hat{\tau}_W$ *and* $\hat{\tau}_{SW}$ *on the basis of* 5000 *simulated fractional Gaussian noise time series with Hurst parameter* H *and a change in the mean of height* $h = 1$ *after a proportion* $\tau = 0.25$.

2.2. Convergence rate

Having shown the consistency of the change-point estimator $\hat{\tau}_W$, we are interested in deriving its rate of convergence. Figure 2.2 suggests that the rate depends on the value of the parameter H, i.e. on the intensity of dependence in the data. The following theorem confirms that under local changes the rate of convergence is determined by the height of the level shift as well as the covariance structure of the data-generating process. In particular, it follows that the smaller the correlation, i.e. the higher the value of D, the faster the convergence of the estimator.

Theorem 5 (Betken (2017)). *Let X_n, $n \in \mathbb{N}$, denote a time series with a change in the mean in $k_0 = \lfloor n\tau \rfloor$, $0 < \tau < 1$, with shift height h_n. Suppose that X_n, $n \in \mathbb{N}$, satisfies Assumption 2. Then, as $n \to \infty$,*

$$\left|\hat{k}_W - k_0\right| = \mathcal{O}_P\left(g^{-}_{D,r}(h_n^{-1})\right)$$

if either

- $h_n = h$ *with* $h \neq 0$

or

- $\lim_{n\to\infty} h_n = 0$, $h_n^{-1} = o(g_{D,r}(n))$ *with D and r as in Assumption 2, and F has a bounded density f.*

Remark 8. For $h_n^{-1} = o(g_{D,r}(n))$ and $m_n = m_{n,D,r} := g^{-}_{D,r}(h_n^{-1})$,

1. $m_n \longrightarrow \infty$, as $n \longrightarrow \infty$,
2. $\frac{m_n}{n} \longrightarrow 0$, as $n \longrightarrow \infty$,
3. $m_n \sim \frac{d_{m_n,r}}{h_n}$, as $n \longrightarrow \infty$.

The first and second relation show that the convergence rate that has been achieved in proving consistency of the estimator is improved by Theorem 5. The third relation confirms that the rate of convergence depends on the intensity of dependence in the data and the change-point height.

Under the additional assumption that the considered data is generated by Gaussian processes, Horváth and Kokoszka (1997) derive the same convergence rate for the CUSUM-based change-point estimator $\hat{k}_{C,0}$. It is shown by Ben Hariz and Wylie (2005) that under local changes this convergence rate, derived under the assumption of Gaussianity, can also be established under general, non-restrictive conditions on the data-generating sequences. In addition, these authors note that the convergence rate, obtained for the CUSUM-based change-point estimator under long-range dependence, is the same as in the case of independent or short-range dependent data, and, in particular does not depend on the value of the LRD parameter D if the change-point height is fixed. This result corresponds to the convergence rate derived by Lavielle and Moulines (2000) for the least-squares estimate of the change-point location and is confirmed by Theorem 5 since, under fixed changes, $|\hat{k}_W - k_0| = \mathcal{O}_P(1)$. An explanation for this phenomenon might be the occurrence of two opposing effects associated with the behavior of long-range dependent processes (see Ben Hariz and Wylie (2005)): long-range dependence usually leads to a higher variance of observations, i.e. the vertical fluctuations become bigger, which is, in the considered case, reflected by the need of a stronger standardization and a slower convergence of the statistic $W_{k,n}$. For this reason, one may expect estimation to be more difficult if the correlation between observations is high. At the same time, the behavior of the increments of $W_{k,n}$, $k \in \{1, \ldots, n-1\}$, becomes more regular when the correlation increases, meaning that the horizontal fluctuations become smaller, therefore making estimation seem easier.

Since fractional Brownian motion processes with parameter $H \in \left(\frac{1}{1}, 1\right)$ typically characterize the asymptotic distribution in limit theorems for long-range dependent time series, these effects are also reflected by properties of the fractional Brownian motion B_H: as $\operatorname{Var} B_H(t) = |t|^{2H}$, increasing values of the Hurst parameter H (corresponding to a higher intensity of dependence) go along with a higher variance. Moreover, the trajectories of a fractional Brownian motion are Hölder continuous of any order strictly smaller than H, so that the sample paths become smoother when H increases. The influence of vertical fluctuations is more pronounced for relatively small values of the change-point height h_n, while for a fixed height the effect of a higher variability is canceled out by the effects of a more regular local path behavior.

This observation is illustrated by simulations of the mean absolute error, defined by

$$\mathrm{MAE} := \frac{1}{m}\sum_{i=1}^{m}\left|\hat{k}_{W,i} - k_0\right|$$

for estimates $\hat{k}_{W,i}$, $i = 1, \ldots, m$, computed on the basis of $m = 5000$ different sequences of fractional Gaussian noise time series; see Figure 2.3.

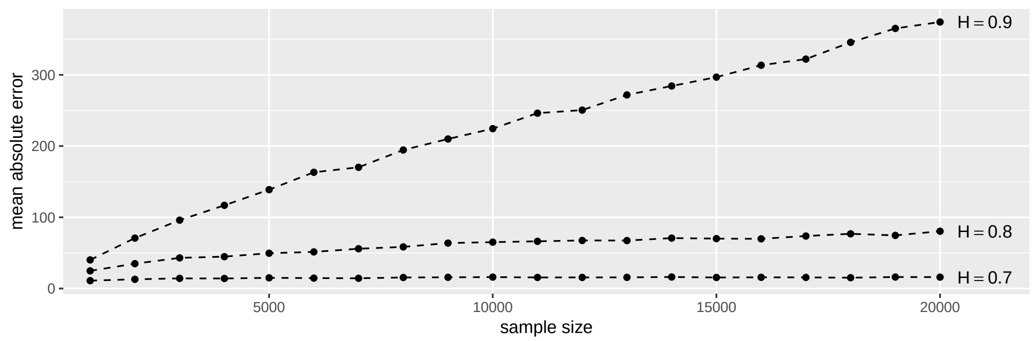

Figure 2.3.: *Mean absolute error of $\hat{k}_W$ computed on the basis of 5000 sequences of fractional Gaussian noise time series for different values of H and a shift in the mean of height $h = 0.5$ after a proportion $\tau = 0.5$.*

Since $\hat{k}_W - k_0 = \mathcal{O}_P(1)$ due to Theorem 5, we expect the mean absolute error to approach a constant as $n \to \infty$. This can be clearly seen in Figure 2.3 for $H \in \{0.7, 0.8\}$. For strongly correlated data, characterized by $H = 0.9$, the convergence seems to be rather slow, though.

The proof of Theorem 5 requires the following result:

Lemma 2 (Betken (2017)). *Let $Y_n = G(\xi_n)$, $n \in \mathbb{N}$, be a mean-zero subordinated Gaussian sequence according to Model 1 and let h_n, $n \in \mathbb{N}$, be a sequence of real numbers with* $\lim_{n\to\infty} h_n = h$.

1. *As $n \to \infty$, the process*

$$\frac{1}{nd_{n,r}} \sum_{i=1}^{\lfloor nt \rfloor} \sum_{j=\lfloor n\tau \rfloor+1}^{n} \left(1_{\{Y_i \leq Y_j + h_n\}} - \int_{\mathbb{R}} F(x+h_n) dF(x) \right)$$

converges in distribution to

$$(1-\tau)\frac{1}{r!} Z_{r,H}(t) \int_{\mathbb{R}} J_r(x+h) dF(x) - t\frac{1}{r!} \left(Z_{r,H}(1) - Z_{r,H}(\tau) \right) \int_{\mathbb{R}} J_r(x) dF(x+h)$$

uniformly in $t \in [0, \tau]$.

2. *As $n \to \infty$, the process*

$$\frac{1}{nd_{n,r}} \sum_{i=1}^{\lfloor n\tau \rfloor} \sum_{j=\lfloor nt \rfloor+1}^{n} \left(1_{\{Y_i \leq Y_j + h_n\}} - \int_{\mathbb{R}} F(x+h_n) dF(x) \right)$$

converges in distribution to

$$(1-t)\frac{1}{r!} Z_{r,H}(\tau) \int_{\mathbb{R}} J_r(x+h) dF(x) - \tau\frac{1}{r!} \left(Z_{r,H}(1) - Z_{r,H}(t) \right) \int_{\mathbb{R}} J_r(x) dF(x+h)$$

uniformly in $t \in [\tau, 1]$.

Proof. A proof is given for the first assertion only, as the second assertion can be derived by an analogous argument. The line of argument follows the proof of Theorem 1.1 in Dehling et al. (2013).
Let F_k and $F_{k+1,n}$ denote the empirical distribution functions of the first k and last $n-k$ realizations of $Y_1, \ldots, Y_n$, i.e.

$$F_k(x) := \frac{1}{k} \sum_{i=1}^{k} 1_{\{Y_i \leq x\}},$$

$$F_{k+1,n}(x) := \frac{1}{n-k} \sum_{i=k+1}^{n} 1_{\{Y_i \leq x\}}.$$

Given this notation, it follows that

$$\sum_{i=1}^{\lfloor nt\rfloor}\sum_{j=\lfloor n\tau\rfloor+1}^{n} 1_{\{Y_i\leq Y_j+h_n\}} = (n-\lfloor n\tau\rfloor)\lfloor nt\rfloor \int F_{\lfloor nt\rfloor}(x+h_n)dF_{\lfloor n\tau\rfloor+1,n}(x)$$

for $t\leq\tau$. This yields the following decomposition:

$$\begin{aligned}
&\frac{1}{nd_{n,r}}\sum_{i=1}^{\lfloor nt\rfloor}\sum_{j=\lfloor n\tau\rfloor+1}^{n}\left(1_{\{Y_i\leq Y_j+h_n\}}-\int F(x+h_n)dF(x)\right) \qquad (2.4)\\
=&\frac{n-\lfloor n\tau\rfloor}{n}d_{n,r}^{-1}\lfloor nt\rfloor\int\left(F_{\lfloor nt\rfloor}(x+h_n)-F(x+h_n)\right)dF_{\lfloor n\tau\rfloor+1,n}(x)\\
+&\frac{n-\lfloor n\tau\rfloor}{n}d_{n,r}^{-1}\lfloor nt\rfloor\int F(x+h_n)d\left(F_{\lfloor n\tau\rfloor+1,n}-F\right)(x).
\end{aligned}$$

For the first term of the sum on the right-hand side, we have

$$\begin{aligned}
&\sup_{t\in[0,\tau]}\Big|d_{n,r}^{-1}\lfloor nt\rfloor\int\left(F_{\lfloor nt\rfloor}(x+h_n)-F(x+h_n)\right)dF_{\lfloor n\tau\rfloor+1,n}(x)\\
&\qquad-\frac{1}{r!}Z_{r,H}(t)\int J_r(x+h)dF(x)\Big|\\
\leq&\sup_{t\in[0,\tau]}\Big|\int d_{n,r}^{-1}\lfloor nt\rfloor\left(F_{\lfloor nt\rfloor}(x+h_n)-F(x+h_n)\right)\\
&\qquad-\frac{1}{r!}Z_{r,H}(t)J_r(x+h_n)dF_{\lfloor n\tau\rfloor+1,n}(x)\Big|\\
&+\frac{1}{r!}\sup_{t\in[0,\tau]}|Z_{r,H}(t)|\left|\int\left(J_r(x+h_n)-J_r(x+h)\right)dF_{\lfloor n\tau\rfloor+1,n}(x)\right|\\
&+\frac{1}{r!}\sup_{t\in[0,\tau]}|Z_{r,H}(t)|\left|\int J_r(x+h)d\left(F_{\lfloor n\tau\rfloor+1,n}-F\right)(x)\right|.
\end{aligned}$$

In the following, it is shown that each of the summands on the right-hand side of the above inequality converges to 0.
The first summand converges to 0 because of the empirical process non-central limit theorem in Dehling and Taqqu (1989). In order to show that the second and third summand vanish as well, note that $\sup_{t\in[0,\tau]}|Z_{r,H}(t)|<\infty$ almost surely, since the sample paths of Hermite processes are almost surely continuous; see Mikosch (1998). Furthermore, we have

$$\begin{aligned}
\int J_r(x+h)dF_{\lfloor n\tau\rfloor+1,n}(x)&=-\int\int 1_{\{x+h\leq G(y)\}}H_r(y)\varphi(y)dydF_{\lfloor n\tau\rfloor+1,n}(x)\\
&=-\int\int 1_{\{x\leq G(y)-h\}}dF_{\lfloor n\tau\rfloor+1,n}(x)H_r(y)\varphi(y)dy\\
&=-\int F_{\lfloor n\tau\rfloor+1,n}(G(y)-h)H_r(y)\varphi(y)dy.
\end{aligned}$$

Analogously, it follows that

$$\int J_r(x+h_n)dF_{\lfloor n\tau\rfloor+1,n}(x) = -\int F_{\lfloor n\tau\rfloor+1,n}(G(y)-h_n)H_r(y)\varphi(y)dy.$$

Therefore, we may conclude that

$$\begin{aligned}
&\left|\int \left(J_r(x+h_n)-J_r(x+h)\right)dF_{\lfloor n\tau\rfloor+1,n}(x)\right| \\
&\leq 2\sup_{x\in\mathbb{R}}\left|F_{\lfloor n\tau\rfloor+1,n}(x)-F(x)\right|\int |H_r(y)|\,\varphi(y)dy \\
&\qquad\qquad + \int |F(G(y)-h_n)-F(G(y)-h)|\,|H_r(y)|\,\varphi(y)dy.
\end{aligned}$$

Since $\int |H_r(y)|\,\varphi(y)dy<\infty$, the first expression on the right-hand side converges to 0 by the Glivenko-Cantelli theorem. The second expression converges to 0 due to continuity of F and the dominated convergence theorem.
To show convergence of the third summand, note that

$$\begin{aligned}
&\left|\int J_r(x+h)d\left(F_{\lfloor n\tau\rfloor+1,n}(x)-F(x)\right)\right| \\
&= \frac{1}{n-\lfloor n\tau\rfloor}\left|\sum_{i=\lfloor n\tau\rfloor+1}^{n}\left(J_r(Y_i+h)-\operatorname{E}J_r(Y_i+h)\right)\right| \\
&\leq \frac{n}{n-\lfloor n\tau\rfloor}\frac{1}{n}\left|\sum_{i=1}^{n}\left(J_r(Y_i+h)-\operatorname{E}J_r(Y_i+h)\right)\right| \\
&+ \frac{\lfloor n\tau\rfloor}{n-\lfloor n\tau\rfloor}\frac{1}{\lfloor n\tau\rfloor}\left|\sum_{i=1}^{\lfloor n\tau\rfloor}\left(J_r(Y_i+h)-\operatorname{E}J_r(Y_i+h)\right)\right|.
\end{aligned}$$

For both summands on the right-hand side of the above inequality, the ergodic theorem implies almost sure convergence to 0.
For the second summand on the right-hand side of (2.4), we have

$$\begin{aligned}
&\frac{n-\lfloor n\tau\rfloor}{n}d_{n,r}^{-1}\lfloor nt\rfloor\int F(x+h_n)d\left(F_{\lfloor n\tau\rfloor+1,n}-F\right)(x) \\
&= -\frac{\lfloor nt\rfloor}{n}d_{n,r}^{-1}(n-\lfloor n\tau\rfloor)\int\left(F_{\lfloor n\tau\rfloor+1,n}(x)-F(x)\right)dF(x+h_n).
\end{aligned}$$

Since $\frac{\lfloor nt\rfloor}{n}\longrightarrow t$ uniformly in t, consider

$$\begin{aligned}
&\Bigg|d_{n,r}^{-1}(n-\lfloor n\tau\rfloor)\int\left(F_{\lfloor n\tau\rfloor+1,n}(x)-F(x)\right)dF(x+h_n) \\
&\qquad -\frac{1}{r!}(Z_{r,H}(1)-Z_{r,H}(\tau))\int J_r(x)dF(x+h_n)\Bigg|.
\end{aligned}$$

By an application of the triangular inequality, it follows that this expression is bounded from above by

$$\begin{aligned}
&\left|\int d_{n,r}^{-1} n \left(F_n(x) - F(x)\right) - \frac{1}{r!} Z_{r,H}(1) J_r(x) dF(x+h)\right| \\
&+ \left|\int d_{n,r}^{-1} \lfloor n\tau \rfloor \left(F_{\lfloor n\tau \rfloor}(x) - F(x)\right) - \frac{1}{r!} Z_{r,H}(\tau) J_r(x) dF(x+h_n)\right| \\
&+ \frac{1}{r!} \left|Z_{r,H}(1) - Z_{r,H}(\tau)\right| \left|\int J_r(x) d\left(F(x+h_n) - F(x+h)\right)\right|.
\end{aligned}$$

The first and second of the above summands converge to 0 because of the empirical process non-central limit theorem. For the third summand, we have

$$\left|\int J_r(x) d\left(F(x+h_n) - F(x+h)\right)\right| = \left|\int \left(J_r(x-h_n) - J_r(x-h)\right) dF(x)\right|.$$

As shown before in this proof, convergence to 0 follows by the Glivenko-Cantelli theorem and the dominated convergence theorem. □

Proof of Theorem 5. For the proof of Theorem 5, we write $\hat{k}$ instead of $\hat{k}_W$ and $W_n(k)$ instead of $W_{k,n}$. We assume that $h > 0$ under fixed changes and that for some $n_0 \in \mathbb{N}$ $h_n > 0$ for all $n \geq n_0$ under local changes. Furthermore, we subsume fixed and local changes under the general assumption that $\lim_{n\to\infty} h_n = h$ (under fixed changes $h_n = h$ for all $n \in \mathbb{N}$, under local changes $h = 0$).

In order to prove Theorem 5, we need to show that for all $\varepsilon > 0$ there exists an $n_\varepsilon \in \mathbb{N}$ and an $M > 0$ such that

$$P\left(\left|\hat{k} - k_0\right| > M m_n\right) < \varepsilon$$

for all $n \geq n_\varepsilon$.

For $M \in \mathbb{R}^+$, define $D_{n,M} := \{k \in \{1, \ldots, n-1\} \mid |k - k_0| > M m_n\}$, so that

$$P\left(\left|\hat{k} - k_0\right| > M m_n\right) \leq P\left(\sup_{k \in D_{n,M}} |W_n(k)| \geq |W_n(k_0)|\right) \leq P_1 + P_2$$

with

$$\begin{aligned}
P_1 &:= P\left(\sup_{k \in D_{n,M}} \left(W_n(k) - W_n(k_0)\right) \geq 0\right), \\
P_2 &:= P\left(\sup_{k \in D_{n,M}} \left(-W_n(k) - W_n(k_0)\right) \geq 0\right).
\end{aligned}$$

Note that $D_{n,M} = D_{n,M,1} \cup D_{n,M,2}$, where

$$D_{n,M,1} := \{k \in \{1, \ldots, n-1\} \mid k_0 - k > Mm_n\},$$
$$D_{n,M,2} := \{k \in \{1, \ldots, n-1\} \mid k - k_0 > Mm_n\}.$$

Therefore, $P_2 \leq P_{2,1} + P_{2,2}$, where

$$P_{2,1} := P\left(\sup_{k \in D_{n,M,1}} (-W_n(k) - W_n(k_0)) \geq 0\right),$$
$$P_{2,2} := P\left(\sup_{k \in D_{n,M,2}} (-W_n(k) - W_n(k_0)) \geq 0\right).$$

In the following, we will consider the first summand only, since for the second summand analogous implications result from the same argument.
We define

$$\widehat{W}_n(k) := \delta_n(k)\Delta(h_n),$$

where

$$\delta_n(k) := \begin{cases} k(n-k_0) & \text{for } k \leq k_0 \\ k_0(n-k) & \text{for } k > k_0 \end{cases} \quad \text{and} \quad \Delta(h_n) := \int (F(x+h_n) - F(x))\, dF(x).$$

Note that

$$P_{2,1} \leq P\left(\sup_{k \in D_{n,M,1}} \left(\widehat{W}_n(k) - W_n(k) + \widehat{W}_n(k_0) - W_n(k_0)\right) \geq \widehat{W}_n(k_0)\right)$$
$$\leq P\left(2 \sup_{t \in [0,\tau]} \left|W_n(\lfloor nt \rfloor) - \widehat{W}_n(\lfloor nt \rfloor)\right| \geq k_0(n-k_0)\Delta(h_n)\right).$$

We have

$$\sup_{t \in [0,\tau]} \left|W_n(\lfloor nt \rfloor) - \widehat{W}_n(\lfloor nt \rfloor)\right|$$
$$= \sup_{t \in [0,\tau]} \left| \sum_{i=1}^{\lfloor nt \rfloor} \sum_{j=\lfloor n\tau \rfloor + 1}^{n} \left(1_{\{Y_i \leq Y_j + h_n\}} - \int F(x+h_n) dF(x)\right) + \sum_{i=1}^{\lfloor nt \rfloor} \sum_{j=\lfloor nt \rfloor + 1}^{\lfloor n\tau \rfloor} \left(1_{\{Y_i \leq Y_j\}} - \frac{1}{2}\right) \right|.$$

Due to Lemma 2 and Theorem 1.1 in Dehling et al. (2013),

$$2 \sup_{t \in [0,\tau]} \left|W_n(\lfloor nt \rfloor) - \widehat{W}_n(\lfloor nt \rfloor)\right| = \mathcal{O}_P(nd_{n,r}),$$

i.e. for all $\varepsilon > 0$ there exists a $K > 0$ such that

$$P\left(2\sup_{t\in[0,\tau]}\left|W_n(\lfloor nt\rfloor) - \widehat{W}_n(\lfloor nt\rfloor)\right| \geq Knd_{n,r}\right) < \varepsilon$$

for all n. Furthermore, $k_0(n-k_0)\Delta(h_n) \sim Cn^2h_n$ for some constant C. Note that $Knd_{n,r} \leq k_0(n-k_0)\Delta(h_n)$ if and only if

$$K \leq \frac{k_0}{n}\frac{n-k_0}{n}\frac{\Delta(h_n)}{h_n}\frac{nh_n}{d_{n,r}}.$$

The right-hand side of the inequality diverges if $h_n = h$ is fixed or if $h_n^{-1} = o(g_{D,r}(n))$. Therefore, it is possible to find an $n_\varepsilon \in \mathbb{N}$ such that

$$P_{2,1} \leq P\left(2\sup_{t\in[0,\tau]}\left|W_n(\lfloor nt\rfloor) - \widehat{W}_n(\lfloor nt\rfloor)\right| \geq Knd_{n,r}\right) < \varepsilon$$

for all $n \geq n_\varepsilon$.
We will now turn to the summand P_1. Note that $P_1 \leq P_{1,1} + P_{1,2}$, where

$$P_{1,1} := P\left(\max_{k\in D_{n,M,1}} W_n(k) - W_n(k_0) \geq 0\right),$$
$$P_{1,2} := P\left(\max_{k\in D_{n,M,2}} W_n(k) - W_n(k_0) \geq 0\right).$$

In the following, we will consider the first summand only, since for the second summand analogous implications result from the same argument. We define a random sequence κ_n, $n \in \mathbb{N}$, by choosing $\kappa_n \in D_{n,M,1}$ such that

$$\max_{k\in D_{n,M,1}}\left(W_n(k) - \widehat{W}_n(k) + \widehat{W}_n(k_0) - W_n(k_0)\right)$$
$$= W_n(\kappa_n) - \widehat{W}_n(\kappa_n) + \widehat{W}_n(k_0) - W_n(k_0).$$

Note that for any sequence k_n, $n \in \mathbb{N}$, with $k_n \in D_{n,M,1}$

$$\widehat{W}_n(k_0) - \widehat{W}_n(k_n) = (n-k_0)l_n\Delta(h_n),$$

where $l_n := k_0 - k_n$. Since $\kappa_n \in D_{n,M,1}$ and since m_n tends to ∞, we have

$$l_n d_{l_n,r}^{-1} = l_n^{1-H}L_\gamma^{-\frac{r}{2}}(l_n) \geq (Mm_n)^{1-H}L_\gamma^{-\frac{r}{2}}(Mm_n)$$

for n sufficiently large. Thus, it follows that

$$\frac{1}{nd_{l_n,r}}\left(\widehat{W}_n(k_0) - \widehat{W}_n(\kappa_n)\right) \geq \frac{n-k_0}{n}\frac{m_n}{d_{m_n,r}}M^{1-H}\frac{L_\gamma^{\frac{r}{2}}(m_n)}{L_\gamma^{\frac{r}{2}}(Mm_n)}\Delta(h_n).$$

If h_n is fixed, the right-hand side of the inequality diverges. Under local changes, the right-hand side asymptotically behaves like

$$(1-\tau)M^{1-H}\int f^2(x)dx,$$

since, in this case, $h_n \sim \frac{d_{m_n,r}}{m_n}$ due to the assumptions of Theorem 5. In the latter case, for any $\delta > 0$, it is possible to find an $n_\delta \in \mathbb{N}$ such that

$$\frac{1}{nd_{l_n,r}}\left(\widehat{W}_n(k_0) - \widehat{W}_n(k_n)\right) \geq M^{1-H}(1-\tau)\int f^2(x)dx - \delta$$

for all $n \geq n_\delta$.
All in all, the previous considerations show that there exists an $n_0 \in \mathbb{N}$ and a constant K such that for all $n \geq n_0$

$$P_{1,1} \leq P\left(\max_{k\in D_{n,M,1}} \frac{1}{nd_{k_0-k,r}}\left(W_n(k) - \widehat{W}_n(k) + \widehat{W}_n(k_0) - W_n(k_0)\right) \geq b(M)\right),$$

where $b(M) := KM^{1-H} - \delta$ with $\delta > 0$ fixed.

Elementary calculations show that for $k \leq k_0$

$$W_n(k) - \widehat{W}_n(k) + \widehat{W}_n(k_0) - W_n(k_0) = A_{n,1}(k) + A_{n,2}(k) + A_{n,3}(k) + A_{n,4}(k),$$

where

$$\begin{aligned}
A_{n,1}(k) &:= -(n-k_0)(k_0-k)\int \left(F_{k+1,k_0}(x+h_n) - F(x+h_n)\right) dF_{k_0+1,n}(x),\\
A_{n,2}(k) &:= -(n-k_0)(k_0-k)\int \left(F_{k_0+1,n}(x) - F(x)\right) dF(x+h_n),\\
A_{n,3}(k) &:= (k_0-k)k\int \left(F_k(x) - F(x)\right) dF_{k+1,k_0}(x),\\
A_{n,4}(k) &:= -k(k_0-k)\int \left(F_{k+1,k_0}(x) - F(x)\right) dF(x).
\end{aligned}$$

Thus, for $n \geq n_0$

$$\begin{aligned}
P_{1,1} \leq& P\left(\max_{k\in D_{n,M,1}} \frac{1}{nd_{k_0-k,r}}\sum_{i=1}^{4}|A_{n,i}(k)| \geq b(M)\right)\\
\leq& \sum_{i=1}^{4} P\left(\max_{k\in D_{n,M,1}} \frac{1}{nd_{k_0-k,r}}|A_{n,i}(k)| \geq \frac{1}{4}b(M)\right).
\end{aligned}$$

In the following, it is shown that for each $i \in \{1,2,3,4\}$

$$P\left(\max_{k\in D_{n,M,1}} \frac{1}{nd_{k_0-k,r}}|A_{n,i}(k)| \geq \frac{1}{4}b(M)\right) < \frac{\varepsilon}{4}$$

for n and M sufficiently large.

1. Note that

$$
\begin{aligned}
&\max_{k\in D_{n,M,1}} \frac{1}{nd_{k_0-k,r}} |A_{n,1}(k)| \\
\leq &\max_{k\in D_{n,M,1}} \sup_{x\in\mathbb{R}} \left| d_{k_0-k,r}^{-1}(k_0-k)\left(F_{k+1,k_0}(x)-F(x)\right)\right|.
\end{aligned}
$$

Due to stationarity,

$$
\begin{aligned}
&\max_{k\in D_{n,M,1}} \sup_{x\in\mathbb{R}} \left| d_{k_0-k,r}^{-1}(k_0-k)\left(F_{k+1,k_0}(x)-F(x)\right)\right| \\
\overset{\mathcal{D}}{=} &\max_{k\in D_{n,M,1}} \sup_{x\in\mathbb{R}} \left| d_{k_0-k,r}^{-1}(k_0-k)\left(F_{k_0-k}(x)-F(x)\right)\right|.
\end{aligned}
$$

The triangle inequality yields

$$
\begin{aligned}
&\max_{k\in D_{n,M,1}} \sup_{x\in\mathbb{R}} \left| d_{k_0-k,r}^{-1}(k_0-k)\left(F_{k_0-k}(x)-F(x)\right)\right| \\
\leq &\max_{k\in D_{n,M,1}} \sup_{x\in\mathbb{R}} \left| d_{k_0-k,r}^{-1}(k_0-k)\left(F_{k_0-k}(x)-F(x)\right) - \frac{1}{r!}Z_{r,H}(1)J_r(x)\right| \\
&+\frac{1}{r!}\left|Z_{r,H}(1)\right| \sup_{x\in\mathbb{R}} \left|J_r(x)\right|.
\end{aligned}
$$

Since

$$
\sup_{x\in\mathbb{R}} \left| d_{n,r}^{-1} n\left(F_n(x)-F(x)\right) - \frac{1}{r!}Z_{r,H}(1)J_r(x)\right| \xrightarrow{\text{a.s.}} 0, \text{ as } n\to\infty,
$$

and as $k_0 - k \geq Mm_n$ with m_n approaching ∞, it follows that

$$
\max_{k\in D_{n,M,1}} \sup_{x\in\mathbb{R}} \left| d_{k_0-k,r}^{-1}(k_0-k)\left(F_{k_0-k}(x)-F(x)\right) - \frac{1}{r!}Z_{r,H}(1)J_r(x)\right|
$$

converges to 0 almost surely. Therefore,

$$
\begin{aligned}
&P\left(\max_{k\in D_{n,M,1}} \frac{1}{nd_{k_0-k,r}} |A_{n,1}(k)| \geq \frac{1}{4}b(M)\right) \\
\leq &P\left(\max_{k\in D_{n,M,1}} \sup_{x\in\mathbb{R}} \left| d_{k_0-k,r}^{-1}(k_0-k)\left(F_{k+1,k_0}(x)-F(x)\right)\right| \geq \frac{1}{4}b(M)\right) \\
\leq &P\left(\frac{1}{r!}\left|Z_{r,H}(1)\right| \sup_{x\in\mathbb{R}} \left|J_r(x)\right| \geq \frac{1}{4}b(M)\right) + \frac{\varepsilon}{8}
\end{aligned}
$$

for n sufficiently large.

Furthermore, it is well-known that all moments of Hermite processes are finite; see Pipiras and Taqqu (2017). As a result, $\sup_{x\in\mathbb{R}} |J_r(x)| < \infty$. It therefore follows by Markov's inequality that for some $M_\varepsilon \in \mathbb{R}$

$$
P\left(\frac{1}{r!}\left|Z_{r,H}(1)\right| \sup_{x\in\mathbb{R}} \left|J_r(x)\right| \geq \frac{1}{4}b(M)\right) \leq \mathrm{E}\left|Z_{r,H}(1)\right| \frac{4\sup_{x\in\mathbb{R}}|J_r(x)|}{b(M)r!} < \frac{\varepsilon}{8}
$$

for all $M \geq M_\varepsilon$.

2. We have

$$\max_{k\in D_{n,M,1}} \frac{1}{nd_{k_0-k,r}} |A_{n,2}(k)| \leq \left| d_{n,r}^{-1}(n-k_0) \int (F_{k_0+1,n}(x) - F(x)) \, dF(x+h_n) \right|$$

for n sufficiently large. As a result,

$$\max_{k\in D_{n,M,1}} \frac{1}{nd_{k_0-k,r}} |A_{n,2}(k)| \leq \sup_{x\in\mathbb{R}} \left| d_{n,r}^{-1}(n-k_0) \left(F_{k_0+1,n}(x) - F(x)\right) \right| .$$

Due to the empirical process limit theorem of Dehling and Taqqu (1989),

$$\sup_{x\in\mathbb{R}} \left| d_{n,r}^{-1}(n-k_0) \left(F_{k_0+1,n}(x) - F(x)\right) \right| \xrightarrow{\mathcal{D}} \frac{1}{r!} |Z_{r,H}(1) - Z_{r,H}(\tau)| \sup_{x\in\mathbb{R}} |J_r(x)| .$$

Moreover,

$$\frac{1}{r!} |Z_{r,H}(1) - Z_{r,H}(\tau)| \sup_{x\in\mathbb{R}} |J_r(x)| \stackrel{\mathcal{D}}{=} \frac{1}{r!} (1-\tau)^H |Z_{r,H}(1)| \sup_{x\in\mathbb{R}} |J_r(x)|$$

since $Z_{r,H}$ is an H-self-similar process with stationary increments. Thus, we have

$$\begin{aligned} &P\left(\max_{k\in D_{n,M,1}} \frac{1}{nd_{k_0-k,r}} |A_{n,2}(k)| \geq \frac{1}{4} b(M) \right) \\ \leq &P\left(\frac{1}{r!} (1-\tau)^H |Z_{r,H}(1)| \sup_{x\in\mathbb{R}} |J_r(x)| \geq \frac{1}{4} b(M) \right) + \frac{\varepsilon}{8} \end{aligned}$$

for n sufficiently large. Again, it follows by Markov's inequality that

$$P\left(\frac{1}{r!} (1-\tau)^H |Z_{r,H}(1)| \sup_{x\in\mathbb{R}} |J_r(x)| \geq \frac{1}{4} b(M) \right) < \frac{\varepsilon}{8}$$

for M sufficiently large.

3. Note that

$$\frac{1}{nd_{k_0-k,r}} |A_{n,3}(k)| \leq \left| d_{n,r}^{-1} k \int (F_k(x) - F(x)) \, dF_{k+1,k_0}(x) \right|$$

for n sufficiently large. Therefore,

$$\max_{k\in D_{n,M,1}} \frac{1}{nd_{k_0-k,r}} |A_{n,3}(k)| \leq \sup_{x\in\mathbb{R}, t\in[0,1]} \left| d_{n,r}^{-1} \lfloor nt \rfloor \left(F_{\lfloor nt \rfloor}(x) - F(x) \right) \right| .$$

The expression on the right-hand side converges in distribution to

$$\frac{1}{r!} \sup_{t\in[0,1]} |Z_{r,H}(t)| \sup_{x\in\mathbb{R}} |J_r(x)|$$

due to the empirical process non-central limit theorem.

Since $Z_{r,H}$ is an H-self-similar process,

$$\{Z_{r,H}(t),\ 0 \leq t \leq 1\} \stackrel{\mathcal{D}}{=} \left\{t^H Z_{r,H}(1),\ 0 \leq t \leq 1\right\},$$

so that

$$\sup_{t\in[0,1]} |Z_{r,H}(t)| \stackrel{\mathcal{D}}{=} |Z_{r,H}(1)|.$$

As a result, the aforementioned argument yields

$$\begin{aligned}
&P\left(\max_{k\in D_{n,M,1}} \frac{1}{nd_{k_0-k,r}} |A_{n,3}(k)| \geq \frac{1}{4}b(M)\right)\\
\leq &P\left(\frac{1}{r!} |Z_{r,H}(1)| \sup_{x\in\mathbb{R}} |J_r(x)| \geq \frac{1}{4}b(M)\right) + \frac{\varepsilon}{8}\\
< &\frac{\varepsilon}{4}
\end{aligned}$$

for n and M sufficiently large.

4. We have

$$\begin{aligned}
&\max_{k\in D_{n,M,1}} \frac{1}{nd_{k_0-k,r}} |A_{n,4}(k)|\\
\leq &\max_{k\in D_{n,M,1}} \sup_{x\in\mathbb{R}} \left|d^{-1}_{k_0-k,r}(k_0-k)\left(F_{k+1,k_0}(x) - F(x)\right)\right|.
\end{aligned}$$

Hence, the same argument that has been used to obtain an analogous result for $A_{n,1}$ can be applied to conclude that

$$P\left(\max_{k\in D_{n,M,1}} \frac{1}{nd_{k_0-k,r}} |A_{n,4}(k)| \geq \frac{1}{4}b(M)\right) < \frac{\varepsilon}{4}$$

for n and M sufficiently large.

In total, it follows that for all $\varepsilon > 0$ there exists an $n_\varepsilon \in \mathbb{N}$ and an $M > 0$ such that

$$P\left(\left|\hat{k} - k_0\right| > Mm_n\right) < \varepsilon$$

for all $n \geq n_\varepsilon$. This proves Theorem 5. □

2.3. Asymptotic distribution

Theorem 5 implies uniform tightness of the sequence $m_n^{-1}(\hat{k}_W - k_0)$, $n \in \mathbb{N}$. Thus, it seems natural to wonder whether $m_n^{-1}(\hat{k}_W - k_0)$ converges in distribution to a non-degenerate random variable, so that the rate of convergence achieved by Theorem 5 can be considered optimal. The following theorem provides an answer to this question:

Theorem 6 (Betken (2017))**.** *Let X_n, $n \in \mathbb{N}$, denote a time series with a change in the mean in $k_0 = \lfloor n\tau \rfloor$, $0 < \tau < 1$, with shift height h_n. Suppose that X_n, $n \in \mathbb{N}$, satisfies Assumption 2. In particular, $X_n = Y_n + h_n$ for some mean-zero subordinated Gaussian sequence Y_n, $n \in \mathbb{N}$, with marginal distribution function F. Assume that F has a bounded density f and that the Hermite rank of the class of functions $1_{\{Y_1 \leq x\}} - F(x)$, $x \in \mathbb{R}$, equals 1. Let $m_n := g_{D,1}^{-}(h_n^{-1})$ with D as in Assumption 2, define $h(s;\tau)$ by*

$$h(s;\tau) := \begin{cases} s(1-\tau)\int_{\mathbb{R}} f^2(x)dx & \text{if } s \leq 0 \\ -s\tau \int_{\mathbb{R}} f^2(x)dx & \text{if } s > 0 \end{cases}$$

and let $B_H(t)$, $t \in \mathbb{R}$, with $H = 1 - \frac{D}{2}$ be a (standard) fractional Brownian motion. If $h_n^{-1} = o(g_{D,r}(n))$, as $n \to \infty$, then, for all $M > 0$,

$$V_n(s) := \frac{1}{n^3 h_n d_{m_n,1}} \left(W^2_{k_0+\lfloor m_n s \rfloor,n} - W^2_{k_0,n} \right), \; -M \leq s \leq M,$$

converges in distribution to

$$2\tau(1-\tau)\int_{\mathbb{R}} f^2(x)dx \left(B_H(s)\int_{\mathbb{R}} J_1(x)dF(x) + h(s;\tau) \right), \; -M \leq s \leq M,$$

in the Skorohod space $D\,[-M, M]$.

Given an interval I and a right-continuous function f, we write $\operatorname{sargmax}_{s \in I} f(s)$ for the maximizer of f with the smallest value. With this notation,

$$m_n^{-1}\left(\hat{k}_W - k_0\right) = \operatorname*{sargmax}_{s \in (-\infty,\infty)} V_n(s),$$

so that the asymptotic distribution of $m_n^{-1}(\hat{k}_W - k_0)$ can be deduced from Theorem 6 by showing

1. that

$$\operatorname*{sargmax}_{s \in [-M,M]} V_n(s) \xrightarrow{\mathcal{D}} \operatorname*{argmax}_{s \in [-M,M]} G_{H,\tau}(s),$$

   where

$$G_{H,\tau}(s) := B_H(s)\int_{\mathbb{R}} J_1(x)dF(x) + h(s;\tau);$$

2. that

$$\left| \underset{s\in[-M,M]}{\text{sargmax}}\, V_n(s) - \underset{s\in(-\infty,\infty)}{\text{sargmax}}\, V_n(s) \right| \xrightarrow{P} 0;$$

3. that for sufficiently large M

$$\underset{s\in[-M,M]}{\text{argmax}}\ G_{H,\tau}(s) = \underset{s\in(-\infty,\infty)}{\text{argmax}}\ G_{H,\tau}(s).$$

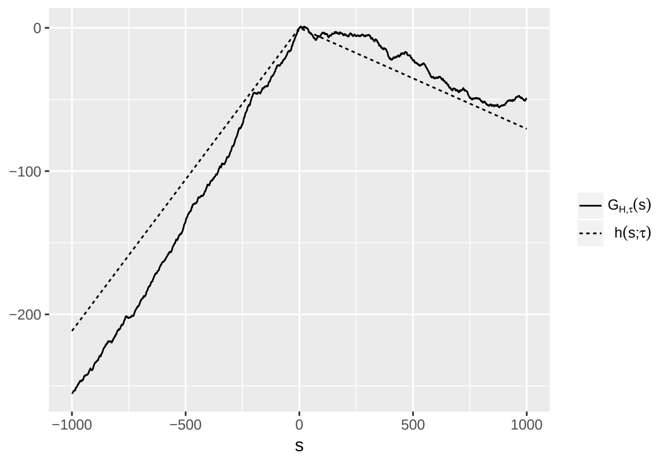

Figure 2.4.: *Realization of* $G_{H,\tau}(s)$*,* $s \in [-M, M]$*, with* $M = 1000$*,* $H = 0.7$*, and* $\tau = 0.25$ *under the assumption of Gaussian time series.*

Corollary 1 (Betken (2017))**.** *Under the assumptions of Theorem 6,* $m_n^{-1}(\hat{k}_W - k_0)$ *converges in distribution to*

$$\underset{s\in(-\infty,\infty)}{\text{argmax}} \left(B_H(s) \int_{\mathbb{R}} J_1(x) dF(x) + h(s;\tau) \right) \tag{2.5}$$

if $h_n^{-1} = o\left(g_{D,r}(n)\right)$*.*

The limit in formula (2.5) closely resembles the limit of the CUSUM-based change-point estimator considered in Horváth and Kokoszka (1997). Moreover, the condition $h_n^{-1} = o\left(g_{D,r}(n)\right)$ is equivalent to Assumption C.5 (i) in that article.

Remark 9. The proof of Theorem 6 is mainly based on the empirical process non-central limit theorem for subordinated Gaussian sequences in Dehling and Taqqu (1989). The sequential empirical process has also been studied by many other authors in the context of different models; see, among others, the following: Müller (1970) and Kiefer (1972) for independent and identically distributed data, Berkes and Philipp (1977) and Philipp and

Pinzur (1980) for strongly mixing processes, Berkes et al. (2009) for S-mixing processes, Giraitis and Surgailis (1999) for long memory linear (or moving average) processes, Dehling et al. (2014) for multiple mixing processes. Presumably, in these situations the asymptotic distribution of $\hat{k}_W$ can be derived by the same argument as in the proofs of Theorem 6 and Corollary 1. In particular, Theorem 1 in Giraitis and Surgailis (1999) can be considered as a generalization of Theorem 1.1 in Dehling and Taqqu (1989), i.e. with an appropriate normalization, the change-point estimator $\hat{k}_W$, computed with respect to long-range dependent linear processes as defined in Giraitis and Surgailis (1999), would converge in distribution to a limit that corresponds to (2.5).

The first assertion of Theorem 6 can be proved by an application of the following lemma, which establishes a condition under which convergence in distribution of a sequence of random variables with values in a càdlàg space entails convergence of the smallest argmax.

Lemma 3 (Betken (2017)). *Let K be a compact interval and denote by $D(K)$ the corresponding Skorohod space. Assume that Z_n, $n \in \mathbb{N}$, are random variables taking values in $D(K)$ and that Z_n converges in distribution to a random variable Z, where (almost surely) Z is continuous and has a unique maximizer. Then, the* sargmax *of Z_n converges in distribution to the* argmax *of Z as n tends to ∞.*

Proof. Due to Skorohod's representation theorem, there exist random variables $\tilde{Z}$ and $\tilde{Z}_n$, $n \in \mathbb{N}$, defined on a common probability space $(\tilde{\Omega}, \tilde{F}, \tilde{P})$, such that

$$\tilde{Z}_n \overset{\mathcal{D}}{=} Z_n, \quad \tilde{Z} \overset{\mathcal{D}}{=} Z \quad \text{with} \quad \tilde{Z}_n \xrightarrow{\text{a.s.}} \tilde{Z}, \text{ as } n \to \infty.$$

Due to Lemma 2.9 in Seijo and Sen (2011), the smallest argmax functional is continuous at W (with respect to the Skorohod metric and the uniform metric) if $W \in D(K)$ is a continuous function which has a unique maximizer. Since (almost surely) Z is continuous with unique maximizer, $\text{sargmax}(\tilde{Z}_n)$ converges to $\text{argmax}(\tilde{Z})$ almost surely. As almost sure convergence implies convergence in distribution, $\text{sargmax}(\tilde{Z}_n)$ converges in distribution to $\text{argmax}(\tilde{Z})$. As a result, $\text{sargmax}(Z_n)$ converges in distribution to $\text{argmax}(Z)$. □

Note that in order to justify an application of Lemma 3, it remains to be shown that the limit attains its maximal value at a unique point since the sample paths of a fractional Brownian motion are almost surely continuous. In the considered case, uniqueness of the maximum can be derived by the following criterion established by Lifshits (1982).

Theorem 7 (Ferger (1999)). *Let $Y(t)$, $t \in T$, denote a Gaussian process indexed by a compact metric space T with almost surely continuous trajectories. If*

$$\mathrm{E}\left((Y(s) - Y(t))^2\right) \neq 0 \quad \textit{for all } s \neq t,$$

Y attains its maximal value at a unique point almost surely. If the parameter set T is only σ-compact (i.e. a countable union of compact sets), the assertion remains valid under the additional assumption that the set of maximizers of Y is nonempty and bounded almost surely.

In addition to Lemma 3 and Theorem 7, the following Lemma is needed for the proof of Theorem 6.

Lemma 4 (Betken (2017))**.** *Suppose that Assumption 2 holds and let* l_n*,* $n \in \mathbb{N}$*, and* h_n*,* $n \in \mathbb{N}$*, be two sequences with* $\lim_{n\to\infty} h_n = h$*,* $\lim_{n\to\infty} l_n = \infty$ *and* $l_n = \mathcal{O}(n)$*. Then, as* $n \to \infty$*,*

$$\sup_{s\in[0,1]}\Bigg| d_{l_n,r}^{-1}\lfloor l_n s\rfloor \int_{\mathbb{R}} \left(F_{\lfloor l_n s\rfloor}(x+h_n) - F_{\lfloor l_n s\rfloor}(x+h)\right) dF_n(x) \\ - d_{l_n,r}^{-1}\lfloor l_n s\rfloor \int_{\mathbb{R}} \left(F(x+h_n) - F(x+h)\right) dF(x)\Bigg| \tag{2.6}$$

and

$$\sup_{s\in[0,1]}\Bigg| d_{l_n,r}^{-1}\lfloor l_n s\rfloor \int_{\mathbb{R}} \left(F_n(x+h_n) - F_n(x+h)\right) dF_{\lfloor l_n s\rfloor}(x) \\ - d_{l_n,r}^{-1}\lfloor l_n s\rfloor \int_{\mathbb{R}} \left(F(x+h_n) - F(x+h)\right) dF(x)\Bigg| \tag{2.7}$$

converge to 0 *almost surely.*

Proof. For the expression in formula (2.6), the triangle inequality yields

$$\begin{aligned}
&\sup_{s\in[0,1]}\Bigg| d_{l_n,r}^{-1}\lfloor l_n s\rfloor \int \left(F_{\lfloor l_n s\rfloor}(x+h_n) - F_{\lfloor l_n s\rfloor}(x+h)\right) dF_n(x) \\
&\quad - d_{l_n,r}^{-1}\lfloor l_n s\rfloor \int \left(F(x+h_n) - F(x+h)\right) dF(x)\Bigg| \\
&\leq 2 \sup_{s\in[0,1], x\in\mathbb{R}} \left| d_{l_n,r}^{-1}\lfloor l_n s\rfloor \left(F_{\lfloor l_n s\rfloor}(x) - F(x)\right) - \frac{1}{r!} Z_{r,H}(s) J_r(x)\right| \\
&\quad + \frac{1}{r!}\sup_{s\in[0,1]} |Z_{r,H}(s)| \left|\int (J_r(x+h_n) - J_r(x+h)) dF_n(x)\right| \\
&\quad + \left| d_{l_n,r}^{-1} l_n \int \left(F(x+h_n) - F(x+h)\right) d\left(F_n - F\right)(x)\right|.
\end{aligned}$$

The first summand on the right-hand side converges to 0 because of the empirical process non-central limit theorem.
Since $Z_{r,H}$ is almost surely continuous, $\sup_{s\in[0,1]} |Z_{r,H}(s)| < \infty$ almost surely. Moreover, it is shown in the proof of Lemma 2 that

$$\left|\int (J_r(x+h_n) - J_r(x+h)) dF_n(x)\right|$$

converges to 0. As a result, the second summand vanishes as n tends to ∞.

Since $l_n = \mathcal{O}(n)$,

$$\begin{aligned}
&\left|d_{l_n,r}^{-1} l_n \int \left(F(x+h_n) - F(x+h)\right) d\left(F_n - F\right)(x)\right| \\
\leq &K \left|\int \left(d_{n,r}^{-1} n \left(F_n(x) - F(x)\right) - \frac{1}{r!} Z_{r,H}(1) J_r(x)\right) dF(x+h_n)\right| \\
+&K \left|\int \left(d_{n,r}^{-1} n \left(F_n(x) - F(x)\right) - \frac{1}{r!} Z_{r,H}(1) J_r(x)\right) dF(x+h)\right| \\
+&K \frac{1}{r!} \left|Z_{r,H}(1)\right| \left|\int J_r(x) d\left(F(x+h_n) - F(x+h)\right)\right|
\end{aligned}$$

for some constant K and n sufficiently large. The first and second summand on the right-hand side of the above inequality converge to 0 due to the empirical process non-central limit theorem. In addition, we have

$$\left|\int J_r(x) d\left(F(x+h_n) - F(x+h)\right)\right| = \left|\int \left(J_r(x-h_n) - J_r(x-h)\right) dF(x)\right|.$$

Therefore, it follows by the same argument as in the proof of Lemma 2 that the third summand converges to 0.

Considering the term in formula (2.7), note that

$$\begin{aligned}
&\sup_{s\in[0,1]} \left|d_{l_n,r}^{-1} \lfloor l_n s\rfloor \int \left(F_n(x+h_n) - F_n(x+h)\right) dF_{\lfloor l_n s\rfloor}(x)\right. \\
&\qquad \left. - d_{l_n,r}^{-1} \lfloor l_n s\rfloor \int \left(F(x+h_n) - F(x+h)\right) dF(x)\right| \\
&\leq 2 \sup_{s\in[0,1], x\in\mathbb{R}} \left|d_{l_n,r}^{-1} \lfloor l_n s\rfloor \left(F_{\lfloor l_n s\rfloor}(x) - F(x)\right) - \frac{1}{r!} Z_{r,H}(s) J_r(x)\right| \\
&\quad + \frac{1}{r!} \sup_{s\in[0,1]} \left|Z_{r,H}(s)\right| \left|\int J_r(x) d\left(F_n(x+h_n) - F_n(x+h)\right)\right| \\
&\quad + 2K \sup_{x\in\mathbb{R}} \left|d_{n,r}^{-1} n \left(F_n(x) - F(x)\right) - \frac{1}{r!} Z_{r,H}(1) J_r(x)\right| \\
&\quad + \frac{1}{r!} \left|Z_{r,H}(1)\right| \int \left|J_r(x+h_n) - J_r(x+h)\right| dF(x)
\end{aligned}$$

for some constant K and n sufficiently large. The first and third summand on the right-hand side of the above inequality converge to 0 due to the empirical process non-central limit theorem. The last summand converges to 0 due to the corresponding argument in the proof of Lemma 2.

By definition of the Hermite coefficient $J_r(x)$, integration by parts, and applications of the triangle inequality, it follows that

$$\begin{aligned}
&\left|\int J_r(x)d\left(F_n(x+h_n)-F_n(x+h)\right)\right| \\
=&\left|\int \left(F_n(G(y)-h_n)-F_n(G(y)-h)\right)H_r(y)\varphi(y)dy\right| \\
\leq&\left(2\sup_{x\in\mathbb{R}}|F_n(x)-F(x)|+\sup_{x\in\mathbb{R}}|F(x-h_n)-F(x-h)|\right)\int|H_r(y)|\,\varphi(y)dy.
\end{aligned}$$

The right-hand side of the above inequality converges to 0 almost surely due to the Glivenko-Cantelli theorem and because F is uniformly continuous. As a result, the second summand vanishes as n tends to ∞ as well. □

Proof of Theorem 6 and Corollary 1. Note that

$$\begin{aligned}
&W_n^2(k_0+\lfloor m_n s\rfloor)-W_n^2(k_0) \\
=&\left(W_n(k_0+\lfloor m_n s\rfloor)-W_n(k_0)\right)\left(W_n(k_0+\lfloor m_n s\rfloor)+W_n(k_0)\right).
\end{aligned}$$

We will see that (with an appropriate normalization) $W_n(k_0+\lfloor m_n s\rfloor)-W_n(k_0)$ converges in distribution to a non-deterministic limit process, whereas $W_n(k_0+\lfloor m_n s\rfloor)+W_n(k_0)$ (with stronger normalization) converges in probability to a deterministic expression.
For notational convenience, we write d_{m_n} instead of $d_{m_n,1}$, J instead of J_1, $\hat{k}$ instead of $\hat{k}_W$ and we define $l_n(s) := k_0+\lfloor m_n s\rfloor$.
Note that

$$W_n(k_0+\lfloor m_n s\rfloor)-W_n(k_0)=\tilde{V}_n(l_n(s))+V_n(l_n(s)),$$

where

$$\tilde{V}_n(l)=\begin{cases}-\sum_{i=l+1}^{k_0}\sum_{j=k_0+1}^{n}\left(1_{\{Y_i\leq Y_j+h_n\}}-1_{\{Y_i\leq Y_j\}}\right) & \text{if } s<0\\ -\sum_{i=1}^{k_0}\sum_{j=k_0+1}^{l}\left(1_{\{Y_i\leq Y_j+h_n\}}-1_{\{Y_i\leq Y_j\}}\right) & \text{if } s>0\end{cases}$$

and

$$V_n(l)=\begin{cases}\sum_{i=1}^{l}\sum_{j=l+1}^{k_0}\left(1_{\{Y_i\leq Y_j\}}-\frac{1}{2}\right)-\sum_{i=l+1}^{k_0}\sum_{j=k_0+1}^{n}\left(1_{\{Y_i\leq Y_j\}}-\frac{1}{2}\right) & \text{if } s<0\\ \sum_{i=k_0+1}^{l}\sum_{j=l+1}^{n}\left(1_{\{Y_i\leq Y_j\}}-\frac{1}{2}\right)-\sum_{i=1}^{k_0}\sum_{j=k_0+1}^{l}\left(1_{\{Y_i\leq Y_j\}}-\frac{1}{2}\right) & \text{if } s>0.\end{cases}$$

In the following, it is shown that $\frac{1}{nd_{m_n}}\tilde{V}_n(l_n(s))$ converges to $h(s;\tau)$ in probability and that $\frac{1}{nd_{m_n}}V_n(l_n(s))$ converges in distribution to $B_H(s)\int J(x)dF(x)$ in $D\left[-M,M\right]$.

For this purpose, rewrite $\tilde{V}_n(l_n(s))$ as follows:

$$\tilde{V}_n(l_n(s)) = -(k_0 - l_n(s))(n - k_0) \int \left(F_{l_n(s)+1,k_0}(x + h_n) - F_{l_n(s)+1,k_0}(x)\right) dF_{k_0+1,n}(x)$$

if $s < 0$, and

$$\tilde{V}_n(l_n(s)) = -k_0(l_n(s) - k_0) \int (F_{k_0}(x + h_n) - F_{k_0}(x)) \, dF_{k_0+1,l_n(s)}(x)$$

if $s > 0$.
For $s < 0$, the limit of $\frac{1}{nd_{m_n}} \tilde{V}_n(l_n(s))$ corresponds to the limit of

$$-(1 - \tau)d_{m_n}^{-1}(k_0 - l_n(s)) \int (F(x + h_n) - F(x)) \, dF(x)$$

due to Lemma 4 and stationarity of the random sequence Y_n, $n \in \mathbb{N}$. Note that

$$-d_{m_n}^{-1} \lfloor m_n s \rfloor h_n \int \frac{1}{h_n} (F(x + h_n) - F(x)) \, dF(x)$$

converges to $-s \int f^2(x)dx$ since $h_n \sim \frac{d_{m_n}}{m_n}$.
For $s > 0$, the limit of $\frac{1}{nd_{m_n}} \tilde{V}_n(l_n(s))$ corresponds to the limit of

$$-\tau d_{m_n}^{-1}(l_n(s) - k_0) \int (F(x + h_n) - F(x)) \, dF(x)$$

due to Lemma 4 and stationarity of the random sequence Y_n, $n \in \mathbb{N}$. Note that

$$d_{m_n}^{-1} \lfloor m_n s \rfloor h_n \int \frac{1}{h_n} (F(x + h_n) - F(x)) \, dF(x)$$

converges to $s \int f^2(x)dx$ since $h_n \sim \frac{d_{m_n}}{m_n}$.
All in all, it follows that $\frac{1}{nd_{m_n}} \tilde{V}_n(l_n(s))$ converges to $h(s;\tau)$ defined by

$$h(s;\tau) = \begin{cases} s(1-\tau) \int f^2(x)dx & \text{if } s \leq 0 \\ -s\tau \int f^2(x)dx & \text{if } s > 0. \end{cases}$$

In the following, it is shown that $\frac{1}{nd_{m_n}} V_n(l_n(s))$ converges in distribution to

$$B_H(s) \int J(x)dF(x), \; -M \leq s \leq M.$$

For this purpose, we show convergence of the correspondingly restricted process in $D[-M, 0]$ and $D[0, M]$. Following this, the proof is extended to weak convergence in $D[-M, M]$.

Note that if $s < 0$,

$$\begin{aligned} V_n(l_n(s)) = & - l_n(s)(k_0 - l_n(s)) \int \big(F_{l_n(s)+1,k_0}(x) - F(x)\big)\, dF_{l_n(s)}(x) \\ & - (k_0 - l_n(s))(n - k_0) \int \big(F_{l_n(s)+1,k_0}(x) - F(x)\big)\, dF_{k_0+1,n}(x) \\ & + l_n(s)(k_0 - l_n(s)) \int (F_{l_n(s)}(x) - F(x)) dF(x) \\ & + (k_0 - l_n(s))(n - k_0) \int (F_{k_0+1,n}(x) - F(x))\, dF(x). \end{aligned}$$

If $s > 0$, we have

$$\begin{aligned} V_n(l_n(s)) = & (l_n(s) - k_0)(n - l_n(s)) \int \big(F_{k_0+1,l_n(s)}(x) - F(x)\big)\, dF_{l_n(s)+1,n}(x) \\ & + k_0(l_n(s) - k_0) \int \big(F_{k_0+1,l_n(s)}(x) - F(x)\big)\, dF_{k_0}(x) \\ & - (l_n(s) - k_0)(n - l_n(s)) \int \big(F_{l_n(s)+1,n}(x) - F(x)\big)\, dF(x) \\ & - k_0(l_n(s) - k_0) \int (F_{k_0}(x) - F(x))\, dF(x). \end{aligned}$$

By the line of argument in the proof of Lemma 4, it follows that the limit of $\frac{1}{nd_{m_n}} V_n(l_n(s))$ corresponds to the limit of

$$\frac{1}{nd_{m_n}} \left(A_{1,n}(s) + A_{2,n}(s) + A_{3,n}(s)\right),$$

where

$$\begin{aligned} A_{1,n}(s) &:= (-l_n(s) - n + k_0)(k_0 - l_n(s)) \int \big(F_{l_n(s)+1,k_0}(x) - F(x)\big)\, dF(x), \\ A_{2,n}(s) &:= (k_0 - l_n(s)) l_n(s) \int (F_{l_n(s)}(x) - F(x)) dF(x), \\ A_{3,n}(s) &:= (k_0 - l_n(s))(n - k_0) \int (F_{k_0+1,n}(x) - F(x))\, dF(x), \end{aligned}$$

if $s < 0$, and

$$\begin{aligned} A_{1,n}(s) &:= (n - l_n(s) + k_0)(l_n(s) - k_0) \int \big(F_{k_0+1,l_n(s)}(x) - F(x)\big)\, dF(x), \\ A_{2,n}(s) &:= -(l_n(s) - k_0)(n - l_n(s)) \int \big(F_{l_n(s)+1,n}(x) - F(x)\big)\, dF(x), \\ A_{3,n}(s) &:= -(l_n(s) - k_0) k_0 \int (F_{k_0}(x) - F(x))\, dF(x), \end{aligned}$$

if $s > 0$.

Note that for $s < 0$

$$\frac{1}{nd_{m_n}} A_{2,n}(s) = -\frac{1}{nd_{m_n}} \lfloor m_n s \rfloor l_n(s) \int (F_{l_n(s)}(x) - F(x)) dF(x). \tag{2.8}$$

As $n \to \infty$, the above expression converges to 0 uniformly in $s \in [-M, 0]$, since

$$\begin{aligned}
&\sup_{s \in [-M,0]} \left| d_n^{-1} l_n(s) \int (F_{l_n(s)}(x) - F(x)) dF(x) \right| \\
&\leq \sup_{x \in \mathbb{R}, t \in [0,1]} \left| d_n^{-1} \lfloor nt \rfloor (F_{\lfloor nt \rfloor}(x) - F(x)) - B_H(t) J(x) \right| \\
&\quad + \sup_{t \in [0,1]} |B_H(t)| \left| \int J(x) dF(x) \right|,
\end{aligned}$$

and since $m_n d_{m_n}^{-1} = o(n d_n^{-1})$, i.e. the expression in (2.8) is bounded in probability. Analogously, it follows that $\frac{1}{nd_{m_n}} A_{2,n}(s)$ converges to 0 uniformly in $s \in [0, M]$. Moreover, it can be shown by an analogous argument that $\frac{1}{nd_{m_n}} A_{3,n}(s)$ converges to 0, uniformly in $s \in [-M, M]$, if n tends to ∞.
Therefore, it remains to be shown that $\frac{1}{nd_{m_n}} A_{1,n}$ converges in distribution to a non-deterministic expression. Due to stationarity,

$$\frac{1}{nd_{m_n}} A_{1,n}(s) \overset{\mathcal{D}}{=} \frac{n - \lfloor m_n s \rfloor}{n} d_{m_n}^{-1} \left(\lfloor m_n s \rfloor \right) \int \left(F_{\lfloor m_n s \rfloor}(x) - F(x) \right) dF(x), \;\; s \in [0, M].$$

As a result, $\frac{1}{nd_{m_n}} A_{1,n}(s)$ converges in distribution to $B_H(s) \int J(x) dF(x)$ in $D[0, M]$. Furthermore, we have

$$\frac{1}{nd_{m_n}} A_{1,n}(s) \overset{\mathcal{D}}{=} \frac{n + \lfloor m_n s \rfloor}{n} d_{m_n}^{-1} \lfloor m_n s \rfloor \int \left(F_{-\lfloor m_n s \rfloor}(x) - F(x) \right) dF(x), \;\; s \in [-M, 0].$$

Note that

$$\begin{aligned}
&- \frac{n + \lfloor m_n s \rfloor}{n} d_{m_n}^{-1} \left(-\lfloor m_n s \rfloor \right) \int \left(F_{-\lfloor m_n s \rfloor}(x) - F(x) \right) dF(x) \\
&= - \frac{n + \lfloor m_n s \rfloor}{n} d_{m_n}^{-1} \left(\lceil m_n(-s) \rceil \right) \int \left(F_{\lceil m_n(-s) \rceil}(x) - F(x) \right) dF(x).
\end{aligned}$$

Therefore, $\frac{1}{nd_{m_n}} A_{1,n}(s)$ converges in distribution to $-B_H(-s) \int J(x) dF(x)$ in $D[-M, 0]$. Considering $\frac{1}{nd_{m_n}} A_{1,n}(s)$ as a stochastic process with path space $D[-M, M]$, we note that for $s \in [0, M]$ and $t \in [-M, 0]$

$$\left(\frac{1}{nd_{m_n}} A_{1,n}(s), \frac{1}{nd_{m_n}} A_{1,n}(t) \right)' \overset{\mathcal{D}}{=} \begin{pmatrix} w_n(s-t) - w_n(-t) \\ -w_n(-t) \end{pmatrix} + \mathcal{O}_P(1),$$

where

$$w_n(t) := \int d_{m_n}^{-1} \lfloor m_n t \rfloor \left(F_{\lfloor m_n t \rfloor}(x) - F(x) \right) dF(x).$$

It then follows from an application of the continuous mapping theorem and the empirical process non-central limit theorem of Dehling and Taqqu (1989) that

$$\left(\frac{1}{nd_{m_n}}A_{1,n}(s), \frac{1}{nd_{m_n}}A_{1,n}(t)\right)' \xrightarrow{\mathcal{D}} (B_H(s-t)-B_H(-t), -B_H(-t))' \int J(x)dF(x).$$

The limit is Gaussian with mean 0 and covariances

$$\text{Cov}(B_H(s-t)-B_H(-t), -B_H(-t)) = \frac{1}{2}\left(|s|^{2H}+|t|^{2H}-|s-t|^{2H}\right),$$

i.e. the covariance function of the limit variable corresponds to the covariances of a (standard) fractional Brownian motion with index set $\mathbb{R}$ as defined in Theorem 6. By an extension of the argument to

$$\left(\frac{1}{nd_{m_n}}A_{1,n}(t_1), \frac{1}{nd_{m_n}}A_{1,n}(t_2), \ldots, \frac{1}{nd_{m_n}}A_{1,n}(t_k)\right)'$$

with $k \in \mathbb{N}$ and $t_1, t_2, \ldots, t_k \in [-M, M]$, $t_1 < t_2 < \ldots < t_k$, the marginal distributions of the limit variable correspond to the marginal distributions of $B_H(s)\int J(x)dF(x)$, $s \in [-M, M]$. Moreover, tightness of $\frac{1}{nd_{m_n}}A_{1,n}$ in $D[-M, 0]$ and in $D[0, M]$ implies that $\frac{1}{nd_{m_n}}A_{1,n}$ is tight in $D[-M, M]$.
All in all, it follows that

$$\frac{1}{nd_{m_n}}\left(W_n(k_0+\lfloor m_n s\rfloor)-W_n(k_0)\right) \xrightarrow{\mathcal{D}} B_H(s)\int J(x)dF(x)+h(s;\tau)$$

in $D[-M, M]$.

With the stronger normalization $h_n n^2$, the limit of $\frac{1}{h_n n^2}W_n(k_0+\lfloor m_n s\rfloor)$ corresponds to the limit of $\frac{1}{h_n n^2}W_n(k_0)$. Furthermore, we have

$$\begin{aligned}\frac{1}{h_n n^2}W_n(k_0) =& \frac{1}{h_n n^2}k_0(n-k_0)\int \left(F_{k_0}(x+h_n)-F_{k_0}(x)\right)dF_{k_0+1,n}(x)\\ &+\frac{1}{h_n n^2}\sum_{i=1}^{k_0}\sum_{j=k_0+1}^{n}\left(1_{\{Y_i\leq Y_j\}}-\frac{1}{2}\right).\end{aligned}$$

The second summand on the right-hand side of the above equation vanishes as n tends to ∞ since $h_n^{-1} = o(nd_n^{-1})$. According to Lemma 4, the limit of

$$d_n^{-1}k_0\int \left(F_{k_0}(x+h_n)-F_{k_0}(x)\right)dF_{k_0+1,n}(x)$$

corresponds to the limit of $d_n^{-1}k_0\int\left(F(x+h_n)-F(x)\right)dF(x)$. Therefore,

$$h_n^{-1}\int \left(F_{k_0}(x+h_n)-F_{k_0}(x)\right)dF_{k_0+1,n}(x) \xrightarrow{\text{a.s.}} \int f^2(x)dx, \text{ as } n\to\infty.$$

It follows that

$$\frac{1}{h_n n^2}\left(W_n(k_0+m_n s)+W_n(k_0)\right) \xrightarrow{P} 2\tau(1-\tau)\int f^2(x)dx, \text{ as } n\to\infty,$$

in $D[-M,M]$. This completes the proof of the first assertion in Theorem 6.

In order to show that

$$m_n^{-1}(\hat{k}-k_0) \xrightarrow{\mathcal{D}} \underset{s\in(-\infty,\infty)}{\operatorname{argmax}}\left(B_H(s)\int J(x)dF(x)+h(s;\tau)\right), \text{ as } n\to\infty,$$

we make use of Lemma 3.
For this purpose, note that according to Theorem 7, Lifshits' criterion for unimodality of Gaussian processes, the random function $G_{H,\tau}(s)=B_H(s)\int J(x)dF(x)+h(s;\tau)$ almost surely attains its maximal value in $[-M,M]$ at a unique point for every $M>0$. Hence, an application of Lemma 3 yields

$$\underset{s\in[-M,M]}{\operatorname{sargmax}}\frac{1}{n^3 h_n d_{m_n,1}}\left(W_n^2(k_0+\lfloor m_n s\rfloor)-W_n^2(k_0)\right) \xrightarrow{\mathcal{D}} \underset{s\in[-M,M]}{\operatorname{argmax}}\, G_{H,\tau}(s). \tag{2.9}$$

It remains to be shown that instead of considering the sargmax in $[-M,M]$, we may as well consider the smallest argmax in $\mathbb{R}$. By the law of the iterated logarithm, i.e. Theorem 1 in Section 1.1.1, $\lim_{|s|\to\infty} s^{-1}B_H(s)=0$ almost surely, so that

$$B_H(s)\int J(x)dF(x)+h(s;\tau) \xrightarrow{\text{a.s.}} -\infty$$

if $|s|\to\infty$. As a result, the limit in formula (2.9) equals $\operatorname{argmax}_{s\in(-\infty,\infty)} G_{H,\tau}(s)$ if M is sufficiently large.
For $M>0$, define

$$\hat{\hat{k}}(M) := \min\left\{k : |k_0-k|\leq Mm_n,\ |W_n(k)|=\max_{|k_0-i|\leq Mm_n}|W_n(i)|\right\}.$$

By definition of $\hat{\hat{k}}(M)$, it follows that

$$\begin{aligned}&\left|\underset{s\in[-M,M]}{\operatorname{sargmax}}\left(W_n^2(k_0+\lfloor m_n s\rfloor)-W_n^2(k_0)\right)-\underset{s\in(-\infty,\infty)}{\operatorname{sargmax}}\left(W_n^2(k_0+\lfloor m_n s\rfloor)-W_n^2(k_0)\right)\right|\\&=m_n^{-1}\left|\hat{\hat{k}}(M)-\hat{k}\right|+\mathcal{O}_P(1).\end{aligned}$$

We have to show that for some $M\in\mathbb{R}$,

$$m_n^{-1}\left|\hat{\hat{k}}(M)-\hat{k}\right| \xrightarrow{P} 0$$

as n tends to ∞. Note that

$$P\left(\hat{k}=\hat{\hat{k}}(M)\right)=P\left(\left|\hat{k}-k_0\right|\leq Mm_n\right)=1-P\left(\left|\hat{k}-k_0\right|>Mm_n\right).$$

Furthermore, we have

$$\lim_{M\to\infty} \liminf_{n\to\infty} \left(1 - P\left(|\hat{k} - k_0| > Mm_n\right)\right) = 1$$

because $|\hat{k} - k_0| = O_P(m_n)$ by Theorem 5. As a result,

$$\lim_{M\to\infty} \liminf_{n\to\infty} P\left(\hat{k} = \hat{\hat{k}}(M)\right) = 1.$$

Hence, for all $\varepsilon > 0$ there exists an $M_\varepsilon \in \mathbb{R}$ and an $n_\varepsilon \in \mathbb{N}$ such that

$$P\left(\hat{k} \neq \hat{\hat{k}}(M)\right) < \varepsilon$$

for all $n \geq n_\varepsilon$ and all $M \geq M_\varepsilon$. This concludes the proof of Theorem 6. □

2.4. Simulations

We will now investigate the finite sample performance of the change-point estimator $\hat{k}_W$ and compare it to corresponding simulation results for the estimators $\hat{k}_{SW}$ (based on the self-normalized Wilcoxon test statistic) and $\hat{k}_{C,0}$ (based on the CUSUM test statistic with parameter $\beta = 0$). For this purpose, we simulate subordinated Gaussian time series $G(\xi_i)$, $i \geq 1$, where ξ_i, $i \geq 1$, is a fractional Gaussian noise sequence generated by the function `fgnSim` (`fArma` package in `R`) with Hurst parameter H. Two different scenarios are considered:

1. Normal margins: We choose $G(t) = t$. Note that in this case the Hermite coefficient $J_1(x)$ is not equal to 0 for all $x \in \mathbb{R}$ (see Dehling et al. (2013)), so that $r = 1$, where r denotes the Hermite rank of $1_{\{G(\xi_i)\leq x\}} - F(x), x \in \mathbb{R}$. Consequently, Assumption 2 holds for all values of $D \in (0, 1)$.

2. Pareto margins: In order to get standardized Pareto-distributed data which has a representation as a functional of a Gaussian process, we consider the transformation

$$G(t) = \left(\frac{\alpha k^2}{(\alpha-1)^2(\alpha-2)}\right)^{-\frac{1}{2}} \left(k(\Phi(t))^{-\frac{1}{\alpha}} - \frac{\alpha k}{\alpha - 1}\right)$$

with parameters $k, \alpha > 0$ and with Φ denoting the standard normal distribution function; see Example 1. Since G is a strictly decreasing function, it follows by Theorem 2 in Dehling et al. (2013) that $r = 1$, where r denotes the Hermite rank of $1_{\{G(\xi_i)\leq x\}} - F(x), x \in \mathbb{R}$. As a result, Assumption 2 holds for all values of $D \in (0, 1)$.

In order to compare the finite sample behavior of the change-point estimators, we consider mean, sample standard deviation (S.D.) and quartiles of $\hat{k}_W$, $\hat{k}_{SW}$ and $\hat{k}_{C,0}$, computed with respect to 500 simulated time series of length 600 for different shift heights h and different change-point locations τ. The simulation results are reported in Tables 2.1, 2.2, and 2.3.

Correspondent to the expected behavior of consistent change-point estimators, the following observations can be made on the basis of the simulation study:

- Bias and variance of the estimated change-point location decrease when the height of the level shift increases.
- Estimation of the time of change is more accurate for breakpoints located in the middle of the sample than estimation of change-point locations that lie close to the boundary of the testing region.
- High values of H go along with an increase of bias and variance. This seems natural since the variance of the time series increases when the correlation between observations, characterized by the value of H, increases.

A comparison of the descriptive statistics of the estimators $\hat{k}_W$ (based on the Wilcoxon statistic) and $\hat{k}_{SW}$ (based on the self-normalized Wilcoxon statistic) shows that:

- In most cases, the estimator $\hat{k}_W$ has a higher bias, especially for an early change-point location. Nevertheless, the difference between the biases of $\hat{k}_{SW}$ and $\hat{k}_W$ is relatively small.
- In general, the sample standard deviation of $\hat{k}_W$ is smaller than that of $\hat{k}_{SW}$. Indeed, it is only slightly better when $\tau = 0.25$, but there is a clear difference when $\tau = 0.5$.

All in all, our simulations do not give rise to choosing $\hat{k}_{SW}$ over $\hat{k}_W$. In particular, better standard deviations of $\hat{k}_W$ compensate for smaller biases of $\hat{k}_{SW}$.

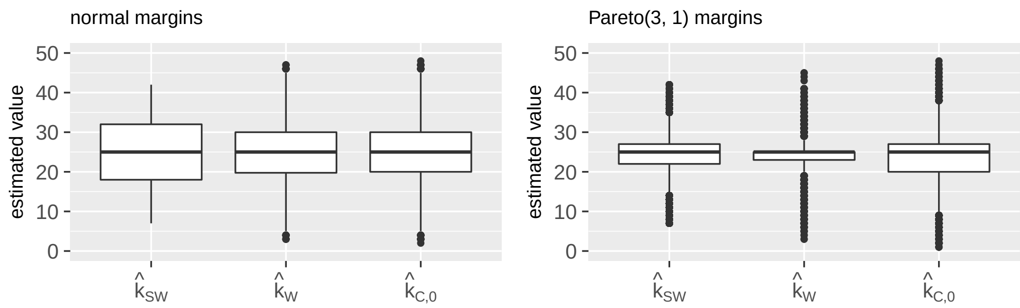

Figure 2.5.: *Boxplots of the estimators* $\hat{k}_{SW}$*,* $\hat{k}_W$ *and* $\hat{k}_{C,0}$ *on the basis of* 5000 *simulated fractional Gaussian noise and Pareto(*3, 1*) time series of length* 50 *with Hurst parameter* $H = 0.7$ *and a change in the mean of height* $h = 0.5$ *after a proportion* $\tau = 0.5$*.*

Comparing the finite sample performances of $\hat{k}_W$ and the CUSUM-based change-point estimator $\hat{k}_{C,0}$, we make the following observations:

- For fractional Gaussian noise time series, bias and variance of $\hat{k}_{C,0}$ tend to be smaller; at least when $\tau = 0.25$ and especially for relatively high level shifts. Nonetheless, the deviations are in most cases negligible.
- If there is a level shift after a proportion $\tau = 0.5$ in a time series with normal margins, bias and variance of $\hat{k}_W$ tend to be smaller, especially for relatively high level shifts. Again, in most cases the deviations are negligible.
- For Pareto-distributed time series $\hat{k}_W$ clearly outperforms $\hat{k}_{C,0}$ by yielding smaller biases and decisively smaller standard deviations for almost every combination of parameters that has been considered. The performance of the estimator $\hat{k}_{C,0}$ surpasses the performance of $\hat{k}_W$ only for high values of the jump height h.

It is well-known that Wilcoxon-based testing procedures are more robust against outliers in data sets than CUSUM-like change-point tests, i.e. Wilcoxon-based tests outperform CUSUM-like tests if heavy-tailed time series are considered. Our simulations confirm that this observation is also reflected by the finite sample behavior of the corresponding change-point estimators; see Figure 2.5.

	τ	h		$H = 0.6$	$H = 0.7$	$H = 0.8$	$H = 0.9$
normal margins	0.25	0.5	mean (S.D.)	193.840 (64.020)	227.590 (99.788)	252.408 (110.084)	270.646 (113.720)
			quartiles	(150, 168, 217.25)	(150, 191, 284.25)	(157, 226.5, 335.25)	(172.75, 250, 353)
		1	mean (S.D.)	164.244 (27.156)	176.362 (42.059)	188.328 (63.751)	215.108 (88.621)
			quartiles	(150, 153.5, 167)	(150, 158, 190)	(150, 159.5, 206.25)	(150 176 256)
		2	mean (S.D.)	153.604 (8.255)	156.656 (12.393)	164.338 (29.570)	173.610 (41.514)
			quartiles	(150, 151, 154)	(150, 151, 158)	(150, 151, 164)	(150, 152, 180.25)
	0.5	0.5	mean (S.D.)	299.506 (30.586)	301.870 (61.392)	300.774 (82.610)	298.930 (98.368)
			quartiles	(291, 300, 309)	(274.75, 300.5, 320.25)	(264, 299, 339.25)	(233, 299, 353)
		1	mean (S.D.)	300.014 (9.141)	300.438 (18.695)	302.592 (42.213)	300.902 (50.487)
			quartiles	(298, 300, 302)	(297, 300, 304)	(293, 300 307)	(290, 300, 311)
		2	mean (S.D.)	300.064 (1.294)	299.922 (3.215)	299.504 (5.520)	300.282 (7.494)
			quartiles	(300, 300, 300)	(300, 300, 300)	(300, 300, 300)	(300, 300, 300)
Pareto(3, 1) margins	0.25	0.5	mean (S.D.)	158.166 (17.762)	164.080 (31.219)	179.512 (58.871)	194.126 (74.767)
			quartiles	(150, 151, 159.25)	(150, 152, 168)	(150, 154, 191.25)	(150, 159, 218.25)
		1	mean (S.D.)	154.160 (8.765)	156.090 (13.516)	164.712 (28.774)	178.174 (54.429)
			quartiles	(150, 151, 155)	(150, 151, 157)	(150, 152, 168)	(150, 152, 186)
		2	mean (S.D.)	152.256 (4.852)	155.592 (11.092)	160.686 (24.599)	169.374 (38.197)
			quartiles	(150, 150, 152)	(150, 151, 155.25)	(150, 151, 159)	(150, 150, 172)
	0.5	0.5	mean (S.D.)	298.072 (6.008)	296.432 (13.441)	293.060 (26.221)	289.946 (45.739)
			quartiles	(297, 300, 300)	(296, 300, 300)	(294, 300, 301)	(291, 300, 301)
		1	mean (S.D.)	299.178 (2.712)	298.744 (4.587)	296.674 (11.585)	296.168 (20.424)
			quartiles	(299, 300, 300)	(299, 300, 300)	(298, 300, 300)	(300, 300, 300)
		2	mean (S.D.)	299.798 (1.008)	299.716 (1.543)	299.384 (3.070)	298.896 (6.560)
			quartiles	(300, 300, 300)	(300, 300, 300)	(300, 300, 300)	(300, 300, 300)

Table 2.1.: *Descriptive statistics of the sampling distribution of* $\hat{k}_W$ *for a change in the mean based on* 500 *simulated time series of length* 600 *with Hurst parameter* H *and a level shift in* τ *of height* h.

margins	τ	h		$H = 0.6$	$H = 0.7$	$H = 0.8$	$H = 0.9$
normal margins	0.25	0.5	mean (S.D.)	172.288 (63.639)	216.934 (110.934)	242.202 (119.655)	268.878 (122.615)
			quartiles	(135, 153, 183.25)	(138, 171, 272.5)	(143, 207.5, 333.5)	(157, 243.5, 370.25)
		1	mean (S.D.)	152.406 (24.840)	160.618 (39.834)	174.424 (70.673)	204.906 (99.648)
			quartiles	(140, 149, 158)	(139, 150.5, 172.25)	(136, 150, 188.25)	(139.75, 161.5, 243.75)
		2	mean (S.D.)	148.836 (9.007)	150.208 (13.575)	153.194 (28.251)	160.026 (40.979)
			quartiles	(144, 150, 152)	(142.75, 150, 154)	(138, 150, 158)	(137.75, 150, 165)
	0.5	0.5	mean (S.D.)	297.712 (43.291)	302.204 (77.719)	302.866 (96.511)	297.662 (110.175)
			quartiles	(277, 297, 320)	(262, 300, 337)	(248, 298.5, 369.5)	(215, 301, 369.5)
		1	mean (S.D.)	299.052 (16.132)	299.910 (28.907)	302.386 (55.267)	300.956 (62.821)
			quartiles	(290, 299, 308)	(288, 300, 313)	(277, 300, 324.25)	(270, 300, 329)
		2	mean (S.D.)	300.010 (6.054)	299.612 (10.079)	298.844 (14.059)	301.424 (21.022)
			quartiles	(297, 300, 303.25)	(294, 300, 305)	(291, 300, 307)	(289, 300, 312)
Pareto(3, 1) margins	0.25	0.5	mean (S.D.)	151.562 (18.392)	155.034 (32.505)	165.260 (58.363)	182.706 (83.268)
			quartiles	(142, 150, 157)	(140, 150, 163)	(136, 150, 173)	(136.75, 150, 196.25)
		1	mean (S.D.)	150.206 (9.116)	150.272 (15.405)	152.824 (25.074)	166.602 (58.982)
			quartiles	(145, 150, 154)	(143, 150, 156)	(140, 150, 159.25)	(136, 150, 174.25)
		2	mean (S.D.)	149.210 (6.201)	149.934 (11.821)	151.946 (21.426)	156.836 (39.311)
			quartiles	(146, 150, 152)	(143, 150, 153)	(140, 150, 156)	(136, 150, 160.25)
	0.5	0.5	mean (S.D.)	300.524 (11.841)	299.488 (21.317)	299.664 (37.136)	295.048 (55.000)
			quartiles	(294, 300, 307)	(290, 300, 310)	(287, 300, 317)	(280.75, 300, 318)
		1	mean (S.D.)	300.498 (6.600)	300.560 (10.383)	299.520 (18.862)	297.766 (28.308)
			quartiles	(297, 300, 304)	(296, 300, 306)	(292, 300, 309.25)	(289, 300, 312.25)
		2	mean (S.D.)	300.444 (4.411)	300.234 (7.517)	300.524 (11.122)	298.840 (16.004)
			quartiles	(298, 300, 303)	(296, 300, 304)	(295.75, 300, 307)	(292, 300, 308)

Table 2.2.: *Descriptive statistics of the sampling distribution of* $\hat{k}_{SW}$ *for a change in the mean based on* 500 *simulated time series of length* 600 *with Hurst parameter* H *and a level shift in* τ *of height* h.

	τ	h		$H = 0.6$	$H = 0.7$	$H = 0.8$	$H = 0.9$
normal margins	0.25	0.5	mean (S.D.)	193.060 (64.917)	228.948 (101.442)	253.114 (111.182)	271.380 (114.590)
			quartiles	(150, 166.5, 222)	(151, 191.5, 286.75)	(156.75, 226, 341.5)	(172.75, 249.5, 354.25)
		1	mean (S.D.)	162.028 (22.948)	173.838 (39.845)	187.386 (63.865)	213.114 (87.356)
			quartiles	(150, 153, 164)	(150, 156.5, 187.25)	(150, 158, 206)	(150, 173, 254.25)
		2	mean (S.D.)	152.374 (6.249)	154.878 (10.395)	159.700 (22.064)	165.940 (33.124)
			quartiles	(150, 150, 152)	(150, 150, 156)	(150, 151, 158)	(150, 150, 165)
	0.5	0.5	mean (S.D.)	297.840 (30.249)	302.060 (63.878)	300.246 (84.346)	298.910 (97.904)
			quartiles	(290, 299, 308)	(276, 301, 322)	(261.75, 300, 340)	(236.25, 299, 353.25)
		1	mean (S.D.)	299.870 (9.356)	299.662 (21.281)	303.646 (42.245)	299.762 (52.492)
			quartiles	(298, 300, 302)	(297, 300, 304)	(293, 300, 307)	(290, 300, 311)
		2	mean (S.D.)	300.060 (1.473)	299.916 (3.199)	299.442 (5.234)	300.460 (8.179)
			quartiles	(300, 300, 300)	(300, 300, 300)	(300, 300, 300)	(300, 300, 300)
Pareto(3, 1) margins	0.25	0.5	mean (S.D.)	175.632 (48.517)	198.452 (79.303)	205.506 (88.482)	210.444(93.831)
			quartiles	(150, 159, 185)	(150, 168, 223.75)	(150, 173, 251.25)	(150, 167, 259.5)
		1	mean (S.D.)	156.586 (14.133)	160.350 (27.204)	170.278 (45.402)	177.278 (66.661)
			quartiles	(150, 152, 159)	(150, 152, 161)	(150, 153, 171)	(150, 150, 174)
		2	mean (S.D.)	150.314 (1.349)	150.566 (3.984)	152.474 (18.578)	155.496 (29.408)
			quartiles	(150, 150, 150)	(150, 150, 150)	(150, 150, 150)	(150, 150, 150)
	0.5	0.5	mean (S.D.)	296.260 (22.306)	292.904 (43.471)	289.192 (64.033)	287.966 (64.827)
			quartiles	(292, 300, 303.25)	(288.75, 300, 305)	(273.75, 300, 308.25)	(285, 300, 303)
		1	mean (S.D.)	298.240 (6.104)	297.306 (9.361)	293.116 (26.614)	292.864 (37.601)
			quartiles	(299, 300, 300)	(299, 300, 300)	(298, 300, 300)	(300, 300, 300)
		2	mean (S.D.)	299.604 (1.843)	299.228 (3.385)	298.350 (8.354)	297.632 (14.525)
			quartiles	(300, 300, 300)	(300, 300, 300)	(300, 300, 300)	(300, 300, 300)

Table 2.3.: *Descriptive statistics of the sampling distribution of* $\hat{k}_{C,0}$ *for a change in the mean based on* 500 *simulated time series of length* 600 *with Hurst parameter* H *and a level shift in* τ *of height* h.

3. Subsampling for long-range dependent time series

Given observations $X_1, \dots, X_n$ and a series of statistics $T_n = T_n(X_1, \dots, X_n)$, $n \in \mathbb{N}$, many issues in statistics are related to an approximation of the distribution function F_{T_n} of T_n. If T_n converges in distribution to some non-degenerate random variable T, the distribution function F_T of T is often considered as a suitable estimate for F_{T_n}. However, in many cases F_T itself is unknown, so that confidence intervals or critical values for test decisions are usually based on simulations of the required quantiles. The self-normalized change-point tests considered in Section 1.2.2 serve as examples since the corresponding test statistics converge in distribution to functionals of stochastic processes which depend on unknown parameters (the Hurst index H and the Hermite rank r). In this case, test decisions are generated by comparing the value of the particular test statistic to critical values attained by simulations of the limit variable with respect to estimated parameter values. To overcome the problem of an unknown limit distribution and to avoid the estimation of nuisance parameters, an application of non-parametric methods to estimate F_{T_n} can be considered as an alternative to approximating the asymptotic distribution of the test statistic by simulations. Since we typically just have one sample, so that only a single realization of T_n is observed, the basic idea of subsampling procedures is to approximate the distribution of T_n by the empirical distribution of values of the statistic computed over subsets of the original sample.

3.1. Sampling-window method

The so-called *sampling-window method*, studied by Politis and Romano (1994), Hall and Jing (1996), and Sherman and Carlstein (1996), utilizes evaluations of the test statistic in subsamples of successive observations, i.e. for some blocklength $l_n < n$, the realizations $T_{l_n,k} := T_{l_n}(X_k, \dots, X_{k+l_n-1})$, $k = 1, \dots, m_n$, where $m_n := n - l_n + 1$, are considered. As a result, we obtain multiple (though dependent) realizations of the test statistic T_{l_n}. Due to the fact that consecutive observations are chosen, the subsamples retain the dependence structure of the original sample, so that the empirical distribution function of $T_{l_n,1}, \dots, T_{l_n,m_n}$, defined by

$$\widehat{F}_{m_n,l_n}(t) := \frac{1}{m_n} \sum_{k=1}^{m_n} 1_{\{T_{l_n,k} \leq t\}}, \tag{3.1}$$

can be considered as an appropriate estimator for F_{T_n}.

In order to establish the validity of the subsampling procedure, i.e. in order to show that the empirical distribution function of $T_{l_n,1}, \ldots, T_{l_n,m_n}$ can be considered as a suitable approximation of F_{T_n}, we aim at proving that the distance between $\widehat{F}_{m_n,l_n}$ and F_{T_n} vanishes as the number of observations tends to ∞.

Definition 10. The subsampling procedure is said to be *consistent* if

$$\left|\widehat{F}_{m_n,l_n}(t) - F_{T_n}(t)\right| \xrightarrow{\mathcal{P}} 0, \quad \text{as } n \to \infty,$$

for all points of continuity t of F_T.

Remark 10. If the subsampling procedure is consistent in the sense of Definition 10, and if F_T is continuous, the usual Glivenko-Cantelli argument for uniform convergence of empirical distribution functions implies that

$$\sup_{t \in \mathbb{R}} \left|\widehat{F}_{m_n,l_n}(t) - F_{T_n}(t)\right| \xrightarrow{\mathcal{P}} 0, \quad \text{as } n \to \infty;$$

see for example Section 20 in Billingsley (1995).

It is shown in Sherman and Carlstein (1996) that the sampling-window method is consistent for any time series satisfying an α-mixing condition and for any measurable statistic converging in distribution to a non-degenerate limit variable. Thereby, consistency can be derived for an extensive class of short-range dependent processes under the mildest possible assumptions on the blocklength and the considered statistic. In the long-range dependent case, the validity of subsampling has been shown to hold for specific statistics under various model assumptions. Hall et al. (1998) prove consistency of the sampling-window method for the sample mean as well as a studentized version of the sample mean under the assumption of subordinated Gaussian processes. Nordman and Lahiri (2005) attained consistency results with respect to the same statistics for long-range dependent linear processes with possibly non-Gaussian innovations. For this model, an alternative proof for consistency can be found in Beran et al. (2013). Zhang et al. (2013) generalize these results by proving consistency with respect to the sample mean under the assumption of subordinated long-range dependent linear processes with possibly non-Gaussian innovations. It was noted by Fan (2012) that the proof in Hall et al. (1998) can easily be generalized to other statistics than the sample mean. Nonetheless, the assumptions on the subordinated Gaussian processes considered in Hall et al. (1998) are rather restrictive; see McElroy and Politis (2007). Their conditions imply that the underlying Gaussian sequence is completely regular, which might hold for some special cases (see Ibragimov and Rozanov (1978)), but excludes standard examples:

Example 2 (Fractional Gaussian noise)**.** Let

$$\xi_H(n) := B_H(n) - B_H(n-1)$$

for some fractional Brownian motion $B_H(t)$, $t \in [0, \infty)$, i.e. ξ_H is a fractional Gaussian noise process; see Definition 6 in Chapter 1. It follows by self-similarity of B_H that

$$\begin{aligned}\operatorname{Cov}\left(\sum_{i=1}^{n}\xi_H(i), \sum_{j=2n+1}^{3n}\xi_H(j)\right) &= \operatorname{Cov}\left(B_H(n), B_H(3n) - B_H(2n)\right) \\ &= \operatorname{Cov}\left(B_H(1), B_H(3) - B_H(2)\right).\end{aligned}$$

As a result, the correlation of linear combinations of variables does not vanish if the gap between past and future observations grows. Thus, fractional Gaussian noise is not completely regular.

Jach et al. (2012) provide a result on the validity of subsampling for a general class of statistics T_n and certain heavy-tailed long-range dependent time series that follow the long memory stochastic volatility model defined in Section 4.1. However, their results are restricted by the assumption that the transformation σ in Model 2 is invertible, Lipschitz-continuous and that the underlying Gaussian process has a causal representation as a functional of a sequence of independent random variables. These assumptions are difficult to check in practice. Moreover, although not explicitly stated in Jach et al. (2012), the proof of consistency only holds for statistics T_n that are Lipschitz-continuous (uniformly in n). Many robust statistics do not satisfy this assumption. In fact, the change-point test statistics considered in Section 1.2 can be taken as examples for non-Lipschitz-continuous statistics. Nonetheless, combining self-normalized statistics and subsampling results in a testing procedure that, if applied in practice, only requires the choice of the blocklength parameter. For this reason, the main aim of this section is to establish the validity of the sampling-window method for general statistics T_n without any assumptions on the continuity of the statistic and only mild assumptions on the data-generating process.

Given observations $X_1, \ldots, X_n$ generated by subordinated Gaussian time series, the consistency proof for the sampling-window method can be reduced to assessing the maximal correlation between σ-algebras generated by two separate blocks of random variables.

Definition 11. Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $\mathcal{A}$ and $\mathcal{B}$ are sub-σ-fields of $\mathcal{F}$. The *maximal correlation* between $\mathcal{A}$ and $\mathcal{B}$ is defined by

$$\rho(\mathcal{A}, \mathcal{B}) := \sup\left|\operatorname{Corr}(f, g)\right|, \ f \in L^2(\mathcal{A}), \ g \in L^2(\mathcal{B}),$$

where $L^2(\mathcal{A}) := L^2(\Omega, \mathcal{A}, P)$ and $L^2(\mathcal{B}) := L^2(\Omega, \mathcal{B}, P)$ denote the families of all (equivalence classes of) real-valued, $\mathcal{A}$-measurable (resp. $\mathcal{B}$-measurable) random variables on Ω with finite second moments.

In order to establish the validity of the subsampling procedure, note that the triangular inequality yields

$$|\widehat{F}_{m_n,l_n}(t) - F_{T_n}(t)| \leq |\widehat{F}_{m_n,l_n}(t) - F_T(t)| + |F_T(t) - F_{T_n}(t)|. \tag{3.2}$$

The second term on the right-hand side of the inequality converges to 0 for all points of continuity t of F_T if the following assumption holds:

Assumption 3. The statistics T_n, $n \in \mathbb{N}$, are measurable and converge in distribution to a (non-degenerate) random variable T with distribution function F_T.

Assumption 3 is considered as a standard requirement for a proof of consistency; see for example Politis and Romano (1994). If the distribution of T_n does not converge, we cannot expect the distribution of T_{l_n} to be close to the distribution of T_n.
Given Assumption 3, it remains to show that the first summand on the right-hand side of inequality (3.2) converges to 0 as well. As L^2-convergence implies convergence in probability, it suffices to show that $\lim_{n\to\infty} \mathrm{E}\,|\widehat{F}_{m_n,l_n}(t) - F_T(t)|^2 = 0$. For this purpose, we consider the following bias-variance decomposition:

$$\mathrm{E}\left(|\widehat{F}_{m_n,l_n}(t) - F_T(t)|^2\right) = \mathrm{Var}\,\widehat{F}_{m_n,l_n}(t) + \left|\mathrm{E}\,\widehat{F}_{m_n,l_n}(t) - F_T(t)\right|^2.$$

Stationarity of the process X_n, $n \in \mathbb{N}$, implies that $\mathrm{E}\,\widehat{F}_{m_n,l_n}(t) = F_{T_l}(t)$, so that, due to the convergence of T_n, the bias term of the above equation converges to 0 as l_n tends to ∞. As a result, it remains to show that the variance term vanishes as n tends to ∞. Initially, note that

$$\begin{aligned}\mathrm{Var}\,\widehat{F}_{m_n,l_n}(t) &= \frac{1}{m_n}\mathrm{Var}\,1_{\{T_{l_n,1}\le t\}} + \frac{2}{m_n^2}\sum_{k=2}^{m_n}(m_n - i + 1)\,\mathrm{Cov}\left(1_{\{T_{l_n,1}\le t\}}, 1_{\{T_{l_n,k}\le t\}}\right)\\ &\le \frac{2}{m_n}\sum_{k=1}^{m_n}\left|\mathrm{Cov}\left(1_{\{T_{l_n,1}\le t\}}, 1_{\{T_{l_n,k}\le t\}}\right)\right|,\end{aligned}$$

due to stationarity of X_n, $n \in \mathbb{N}$.
Since $T_{l_n,k} = T_{l_n}(G(\xi_k), \ldots, G(\xi_{k+l_n-1}))$ for some measurable function G,

$$\left|\mathrm{Cov}\left(1_{\{T_{l_n,1}\le t\}}, 1_{\{T_{l_n,k}\le t\}}\right)\right| \le \rho\left(\sigma(\xi_i, 1 \le i \le l_n), \sigma(\xi_j, k \le j \le k + l_n - 1)\right),$$

where $\sigma(\xi_i, 1 \le i \le l_n)$ (resp. $\sigma(\xi_j, k \le j \le k + l_n - 1)$) denotes the σ-field generated by the random variables $\xi_1, \ldots, \xi_{l_n}$ (resp. $\xi_k, \ldots, \xi_{k+l_n-1}$).
For $\beta \in (0,1)$, we split the sum of covariances into two parts:

$$\begin{aligned}&\frac{1}{m_n}\sum_{k=1}^{m_n}\left|\mathrm{Cov}\left(1_{\{T_{l_n,1}\le t\}}, 1_{\{T_{l_n,k}\le t\}}\right)\right|\\ =&\frac{1}{m_n}\sum_{k=1}^{\lfloor n^\beta\rfloor}\left|\mathrm{Cov}\left(1_{\{T_{l_n,1}\le t\}}, 1_{\{T_{l_n,k}\le t\}}\right)\right| + \frac{1}{m_n}\sum_{k=\lfloor n^\beta\rfloor+1}^{m_n}\left|\mathrm{Cov}\left(1_{\{T_{l_n,1}\le t\}}, 1_{\{T_{l_n,k}\le t\}}\right)\right|\\ \le&\frac{\lfloor n^\beta\rfloor}{m_n} + \frac{1}{m_n}\sum_{k=\lfloor n^\beta\rfloor+1}^{m_n}\rho\left(\sigma(\xi_i, 1 \le i \le l_n), \sigma(\xi_j, k \le j \le k + l_n - 1)\right)\\ \le&\frac{\lfloor n^\beta\rfloor}{m_n} + \frac{1}{m_n}\sum_{k=\lfloor n^\beta\rfloor-l_n+1}^{m_n-l_n}\rho_{k,l_n},\end{aligned}$$

where

$$\rho_{k,l_n} := \rho\left(\sigma(\xi_i, 1 \leq i \leq l_n), \sigma(\xi_j, k + l_n \leq j \leq k + 2l_n - 1)\right).$$

The first summand on the right-hand side of the inequality converges to 0 if $l_n = o(n)$. In order to show that the second summand converges to 0, a sufficiently good approximation to the sum of maximal correlations is needed. In particular, we have to show that

$$\sum_{k=\lfloor n^\beta \rfloor - l_n + 1}^{m_n - l_n} \rho_{k,l_n} = o(m_n).$$

3.1.1. Auxiliary results

The following results characterize the maximal correlation $\rho_{k,l}$ of σ-algebras generated by two blocks of l subsequent random variables separated by a lag of k time units. An exact representation of the maximal correlation in terms of correlation matrices is given by the following proposition:

Proposition 2. *Suppose that ξ_n, $n \in \mathbb{N}$, is a stationary Gaussian time series with mean 0, variance 1 and autocovariance function γ, $\gamma(k) := \operatorname{Cov}(\xi_1, \xi_{k+1})$, such that $\lim_{k\to\infty} \gamma(k) = 0$. Let $\Sigma_{m,l} := (\gamma_{m+i,j})_{1\leq i,j\leq l}$ with $\gamma_{i,j} := \operatorname{Cov}(\xi_i, \xi_j)$. Then*

$$\rho\left(\sigma(\xi_i, 1 \leq i \leq l), \sigma(\xi_j, k + l \leq j \leq k + 2l - 1)\right) = \|\Sigma_l^{-\frac{1}{2}} \Sigma'_{k+l,l} \Sigma_l^{-\frac{1}{2}}\|_2,$$

where $\|\cdot\|_2$ is the spectral norm and $\Sigma_l^{-\frac{1}{2}}$ denotes the inverse of the (principal) square root of $\Sigma_l := \Sigma_{0,l}$.

Proof. Note that the random vector $(\xi_1, \ldots, \xi_l, \xi_{k+l}, \ldots, \xi_{k+2l-1})'$ follows a multivariate normal distribution with mean 0 and covariance matrix

$$\tilde{\Sigma}_{k,l} = \begin{pmatrix} \Sigma_l & \Sigma_{k+l,l} \\ \Sigma'_{k+l,l} & \Sigma_l \end{pmatrix}.$$

According to Proposition 5.1.1. in Brockwell and Davis (2013), the matrix Σ_l is symmetric and positive definite. It therefore has a symmetric, positive definite inverse, and a symmetric, positive definite (principal) square root with a symmetric, positive definite inverse. Define

$$(\zeta_1, \ldots, \zeta_l)' := \Sigma_l^{-\frac{1}{2}}(\xi_1, \ldots, \xi_l)', \quad (\zeta_{k+l}, \ldots, \zeta_{k+2l-1})' := \Sigma_l^{-\frac{1}{2}}(\xi_{k+l}, \ldots, \xi_{k+2l-1})'.$$

By definition, both random vectors have a multivariate standard normal distribution. With

$$\Gamma := \begin{pmatrix} \Sigma_l^{-\frac{1}{2}} & 0 \\ 0 & \Sigma_l^{-\frac{1}{2}} \end{pmatrix},$$

it follows that $(\zeta_1, \dots, \zeta_l, \zeta_{k+l}, \dots, \zeta_{k+2l-1})' = \Gamma\,(\xi_1, \dots, \xi_l, \xi_{k+l}, \dots, \xi_{k+2l-1})'$ is normally distributed with covariance matrix

$$\Gamma\tilde{\Sigma}_{k,l}\Gamma' = \begin{pmatrix} I_l & \Sigma_l^{-\frac{1}{2}}\Sigma_{k+l,l}\Sigma_l^{-\frac{1}{2}} \\ \Sigma_l^{-\frac{1}{2}}\Sigma_{k+l,l}'\Sigma_l^{-\frac{1}{2}} & I_l \end{pmatrix}.$$

Consider the Hilbert space H of (equivalence classes of) random variables

$$a_1\zeta_1 + a_2\zeta_2 + \cdots + a_l\zeta_l + b_1\zeta_{k+l} + \cdots + b_l\zeta_{k+2l-1},$$

where $a_1, \dots, a_l,\ b_1, \dots, b_l$ range over the real numbers and the inner product of ξ_1, $\xi_2 \in H$ is defined by $\langle \xi_1, \xi_2 \rangle := \mathrm{E}\,\xi_1\xi_2$. Let H_1 (resp. H_2) denote the subspace of H spanned by $\zeta_1, \dots, \zeta_l$ (resp. $\zeta_{k+l}, \dots, \zeta_{k+2l-1}$). It then follows by definition of H_1 (resp. H_2), and independence of $\zeta_1, \dots, \zeta_l$ (resp. $\zeta_{k+l}, \dots, \zeta_{k+2l-1}$) that $\zeta_1, \dots, \zeta_l$ (resp. $\zeta_{k+l}, \dots, \zeta_{k+2l-1}$) is an orthonormal basis of H_1 (resp. H_2). By Theorem A903 in Bradley (2007), there exist mean-zero random variables $V_1, \dots, V_l \in H_1$ and $W_1, \dots, W_l \in H_2$ with the following properties:

- $V_1, \dots, V_l$ is an orthonormal basis of H_1;
- $W_1, \dots, W_l$ is an orthonormal basis of H_2;
- $\langle V_i, W_i \rangle = \mathrm{E}\,V_iW_i \geq 0$ for all $i = 1, 2, \dots, l$;
- $\langle V_i, W_j \rangle = \mathrm{E}\,V_iW_j = 0$ for all $i = 1, 2, \dots, l$, $j = 1, 2, \dots, l$ with $i \neq j$.

It is shown in the proof of Theorem 9.2 in Bradley (2007) that there exists an index $K \in \{1, \dots, l\}$ such that

$$\rho(\sigma(\xi_i, 1 \leq i \leq l), \sigma(\xi_j, k+l \leq j \leq k+2l-1)) = \mathrm{Corr}(V_K, W_K).$$

So far, we specified two different sets of random variables that serve as bases for H_1 (resp. H_2). While the joint distributions of $\zeta_1, \dots, \zeta_l$ and $\zeta_{k+l}, \dots, \zeta_{k+2l-1}$ are known, the orthonormal bases $V_1, \dots, V_l$ and $W_1, \dots, W_l$ yield an explicit description of the maximal correlation. So as to derive an approximation for the maximal correlation, we relate both bases by change of basis matrices. For this purpose, let T denote the orthogonal matrix satisfying

$$(V_1, \dots, V_l)' = T'\,(\zeta_1, \dots, \zeta_l)', \quad (W_1, \dots, W_l)' = T'\,(\zeta_{k+l}, \dots, \zeta_{k+2l-1})'.$$

It follows that

$$(V_1, \dots, V_l, W_1, \dots, W_l)' = S\,(\zeta_1, \dots, \zeta_l, \zeta_{k+l}, \dots, \zeta_{k+2l-1})', \quad \text{where } S := \begin{pmatrix} T' & 0 \\ 0 & T' \end{pmatrix}.$$

All in all, we conclude that the random vector $(V_1, \dots, V_l, W_1, \dots, W_l)'$ is normally distributed with mean 0 and autocovariance matrix

$$S\Gamma\tilde{\Sigma}_{k,l}\Gamma'S' = \begin{pmatrix} I_l & T'\Sigma_l^{-\frac{1}{2}}\Sigma_{k+l,l}\Sigma_l^{-\frac{1}{2}}T \\ T'\Sigma_l^{-\frac{1}{2}}\Sigma'_{k+l,l}\Sigma_l^{-\frac{1}{2}}T & I_l \end{pmatrix}.$$

Note that $\operatorname{Var} V_K = \operatorname{Var} W_K = 1$, so that

$$\rho(\sigma(\xi_i, 1 \le i \le l), \sigma(\xi_j, k+l \le j \le k+2l-1)) = \operatorname{Cov}(V_K, W_K) = D_{K,K}$$

for some $K \in \{1, \dots, l\}$. Due to the choice of $V_1, \dots, V_l$ and $W_1, \dots, W_l$, the matrix

$$D := T'\Sigma_l^{-\frac{1}{2}}\Sigma'_{k+l,l}\Sigma_l^{-\frac{1}{2}}T$$

is diagonal, so that the diagonal entries of D correspond to its eigenvalues. As a result, we have

$$D_{K,K} = \|D\|_2 = \|T'\Sigma_l^{-\frac{1}{2}}\Sigma'_{k+l,l}\Sigma_l^{-\frac{1}{2}}T\|_2 = \|\Sigma_l^{-\frac{1}{2}}\Sigma'_{k+l,l}\Sigma_l^{-\frac{1}{2}}\|_2,$$

where the last identity results from the invariance of $\|\cdot\|_2$ under orthogonal transformations. □

On the basis of Proposition 2, we derive an upper bound for the maximal correlation $\rho_{k,l}$:

Corollary 2. *Under the assumptions of Proposition 2,*

$$\rho(\sigma(\xi_i, 1 \le i \le l), \sigma(\xi_j, k+l \le j \le k+2l-1)) \le l\frac{1}{\lambda_{min}} \max_{1 \le i \le 2l-1} |\gamma_{k+i}|,$$

where λ_{min} denotes the smallest eigenvalue of Σ_l.

Proof. Note that for $A = (a_{ij})_{1 \le i,j \le n}$ it holds that $\|A\|_2 \le n \max_{1 \le i,j \le n} |a_{ij}|$. According to this and the Cauchy–Schwarz inequality,

$$\|\Sigma_l^{-\frac{1}{2}}\Sigma'_{k+l,l}\Sigma_l^{-\frac{1}{2}}\|_2 \le \|\Sigma_l^{-\frac{1}{2}}\|_2^2\|\Sigma'_{k+l,l}\|_2 \le \|\Sigma_l^{-\frac{1}{2}}\|_2^2 l \max_{1 \le i \le 2l-1} |\gamma_{k+i}|.$$

By definition of the spectral norm,

$$\|\Sigma_l^{-\frac{1}{2}}\|_2 = \sqrt{\mu_{\max}} = \frac{1}{\sqrt{\lambda_{\min}}},$$

where $\mu_{\max}$ denotes the largest eigenvalue of Σ_l^{-1}. □

The above corollary meets the expectation that a growing time lag between two blocks of random variables leads to a decrease of the maximal correlation, whereas a growth of the blocklength yields an increase of correlation. Given the following assumption on the spectral density of a stationary Gaussian process, the subsequent Lemma shows that the eigenvalues of the covariance matrix do not contribute to a significant increase of the maximal correlation.

Assumption 4. Let ξ_n, $n \in \mathbb{N}$, denote a stationary, long-range dependent Gaussian process with mean 0, variance 1, LRD parameter D and spectral density $f(\lambda) = |\lambda|^{D-1} L_f(\lambda)$ for a slowly varying function L_f which is bounded away from 0 on $[0, \pi]$.

Lemma 5 (Betken and Wendler (2015))**.** *Suppose that ξ_n, $n \in \mathbb{N}$, is a time series with a spectral density satisfying Assumption 4. Let Σ_l denote the covariance matrix of l consecutive random variables, i.e. $\Sigma_l := (\gamma_{i,j})_{1 \leq i,j \leq l}$, $\gamma_{ij} := \mathrm{Cov}(\xi_i, \xi_j)$. Then there exists a constant $K_D \in (0, \infty)$, such that*

$$\lambda_{min}(\Sigma_l) \geq K_D$$

for all $l \in \mathbb{N}$.

Proof. By definition,

$$\lambda_{\min}(\Sigma_l) = \inf_{x \in \mathcal{X}} (x' \Sigma_l x),$$

where $\mathcal{X} = \left\{ x \in \mathbb{R}^l \mid x'x = 1 \right\}$. We rewrite $x'\Sigma_l x$ in the following way:

$$\begin{aligned} x'\Sigma_l x &= \sum_{j=1}^{l} \sum_{k=1}^{l} x_j x_k \gamma(j-k) \\ &= \sum_{j=1}^{l} \sum_{k=1}^{l} x_j x_k \int_{-\pi}^{\pi} \mathrm{e}^{i(j-k)\lambda} L_f(|\lambda|) |\lambda|^{D-1} d\lambda \\ &= 2 \int_0^{\pi} \sum_{j=1}^{l} \sum_{k=1}^{l} x_j x_k \mathrm{e}^{i(j-k)\lambda} L_f(\lambda) \lambda^{D-1} d\lambda \\ &= 2 \int_0^{\pi} \left| \sum_{j=1}^{l} x_j \mathrm{e}^{-ij\lambda} \right|^2 L_f(\lambda) \lambda^{D-1} d\lambda. \end{aligned}$$

By assumption, there exists a constant $C_{\min} \in (0, \infty)$ with

$$L_f(\lambda) \geq C_{\min}$$

for all $\lambda \in [0, \pi]$, so that

$$x'\Sigma_l x \geq 2C_{\min} \int_0^{\pi} \left| \sum_{j=1}^{l} x_j \mathrm{e}^{-ij\lambda} \right|^2 \lambda^{D-1} d\lambda \geq 2C_{\min} \pi^{D-1} \int_0^{\pi} \left| \sum_{j=1}^{l} x_j \mathrm{e}^{-ij\lambda} \right|^2 d\lambda.$$

We rewrite the integrand as

$$\left| \sum_{j=1}^{l} x_j \mathrm{e}^{-ij\lambda} \right|^2 = \sum_{j=1}^{l} x_j^2 + \sum_{1 \leq j \neq k \leq l} x_j x_k \mathrm{e}^{-i(j-k)\lambda}$$

$$
\begin{aligned}
&= \sum_{j=1}^{l} x_j^2 + \sum_{1\le j<k\le l} x_j x_k \left(\mathrm{e}^{-i(j-k)\lambda} + \mathrm{e}^{-i(k-j)\lambda}\right) \\
&= \sum_{j=1}^{l} x_j^2 + 2 \sum_{1\le j<k\le l} x_j x_k \cos((k-j)\lambda) \\
&= \sum_{j=1}^{l} \sum_{k=1}^{l} x_j x_k \cos((k-j)\lambda).
\end{aligned}
$$

Since $\int_0^\pi \cos((k-j)\lambda)d\lambda = 0$ for all j, k with $j \neq k$, we have

$$
\begin{aligned}
\int_0^\pi \left| \sum_{j=1}^{l} x_j \mathrm{e}^{-ij\lambda} \right|^2 d\lambda &= \sum_{j=1}^{l} \sum_{k=1}^{l} x_j x_k \int_0^\pi \cos((k-j)\lambda)d\lambda \\
&= \sum_{j=1}^{l} x_j^2 \int_0^\pi \cos(0)d\lambda + \sum_{1\le j\neq k\le l} x_j x_k \int_0^\pi \cos((k-j)\lambda)d\lambda \\
&= \pi \sum_{j=1}^{l} x_j^2 \\
&= \pi
\end{aligned}
$$

for $x \in \mathcal{X}$.
All in all, the above calculations yield

$$
x'\Sigma_l x \geq 2C_{\min}\pi^{D-1} \int_0^\pi \left| \sum_{j=1}^{l} x_j \mathrm{e}^{-ij\lambda} \right|^2 d\lambda = 2C_{\min}\pi^{D-1}.
$$

As a result,

$$
\lambda_{\min}(\Sigma_l) = \inf_{x\in\mathcal{X}} (x'\Sigma_l x) \geq K_D
$$

with $K_D := 2C_{\min}\pi^{D-1}$. □

Corollary 2 and Lemma 5 together imply the following result:

Corollary 3. *Given a time series ξ_n, $n \in \mathbb{N}$, satisfying Assumption 4,*

$$
\rho(\sigma(\xi_i, 1 \leq i \leq l), \sigma(\xi_j, k+l \leq j \leq k+2l-1)) \lesssim \frac{1}{K_D} l k^{-D} L_\gamma(k)
$$

with K_D as in Lemma 5 and with $f \lesssim g$ signifying that there exists a function h with $f \leq h$ and $h \sim g$.

We will see that under slightly stronger assumptions on the time series ξ_n, $n \in \mathbb{N}$, the upper bound on the maximal correlation achieved by Corollary 3 can be improved.
For this purpose, we specify Assumption 4:

Assumption 5. Let ξ_n, $n \in \mathbb{N}$, denote a stationary, long-range dependent Gaussian process with mean 0, variance 1, LRD parameter D and spectral density $f(\lambda) = |\lambda|^{D-1}L_f(\lambda)$ for a slowly varying function L_f which is bounded away from 0 on $[0, \pi]$. Moreover, assume that $\lim_{\lambda \to 0} L_f(\lambda) \in (0, \infty]$ exists.

The following Lemma is needed for establishing an upper bound which is smaller than that attained by Corollary 3.

Lemma 6 (Betken and Wendler (2015))**.** *Given a time series ξ_n, $n \in \mathbb{N}$, satisfying Assumption 5, there exist constants $C_D \in (0, \infty)$ and $l_0 \in \mathbb{N}$ such that*

$$\left| \sum_{i=1}^{l} x_i \right| \leq C_D l^{D/2}$$

for all $l \geq l_0$ and for all $x_1, \ldots, x_l \in \mathbb{R}$ with $\operatorname{Var}\left(\sum_{i=1}^{l} x_i \xi_i\right) = 1$.

Proof. The assertion of this lemma is equivalent to the following statement: there exists a constant $\tilde{C}_D \in (0, \infty)$, such that for all $x_1, \ldots, x_l \in \mathbb{R}$ with $\sum_{i=1}^{l} x_i = 1$

$$\operatorname{Var}\left(\sum_{i=1}^{l} x_i \xi_i\right) \geq \tilde{C}_D l^{-D}.$$

Assume that $x_1^\star, \ldots, x_l^\star \in \mathbb{R}$ with $\sum_{i=1}^{l} x_i^\star = 1$ denote the values of $x_1, \ldots, x_l$ minimizing $\operatorname{Var}\left(\sum_{i=1}^{l} x_i \xi_i\right)$. Then

$$\widehat{\mu}(\xi_1, \ldots, \xi_n) := \sum_{i=1}^{l} x_i^\star \xi_i$$

is the best linear unbiased estimator for $\mu := \operatorname{E} \xi_1$. For a process ζ_n, $n \in \mathbb{N}$, with spectral density $f_\zeta(x) = \frac{1}{2\pi}|1 - e^{ix}|^{D-1}$, we have

$$\operatorname{Var}\left(\widehat{\mu}(\zeta_1, \ldots, \zeta_n)\right) \geq C_1 l^{-D}$$

for a constant $C_1 \in (0, \infty)$ by Theorem 5.1 in Adenstedt (1974). We rewrite the spectral density f_ζ of ζ_n, $n \in \mathbb{N}$, with the help of the spectral density f of ξ_n, $n \in \mathbb{N}$, as

$$f_\zeta(\lambda) = f(\lambda) g(\lambda), \quad g(\lambda) := \frac{\left|1 - e^{i\lambda}\right|^{D-1}}{2\pi |\lambda|^{D-1} L_f(\lambda)}.$$

Note that the function g is bounded since L_f is bounded away from 0 by assumption. Hence, we have

$$\operatorname{Var}\left(\widehat{\mu}(\xi_1, \ldots, \xi_n)\right) \geq \frac{1}{g(0)} \operatorname{Var}\left(\widehat{\mu}(\zeta_1, \ldots, \zeta_n)\right) \geq \tilde{C}_D l^{-D}$$

for all $l \geq l_0$ by Lemma 4.4 in Adenstedt (1974). □

Beside a slightly stronger assumption on the spectral density of the Gaussian variables, a restriction on the growth of L_γ guarantees an upper bound on $\rho_{k,l}$ that improves the one achieved by Corollary 3.

Assumption 6. Let ξ_n, $n \in \mathbb{N}$, denote a stationary, long-range dependent Gaussian process with mean 0, variance 1 and covariance function $\gamma(k) := \mathrm{Cov}(\xi_1, \xi_{k+1}) = k^{-D} L_\gamma(k)$ for some parameter $D \in (0,1)$ and some slowly varying function L_γ. Assume that there exists a constant $K \in (0, \infty)$, such that for all $k \in \mathbb{N}$

$$\max_{k+1 \leq j \leq k+2l-2} |L_\gamma(k) - L_\gamma(j)| \leq K \frac{l}{k} \min \{L_\gamma(k), 1\}$$

for $l \in \{l_k, \dots, k\}$.

Lemma 7 (Betken and Wendler (2015))**.** *Given a time series ξ_n, $n \in \mathbb{N}$, satisfying Assumptions 5 and 6, there exist constants $C_1, C_2 \in (0, \infty)$, such that*

$$\rho\left(\sigma(\xi_i, 1 \leq i \leq l), \sigma(\xi_j, k+l \leq j \leq k+2l-1)\right) \\ \leq C_1 l^D k^{-D} L_\gamma(k) + C_2 l^2 k^{-D-1} \max\{L_\gamma(k), 1\}$$

for all $k \in \mathbb{N}$ and all $l \in \{l_k, \dots, k\}$.

Proof. It is shown in Kolmogorov and Rozanov (1960) that there exist real numbers $a_1, \dots, a_l$ and $b_1, \dots, b_l$, such that

$$\rho\left(\sigma(\xi_i, 1 \leq i \leq l), \sigma(\xi_j, k+l \leq j \leq k+2l-1)\right) = \mathrm{Cov}\left(\sum_{i=1}^{l} a_i \xi_i, \sum_{j=1}^{l} b_j \xi_{k+l-1+j}\right)$$

and $\mathrm{Var}\left(\sum_{i=1}^{l} a_i \xi_i\right) = \mathrm{Var}\left(\sum_{j=1}^{l} b_j \xi_{k+l-1+j}\right) = 1$. The triangular inequality yields

$$\left|\mathrm{Cov}\left(\sum_{i=1}^{l} a_i \xi_i, \sum_{j=1}^{l} b_j \xi_{k+l-1+j}\right)\right| \\ \leq \left|\sum_{i=1}^{l} a_i \sum_{j=1}^{l} b_j\right| |\gamma(k)| + \sum_{i=1}^{l} \sum_{j=1}^{l} |a_i||b_j| \left|\gamma(k) - \gamma(k+l-1+j-i)\right|.$$

We will treat the two summands on the right-hand side separately. For the first term, it follows by Lemma 6 that

$$\left|\sum_{i=1}^{l} a_i \sum_{j=1}^{l} b_j\right| |\gamma(k)| = \left|\sum_{i=1}^{l} a_i\right| \left|\sum_{j=1}^{l} b_j\right| |\gamma(k)| \leq C_1 l^D k^{-D} L_\gamma(k)$$

for some constant $C_1 \in (0, \infty)$.

In order to derive an upper bound for the second summand, note that

$$\begin{aligned}|\gamma(k)-\gamma(j)| &\leq L_\gamma(k)\left|k^{-D}-j^{-D}\right|+|L_\gamma(k)-L_\gamma(j)|j^{-D}\\ &\leq L_\gamma(k)\left(k^{-D}-(k+2l-2)^{-D}\right)+|L_\gamma(k)-L_\gamma(j)|k^{-D}\end{aligned}$$

for $j \in \{k+1,\ldots,k+2l-2\}$. By means of Taylor series expansion, we have

$$k^{-D}-(k+2l-2)^{-D}=Dk^{-D-1}(2l-2)-\frac{1}{2}D(D+1)\eta^{-D-2}(2l-2)^2$$

for some $\eta \in \{k,\ldots,k+2l-2\}$. Due to Assumption 6,

$$|L_\gamma(k)-L_\gamma(j)|k^{-D} \leq Klk^{-D-1}$$

for some constant $K \in (0,\infty)$. As a result, we get

$$|\gamma(k)-\gamma(j)| \leq Clk^{-D-1}\max\{L_\gamma(k),1\}$$

for some constant $C \in (0,\infty)$. By Hölder's inequality it then follows that

$$\sum_{i=1}^{l}|a_i| \leq \sqrt{l\sum_{i=1}^{l}a_i^2} \leq \sqrt{l}\sqrt{K_D} \quad\text{and}\quad \sum_{j=1}^{l}|b_j| \leq \sqrt{l\sum_{j=1}^{l}b_j^2} \leq \sqrt{l}\sqrt{K_D}.$$

Combining the previous inequalities, we conclude that

$$\sum_{i=1}^{l}|a_i|\sum_{j=1}^{l}|b_j|\,|\gamma(k)-\gamma(k+l+j-i)| \leq C_2l^2k^{-D-1}\max\{L_\gamma(k),1\}$$

for some constant $C_2 \in (0,\infty)$. □

3.1.2. Consistency

As stated in Section 3.1, the sampling-window method is consistent if

$$\sum_{k=\lfloor n^\beta\rfloor-l+1}^{m_n-l}\rho_{k,l}=o(m_n) \tag{3.3}$$

for some $\beta \in (0,1)$. In this section, (3.3) is shown to hold under two different sets of assumptions.

According to the following theorem, the subsampling procedure is consistent under non-restrictive conditions on the data-generating process, but for a severely limited choice of the blocklength. At the same time, the subsequent theorem takes a loss of generality concerning the data-generating process to the benefit of a less restrictive condition on the blocklength.

Theorem 8. *Suppose that ξ_n, $n \in \mathbb{N}$, is a time series satisfying Assumption 4. Moreover, let l_n, $n \in \mathbb{N}$, be an increasing, divergent sequence of integers. If $l_n = o(n^D)$, then, for some $\beta \in (0,1)$,*

$$\sum_{k=\lfloor n^\beta \rfloor - l_n + 1}^{m_n - l_n} \rho_{k,l_n} = o(m_n).$$

In particular, it follows that the sampling-window method is consistent under these assumptions.

Proof. As shown in the proof of Corollary 3, for some constant $C_D \in (0, \infty)$

$$\sum_{k=\lfloor n^\beta \rfloor - l_n + 1}^{m_n - l_n} \rho_{k,l_n} \lesssim l_n C_D \sum_{k=\lfloor n^\beta \rfloor - l_n + 1}^{m_n - l_n} k^{-D} L_\gamma(k), \text{ as } n \to \infty,$$

with L_γ such that $\operatorname{Cov}(\xi_1, \xi_{k+1}) = k^{-D} L_\gamma(k)$.
According to Proposition 1.1.4. in Pipiras and Taqqu (2017),

$$\sum_{k=1}^{m_n - l_n} k^{-D} L_\gamma(k) \sim \frac{L_\gamma(m_n - l_n)(m_n - l_n)^{1-D}}{1-D}, \text{ as } n \to \infty.$$

Since L_γ is slowly varying, it follows from Potter's bound that for any arbitrary small $\delta > 0$, there exist constants $C_\delta, x_\delta \in (0, \infty)$ such that $L_\gamma(x) \leq C_\delta x^\delta$ for $x \geq x_\delta$; see Theorem 1.5.6 in Bingham et al. (1987). As a result,

$$\frac{L_\gamma(m_n - l_n)(m_n - l_n)^{1-D}}{1-D} \leq \frac{C_\delta (m_n - l_n)^{1-D+\delta}}{1-D}$$

for sufficiently large n. Therefore, (3.3) holds if $l_n = o(n^D)$. □

Finally, the following result shows that under additional assumptions on the slowly varying functions L_f and L_γ, consistency holds for a less limited choice of the blocklength.

Theorem 9 (Betken and Wendler (2015))**.** *Suppose that ξ_n, $n \in \mathbb{N}$, is a time series satisfying Assumptions 5 and 6. Moreover, let l_n, $n \in \mathbb{N}$, be an increasing, divergent sequence of integers. If $l_n = \mathcal{O}\left(n^{(1+D)/2-\varepsilon}\right)$ for some $\varepsilon > 0$, then, for some $\beta \in (0,1)$,*

$$\sum_{k=\lfloor n^\beta \rfloor - l_n + 1}^{m_n - l_n} \rho_{k,l_n} = o(m_n).$$

In particular, it follows that the sampling-window method is consistent under these assumptions.

Remark 11. Theorem 8 and Theorem 9 both hold for Gaussian processes ξ_n, $n \in \mathbb{N}$, with spectral density $f(\lambda) = |\lambda|^{D-1} L_f(\lambda)$ characterized by $D < 1$. If $D > 1$, the process ξ_n, $n \in \mathbb{N}$, is strongly mixing due to Theorem 9.8 in Bradley (2007) and consistency follows from Corollary 3.2 in Politis and Romano (1994) for any blocklength l_n that tends to ∞ and satisfies $l_n = o(n)$.

Proof of Theorem 9. Let $\varepsilon > 0$. By assumption, $l_n \leq C_l n^{\alpha}$ for $\alpha := \frac{1}{2}(1 + D) - \varepsilon$ and some constant $C_l \in (0, \infty)$. As a consequence of Potter's Theorem, for every $\delta > 0$, there exists a constant $C_\delta \in (0, \infty)$ such that $L_\gamma(k) \leq C_\delta k^\delta$ for all $k \in \mathbb{N}$; see Theorem 1.5.6 in Bingham et al. (1987).
Moreover, we choose $\beta > \alpha$ and n large enough such that $l_n < \frac{1}{2}\lfloor n^\beta \rfloor$. According to this, Lemma 7 yields

$$
\begin{aligned}
&\frac{1}{m_n} \sum_{k=\lfloor n^\beta \rfloor - l_n + 1}^{m_n - l_n} \rho_{k, l_n} \\
&\leq C_1 l_n^D \frac{1}{m_n} \sum_{k=\lfloor n^\beta \rfloor - l_n + 1}^{m_n - l_n} k^{-D} L_\gamma(k) + C_2 \frac{l_n^2}{m_n} \sum_{k=\lfloor n^\beta \rfloor - l_n + 1}^{m_n - l_n} k^{-D-1} \max\{L_\gamma(k), 1\} \\
&\leq C_\delta C_1 \frac{l_n^D}{m_n} \sum_{k=\lfloor n^\beta \rfloor / 2}^{m_n - l_n} k^{-D+\delta} + C_\delta C_2 \frac{l_n^2}{m_n} \sum_{k=\lfloor n^\beta \rfloor / 2}^{m_n - l_n} k^{-D-1+\delta} \\
&\leq C \left(n^{D\alpha - \beta D + \beta\delta} + n^{2\alpha - \beta D - \beta + \delta\beta} \right)
\end{aligned}
$$

for some constant $C \in (0, \infty)$. By definition of α and for a suitable choice of β, the right-hand side of the above inequality converges to 0. □

Due to the assumptions on the blocklength in Theorems 8 and 9, an increase of the parameter D implies a stronger restriction on the range of possible values for l_n. A popular choice for the blocklength is $l_n = c\sqrt{n}$ for some constant c; see for example Hall et al. (1998). For long-range dependent time series with LRD parameter D, this choice is compatible with the requirements of Theorem 8 for every $D \in \left(\frac{1}{2}, 1\right)$ and with those of Theorem 9 for every $D \in (0, 1)$.
Neither of both consistency results makes any restrictive demands concerning the statistic T_n or the transformation G, such as finite moments of the data-generating variables, continuity of G or continuity of the considered statistics. As a result, both theorems are applicable to heavy-tailed random variables and robust test statistics. Yet, Theorems 8 and 9 impose conditions on the Gaussian variables that underlie the data-generating processes. In the following, it is shown that these assumptions hold for two standard examples of long-range dependent Gaussian processes.

Example 3 (Fractional Gaussian noise)**.** By means of Taylor series expansion, the covariance function of a fractional Gaussian noise time series ξ_n, $n \in \mathbb{N}$, with Hurst parameter H can be rewritten as

$$\gamma(k) = H(2H - 1) \left(k^{-D} + h(k) k^{-D-1} \right)$$

for $D = 2 - 2H$ and a function h, bounded by a finite constant M. Hence,

$$L_\gamma(k) = H(2H-1)\left(1 + h(k)k^{-1}\right)$$

and for all $j \geq k$

$$|L_\gamma(k) - L_\gamma(j)| \leq H(2H-1)\left|h(k)k^{-1} - h(j)j^{-1}\right| \leq 2H(2H-1)Mk^{-1}.$$

The above inequality implies Assumption 6.
Moreover, note that the spectral density f of a fractional Gaussian noise time series is given by

$$\begin{aligned} f(\lambda) &= C_H(1-\cos(\lambda)) \sum_{k=-\infty}^{\infty} |\lambda + 2k\pi|^{D-3} \\ &= \lambda^{D-1} C_H \frac{1-\cos(\lambda)}{\lambda^2} \frac{\sum_{k=-\infty}^{\infty} |\lambda + 2k\pi|^{D-3}}{\lambda^{D-3}} \end{aligned}$$

for some constant C_H; see Sinai (1976). The slowly varying function

$$L_f(\lambda) := C_H \frac{1-\cos(\lambda)}{\lambda^2} \frac{\sum_{k=-\infty}^{\infty} |\lambda + 2k\pi|^{D-3}}{\lambda^{D-3}}$$

is bounded away from 0, because

$$\frac{\sum_{k=-\infty}^{\infty} |\lambda + 2k\pi|^{D-3}}{\lambda^{D-3}} \geq 1,$$

and since $(1-\cos(\lambda))\lambda^{-2}$ is bounded away from 0. Moreover, $L_f(\lambda)$ decreases monotonically as λ approaches 0. As a result, $\lim_{\lambda \to 0} L_f(\lambda)$ exists due to the monotone convergence theorem. Therefore, Assumption 5 holds as well.

Example 4 (Gaussian FARIMA processes)**.** Let η_k, $k \in \mathbb{Z}$, be a Gaussian white noise process with variance $\sigma^2 = \operatorname{Var} \eta_0$. Then, for $d \in (0, 1/2)$, the process

$$\xi_i = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} \eta_{i-j}, \quad i \geq 1,$$

is a Gaussian FARIMA(0, d, 0) process.
By Corollary 1.3.4 in Pipiras and Taqqu (2017), we have

$$\gamma(k) = \sigma^2 \frac{\Gamma(1-2d)}{\Gamma(1-d)\Gamma(d)} \frac{\Gamma(k+d)}{\Gamma(k-d+1)}.$$

Stirling's formula yieldsf $\Gamma(x) = \left(\frac{2\pi}{x}\right)^{1/2} \left(\frac{x}{e}\right)^x \left(1 + \mathcal{O}(x^{-1})\right)$, so that

$$\log\left(\frac{\Gamma(k+d)}{\Gamma(k-d+1)}\right) = 1 - 2d + \log\left(\frac{(k+d)^{k+d-\frac{1}{2}}}{(k-d+\frac{1}{2})^{k-d+1}}\right) + \log\left(1 + \mathcal{O}\left(\frac{1}{k}\right)\right).$$

It follows that

$$\gamma(k) = k^{2d-1}\mathrm{e}^{1-2d}\left(\frac{k+d}{k}\right)^{k+d-\frac{1}{2}}\left(\frac{k}{k-d+1}\right)^{k-d+\frac{1}{2}}\left(1+\mathcal{O}\left(\frac{1}{k}\right)\right).$$

By means of Taylor series expansion,

$$\log\left[\left(\frac{k+d}{k}\right)^{k+d-\frac{1}{2}}\left(\frac{k}{k-d+1}\right)^{k-d+\frac{1}{2}}\right] = 2d-1+\mathcal{O}\left(\frac{1}{k}\right).$$

All in all, the previous calculations yield

$$\gamma(k) = k^{-D}\left(C+\mathcal{O}\left(\frac{1}{k}\right)\right)$$

for some constant C. Therefore, Assumption 6 follows in the same way as in Example 3. According to Pipiras and Taqqu (2017), the spectral density corresponding to a Gaussian FARIMA(0, d, 0) process is given by

$$f(\lambda) = \frac{\sigma^2}{2\pi}|1-e^{-i\lambda}|^{-2d} = |\lambda|^{D-1}\frac{\sigma^2}{2\pi}\left(\frac{|\lambda|}{|1-e^{-i\lambda}|}\right)^{1-D}$$

with $D = 1-2d \in (0,1)$. The slowly varying function

$$L_f(\lambda) := \frac{\sigma^2}{2\pi}\left(\frac{|\lambda|}{|1-e^{-i\lambda}|}\right)^{1-D}$$

is bounded away from 0 since $|1-e^{-i\lambda}| \leq |\lambda|$. Moreover, $\lim_{\lambda\to 0}\frac{|\lambda|}{|1-e^{-i\lambda}|} = \frac{1}{\sqrt{2}}$, so that $\lim_{\lambda\to 0} L_f(\lambda)$ exists and Assumption 5 holds.

Remark 12. Bai et al. (2016) show that the sampling-window method is consistent for a studentized version of the sample mean under the assumption of subordinated Gaussian processes, without any additional conditions on the slowly varying functions L_γ and L_f, but with a stronger restriction on the blocklength l_n, namely $l_n = o(n^D L_\gamma(n))$. In fact, this result can be easily extended to general statistics. In another article by Bai and Taqqu (2015), the validity of subsampling is shown to hold whenever $l_n = o(n)$, i.e. under the mildest possible assumption on the blocklength. However, in this case, the condition on the spectral density is slightly stronger than Assumption 5; the case $\lim_{\lambda\to 0} L_f(\lambda) = \infty$ is excluded.

3.2. Simulations

In the following, the finite sample performance of the sampling-window method is investigated in the context of testing for changes in the mean of a given set of observations $X_1, \ldots, X_n$. More precisely, we apply the subsampling procedure to decide on the testing problem:

$$H: \ \mathrm{E}\, X_1 = \cdots = \mathrm{E}\, X_n$$

against

$$A: \ \mathrm{E}\, X_1 = \cdots = \mathrm{E}\, X_k \neq \mathrm{E}\, X_{k+1} = \cdots = \mathrm{E}\, X_n \quad \text{for some } k \in \{1, \ldots, n-1\}.$$

For this purpose, recall that the Wilcoxon change-point test, considered in Section 1.2.2, is based on the statistic

$$W_n := \max_{1 \le k \le n-1} |W_{k,n}|\,, \ W_{k,n} = \sum_{i=1}^{k} \sum_{j=k+1}^{n} \left(1_{\{X_i \le X_j\}} - \frac{1}{2} \right),$$

while the self-normalized Wilcoxon change-point test, considered in Section 1.2.3, bases test decisions on an evaluation of the test statistic

$$SW_n(\tau_1, \tau_2) = \max_{\lfloor n\tau_1 \rfloor \le k \le \lfloor n\tau_2 \rfloor} \left| \frac{W_{k,n}}{V_{k,n}} \right|, \quad V_{k,n}^2 := \frac{1}{n} \sum_{t=1}^{k} S_t^2(1,k) + \frac{1}{n} \sum_{t=k+1}^{n} S_t^2(k+1,n)$$

with

$$S_t(j,k) := \sum_{h=j}^{t} \left(R_h - \bar{R}_{j,k} \right), \ \bar{R}_{j,k} := \frac{1}{k-j+1} \sum_{t=j}^{k} R_t,$$

and with $R_1, \ldots, R_n$ denoting the ranks of the observations.
The finite sample performance of the sampling-window method is compared to the performance of the change-point tests that generate test decisions on the basis of critical values obtained from the asymptotic distribution of the test statistics. The rejection rates of both testing procedures are computed for simulated subordinated Gaussian time series X_n, $n \in \mathbb{N}$, $X_n = G(\xi_n)$, where ξ_n, $n \in \mathbb{N}$, is a fractional Gaussian noise sequence generated by the function `fgnSim` from the `fArma` package in `R`.
We consider two different scenarios:

1. Normal margins: We choose $G(t) = t$, so that the variables X_n, $n \in \mathbb{N}$, are standard normally distributed.

2. Pareto margins: We choose

$$G(t) = \left(\frac{\alpha k^2}{(\alpha-1)^2(\alpha-2)} \right)^{-\frac{1}{2}} \left(k(\Phi(t))^{-\frac{1}{\alpha}} - \frac{\alpha k}{\alpha - 1} \right)$$

   with parameters $k, \alpha > 0$ and with Φ denoting the standard normal distribution function, so that the variables X_n, $n \in \mathbb{N}$, are Pareto(α, k)-distributed; see Example 1.

In both cases, the Hermite rank r of $1_{\{G(\xi_i)\leq x\}} - F(x), x \in \mathbb{R}$, equals 1; see Section 2.4. As a result, $\frac{1}{nd_{n,1}} W_n$ converges in distribution to

$$\sup_{t\in[0,1]} |B_H(t) - tB_H(1)| \left| \int_{\mathbb{R}} J_1(x)dF(x) \right|,$$

where, due to the consideration of time series generated by strictly monotone transformations of fractional Gaussian noise, $d_{n,1} \sim n^{1-\frac{D}{2}}$ and $\left|\int_{\mathbb{R}} J_1(x)dF(x)\right| = \frac{1}{2\sqrt{\pi}}$; see Theorem 2 in Dehling et al. (2013).

According to Theorem 1 in Betken (2016), the self-normalized test statistic $SW_n(\tau_1, \tau_2)$ converges in distribution to

$$\sup_{t\in[\tau_1,\tau_2]} \frac{|B_H(t) - tB_H(1)|}{\left\{\int_0^t V_H^2(s; 0, \lambda)ds + \int_t^1 V_H^2(s; t, 1)ds\right\}^{\frac{1}{2}}}$$

with

$$V_H(t; t_1, t_2) = B_H(t) - B_H(t_1) - \frac{t - t_1}{t_2 - t_1} \{B_H(t_2) - B_H(t_1)\}.$$

To generate test decisions based on the asymptotic distributions of the test statistics, critical values can be taken from Table 1 and Table 2 in Betken (2016).

For both testing procedures, the frequencies of rejections are reported in Table 3.1 and Table 3.2 for the self-normalized Wilcoxon change-point test, and in Table 3.3 and Table 3.4 for the non-self-normalized Wilcoxon test. The calculations are based on $5,000$ realizations of time series with sample sizes $n = 300$ and $n = 500$. For the applications of the sampling-window method, block lengths $l_n = \lfloor n^\gamma \rfloor$ with $\gamma \in \{0.4, 0.5, 0.6\}$ are considered. The values of the test statistic are compared to the 95%-quantile of its asymptotic distribution and the 95%-quantile of the empirical distribution function $\widehat{F}_{m_n,l_n}$, i.e. the significance level is chosen to be 5%.

For a comparison of the test statistics with the asymptotic critical values, the estimation of the Hermite rank r, the slowly varying function L_γ and the integral $\int J_1(x)dF(x)$ is neglected. Nevertheless, for every simulated time series, the Hurst parameter H is estimated by the local Whittle estimator $\widehat{H}$ as proposed in Künsch (1987). This estimator is based on an approximation of the spectral density by the periodogram at the Fourier frequencies. It depends on the spectral bandwidth parameter b_n which denotes the number of Fourier frequencies used for the estimation. If the bandwidth parameter satisfies $\frac{1}{b_n} + \frac{b_n}{n} \longrightarrow 0$, as $n \longrightarrow \infty$, the local Whittle estimator is a consistent estimator for H; see Robinson (1995). For convenience, we always chose $b_n = \lfloor n^{2/3} \rfloor$. The critical values corresponding to the estimated values of H are determined by linear interpolation. Under the alternative A, the power of the testing procedures is analyzed by considering different choices for the height of the level shift, denoted by h, and the location of the change-point, denoted by τ. In the tables, the columns that are superscribed by $h = 0$ correspond to the frequency of a type 1 error, i.e. the rejection rate under the hypothesis.

The following observations correspond to the expected behavior of change-point tests and can be made with respect to all four testing procedures and Gaussian as well as Pareto-distributed time series:

- An increasing sample size goes along with an improvement of the finite sample performance, i.e. the empirical size approaches the level of significance and the empirical power increases.
- The empirical power of the testing procedures increases when the height of the level shift increases.
- The empirical power is higher for breakpoints located in the middle of the sample than for change-point locations that lie close to the boundary of the testing region.

A comparison of the testing procedures with respect to the non-self-normalized Wilcoxon statistic shows that:

- For both testing procedures, the empirical size is, in most cases, not close to the nominal level of significance, ranging from 1.1% to 20.8% using subsampling and from 2.6% to 36.0% using asymptotic critical values.
- In general, the sampling-window method becomes more conservative for higher values of the Hurst parameter H, while test decisions based on the asymptotic distribution become more liberal.
- Under the alternative, test decisions which are based on asymptotic critical values yield a higher empirical power than the sampling-window method, especially for high values of H. However, a comparison of the rejection rates under the alternative has to be seen in view of the fact that under the hypothesis the rejection frequencies of the testing procedures differ.

A comparison of the testing procedures with respect to the self-normalized Wilcoxon statistic shows that:

- For Gaussian time series, the empirical size, computed on the basis of the statistic's asymptotic distribution, almost equals the level of significance of 5%. The sampling-window method yields rejection rates that slightly exceed the significance level.
- For Pareto(3, 1)-distributed time series, both testing procedures lead to similar results and tend to reject the hypothesis too often when there is no change.
- For fractional Gaussian noise time series, the sampling-window method yields considerably better power than the test which is based on asymptotic critical values.
- For Pareto(3, 1)-distributed time series, the empirical power that results from an application of the sampling-window method exceeds the empirical power achieved by test decisions based on asymptotic critical values. However, in this case, the deviations of the rejection rates are rather small.

- For both testing procedures and Gaussian as well as Pareto(3, 1)-distributed time series, the empirical size is not much affected by the value of the Hurst parameter H, while the empirical power tends to decrease as H increases.

All in all, the self-normalized Wilcoxon change-point test seems to be more reliable than the non-self-normalized change-point test. This may be due to the fact that in the original scaling of the Wilcoxon statistic, the estimator of the Hurst parameter enters as a power of the sample size n. Thus, a small error in the estimation of H might lead to a large error in the value of the test statistic. By using the sampling-window method for the self-normalized Wilcoxon statistic, an estimation of unknown parameters is avoided, so that the performances of the different testing procedures are similar.
In most cases covered by the simulations, the choice of the block length for the sampling-window method does not have a significant impact on the frequency of a type 1 error. Considering the classical Wilcoxon statistic, an increase of the block length results in a higher frequency of rejections. For applications of the sampling-window method to the self-normalized Wilcoxon change-point test, the choice of the block length has the opposite effect: an increase of the block length tends to go along with a decrease in power, especially for big values of the Hurst parameter H and Pareto-distributed random variables. For smaller values of H, the effect is not pronounced, though.

		sampling-window method						asymptotic distribution				
				$\tau = 0.25$		$\tau = 0.5$			$\tau = 0.25$		$\tau = 0.5$	
fGn	n	l_n	$h = 0$	$h = 0.5$	$h = 1$	$h = 0.5$	$h = 1$	$h = 0$	$h = 0.5$	$h = 1$	$h = 0.5$	$h = 1$
$H = 0.6$	300	9	0.041	0.263	0.700	0.502	0.952					
		17	0.064	0.313	0.742	0.570	0.964	0.044	0.209	0.521	0.424	0.861
		30	0.070	0.322	0.705	0.555	0.943					
	500	12	0.053	0.396	0.859	0.697	0.994					
		22	0.060	0.421	0.861	0.720	0.995	0.049	0.303	0.687	0.577	0.958
		41	0.069	0.411	0.829	0.697	0.991					
$H = 0.7$	300	9	0.057	0.155	0.412	0.291	0.759					
		17	0.070	0.171	0.423	0.313	0.763	0.053	0.108	0.268	0.228	0.611
		30	0.077	0.177	0.403	0.314	0.737					
	500	12	0.056	0.183	0.513	0.382	0.856					
		22	0.059	0.193	0.508	0.382	0.854	0.048	0.133	0.359	0.302	0.730
		41	0.065	0.192	0.476	0.387	0.819					
$H = 0.8$	300	9	0.070	0.126	0.251	0.223	0.526					
		17	0.067	0.117	0.234	0.208	0.494	0.048	0.081	0.144	0.141	0.362
		30	0.073	0.114	0.218	0.201	0.466					
	500	12	0.066	0.121	0.295	0.217	0.591					
		22	0.068	0.114	0.278	0.210	0.567	0.053	0.085	0.198	0.163	0.462
		41	0.069	0.119	0.257	0.205	0.532					
$H = 0.9$	300	9	0.093	0.126	0.208	0.209	0.462					
		17	0.074	0.097	0.161	0.169	0.397	0.057	0.065	0.106	0.125	0.308
		30	0.073	0.095	0.145	0.165	0.367					
	500	12	0.079	0.105	0.194	0.185	0.461					
		22	0.067	0.091	0.166	0.162	0.416	0.051	0.068	0.120	0.128	0.350
		41	0.063	0.087	0.146	0.152	0.391					

Table 3.1.: *Rejection rates of the self-normalized Wilcoxon change-point test obtained by comparison with asymptotic critical values (right) and subsampling (left) with block length* $l_n = \lfloor n^\gamma \rfloor$, $\gamma \in \{0.4, 0.5, 0.6\}$ *for fractional Gaussian noise time series of length* n *with Hurst parameter* H.

		sampling-window method						asymptotic distribution				
				$\tau = 0.25$		$\tau = 0.5$			$\tau = 0.25$		$\tau = 0.5$	
Pareto	n	l_n	$h = 0$	$h = 0.5$	$h = 1$	$h = 0.5$	$h = 1$	$h = 0$	$h = 0.5$	$h = 1$	$h = 0.5$	$h = 1$
$H = 0.6$	300	9	0.041	0.847	0.977	0.990	1.000					
		17	0.067	0.871	0.946	0.990	1.000	0.056	0.820	0.912	0.984	0.999
		30	0.070	0.831	0.946	0.979	1.000					
	500	12	0.055	0.947	0.997	0.999	1.000					
		22	0.066	0.946	0.994	0.999	1.000	0.061	0.920	0.970	0.996	1.000
		41	0.071	0.921	0.976	0.996	1.000					
$H = 0.7$	300	9	0.057	0.571	0.821	0.990	0.994					
		17	0.064	0.527	0.738	0.876	0.990	0.070	0.529	0.702	0.856	0.982
		30	0.077	0.527	0.738	0.842	0.975					
	500	12	0.066	0.693	0.904	0.949	0.999					
		22	0.068	0.684	0.893	0.942	0.998	0.076	0.663	0.820	0.940	0.995
		41	0.072	0.632	0.838	0.921	0.994					
$H = 0.8$	300	9	0.070	0.355	0.574	0.703	0.931					
		17	0.068	0.284	0.454	0.666	0.905	0.072	0.297	0.428	0.640	0.875
		30	0.073	0.284	0.454	0.633	0.857					
	500	12	0.064	0.401	0.609	0.738	0.948					
		22	0.063	0.379	0.581	0.714	0.933	0.069	0.369	0.510	0.715	0.920
		41	0.064	0.345	0.509	0.688	0.903					
$H = 0.9$	300	9	0.093	0.253	0.396	0.597	0.832					
		17	0.071	0.168	0.254	0.532	0.772	0.073	0.165	0.236	0.499	0.738
		30	0.073	0.168	0.254	0.482	0.729					
	500	12	0.073	0.256	0.405	0.585	0.839					
		22	0.064	0.219	0.340	0.547	0.802	0.068	0.199	0.296	0.529	0.782
		41	0.065	0.190	0.296	0.503	0.762					

Table 3.2.: *Rejection rates of the self-normalized Wilcoxon change-point test obtained by comparison with asymptotic critical values (right) and by subsampling (left) with block length* $l_n = \lfloor n^{\gamma} \rfloor$, $\gamma \in \{0.4, 0.5, 0.6\}$ *for Pareto(3, 1)-transformed fractional Gaussian noise time series of length* n *with Hurst parameter* H.

		sampling-window method						asymptotic distribution				
				$\tau = 0.25$		$\tau = 0.5$			$\tau = 0.25$		$\tau = 0.5$	
fGn	n	l_n	$h = 0$	$h = 0.5$	$h = 1$	$h = 0.5$	$h = 1$	$h = 0$	$h = 0.5$	$h = 1$	$h = 0.5$	$h = 1$
$H = 0.6$	300	9	0.066	0.20	0.232	0.386	0.591					
		17	0.054	0.223	0.411	0.439	0.784	0.026	0.096	0.160	0.223	0.727
		30	0.059	0.264	0.529	0.663	0.870					
	500	12	0.063	0.285	0.436	0.569	0.856					
		22	0.058	0.345	0.663	0.627	0.952	0.036	0.148	0.256	0.378	0.897
		41	0.062	0.397	0.789	0.683	0.975					
$H = 0.7$	300	9	0.052	0.080	0.088	0.162	0.302					
		17	0.049	0.095	0.158	0.206	0.466	0.035	0.067	0.228	0.167	0.665
		30	0.051	0.120	0.227	0.267	0.593					
	500	12	0.042	0.104	0.153	0.249	0.539					
		22	0.039	0.131	0.267	0.287	0.689	0.030	0.079	0.259	0.225	0.714
		41	0.046	0.160	0.373	0.343	0.789					
$H = 0.8$	300	9	0.028	0.030	0.031	0.054	0.092					
		17	0.029	0.038	0.048	0.075	0.179	0.077	0.153	0.421	0.245	0.673
		30	0.034	0.057	0.088	0.070	0.272					
	500	12	0.023	0.031	0.036	0.064	0.162					
		22	0.028	0.044	0.070	0.097	0.273	0.050	0.112	0.439	0.226	0.714
		41	0.039	0.071	0.129	0.137	0.391					
$H = 0.9$	300	9	0.009	0.010	0.006	0.016	0.020					
		17	0.009	0.014	0.009	0.021	0.060	0.36	0.484	0.739	0.524	0.830
		30	0.015	0.029	0.028	0.011	0.153					
	500	12	0.008	0.006	0.003	0.015	0.026					
		22	0.011	0.009	0.011	0.029	0.086	0.319	0.439	0.743	0.511	0.845
		41	0.021	0.021	0.032	0.058	0.197					

Table 3.3.: *Rejection rates of the non-self-normalized Wilcoxon change-point test obtained by comparison with asymptotic critical values (right) and by subsampling (left) with block length* $l_n = \lfloor n^{\gamma} \rfloor$, $\gamma \in \{0.4, 0.5, 0.6\}$ *for fractional Gaussian noise time series of length* n *with Hurst parameter* H.

			sampling-window method					asymptotic distribution				
				$\tau = 0.25$		$\tau = 0.5$			$\tau = 0.25$		$\tau = 0.5$	
Pareto	n	l_n	$h = 0$	$h = 0.5$	$h = 1$	$h = 0.5$	$h = 1$	$h = 0$	$h = 0.5$	$h = 1$	$h = 0.5$	$h = 1$
$H = 0.6$	300	9	0.170	0.949	0.742	0.991	0.923					
		17	0.130	0.963	0.861	0.996	0.991	0.108	0.938	0.985	0.998	1.000
		30	0.109	0.962	0.871	0.998	0.998					
	500	12	0.163	0.991	0.916	1.000	0.993					
		22	0.132	0.997	0.976	1.000	0.999	0.128	0.988	0.999	1.000	1.000
		41	0.114	0.997	0.989	1.000	1.000					
$H = 0.7$	300	9	0.224	0.785	0.568	0.939	0.796					
		17	0.175	0.802	0.680	0.955	0.949	0.179	0.833	0.969	0.974	0.999
		30	0.140	0.789	0.708	0.959	0.976					
	500	12	0.208	0.921	0.763	0.989	0.956					
		22	0.167	0.931	0.862	0.992	0.996	0.191	0.940	0.994	0.996	1.000
		41	0.143	0.925	0.891	0.994	0.998					
$H = 0.8$	300	9	0.203	0.508	0.326	0.743	0.565					
		17	0.160	0.496	0.347	0.776	0.808	0.204	0.729	0.925	0.918	0.993
		30	0.137	0.484	0.364	0.791	0.881					
	500	12	0.190	0.639	0.445	0.865	0.770					
		22	0.160	0.649	0.513	0.886	0.929	0.212	0.805	0.963	0.948	0.999
		41	0.137	0.626	0.556	0.890	0.961					
$H = 0.9$	300	9	0.128	0.150	0.077	0.320	0.336					
		17	0.097	0.128	0.071	0.403	0.550	0.309	0.712	0.901	0.848	0.966
		30	0.092	0.125	0.077	0.481	0.677					
	500	12	0.112	0.159	0.089	0.402	0.436					
		22	0.100	0.161	0.101	0.518	0.680	0.27	0.726	0.911	0.851	0.975
		41	0.095	0.170	0.106	0.571	0.771					

Table 3.4.: *Rejection rates of the non-self-normalized Wilcoxon change-point test obtained by comparison with asymptotic critical values (right) and by subsampling (left) with block length $l_n = \lfloor n^{\gamma} \rfloor$, $\gamma \in \{0.4, 0.5, 0.6\}$, for Pareto(3, 1)-transformed fractional Gaussian noise time series of length n with Hurst parameter H.*

4. Testing for change-points in LMSV time series

The analysis of financial time series, such as stock market prices, usually focuses on log-returns instead of the observed data itself. One of the reason why is that in general price data cannot be assumed to stem from stationary processes, whereas the log-returns display features of stationary time series. As an example, we consider the daily closing indices of Standard & Poor's 500 (S&P 500, in short) and its log-returns, defined by

$$L_t := \log R_t, \; R_t := \frac{P_t}{P_{t-1}},$$

where P_t denotes the value of the index on day t, in the period from January 2005 to December 2010; see Figure 4.1. The plots show that the considered time series exhibits *volatility clustering*, meaning that large price changes, i.e. log-returns with relatively large absolute values, tend to cluster.

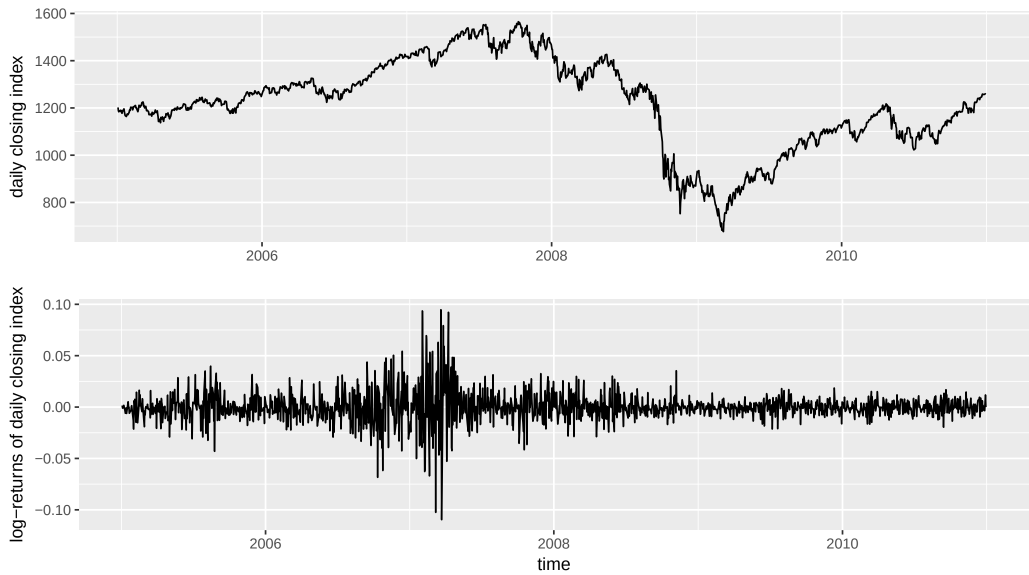

Figure 4.1.: *Daily closing index of Standard & Poor's 500 and its log-returns from January 2005 to December 2010. The data has been obtained from Google Finance.*

Comparing the plots of the sample autocorrelation function of the log-returns and the sample autocorrelation function of their absolute values in Figure 4.2, we observe a

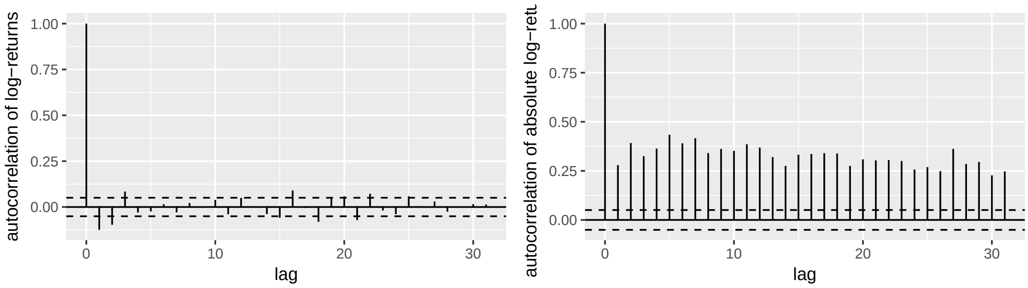

Figure 4.2.: *Sample autocorrelation of the log-returns and the absolute log-returns of Standard & Poor's 500 daily closing index from January 2005 to December 2010. The two dashed horizontal lines mark the bounds for the 95% confidence interval of the autocovariances under the assumption of data generated by white noise.*

phenomenon that is often encountered in the context of financial data: the log-returns of the index appear to be uncorrelated, whereas the absolute log-returns tend to be highly correlated.

Another characteristic of financial time series is the occurrence of heavy tails. In particular, probability distributions of log-returns often exhibit tails which are heavier than those of a normal distribution. For the S&P 500 data, this property is highlighted by the Q-Q plot in Figure 4.3.

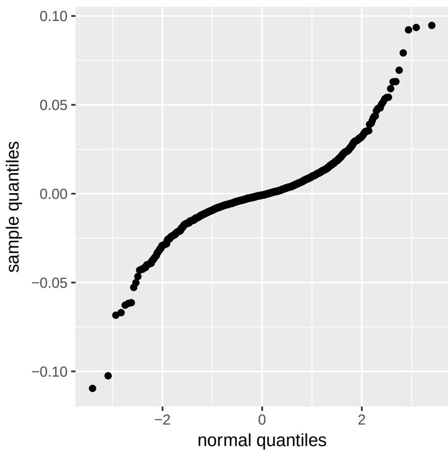

Figure 4.3.: *Q-Q plot for the log-returns of Standard & Poor's 500 daily closing index from January 2005 to December 2010.*

4.1. Long Memory Stochastic Volatility model

All of the previously described features of financial data can be covered by stochastic volatility models. While these models are often restricted to modeling a relatively fast decay of dependence in the data, the so-called long memory stochastic volatility (LMSV) model allows for long-range dependence. In this sense, the LMSV model can be considered as a generalization of stochastic volatility models considered, for example, in Taylor (1986). Initially, the LMSV model had been introduced by Breidt et al. (1998) and, independently, by Harvey (2002). An overview of stochastic volatility models with long-range dependence and their basic properties is given in Deo et al. (2006) and in Hurvich and Soulier (2009).

Definitions of LMSV time series X_n, $n \in \mathbb{N}$, are typically based on the assumption that

$$X_n = Y_n \varepsilon_n \quad \text{with} \quad Y_n = \exp\left(\frac{1}{2}\xi_n\right),$$

where ε_n, $n \in \mathbb{N}$, is an independent, identically distributed sequence of random variables with mean 0, and ξ_n, $n \in \mathbb{N}$, is a Gaussian process, independent of ε_n, $n \in \mathbb{N}$.

The LMSV model considered in this thesis generalizes the preceding concepts of stochastic volatility models with long-range dependence by allowing for general subordinated Gaussian sequences Y_n, $n \in \mathbb{N}$, and dependence between ξ_n, $n \in \mathbb{N}$, and ε_n, $n \in \mathbb{N}$.

Model 2. Let the data generating process X_n, $n \in \mathbb{N}$, satisfy

$$X_n = Y_n \varepsilon_n, \quad n \in \mathbb{N},$$

where ε_n, $n \in \mathbb{N}$, is an independent, identically distributed sequence of random variables with mean 0, and Y_n, $n \in \mathbb{N}$, is a long-range dependent subordinated Gaussian process according to Definition 7 in Chapter 1 with $Y_n = \sigma(\xi_n)$, $n \in \mathbb{N}$, for some stationary, long-range dependent Gaussian process ξ_n, $n \in \mathbb{N}$, with LRD parameter D and a non-negative measurable function σ (not equal to 0). More precisely, assume that ξ_n, $n \in \mathbb{N}$, admits a linear representation with respect to an independent, standard normally distributed sequence η_k, $k \in \mathbb{Z}$, i.e.

$$\xi_n = \sum_{k=1}^{\infty} c_k \eta_{n-k}, \quad n \in \mathbb{N},$$

with $\sum_{k=1}^{\infty} c_k^2 = 1$. Furthermore, suppose that (ε_n, η_n) is a sequence of independent, identically distributed random . A sequence of random variables X_n, $n \in \mathbb{N}$, which satisfies the previous assumption is called a *long memory stochastic volatility time series*.

Remark 13. Although the usage of the term *LMSV* often presupposes that the sequences ξ_n, $n \in \mathbb{N}$, and ε_n, $n \in \mathbb{N}$, are independent, less restrictive assumptions are imposed by Model 2: instead of claiming mutual independence of ξ_n, $n \in \mathbb{N}$, and ε_n, $n \in \mathbb{N}$, the sequence of random vectors (η_n, ε_n) is assumed to be independent.

In particular, this implies that for a fixed index j, the random variables ξ_j and ε_j are independent, while ξ_j may depend on ε_i, $i < j$. Except for this so-called *leverage effect*, Model 2 corresponds to the LMSV model considered in Kulik and Soulier (2011). In many cases, this version of the LMSV model is also referred to as *LMSV with leverage*. Note that, given an LMSV time series X_n, $n \in \mathbb{N}$,

$$\gamma(k) := \mathrm{Cov}(X_1, X_{k+1}) = 0, \ k \geq 1.$$

Assuming mutual independence of ξ_n, $n \in \mathbb{N}$, and ε_n, $n \in \mathbb{N}$, it follows that

$$\mathrm{Cov}(X_1^2, X_{k+1}^2) = (\mathrm{Var}\,\varepsilon_1)^2 \,\mathrm{Cov}(Y_1^2, Y_{k+1}^2).$$

Accordingly, the random variables X_n, $n \in \mathbb{N}$, are uncorrelated, while their squares inherit the dependence structure from the subordinated Gaussian sequence Y_n, $n \in \mathbb{N}$. Under the assumption that the marginal distribution of the random variables ε_n, $n \in \mathbb{N}$, has a regularly varying right tail, i.e. $\bar{F}_\varepsilon(x) := P(\varepsilon_1 > x) = x^{-\alpha}L(x)$ for some $\alpha > 0$ and a slowly varying function L, and that $\mathrm{E}\,\sigma^{\alpha+\delta}(\xi_1) < \infty$ for some $\delta > 0$, the tail behavior of the sequence X_n, $n \in \mathbb{N}$, can be related to the tail behavior of ε_n, $n \in \mathbb{N}$, by the following asymptotic equivalence:

$$P(X_1 > x) \sim E\sigma^\alpha(\xi_1)P(\varepsilon_1 > x), \text{ as } x \to \infty,$$

i.e. the variables X_n, $n \in \mathbb{N}$, inherit the tail behavior from the sequence ε_n, $n \in \mathbb{N}$. This result is known as Breiman's Lemma; see Breiman (1965).

4.2. The sequential empirical process of subordinated LMSV time series

In this section, we study the sequential empirical process of subordinated LMSV time series, i.e. given a time series X_n, $n \in \mathbb{N}$, satisfying the conditions specified by Model 2 and a measurable function ψ, we consider the two-parameter empirical process $e_n(x,t)$, $x \in [-\infty, \infty]$, $t \in [0,1]$, defined by

$$e_n(x,t) := \sum_{j=1}^{\lfloor nt \rfloor} \left(1_{\{\psi(X_j) \leq x\}} - F_{\psi(X_1)}(x) \right), \quad x \in [-\infty, \infty], \ t \in [0,1],$$

with $F_{\psi(X_1)}$ denoting the distribution function of $\psi(X_1)$.
As noted in Section 1.2.2, the asymptotic distribution of the Wilcoxon change-point test statistic can be derived from the limit behavior of the two-parameter empirical process. Beyond change-point analysis, the theory of empirical processes has many other applications in non-parametric statistics. For instance, the asymptotic distribution of certain classes of statistics such as V- or U-statistics can be derived from empirical process limit theorems. For this reason, the theoretical results in this section are of independent interest.
For subordinated Gaussian time series, the asymptotic behavior of the empirical process has been characterized in Dehling et al. (2013); see also Theorem 3 in Section 1.2.2. For subordinated long memory stochastic volatility time series, the following theorem constitutes an analogous result.

Theorem 10 (Betken and Kulik (2017)). *Suppose X_n, $n \in \mathbb{N}$, satisfies the conditions specified by Model 2 and let ψ be a measurable function. Moreover, assume that*

$$\int_{\mathbb{R}} \frac{d}{dy} \Psi_x(y) dy < \infty, \tag{4.1}$$

where $\Psi_x(y) := P\left(\psi(y\varepsilon_1) \leq x\right)$. Let r denote the Hermite rank of the class of functions $1_{\{\sigma(\xi_1)\leq x\}} - F_{\sigma(\xi_1)}(x)$, $x \in \mathbb{R}$, with $F_{\sigma(\xi_1)}$ denoting the distribution function of $\sigma(\xi_1)$, and let $d_{n,r}$ denote the normalizing sequence defined by (1.5) *in Section 1.2.2. If $rD < 1$, then*

$$\frac{1}{d_{n,r}} e_n(x,t) \xrightarrow{\mathcal{D}} \frac{1}{r!} J_r(\Psi_x \circ \sigma) Z_{r,H}(t), \ x \in [-\infty, \infty], \ t \in [0,1], \tag{4.2}$$

with $Z_{r,H}$ denoting an r-th order Hermite process, $H = 1 - \frac{rD}{2}$, and $\xrightarrow{\mathcal{D}}$ convergence in distribution with respect to the σ-field generated by the open balls in $D\left([-\infty,\infty] \times [0,1]\right)$ equipped with the supremum norm.

To prove Theorem 10, define a sequence of σ-fields $\mathcal{F}_j$, $j \in \mathbb{N}$, by

$$\mathcal{F}_j := \sigma\left(\varepsilon_j, \varepsilon_{j-1}, \ldots, \eta_j, \eta_{j-1}, \ldots\right),$$

i.e. $\mathcal{F}_j$ denotes the σ-field generated by the random variables $\varepsilon_j, \varepsilon_{j-1}, \ldots, \eta_j, \eta_{j-1}, \ldots$. Due to this construction, ε_j is independent of $\mathcal{F}_{j-1}$ and Y_n is $\mathcal{F}_{j-1}$-measurable.

To prove Theorem 10, we consider the following decomposition:

$$e_n(x,t) = M_n(x,t) + R_n(x,t),$$

where

$$M_n(x,t) := \sum_{j=1}^{\lfloor nt \rfloor} \left(1_{\{\psi(X_j)\leq x\}} - \mathrm{E}\left(1_{\{\psi(X_j)\leq x\}} \mid \mathcal{F}_{j-1}\right)\right),$$
$$R_n(x,t) := \sum_{j=1}^{\lfloor nt \rfloor} \left(\mathrm{E}\left(1_{\{\psi(X_j)\leq x\}} \mid \mathcal{F}_{j-1}\right) - F_{\psi(X_1)}(x)\right).$$

For fixed $x \in \mathbb{R}$, we write

$$M_n(t) := M_n(x,t) = \sum_{j=1}^{\lfloor nt \rfloor} \zeta_j(x)$$

with

$$\zeta_j(x) = 1_{\{\psi(X_j)\leq x\}} - \mathrm{E}\left(1_{\{\psi(X_j)\leq x\}} \mid \mathcal{F}_{j-1}\right).$$

Note that $M_n(t)$ is a martingale with respect to the filtration $\mathcal{F}_{\lfloor nt \rfloor}$, $t \geq 0$. For this reason, M_n is called the *Martingale part*. Since the asymptotic behavior of R_n is determined by the dependence structure of the long-range dependent subordinated Gaussian sequence $\sigma(\xi_n)$, $n \in \mathbb{N}$, we refer to R_n as the *long-range dependent part*.
The following two subsections show that $n^{-1/2} M_n(x,t) = \mathcal{O}_P(1)$ uniformly in x and t, while $d_{n,r}^{-1} R_n(x,t)$ converges in distribution to the limit process in formula (4.2). Theorem 10 then follows because $\sqrt{n} = \mathrm{o}(d_{n,r})$.

4.2.1. Martingale part

For fixed $x \in \mathbb{R}$, the following lemma characterizes the asymptotic distribution of the martingale part $M_n(x,t)$.

Lemma 8 (Betken and Kulik (2017))**.** *Under the conditions of Theorem 10,*

$$\frac{1}{\sqrt{n}} M_n(x,t) \xrightarrow{\mathcal{D}} \beta(x) B(t), \quad t \in [0,1],$$

where $B(t)$, $t \in [0,1]$, denotes a Brownian motion, $\beta^2(x) := \mathrm{E}\,\zeta_1^2(x)$ and convergence holds in $D[0,1]$ for every $x \in \mathbb{R}$.

Proof. Define

$$\zeta_{n,j} := n^{-\frac{1}{2}} \zeta_j(x) = X_{n,j}(x) - \mathrm{E}(X_{n,j}(x) \mid \mathcal{F}_{j-1})$$

with $X_{n,j}(x) := n^{-\frac{1}{2}} 1_{\{\psi(X_j) \leq x\}}$. To show convergence in $D[0,1]$, we apply the functional martingale central limit theorem stated in Theorem 18.2 of Billingsley (1999). For this, we have to show that

$$\sum_{j=1}^{\lfloor nt \rfloor} \mathrm{E}\left(\zeta_{n,j}^2 \mid \mathcal{F}_{j-1}\right) \xrightarrow{\mathcal{D}} \beta(x) t$$

for every $t \in [0,1]$, and that

$$\lim_{n \to \infty} \sum_{j=1}^{\lfloor nt \rfloor} \mathrm{E}\left(\zeta_{n,j}^2 1_{\{|\zeta_{n,j}| \geq \varepsilon\}}\right) = 0$$

for every $t \in [0,1]$ and every $\varepsilon > 0$. The latter requirement is known as *Lindeberg condition.* Due to Lemma 3.3 in Dvoretzky (1972), it suffices to prove that

$$\lim_{n \to \infty} \sum_{j=1}^{\lfloor nt \rfloor} \mathrm{E}\left(X_{n,j}^2(x) 1_{\left\{|X_{n,j}(x)| \geq \frac{\varepsilon}{2}\right\}}\right) = 0, \tag{4.3}$$

in order to show that the Lindeberg condition holds. As the indicator function is bounded, the summands on the left-hand side of (4.3) vanish for sufficiently large n

and hence convergence to 0 follows. Furthermore, the random variable $\mathrm{E}(\zeta_j^2(x) \mid \mathcal{F}_{j-1})$ can be considered as a measurable function of the random variable Y_j and therefore as a function of η_{j-n}, $n \in \mathbb{N}$. As a result, $\mathrm{E}\left(\zeta_n^2(x) \mid \mathcal{F}_{n-1}\right)$, $n \in \mathbb{N}$, is an ergodic sequence and it follows by the ergodic theorem that

$$\sum_{j=1}^{\lfloor nt \rfloor} \mathrm{E}\left(\zeta_{n,j}^2 \mid \mathcal{F}_{j-1}\right) = \frac{\lfloor nt \rfloor}{n} \frac{1}{\lfloor nt \rfloor} \sum_{j=1}^{\lfloor nt \rfloor} \mathrm{E}\left(\zeta_j^2(x) \mid \mathcal{F}_{j-1}\right) \xrightarrow{P} t\,\mathrm{E}\,\zeta_1^2(x)$$

for every $t \in [0,1]$. □

Lemma 8 implies tightness of $n^{-\frac{1}{2}} M_n(x,t)$ in $D[0,1]$ for fixed $x \in \mathbb{R}$. However, to prove Theorem 10, we have to verify tightness in $D([-\infty,\infty] \times [0,1])$. For this purpose, the notion of canonical stopping times for two-parameter processes is introduced:

Definition 12 (Ivanoff (1983))**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space, X a random element with values in $D([0,1] \times [0,1])$ and $\{\mathcal{F}(s,t) \mid (s,t) \in [0,1] \times [0,1]\}$ an increasing, right-continuous family of sub-σ-fields of $\mathcal{F}$ with $X(s,t)$ being adapted to $\mathcal{F}(s,t)$. A random variable S is called a *1-stopping time* relative to $\mathcal{F}(s,t)$ if $\{S \leq s\}$ is measurable with respect to $\mathcal{F}(s,1)$. A random variable T is called a *2-stopping time* relative to $\mathcal{F}(s,t)$ if $\{T \leq t\}$ is measurable with respect to $\mathcal{F}(1,t)$. S is a *canonical 1-stopping time* for X if S is a 1-stopping time belonging to a set of the form

$$S_\varepsilon := \left\{ S_0 = 0, S_i = \inf\left\{ s : \sup_{t \in B} |X(s,t) - X(S_{i-1},t)| > \varepsilon \right\}, i = 1, 2, \ldots \right\},$$

where $B \subseteq [0,1]$ is a closed set. T is a *canonical 2-stopping time* for X if T is a 2-stopping time belonging to a set of the form

$$T_\varepsilon := \left\{ T_0 = 0, T_i = \inf\left\{ t : \sup_{s \in B} |X(s,t) - X(s,T_{i-1})| > \varepsilon \right\}, i = 1, 2, \ldots \right\},$$

where $B \subseteq [0,1]$ is a closed set.

For the proof of two-parameter tightness, Theorem 3.1 in Ivanoff (1983) is needed:

Theorem 11 (Ivanoff (1983))**.** *Let $(\Omega_n, \mathcal{F}_n, P_n)$, $n \in \mathbb{N}$, be a sequence of probability spaces such that X_n is a random element with values in $D([0,1] \times [0,1])$ for each n, and the process $X_n(x,t)$ is adapted to a complete, right-continuous filtration $\mathcal{F}_n(x,t) \subseteq \mathcal{F}_n$. If $X_n(x,t)$ is tight for each $(x,t) \in [0,1] \times [0,1]$, and if for all sequences δ_n, $n \in \mathbb{N}$, $\delta_n \searrow 0$, each canonical 1-stopping time S_n and each canonical 2-stopping time T_n*

$$\sup_{t \in [0,1]} |X_n(S_n + \delta_n, t) - X_n(S_n, t)| \xrightarrow{P} 0, \quad \text{as} \quad n \to \infty, \tag{4.4}$$

$$\sup_{x \in [0,1]} |X_n(x, T_n + \delta_n) - X_n(x, T_n)| \xrightarrow{P} 0, \quad \text{as } n \to \infty, \tag{4.5}$$

then X_n, $n \in \mathbb{N}$, is tight in $D([0,1] \times [0,1])$.

Based on an application of Theorem 11, it is possible to give a proof for the following lemma which establishes tightness of $n^{-\frac{1}{2}}M_n(x,t)$ as a process with path space $D([-\infty,\infty]\times[0,1])$.

Lemma 9 (Betken and Kulik (2017)). *Under the conditions of Theorem 10, we have*

$$\frac{1}{\sqrt{n}}M_n(x,t)=\mathcal{O}_P(1)$$

in $D([-\infty,\infty]\times[0,1])$.

Proof. Let $H:[0,1]\times[0,1]\longrightarrow[-\infty,\infty]\times[0,1]$ be defined by $H(x,t):=(h(x),t)$ for some increasing isomorphism $h:[0,1]\longrightarrow[-\infty,\infty]$. In the following, it is shown that the conditions of Theorem 11 hold for the random process $X_n(x,t):=n^{-1/2}M_n\left(H(x,t)\right)$, $x,t\in[0,1]$.

Initially, note that $n^{-1/2}\left(M_n\circ H\right)(x,t)$ is tight for fixed $(x,t)\in[0,1]\times[0,1]$ due to Lemma 8. Recall that $\mathcal{F}_j=\sigma\left(\varepsilon_j,\varepsilon_{j-1},\ldots,\eta_j,\eta_{j-1},\ldots\right)$ and define

$$\mathcal{F}_n(x,t):=\mathcal{F}_{\lfloor nt\rfloor}$$

for all $x\in[0,1]$. Then, $X_n(x,t)$ is adapted to $\mathcal{F}_n(x,t)$. Moreover, the corresponding filtration is right-continuous.

Let T_n denote a canonical 2-stopping time for $X_n(x,t)$, measurable with respect to $\mathcal{F}_n(1,t)=\mathcal{F}_{\lfloor nt\rfloor}$, and define $\tau_n:=\lfloor nT_n\rfloor$. Note that

$$\begin{aligned}\left|X_n\left(x,T_n+\delta_n\right)-X_n\left(x,T_n\right)\right|&=\left|\frac{1}{\sqrt{n}}M_n(h(x),T_n+\delta_n)-\frac{1}{\sqrt{n}}M_n(h(x),T_n)\right|\\&=\left|\frac{1}{\sqrt{n}}\sum_{j=\tau_n+1}^{\tau_n+\lfloor n\delta_n\rfloor}\zeta_j(h(x))\right|\end{aligned}$$

with

$$\zeta_j(h(x))=1_{\{\psi(X_j)\leq h(x)\}}-\mathrm{E}\left(1_{\{\psi(X_j)\leq h(x)\}}\mid\mathcal{F}_{j-1}\right).$$

For $0\leq x\leq y\leq 1$, define

$$\begin{aligned}\Lambda_{1,n}(y,x)&:=\frac{1}{\sqrt{n}}\sum_{j=\tau_n+1}^{\tau_n+\lfloor n\delta_n\rfloor}\left(1_{\{h(x)<\psi(X_j)\leq h(y)\}}-\mathrm{E}\left(1_{\{h(x)<\psi(X_j)\leq h(y)\}}\mid\mathcal{F}_{j-1}\right)\right),\\\Lambda_{1,n}(y)&:=\Lambda_{1,n}(y,0).\end{aligned}$$

In order to show (4.5), we have to prove that

$$\sup_{y\in[0,1]}\left|\Lambda_{1,n}(y)\right|\xrightarrow{P}0,\quad\text{as } n\to\infty.$$

Prior to the proof, we establish the following result:

Lemma 10 (Betken and Kulik (2017)). *Under the conditions of Theorem 10,*

$$\operatorname{Var}\Lambda_{1,n}(y,x) = \operatorname{E}\Lambda_{1,n}^2(y,x) \leq \frac{1}{n}\sum_{j=1}^{\lfloor n\delta_n\rfloor}\left(F_{\phi(X_{j+\tau_n})}(y) - F_{\phi(X_{j+\tau_n})}(x)\right)$$

with $F_{\phi(X_{j+\tau_n})}(x) := P\left(\phi(X_{j+\tau_n}) \leq x\right)$, $\phi := h^{-1}\circ\psi$, *and for* $x \leq y$.

Proof. Write

$$\Lambda_{1,n}(y,x) = \frac{1}{\sqrt{n}}\sum_{j=\tau_n+1}^{\tau_n+\lfloor n\delta_n\rfloor}\lambda_{1,j}(y,x)$$

with

$$\lambda_{1,j}(y,x) := 1_{\{h(x)<\psi(X_j)\leq h(y)\}} - \operatorname{E}\left(1_{\{h(x)<\psi(X_j)\leq h(y)\}} \mid \mathcal{F}_{j-1}\right).$$

Since $\{\tau_n = k\}$ is measurable with respect to $\mathcal{F}_j$ for all $j \geq k$, it follows that

$$\begin{aligned}
&\operatorname{E}\left(\sum_{j=\tau_n+1}^{\tau_n+\lfloor n\delta_n\rfloor}\lambda_{1,j}(y,x)\right) = \sum_{k=1}^{n}\sum_{j=1}^{\lfloor n\delta_n\rfloor}\operatorname{E}\left(\lambda_{1,j+k}(y,x)1_{\{\tau_n=k\}}\right)\\
&= \sum_{k=1}^{n}\sum_{j=1}^{\lfloor n\delta_n\rfloor}\operatorname{E}\left(1_{\left\{x<\phi(X_{j+k})\leq y\right\}}1_{\{\tau_n=k\}} - \operatorname{E}\left(1_{\left\{x<\phi(X_{j+k})\leq y\right\}}1_{\{\tau_n=k\}} \mid \mathcal{F}_{j+k-1}\right)\right)\\
&= 0.
\end{aligned}$$

Furthermore,

$$\operatorname{E}\left[\left(\sum_{j=\tau_n+1}^{\tau_n+\lfloor n\delta_n\rfloor}\lambda_{1,j}(y,x)\right)^2\right] = \sum_{k=1}^{n}\sum_{i=1}^{\lfloor n\delta_n\rfloor}\sum_{j=1}^{\lfloor n\delta_n\rfloor}\operatorname{E}\left(\lambda_{1,i+k}(y,x)\lambda_{1,j+k}(y,x)1_{\{\tau_n=k\}}\right).$$

For $i < j$,

$$\begin{aligned}
&\lambda_{1,i+k}(y,x)\lambda_{1,j+k}(y,x)1_{\{\tau_n=k\}}\\
&= 1_{\{x<\phi(X_{i+k})\leq y\}}1_{\left\{x<\phi(X_{j+k})\leq y\right\}}1_{\{\tau_n=k\}}\\
&\quad + \operatorname{E}\left(1_{\left\{x<\phi(X_{j+k})\leq y\right\}}1_{\{\tau_n=k\}}\operatorname{E}\left(1_{\{x<\phi(X_{i+k})\leq y\}} \mid \mathcal{F}_{i+k-1}\right) \mid \mathcal{F}_{j+k-1}\right)\\
&\quad - \operatorname{E}\left(1_{\{x<\phi(X_{i+k})\leq y\}}1_{\left\{x<\phi(X_{j+k})\leq y\right\}}1_{\{\tau_n=k\}} \mid \mathcal{F}_{j+k-1}\right)\\
&\quad - 1_{\left\{x<\phi(X_{j+k})\leq y\right\}}1_{\{\tau_n=k\}}\operatorname{E}\left(1_{\{x<\phi(X_{i+k})\leq y\}} \mid \mathcal{F}_{i+k-1}\right).
\end{aligned}$$

As a result,

$$\operatorname{E}\left(\lambda_{1,i+k}(y,x)\lambda_{1,j+k}(y,x)1_{\{\tau_n=k\}}\right) = 0.$$

Moreover, we have

$$
\begin{aligned}
&\sum_{k=1}^{n} \mathrm{E}\left(\lambda_{1,i+k}^2(y,x) 1_{\{\tau_n=k\}}\right) \\
=&\sum_{k=1}^{n} \mathrm{E}\left[\left(1_{\{x<\phi(X_{i+k})\le y\}} - \mathrm{E}\left(1_{\{x<\phi(X_{i+k})\le y\}} \mid \mathcal{F}_{i+k-1}\right)\right)^2 1_{\{\tau_n=k\}}\right] \\
=&\sum_{k=1}^{n} \mathrm{E}\Big[\mathrm{E}\left(1_{\{x<\phi(X_{i+k})\le y\}}\, \mathrm{E}\left(1_{\{x<\phi(X_{i+k})\le y\}} 1_{\{\tau_n=k\}} \mid \mathcal{F}_{i+k-1}\right) \mid \mathcal{F}_{i+k-1}\right) \\
&\quad + 1_{\{x<\phi(X_{i+k})\le y\}} 1_{\{\tau_n=k\}} - 2 1_{\{x<\phi(X_{i+k})\le y\}}\, \mathrm{E}\left(1_{\{x<\phi(X_{i+k})\le y\}} 1_{\{\tau_n=k\}} \mid \mathcal{F}_{i+k-1}\right)\Big] \\
\le&\sum_{k=1}^{n} \mathrm{E}\left(1_{\{x<\phi(X_{i+k})\le y\}} 1_{\{\tau_n=k\}}\right) = \sum_{k=1}^{n} \mathrm{E}\left(1_{\{x<\phi(X_{i+\tau_n})\le y\}} 1_{\{\tau_n=k\}}\right) \\
=&\,\mathrm{E}\left(1_{\{x<\phi(X_{i+\tau_n})\le y\}}\right) = F_{\phi(X_{i+\tau_n})}(y) - F_{\phi(X_{i+\tau_n})}(x).
\end{aligned}
$$

Note that $\mathrm{E}(\Lambda_{1,n}(y)) = 0$. As a result, it follows that

$$
\mathrm{Var}\, \Lambda_{1,n}(y,x) = \mathrm{E}\, \Lambda_{1,n}^2(y,x) \le \frac{1}{n} \sum_{j=1}^{\lfloor n\delta_n \rfloor} \left(F_{\phi(X_{j+\tau_n})}(y) - F_{\phi(X_{j+\tau_n})}(x)\right).
$$

□

Due to Lemma 10,

$$
\mathrm{Var}\, \Lambda_{1,n}(y) \le \frac{1}{n} \sum_{i=1}^{\lfloor n\delta_n \rfloor} F_{\phi(X_{i+\tau_n})}(y) \le \delta_n \longrightarrow 0.
$$

In addition, $\mathrm{E}(\Lambda_{1,n}(y)) = 0$, so that $\Lambda_{1,n}(y)$ converges to 0 in probability. This implies convergence of the finite-dimensional distributions of $\Lambda_{1,n}(y)$, $y \in [0,1]$, considered as a process with values in $D\,[0,1]$.

Proof of (4.5)**.** In order to establish (4.5), it remains to show tightness of $\Lambda_{1,n}(y)$, $y \in [0,1]$. For this, the argument that proves Theorem 15.6 in Billingsley (1968) is adopted. For any function v in $D[0,1]$, define the modulus $\omega_v(\delta)$ by

$$
\omega_v(\delta) = \sup \min \left\{|v(x) - v(x_1)|, |v(x_2) - v(x)|\right\},
$$

where the supremum extends over $x_1, x, x_2 \in [0,1]$ with $x_1 \le x \le x_2$, $x_2 - x_1 \le \delta$. Given the convergence of the finite-dimensional distributions, Theorem 15.4 in Billingsley (1968) implies that it suffices to show that for each $\varepsilon, \eta > 0$, there exists a δ, $0 < \delta < 1$, and an $n_0 \in \mathbb{N}$ such that

$$
P\left(\omega_{\Lambda_{1,n}}(\delta) \ge \varepsilon\right) \le \eta \quad \text{for all } n \ge n_0.
$$

Define

$$M_m := \max_{0 \le i \le j \le k \le m} \min \{|S_j - S_i|, |S_k - S_j|\}, \quad \text{where } S_i = \Lambda_{1,n}(\tau + \frac{i}{m}\delta).$$

Following the proof of Theorem 12.5 in Billingsley (1968), there exists an $n_0 \in \mathbb{N}$ such that

$$P(M_m \ge \lambda) \le \frac{K}{\lambda^2} \sum_{i=1}^{m} u_i$$

holds for all positive λ, some constant K and $n \ge n_0$, if

$$P(|S_j - S_i| \ge \lambda, |S_k - S_j| \ge \lambda) \le \frac{\varepsilon_n}{\lambda^2} \sum_{i<l\le k} u_l, \; 0 \le i \le j \le k \le m,$$

for some sequence ε_n, $n \in \mathbb{N}$, converging to 0.
For $x_1 \le x \le x_2$, it follows by the Cauchy - Schwarz inequality for expected values and Lemma 10 that

$$\begin{aligned}
&\mathrm{E}\,|\Lambda_{1,n}(x_2, x)|\,|\Lambda_{1,n}(x, x_1)| \\
=&\mathrm{E}\left(\frac{1}{\sqrt{n}}\left|\sum_{j=\tau_n+1}^{\tau_n+\lfloor n\delta_n\rfloor} \lambda_{1,j}(x_2, x)\right| \frac{1}{\sqrt{n}}\left|\sum_{j=\tau_n+1}^{\tau_n+\lfloor n\delta_n\rfloor} \lambda_{1,j}(x, x_1)\right|\right) \\
\le&\sqrt{\frac{1}{n}\mathrm{E}\left[\left(\sum_{j=\tau_n+1}^{\tau_n+\lfloor n\delta_n\rfloor} \lambda_{1,j}(x_2, x)\right)^2\right] \frac{1}{n}\mathrm{E}\left[\left(\sum_{j=\tau_n+1}^{\tau_n+\lfloor n\delta_n\rfloor} \lambda_{1,j}(x, x_1)\right)^2\right]} \\
\le&\frac{1}{n}\sqrt{\sum_{i=1}^{\lfloor n\delta_n\rfloor}\left(F_{\phi(X_{i+\tau_n})}(x_2) - F_{\phi(X_{i+\tau_n})}(x)\right)}\sqrt{\sum_{i=1}^{\lfloor n\delta_n\rfloor}\left(F_{\phi(X_{i+\tau_n})}(x) - F_{\phi(X_{i+\tau_n})}(x_1)\right)} \\
\le&\frac{1}{n}\sum_{i=1}^{\lfloor n\delta_n\rfloor}\left(F_{\phi(X_{i+\tau_n})}(x_2) - F_{\phi(X_{i+\tau_n})}(x_1)\right). \qquad (4.6)
\end{aligned}$$

The Markov inequality yields

$$\begin{aligned}
P(|S_j - S_i| \ge \lambda, |S_k - S_j| \ge \lambda) &\le P\left(|S_j - S_i|\,|S_k - S_j| \ge \lambda^2\right) \\
&\le \frac{1}{\lambda^2}\mathrm{E}\,|S_j - S_i|\,|S_k - S_j|.
\end{aligned}$$

Therefore, it follows by (4.6) that for some $\gamma \in (0, 1)$,

$$P(|S_j - S_i| \ge \lambda, |S_k - S_j| \ge \lambda) \le \frac{\delta_n^{\gamma}}{\lambda^2} \sum_{i<l\le k} u_l,$$

where

$$u_l := \frac{\delta_n^{-\gamma}}{n} \sum_{j=1}^{\lfloor n\delta_n \rfloor} \left(F_{\phi(X_{j+\tau_n})}\left(\tau + \frac{l}{m}\delta\right) - F_{\phi(X_{j+\tau_n})}\left(\tau + \frac{l-1}{m}\delta\right)\right).$$

As a result, we get

$$P\left(M_m \geq \varepsilon\right) \leq \frac{K}{\varepsilon^2} \frac{\delta_n^{-\gamma}}{n} \sum_{j=1}^{\lfloor n\delta_n \rfloor} \left(F_{\phi(X_{j+\tau_n})}(\tau + \delta) - F_{\phi(X_{j+\tau_n})}(\tau)\right). \tag{4.7}$$

Define

$$\omega(\Lambda_{1,n}, [\tau, \tau+\delta]) := \sup \min \left\{ |\Lambda_{1,n}(x, x_1)|, |\Lambda_{1,n}(x_2, x)| \right\},$$

where the supremum extends over x_1, x, x_2 satisfying $\tau \leq x_1 \leq x \leq x_2 \leq \tau + \delta$. Letting m tend to ∞ in (4.7) yields

$$P\left(\omega(\Lambda_{1,n}, [\tau, \tau+\delta]) \geq \varepsilon\right) \leq \frac{K}{\varepsilon^2} \frac{\delta_n^{-\gamma}}{n} \sum_{j=1}^{\lfloor n\delta_n \rfloor} \left(F_{\phi(X_{j+\tau_n})}(\tau + \delta) - F_{\phi(X_{j+\tau_n})}(\tau)\right) \tag{4.8}$$

due to right-continuity of $\Lambda_{1,n}(y)$, $y \in [0,1]$.
Suppose that $\delta = \frac{1}{u}$ for some integer u and assume that

$$\omega(\Lambda_{1,n}, [2i\delta, (2i+2)\delta]) \leq \varepsilon, \qquad 0 \leq i \leq u-1, \tag{4.9}$$

$$\omega(\Lambda_{1,n}, [(2i+1)\delta, (2i+3)\delta]) \leq \varepsilon, \qquad 0 \leq i \leq u-2. \tag{4.10}$$

If $x_1 \leq x \leq x_2$ and $x_2 - x_1 \leq \delta$, then x_1 and x_2 both lie in one of the $2u-1$ intervals $[2i\delta, (2i+2)\delta]$, $0 \leq i \leq u-1$, $[(2i+1)\delta, (2i+3)\delta]$, $0 \leq i \leq u-2$, so that

$$\min \left\{ \left|\Lambda_{1,n}(x, x_1)\right|, \left|\Lambda_{1,n}(x_2, x)\right| \right\} \leq \varepsilon.$$

Thus, (4.9) and (4.10) together imply $\omega_{\Lambda_{1,n}}(\delta) \leq \varepsilon$. It now follows by (4.8) that

$$P\left(\omega_{\Lambda_{1,n}}(\delta) \geq \varepsilon\right) \leq \frac{K}{\varepsilon^2}\left(\Sigma' + \Sigma''\right),$$

where each of Σ' and Σ'' is a sum of the form

$$\begin{aligned} &\sum_{k=1}^{l} \frac{\delta_n^{-\gamma}}{n} \sum_{j=1}^{\lfloor n\delta_n \rfloor} \left(F_{\phi(X_{j+\tau_n})}(x_k) - F_{\phi(X_{j+\tau_n})}(x_{k-1})\right) \\ =&\frac{\delta_n^{-\gamma}}{n} \sum_{j=1}^{\lfloor n\delta_n \rfloor} \left(F_{\phi(X_{j+\tau_n})}(x_l) - F_{\phi(X_{j+\tau_n})}(x_0)\right) \\ \leq&\delta_n^{1-\gamma} \end{aligned}$$

with $0 \leq x_1 \leq \ldots \leq x_l \leq 1$ and $x_k - x_{k-1} \leq 2\delta$. Hence, we may conclude that

$$P\left(\omega_{\Lambda_{1,n}}(\delta) \geq \varepsilon\right) \leq \frac{2K}{\varepsilon^2}\delta_n^{1-\gamma}.$$

Since the right-hand side of the above inequality converges to 0, (4.5) has been proved.

Proof of (4.4). Let S_n denote a canonical 1-stopping time for $X_n(x,t)$.
For $0 \leq s \leq t \leq 1$, define

$$\Lambda_{2,n}(t,s) := \frac{1}{\sqrt{n}} \sum_{j=\lfloor ns \rfloor+1}^{\lfloor nt \rfloor} \left(1_{\{0<\phi(X_j)-S_n\leq \delta_n\}} - \mathrm{E}\left(1_{\{0<\phi(X_j)-S_n\leq \delta_n\}} \mid \mathcal{F}_{j-1} \right) \right),$$
$$\Lambda_{2,n}(t) := \Lambda_{2,n}(t,0),$$

where $\phi := h^{-1} \circ \psi$. Note that

$$\sup_{t\in[0,1]} |X_n(S_n+\delta_n,t) - X_n(S_n,t)| = \sup_{t\in[0,1]} |\Lambda_{2,n}(t)|.$$

Prior to the proof of (4.4), we establish the following result:

Lemma 11 (Betken and Kulik (2017)). *Under the conditions of Theorem 10,*

$$\mathrm{Var}\, \Lambda_{2,n}(t,s) = \mathrm{E}\, \Lambda_{2,n}^2(t,s) \leq \frac{1}{n} \sum_{i=\lfloor ns \rfloor+1}^{\lfloor nt \rfloor} \left(F_{\phi(X_i)-S_n}(\delta_n) - F_{\phi(X_i)-S_n}(0) \right)$$

with $F_{\phi(X_i)-S_n}(x) := P\left(\phi(X_i) - S_n \leq x\right)$ *and for* $s \leq t$.

Proof. Note that

$$\Lambda_{2,n}(t,s) = \frac{1}{\sqrt{n}} \sum_{j=\lfloor ns \rfloor+1}^{\lfloor nt \rfloor} \lambda_{2,j}(\delta_n, 0),$$

where

$$\lambda_{2,j}(y,x) := 1_{\{x<\phi(X_j)-S_n\leq y\}} - \mathrm{E}\left(1_{\{x<\phi(X_j)-S_n\leq y\}} \mid \mathcal{F}_{j-1} \right).$$

Moreover, we have

$$\mathrm{E}\left(\sum_{j=\lfloor ns \rfloor+1}^{\lfloor nt \rfloor} \lambda_{2,j}(\delta_n, 0) \right) = 0.$$

Furthermore,

$$\mathrm{Var}\left(\sum_{j=\lfloor ns \rfloor+1}^{\lfloor nt \rfloor} \lambda_{2,j}(\delta_n, 0) \right) = \mathrm{E}\left[\left(\sum_{j=\lfloor ns \rfloor+1}^{\lfloor nt \rfloor} \lambda_{2,j}(\delta_n, 0) \right)^2 \right]$$
$$= \sum_{i=\lfloor ns \rfloor+1}^{\lfloor nt \rfloor} \sum_{j=\lfloor ns \rfloor+1}^{\lfloor nt \rfloor} \mathrm{E}\left(\lambda_{2,i}(\delta_n, 0) \lambda_{2,j}(\delta_n, 0) \right).$$

The canonical 1-stopping time S_n takes on only countably many values s_i, $i \in \mathbb{N}$; see Ivanoff (1983). Note that for $i < j$

$$\begin{aligned}
\lambda_{2,i}(\delta_n,0)\lambda_{2,j}(\delta_n,0) &= \sum_{k=1}^{\infty} \lambda_{2,i}(\delta_n,0)\lambda_{2,j}(\delta_n,0) 1_{\{S_n=s_k\}} \\
&= \sum_{k=1}^{\infty} 1_{\{S_n=s_k\}} \Big\{ 1_{\{0<\phi(X_i)-s_k\le\delta_n\}} 1_{\{0<\phi(X_j)-s_k\le\delta_n\}} \\
&\quad + \mathrm{E}\left(1_{\{0<\phi(X_j)-s_k\le\delta_n\}} \mathrm{E}\left(1_{\{0<\phi(X_i)-s_k\le\delta_n\}} \mid \mathcal{F}_{i-1}\right) \mid \mathcal{F}_{j-1}\right) \\
&\quad - \mathrm{E}\left(1_{\{0<\phi(X_i)-s_k\le\delta_n\}} 1_{\{0<\phi(X_j)-s_k\le\delta_n\}} \mid \mathcal{F}_{j-1}\right) \\
&\quad - 1_{\{0<\phi(X_j)-s_k\le\delta_n\}} \mathrm{E}\left(1_{\{0<\phi(X_i)-s_k\le\delta_n\}} \mid \mathcal{F}_{i-1}\right) \Big\} \\
&= 1_{\{0<\phi(X_i)-S_n\le\delta_n\}} 1_{\{0<\phi(X_j)-S_n\le\delta_n\}} \\
&\quad + \mathrm{E}\left(1_{\{0<\phi(X_j)-S_n\le\delta_n\}} \mathrm{E}\left(1_{\{0<\phi(X_i)-S_n\le\delta_n\}} \mid \mathcal{F}_{i-1}\right) \mid \mathcal{F}_{j-1}\right) \\
&\quad - \mathrm{E}\left(1_{\{0<\phi(X_i)-S_n\le\delta_n\}} 1_{\{0<\phi(X_j)-S_n\le\delta_n\}} \mid \mathcal{F}_{j-1}\right) \\
&\quad - 1_{\{0<\phi(X_j)-S_n\le\delta_n\}} \mathrm{E}\left(1_{\{0<\phi(X_i)-S_n\le\delta_n\}} \mid \mathcal{F}_{i-1}\right).
\end{aligned}$$

As a result, $\mathrm{E}\left(\lambda_{2,i}(\delta_n,0)\lambda_{2,j}(\delta_n,0)\right) = 0$. Moreover,

$$\begin{aligned}
&\mathrm{E}\,\lambda_{2,j}^2(\delta_n,0) \\
=& \mathrm{E}\left(1_{\{0<\phi(X_j)-S_n\le\delta_n\}}\right) - 2\,\mathrm{E}\left(1_{\{0<\phi(X_j)-S_n\le\delta_n\}} \mathrm{E}\left(1_{\{0<\phi(X_j)-S_n\le\delta_n\}} \mid \mathcal{F}_{j-1}\right)\right) \\
+& \mathrm{E}\left(\mathrm{E}\left(1_{\{0<\phi(X_j)-S_n\le\delta_n\}} \mathrm{E}\left(1_{\{0<\phi(X_j)-S_n\le\delta_n\}} \mid \mathcal{F}_{j-1}\right) \mid \mathcal{F}_{j-1}\right)\right) \\
=& \mathrm{E}\left(1_{\{0<\phi(X_j)-S_n\le\delta_n\}}\right) - \mathrm{E}\left(1_{\{0<\phi(X_j)-S_n\le\delta_n\}} \mathrm{E}\left(1_{\{0<\phi(X_j)-S_n\le\delta_n\}} \mid \mathcal{F}_{j-1}\right)\right) \\
\le& \mathrm{E}\left(1_{\{0<\phi(X_j)-S_n\le\delta_n\}}\right) = F_{\phi(X_j)-S_n}(\delta_n) - F_{\phi(X_j)-S_n}(0).
\end{aligned}$$

Therefore,

$$\mathrm{Var}\,\Lambda_{2,n}(t,s) = \mathrm{E}\,\Lambda_{2,n}^2(t,s) \le \frac{1}{n} \sum_{j=\lfloor ns\rfloor+1}^{\lfloor nt\rfloor} \left(F_{\phi(X_j)-S_n}(\delta_n) - F_{\phi(X_j)-S_n}(0)\right).$$

□

Lemma 11 yields

$$\mathrm{Var}\,\Lambda_{2,n}(t) \le t\,\mathrm{E}\left(\sup_{x\in[0,1]} \left(F_{\lfloor nt\rfloor}(\delta_n + x) - F_{\lfloor nt\rfloor}(x)\right)\right), \tag{4.11}$$

where $F_{\lfloor nt \rfloor}$ denotes the empirical distribution function of $\phi(X_i)$, $i \geq 1$, i.e.

$$F_l(x) := \frac{1}{l} \sum_{i=1}^{l} 1_{\{\phi(X_i) \leq x\}}.$$

Note that $\sup_{x \in [0,1]} \left(F_{\lfloor nt \rfloor}(\delta_n + x) - F_{\lfloor nt \rfloor}(x) \right)$ converges to 0 almost surely due to the Glivenko-Cantelli theorem for stationary, ergodic sequences and since $F_{\phi(X_1)}$ is a continuous distribution function. It follows that its expected value converges to 0, as well. Therefore, the right-hand side of (4.11) converges to 0. Moreover, $\mathrm{E}\,\Lambda_{2,n}(t) = 0$ such that $\Lambda_{2,n}(t)$ converges to 0 in probability. This implies convergence of the finite-dimensional distributions of $\Lambda_{2,n}(t)$, $t \in [0,1]$, as a process with values in $D\,[0,1]$.

Due to convergence of the finite-dimensional distributions, it is again possible to make use of Theorem 15.4 in Billingsley (1968). In order to establish (4.4), it therefore suffices to verify that for each $\varepsilon, \eta > 0$, there exists a δ, $0 < \delta < 1$, and an $n_0 \in \mathbb{N}$ such that

$$P\left(\omega_{\Lambda_{2,n}}(\delta) \geq \varepsilon \right) \leq \eta \quad \text{for all } n \geq n_0.$$

Define

$$M_m := \max_{0 \leq i \leq j \leq k \leq m} \min \left\{ |S_j - S_i|, |S_k - S_j| \right\}, \quad \text{where } S_i = \Lambda_{2,n}(\tau + \frac{i}{m}\delta).$$

For $t_1 \leq t \leq t_2$, we have

$$|\Lambda_{2,n}(t, t_1)| = \frac{1}{\sqrt{n}} \left| \sum_{j=\lfloor nt_1 \rfloor + 1}^{\lfloor nt \rfloor} \alpha_j(\delta_n, 0) \right|, \; |\Lambda_{2,n}(t_2, t)| = \frac{1}{\sqrt{n}} \left| \sum_{j=\lfloor nt \rfloor + 1}^{\lfloor nt_2 \rfloor} \lambda_{2,j}(\delta_n, 0) \right|.$$

The Cauchy-Schwarz inequality yields

$$\begin{aligned}
&\mathrm{E}\,|\Lambda_{2,n}(t, t_1)|\,|\Lambda_{2,n}(t_2, t)| \\
&\leq \sqrt{\mathrm{E}\left[\frac{1}{n} \left(\sum_{j=\lfloor nt_1 \rfloor + 1}^{\lfloor nt \rfloor} \lambda_{2,j}(\delta_n, 0) \right)^2 \right] \mathrm{E}\left[\frac{1}{n} \left(\sum_{j=\lfloor nt \rfloor + 1}^{\lfloor nt_2 \rfloor} \lambda_{2,j}(\delta_n, 0) \right)^2 \right]} \\
&\leq \frac{1}{n} \sum_{i=\lfloor nt_1 \rfloor + 1}^{\lfloor nt_2 \rfloor} \left(F_{\phi(X_i) - S_n}(\delta_n) - F_{\phi(X_i) - S_n}(0) \right).
\end{aligned}$$

By the same argument as in the proof of (4.5), it follows that

$$P\left(M_m \geq \varepsilon \right) \leq \frac{K}{\varepsilon^2} \frac{\gamma_n^{-1}}{n} \sum_{j=\lfloor n\tau \rfloor + 1}^{\lfloor n(\tau + \delta) \rfloor} \left(F_{\phi(X_j) - S_n}(\delta_n) - F_{\phi(X_j) - S_n}(0) \right)$$

for any sequence γ_n, $n \in \mathbb{N}$, converging to 0.

Taking the right-continuity of $\Lambda_{2,n}$ into consideration, we may, as before, conclude that

$$P\left(\omega_{\Lambda_{2,n}}(\delta) \geq \varepsilon\right) \leq \frac{2K}{\varepsilon^2}\gamma_n^{-1}\frac{1}{n}\sum_{j=1}^{n}\left(F_{\phi(X_j)-S_n}(\delta_n) - F_{\phi(X_j)-S_n}(0)\right)$$
$$\leq \frac{2K}{\varepsilon^2}\gamma_n^{-1}\,\mathrm{E}\left(\sup_{x\in[0,1]}\left(F_n(\delta_n + x) - F_n(x)\right)\right).$$

Choosing γ_n, $n \in \mathbb{N}$, such that

$$\gamma_n^{-1}\,\mathrm{E}\left(\sup_{x\in[0,1]}\left(F_n(\delta_n + x) - F_n(x)\right)\right)$$

converges to 0, the right-hand side of the above inequality vanishes as n tends to ∞ due to the choice of γ_n, $n \in \mathbb{N}$. This concludes the proof of (4.4) as well as the proof of Lemma 9. □

4.2.2. Long-range dependent part

Finally, weak convergence of $R_n(x,t)$ is proved.

Lemma 12 (Betken and Kulik (2017))**.** *Under the conditions of Theorem 10,*

$$\frac{1}{d_{n,r}}R_n(x,t) \xrightarrow{\mathcal{D}} \frac{1}{r!}J_r(\Psi_x \circ \sigma)Z_{r,H}(t)$$

in $D\left([-\infty,\infty]\times[0,1]\right)$.

Proof. Note that

$$\mathrm{E}\left(1_{\{\psi(X_j)\leq x\}} \mid \mathcal{F}_{j-1}\right) = \mathrm{E}\left(1_{\{\psi(\sigma(\xi_j)\varepsilon_j)\leq x\}} \mid \mathcal{F}_{j-1}\right) = \Psi_x(\sigma(\xi_j))$$

because Y_j is $\mathcal{F}_{j-1}$-measurable and ε_j is independent of $\mathcal{F}_{j-1}$. Furthermore, it holds that $\mathrm{E}\,\Psi_x(\sigma(\xi_j)) = F_{\psi(X_1)}(x)$, where $F_{\psi(X_1)}$ denotes the distribution function of $\psi(X_1)$. Hence,

$$R_n(x,t) = \lfloor nt\rfloor \int_{\mathbb{R}} \Psi_x(u)d\left(F_{\lfloor nt\rfloor} - \mathrm{E}\,F_{\lfloor nt\rfloor}\right)(u),$$

where F_l denotes the empirical distribution function of the sequence $\sigma(\xi_n)$, $n \in \mathbb{N}$, i.e.

$$F_l(u) := \frac{1}{l}\sum_{j=1}^{l} 1_{\{\sigma(\xi_j)\leq u\}}.$$

We have

$$d_{n,r}^{-1}R_n(x,t) = I_1(x,t) + I_2(x,t)$$

with

$$I_1(x,t) = -\int_{\mathbb{R}} \frac{d}{dy}\Psi_x(y) d_{n,r}^{-1} \left\{ \lfloor nt \rfloor \left(F_{\lfloor nt \rfloor}(y) - \mathrm{E}\, F_{\lfloor nt \rfloor}(y)\right) - \frac{1}{r!} J_r(\sigma; y) \sum_{j=1}^{\lfloor nt \rfloor} H_r(\xi_j) \right\} dy,$$

$$I_2(x,t) = -\int_{\mathbb{R}} \frac{d}{dy}\Psi_x(y) d_{n,r}^{-1} \frac{1}{r!} J_r(\sigma; y) \sum_{j=1}^{\lfloor nt \rfloor} H_r(\xi_j) dy,$$

where r denotes the Hermite rank of the class of functions $1_{\{\sigma(\xi_1)\leq x\}} - F_{\sigma(\xi_1)}(x)$, $x \in \mathbb{R}$, and

$$J_r(\sigma; y) = \mathrm{E}\left(1_{\{\sigma(\xi_1)\leq y\}} H_r(\xi_1)\right).$$

Due to the integrability condition (4.1), it follows from Theorem 3.1 in Dehling and Taqqu (1989) that $I_1(x,t)$ converges to 0 in probability, uniformly in x and t. Furthermore,

$$I_2(x,t) = -d_{n,r}^{-1} \sum_{j=1}^{\lfloor nt \rfloor} H_r(\xi_j) \left\{ \int_{\mathbb{R}} \frac{1}{r!} J_r(\sigma; y) \frac{d}{dy} P\left(\psi(y\varepsilon_1) \leq x\right) dy \right\}.$$

Denoting with φ the standard normal density, integration by parts yields

$$\int_{\mathbb{R}} J_r(\sigma; y) \frac{d}{dy} P\left(\psi(y\varepsilon_1) \leq x\right) dy = -J_r(\Psi_x \circ \sigma).$$

Moreover, Theorem 4.1 in Taqqu (1975) implies

$$d_{n,r}^{-1} \sum_{j=1}^{\lfloor nt \rfloor} H_r(\xi_j) \xrightarrow{\mathcal{D}} Z_{r,H}(t),\ t \in [0,1],$$

in $D[0,1]$. □

4.3. Change-point tests for LMSV time series

As noted in Section 1.2.1, specific change-point problems can be interpreted as testing for changes in the mean of transformed observations. For this reason, the following sections focus on characterizing the limit behavior of change-point tests that are based on CUSUM and Wilcoxon statistics.

4.3.1. CUSUM tests for LMSV time series

Given observations $X_1, \ldots, X_n$, computations of the two-sample CUSUM statistics with respect to the transformed observations $Z_1, \ldots, Z_n$, $Z_i := \psi(X_i)$, $i = 1, \ldots, n$ yield

$$C_{k,n} = \sum_{j=1}^{k} Z_j - \frac{k}{n} \sum_{j=1}^{n} Z_j,$$

$$SC_{k,n} = \frac{C_{k,n}}{V_{k,n}}, \quad V_{k,n} = \left\{ \frac{1}{n} \sum_{t=1}^{k} S_t^2(1,k) + \frac{1}{n} \sum_{t=k+1}^{n} S_t^2(k+1,n) \right\}^{\frac{1}{2}},$$

where

$$S_t(j,k) = \sum_{h=j}^{t} \left(Z_h - \bar{Z}_{j,k} \right) \quad \text{and} \quad \bar{Z}_{j,k} = \frac{1}{k-j+1} \sum_{t=j}^{k} Z_t.$$

As noted in Section 1.2.3, the limit of the self-normalized statistic can be described by means of the function $G_f \in D\,[0,1]$, defined by

$$G_f(t) := \frac{f(t)}{V_f(t)}, \; V_f(t) := \left\{ \int_0^t \left(f(s) - \frac{s}{t} f(t) \right)^2 ds + \int_t^1 \left(f(s) - \frac{1-s}{1-t} f(t) \right)^2 ds \right\}^{\frac{1}{2}}.$$

In order to determine the limit distributions of the test statistics, we consider the partial sum process

$$\sum_{j=1}^{\lfloor nt \rfloor} \left(\psi(X_j) - \mathrm{E}\, \psi(X_j) \right), \; 0 \leq t \leq 1.$$

Dependent on the function ψ, the considered observations are either uncorrelated or display an autocovariance structure that relates to the subordinated Gaussian sequence Y_n, $n \in \mathbb{N}$. To specify this assertion, recall that $\mathcal{F}_j$ denotes the σ-field generated by the random variables $\varepsilon_j, \varepsilon_{j-1}, \ldots, \eta_j, \eta_{j-1}, \ldots$. Due to this construction, ε_j is independent of $\mathcal{F}_{j-1}$ and Y_j is $\mathcal{F}_{j-1}$-measurable.

The asymptotic behavior of the partial sum process is described by Theorem 4.10 in Beran et al. (2013):

Theorem 12 (Beran et al. (2013))**.** *Suppose that X_n, $n \in \mathbb{N}$, is a time series satisfying the conditions specified by Model 2. Moreover, assume that ψ is a measurable function and that $\mathrm{E}\, \psi^2(X_1) < \infty$.*

1. *If $\mathrm{E}(\psi(X_1) \mid \mathcal{F}_0) \neq 0$ and $rD < 1$, where r denotes the Hermite rank of the function Ψ with $\Psi(z) := \mathrm{E}\, \psi(\sigma(z)\varepsilon_1)$, then*

$$\frac{1}{d_{n,r}} \sum_{j=1}^{\lfloor nt \rfloor} \left(\psi(X_j) - \mathrm{E}\, \psi(X_j) \right) \xrightarrow{\mathcal{D}} \frac{1}{r!} J_r(\Psi) Z_{r,H}(t), \quad t \in [0,1],$$

in $D[0,1]$, where $Z_{r,H}$ is an r-th order Hermite process with parameter $H = 1 - \frac{rD}{2}$, $J_r(\Psi)$ denotes the r-th Hermite coefficient in the Hermite expansion of Ψ, and $d_{n,r}$ the normalizing sequence defined by (1.5) *in Section 1.2.2.*

2. *If* $\mathrm{E}(\psi(X_1)\,|\,\mathcal{F}_0)=0$, *then*

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor nt\rfloor}\psi(X_j)\xrightarrow{\mathcal{D}}\sigma B(t),\quad t\in[0,1],$$

in $D[0,1]$, *where* $B(t)$, $t\in[0,1]$, *denotes a Brownian motion and* $\sigma^2=\mathrm{E}\,\psi^2(X_1)$.

As an immediate consequence of Theorem 12, we obtain the asymptotic distribution of the CUSUM and the self-normalized CUSUM statistic:

Corollary 4 (Betken and Kulik (2017)). *Let* $0<\tau_1<\tau_2<1$. *Given the assumptions and notations of Theorem 12, the following assertions hold:*

1. *If* $\mathrm{E}(\psi(X_1)\,|\,\mathcal{F}_0)\neq 0$ *and* $rD<1$,

$$\frac{1}{d_{n,r}}\max_{1\leq k\leq n}C_{k,n}\xrightarrow{\mathcal{D}}\frac{1}{r!}\left|J_r(\Psi)\right|\sup_{t\in[0,1]}\left|W_{r,H}(t)\right|,$$

$$\max_{\lfloor n\tau_1\rfloor\leq k\leq\lfloor n\tau_2\rfloor}\left|SC_{k,n}\right|\xrightarrow{\mathcal{D}}\sup_{t\in[\tau_1,\tau_2]}\left|G_{W_{r,H}}(t)\right|,$$

where $W_{r,H}(t):=Z_{r,H}(t)-tZ_{r,H}(1)$.

2. *If* $\mathrm{E}(\psi(X_1)\,|\,\mathcal{F}_0)=0$,

$$\frac{1}{\sqrt{n}}\max_{1\leq k\leq n}C_{k,n}\xrightarrow{\mathcal{D}}\sigma\sup_{t\in[0,1]}\left|W(t)\right|,$$

$$\max_{\lfloor n\tau_1\rfloor\leq k\leq\lfloor n\tau_2\rfloor}\left|SC_{k,n}\right|\xrightarrow{\mathcal{D}}\sup_{t\in[\tau_1,\tau_2]}\left|G_W(t)\right|,$$

where $W(t):=B(t)-tB(1)$.

According to Corollary 4, the asymptotic behavior of CUSUM-based test statistics essentially depends on the specific change-point problem that is considered. Dependent on the choice of ψ, it may or may not be affected by long-range dependence. If $\psi(x)=x$, i.e. when testing for a change in the mean, $\mathrm{E}(\psi(X_1)\,|\,\mathcal{F}_0)=\sigma(\xi_1)\,\mathrm{E}\,\varepsilon_1=0$. Hence, the asymptotic distribution of the CUSUM statistics is determined by a Brownian motion. If $\psi(x)=x^2$, i.e. when testing for a change in the variance of LMSV time series, $\mathrm{E}(\psi(X_1)\,|\,\mathcal{F}_0)=\sigma^2(\xi_1)\,\mathrm{E}\,\varepsilon_1^2\neq 0$. In this case, the limit of CUSUM-based statistics is a function of a fractional Brownian motion characterized by the parameter $H=1-\frac{rD}{2}$ with D denoting the LRD parameter of the Gaussian sequence ξ_n, $n\in\mathbb{N}$. As a result, the limit distributions are affected by the intensity of dependence in the data. Nevertheless, even if the limit distribution is affected by long-range dependence in the data, the Hurst parameter H in the limit does not necessarily correspond to the LRD parameter characterizing the dependence in the subordinated Gaussian sequence $\sigma(\xi_n)$, $n\in\mathbb{N}$, since the Hermite ranks of Ψ and σ may differ.

4.3.2. Wilcoxon tests for LMSV time series

Given observations $X_1, \ldots, X_n$ and a transformation ψ, let R_i, $i = 1, \ldots, n$ denote the rank statistics of the transformed observations $\psi(X_1), \ldots, \psi(X_n)$, i.e.

$$R_i = \sum_{j=1}^{n} 1_{\{\psi(X_j) \leq \psi(X_i)\}}.$$

Recall that the two-sample Wilcoxon statistics are given by

$$W_{k,n} = \sum_{j=1}^{k} R_j - \frac{k}{n} \sum_{j=1}^{n} R_j,$$

$$SW_{k,n} = \frac{W_{k,n}}{V_{k,n}}, \quad V_{k,n} = \left\{ \frac{1}{n} \sum_{t=1}^{k} S_t^2(1,k) + \frac{1}{n} \sum_{t=k+1}^{n} S_t^2(k+1,n) \right\}^{\frac{1}{2}},$$

where

$$S_t(j,k) = \sum_{h=j}^{t} \left(R_h - \bar{R}_{j,k} \right) \quad \text{and} \quad \bar{R}_{j,k} = \frac{1}{k-j+1} \sum_{t=j}^{k} R_t.$$

As noted in Section 1.2.2, the asymptotic distribution of the Wilcoxon statistics can be derived from the limit behavior of the two-parameter empirical process. Following the proof of Theorem 1 in Dehling et al. (2013), the asymptotic distribution of the Wilcoxon statistic W_n, which corresponds to the maximum of the two sample statistics $W_{k,n}$, $1 \leq k \leq n$, see (1.4) in Chapter 1, can be derived directly from Theorem 10 if the data-generating sequence X_n, $n \in \mathbb{N}$, is ergodic. Under the assumption of LMSV time series satisfying Model 2, ergodicity follows from the fact that every random variable X_j has a representation as a measurable function of the independent, identically distributed random vectors (η_i, ε_i), $i \leq j$. This proves the following Corollary:

Corollary 5 (Betken and Kulik (2017)). *Suppose that X_n, $n \in \mathbb{N}$, is a time series satisfying the conditions specified by Model 2. Assume that ψ is a measurable function and that*

$$\int_{\mathbb{R}} \frac{d}{dy} \Psi_x(y) dy < \infty,$$

where $\Psi_x(y) := P\left(\psi(y\varepsilon_1) \leq x\right)$. Let r denote the Hermite rank of the class of functions $1_{\{\sigma(\xi_1) \leq x\}} - F_{\sigma(\xi_1)}(x)$, $x \in \mathbb{R}$. If $rD < 1$,

$$\frac{1}{n d_{n,r}} \max_{1 \leq k \leq n} W_{k,n} \xrightarrow{\mathcal{D}} \frac{1}{r!} \left| \int_{\mathbb{R}} J_r(\Psi_x \circ \sigma) dF_{\psi(X_1)}(x) \right| \sup_{t \in [0,1]} \left| W_{r,H}(t) \right|,$$

$$\max_{\lfloor n\tau_1 \rfloor \leq k \leq \lfloor n\tau_2 \rfloor} \left| SW_{k,n} \right| \xrightarrow{\mathcal{D}} \sup_{t \in [\tau_1, \tau_2]} \left| G_{W_{r,H}}(t) \right|$$

with $W_{r,H}(t) := Z_{r,H}(t) - tZ_{r,H}(1)$, where $Z_{r,H}$ is an r-th order Hermite process with parameter $H = 1 - \frac{rD}{2}$, $J_r(\Psi_x \circ \sigma)$ denotes the r-th Hermite coefficient in the Hermite expansion of $\Psi_x \circ \sigma$, and $d_{n,r}$ the normalizing sequence defined by (1.5) *in Section 1.2.2.*

Remark 14. Note that the Hermite rank of $\Psi_x \circ \sigma$ does not necessarily correspond to the Hermite rank of $1_{\{\sigma(\xi_1)\leq x\}} - F_{\sigma(\xi_1)}(x)$, $x \in \mathbb{R}$. However, if ψ has an inverse ψ^{-1},

$$J_r(\Psi_x \circ \sigma) = \int_{\mathbb{R}} \int_{\mathbb{R}} 1_{\{\sigma(z)u \leq \psi^{-1}(x)\}} H_r(z)\varphi(z) dz dF_\varepsilon(u),$$

where F_ε denotes the distribution function of ε_1. It follows that $J_r(\Psi_x \circ \sigma) \neq 0$ for some $x \in \mathbb{R}$ implies that $J_r(\sigma; y) \neq 0$ for some $y \in \mathbb{R}$. As a result, the Hermite rank of $\Psi_x \circ \sigma$ is smaller or equal to the Hermite rank of $1_{\{\sigma(\xi_1)\leq x\}} - F_{\sigma(\xi_1)}(x)$, $x \in \mathbb{R}$.

4.4. Simulations

For all simulations, the following specifications are made:

$$X_n = \sigma(\xi_n)\varepsilon_n, \quad n \in \mathbb{N}, \tag{4.12}$$

where

- ε_n, $n \in \mathbb{N}$, is an independent, identically Pareto distributed sequence with scale parameter $k = 1$ and shape parameter α, either
  a) non-centered, i.e. ε_1 is Pareto$(\alpha, 1)$ distributed according to Example 1 in Chapter 1, generated by the function `rgpd` (`fExtremes` package in `R`), or
  b) centered, i.e. $\varepsilon_1 = \tilde{\varepsilon}_1 - \mathrm{E}\,\tilde{\varepsilon}_1$, where $\tilde{\varepsilon}_1$ is Pareto$(\alpha, 1)$ distributed;
- ξ_n, $n \in \mathbb{N}$, is a fractional Gaussian noise sequence generated by the function `fgnSim` (`fArma` package in `R`) with Hurst parameter H;
- $\sigma(z) = \exp(z)$.

As noted in Chapter 1, different change-point problems can be reduced to the identification of changes in the mean of suitably transformed observations $\psi(X_1), \ldots, \psi(X_n)$. In the following, three different change-point alternatives are considered: changes in location, changes in volatility, and changes in the tail index. In each situation, the asymptotic distributions of the CUSUM and Wilcoxon test statistics considered in Section 4.3 are derived from Corollaries 4 and 5. Accordingly, definitions and notations are adopted from the previous Section.

Change in location

We investigate the finite sample performance of the CUSUM and Wilcoxon change-point tests for detecting changes in the mean of LMSV time series X_n, $n \in \mathbb{N}$, i.e. we choose $\psi(x) = x$ for the test statistics described in Section 4.3. For this purpose, we simulate observations $X_1, \ldots, X_n$ which satisfy (4.12) with ε_j, $j = 1, \ldots, n$, that follow a centered Pareto$(\alpha, 1)$ distribution.

CUSUM: According to Corollary 4, the CUSUM statistic converges in distribution to

$$\sigma \sup_{t\in[0,1]} |B(t) - tB(1)| \quad \text{with } \sigma^2 = \exp(2)\frac{\alpha}{\alpha-2}\left(\frac{1}{\alpha-1}\right)^2,$$

while the self-normalized CUSUM statistic converges to

$$\sup_{t\in[\tau_1,\tau_2]} \frac{|B(t) - tB(1)|}{\left\{\int_0^t V^2(s;0,t)ds + \int_t^1 V^2(s;t,1)ds\right\}^{\frac{1}{2}}}$$

with

$$V(t;t_1,t_2) := B(t) - B(t_1) - \frac{t-t_1}{t_2-t_1}\{B(t_2) - B(t_1)\}$$

for $t \in [t_1, t_2]$, $0 < t_1 < t_2 < 1$.

Wilcoxon: According to Corollary 5, the Wilcoxon statistic converges to

$$\frac{1}{r!}\left|\int_{\mathbb{R}} J_r(\Psi_x \circ \sigma) dF_{\psi(X_1)}(x)\right| \sup_{t\in[0,1]} |Z_{r,H}(t) - tZ_{r,H}(1)|, \tag{4.13}$$

while the self-normalized Wilcoxon statistic converges to

$$\sup_{t\in[\tau_1,\tau_2]} \frac{|Z_{r,H}(t) - tZ_{r,H}(1)|}{\left\{\int_0^t V_{r,H}^2(s;0,t)ds + \int_t^1 V_{r,H}^2(s;t,1)ds\right\}^{\frac{1}{2}}} \tag{4.14}$$

with

$$V_{r,H}(t;t_1,t_2) := Z_{r,H}(t) - Z_{r,H}(t_1) - \frac{t-t_1}{t_2-t_1}\{Z_{r,H}(t_2) - Z_{r,H}(t_1)\}$$

for $t \in [t_1, t_2]$, $0 < t_1 < t_2 < 1$ and with r denoting the Hermite rank of the class of functions $1_{\{\sigma(\xi_1)\leq x\}} - F_{\sigma(\xi_1)}(x)$, $x \in \mathbb{R}$.
In order to determine the critical values needed for an application of the Wilcoxon test, we have to calculate the deterministic factor in formula (4.13). For this purpose, note that $\varphi'(z) = -z\varphi(z)$, so that integration by parts yields

$$J_1(\Psi_x \circ \sigma) = \int_{\mathbb{R}} P\left(\psi(\sigma(z)\varepsilon_1) \leq x\right) z\varphi(z)dz = x\int_{\mathbb{R}} \frac{d}{dz}\left(\frac{1}{\sigma(z)}\right) f_{\varepsilon_1}\left(\frac{x}{\sigma(z)}\right)\varphi(z)dz,$$

where f_{ε_1} denotes the density of ε_1. Under the assumption of centered Pareto$(\alpha, 1)$ distributed random variables ε_n, $n \in \mathbb{N}$, it follows that

$$J_1(\Psi_x \circ \sigma) = -\int_{\mu_\alpha}^{\infty} h_\alpha(x,y)dy 1_{\{x>0\}} + \int_1^{\mu_\alpha} h_\alpha(x,y)dy 1_{\{x<0\}}$$

with

$$h_\alpha(x,y) = \alpha y^{-\alpha-1}\varphi\left(\log\left(\frac{1}{x}(y-\mu_\alpha)\right)\right)1_{\{y>1\}},$$

where μ_α denotes the expected value of $\tilde{\varepsilon}_1$, i.e. $\mu_\alpha := \frac{\alpha}{\alpha-1}$. Moreover, the probability density function f_{X_1} of X_1 corresponds to

$$f_{X_1}(x) = \frac{1}{x}\int_{\mu_\alpha}^{\infty} h_\alpha(x,y)dy 1_{\{x>0\}} - \frac{1}{x}\int_{1}^{\mu_\alpha} h_\alpha(x,y)dy 1_{\{x<0\}}.$$

As a result,

$$\int_{\mathbb{R}} J_1(\Psi_x \circ \sigma)dF_{\psi(X_1)}(x) = \int_0^\infty \frac{1}{x}\left[\left(\int_1^{\mu_\alpha} h_\alpha(x,y)dy\right)^2 - \left(\int_{\mu_\alpha}^{\infty} h_\alpha(x,y)dy\right)^2\right]dx.$$

Since the above integral cannot be computed analytically, critical values for the Wilcoxon test are based on an approximation by numerical integration.

In order to compare the finite sample behavior of the change-point tests, the empirical size and the empirical power of the testing procedures are computed. To determine the finite sample performance under the alternative, simulated time series with a change-point of height h after a proportion τ of the data are considered, i.e. random variables X_j, $j = 1, \ldots, n$, with expected values $\mathrm{E}\, X_j$, $j = 1 \ldots, n$, such that $\mathrm{E}\, X_j = 0$ for $j = 1, \ldots, \lfloor n\tau \rfloor$, while $\mathrm{E}\, X_j = h$ for $j = \lfloor n\tau \rfloor + 1, \ldots, n$. The calculations are based on 5000 realizations of time series with sample sizes 500, 1000 and 2000. The simulation results are reported in Tables 4.1 and 4.2. The frequency of a type 1 error, i.e. the rejection rate under the hypothesis, corresponds to the values in the columns that are superscribed by $h = 0$.

The simulation results reported in Tables 4.1 and 4.2 show that under the hypothesis the behavior of the testing procedures differs:

- In most cases the size of the CUSUM test does not deviate much from the level of significance. However, the rejection rates do not seem to draw closer to the significance level as the sample size increases. An increase in dependence, i.e. an increase of the Hurst parameter H and a decrease of tail thickness, i.e. an increase of the tail parameter α, lead to an increase in the number of rejections.
- The Wilcoxon change-point test suffers from size distortions: the rejection rates under the hypothesis are considerably higher than the level of significance. Although the empirical size of the Wilcoxon change-point test decreases with an increasing sample size, thereby approaching the level of significance, the convergence to the significance level seems to be rather slow. An increase in dependence, i.e. an increase of the Hurst parameter H, and an increase of tail thickness, i.e. a decrease of the tail parameter α, lead to a decrease in the number of rejections.

- The size of the self-normalized CUSUM test almost equals the level of significance (even for relatively small samples). Neither Hurst parameter nor tail index seem to have a significant effect on the rejection rates.
- The self-normalized Wilcoxon change-point test tends to be undersized. A change in H or α does not seem to have a significant impact on its finite sample performance under the hypothesis.

Under the alternative we make the following observations:

- The empirical power of the change-point tests increases when the number of observations or the height of the level shift increases. It is higher for breakpoints located in the middle of the sample than for change-point locations that lie close to the boundary of the testing region.
- Heavier tails, i.e. a decrease of the tail parameter α, go along with a decrease of the rejection rates for all four testing procedures.
- An increase of dependence in the data, i.e. an increase of the value of the Hurst parameter, is followed by a decrease of empirical power for the Wilcoxon-based testing procedures, but a slight increase of rejections for the CUSUM-based change-point tests.

A comparison of the finite sample performance of the testing procedures shows that:

- The finite sample performance of the Wilcoxon-based change-point tests confirms that rank-based testing procedures are robust to the influence of heavy tails. CUSUM-based testing procedures, on the other hand, are more sensitive to the influence of heavy tails.
- As expected, the simulation results show that the Hurst parameter does not seem to have a considerable impact on the performance of the CUSUM-based testing procedures. A pronounced effect on the rejection rates can only be observed for the Wilcoxon test.
- The so-called *better size but less power* phenomenon for self-normalized tests, which has also been observed in Shao (2011), Shao and Zhang (2010) and Betken (2016), arises: While the self-normalized change-point tests have better size properties, the non-self-normalized tests usually yield a higher empirical power. Moreover, the empirical size of the self-normalized tests is not influenced by long-range dependence and does not depend on the values of the parameter α.
- For large sample sizes, the power of the CUSUM test exceeds the power of the self-normalized CUSUM test. The deviation of the rejection rates increases with growing values of α.

- The power of the self-normalized Wilcoxon test is higher than the power of the self-normalized CUSUM test for almost every combination of parameters. The power of the self-normalized CUSUM test only exceeds the power of the self-normalized Wilcoxon test for relatively high level shifts or high value of H. The difference between the rejection rates is especially high for small values of α.

All in all, a comparison of the simulation results gives rise to choosing the self-normalized Wilcoxon test over the other testing procedures when testing for a change in the mean.

Change in volatility

We will now investigate the finite sample performance of the CUSUM and Wilcoxon change-point tests for detecting changes in the variance of LMSV time series X_n, $n \in \mathbb{N}$, i.e. we choose $\psi(x) = x^2$ for the test statistics described in Section 4.3. For this purpose, we simulate observations $X_1, \ldots, X_n$ which satisfy (4.12) with ε_j, $j = 1, \ldots, n$, that follow a centered Pareto$(\alpha, 1)$ distribution. An application of Corollary 4 requires $\mathrm{E}\, X_1^4 < \infty$, i.e. in order to test for changes in the variance, the existence of fourth moments is assumed. For this reason, we choose $\alpha > 4$, since the k-th moment of Pareto distributed variables exists only if k is smaller than the tail index α.

CUSUM: In this case, $\mathrm{E}(\psi(X_1) \,|\, \mathcal{F}_0) = \sigma^2(\xi_1)\, \mathrm{E}\, \varepsilon_1^2 \neq 0$. If $rD < 1$ (with r denoting the Hermite rank of Ψ), it follows from Corollary 4 that the CUSUM statistic converges to

$$\frac{1}{r!}\left|J_r(\Psi)\right| \sup_{t \in [0,1]} \left|Z_{r,H}(t) - tZ_{r,H}(1)\right| \quad \text{with } J_r(\Psi) = \mathrm{E}\, \varepsilon_1^2 J_r(\sigma^2),$$

while the self-normalized CUSUM statistic converges to the limit in formula (4.14), where, in both cases, $H = 1 - \frac{rD}{2}$. In particular, the Hermite rank of Ψ equals the Hermite rank of σ^2, so that for $\sigma(z) = \exp(z)$ and ε_1 centered Pareto distributed, we have

$$J_1(\Psi) = \mathrm{E}\left(\exp(2\xi_1)\xi_1\right) \mathrm{Var}\, \varepsilon_1 = 2\exp(2)\frac{\alpha}{\alpha - 2}\left(\frac{1}{\alpha - 1}\right)^2.$$

Wilcoxon: According to Corollary 5, the limit of the Wilcoxon statistic corresponds to the expression in formula (4.13) while the limit of the self-normalized Wilcoxon statistic is given by (4.14). In order to apply the non-self-normalized Wilcoxon test, the value of the multiplicative factor $\left|\int_{\mathbb{R}} J_1(\Psi_x \circ \sigma) dF_{\psi(X_1)}(x)\right|$ has to be computed under the given assumptions. For $x \geq 0$, integration by parts yields

$$J_1(\Psi_x \circ \sigma) = \int_{\mathbb{R}} \left(\sqrt{x}\frac{d}{dz}\frac{1}{\sigma(z)}\right)\left\{f_{\varepsilon_1}\left(\frac{\sqrt{x}}{\sigma(z)}\right) + f_{\varepsilon_1}\left(-\frac{\sqrt{x}}{\sigma(z)}\right)\right\}\varphi(z)dz,$$

where f_{ε_1} denotes the density of ε_1. Under the considered specifications, integration by substitution yields

$$J_1(\Psi_x \circ \sigma) = -1_{\{x>0\}}\int_1^\infty \alpha y^{-\alpha-1}\varphi\left(\log\left(\frac{1}{\sqrt{x}}(y - \mu_\alpha)\right)\right)dy,$$

where μ_α denotes the expected value of $\tilde{\varepsilon}_1$, i.e. $\mu_\alpha := \mathrm{E}\,\tilde{\varepsilon}_1 = \frac{\alpha}{\alpha-1}$. Moreover, the probability density function $f_{\psi(X_1)}$ of $\psi(X_1)$ corresponds to

$$f_{\psi(X_1)}(x) = \frac{1}{2x} 1_{\{x>0\}} \int_1^\infty \alpha y^{-\alpha-1} \varphi\left(\log\left(\frac{1}{\sqrt{x}}(y-\mu_\alpha)\right)\right) dy.$$

As a result,

$$\left|\int_{\mathbb{R}} J_1(\Psi_x \circ \sigma) dF_{\psi(X_1)}(x)\right| = \int_0^\infty \frac{1}{2x}\left(\int_1^\infty \alpha y^{-\alpha-1} \varphi\left(\log\left(\frac{1}{\sqrt{x}}(y-\mu_\alpha)\right)\right) dy\right)^2 dx.$$

Since the above integral cannot be computed analytically, critical values for the Wilcoxon test are based on an approximation by numerical integration.

In order to compare the finite sample behavior of the change-point tests, the empirical size and the empirical power of the testing procedures is computed. To determine the finite sample performance of the testing procedures under the hypothesis, observations $X_1, \ldots, X_n$ with variance ω^2 are simulated. For the computation of the empirical power, a change-point of height h^2 is added after τ percent of the data, i.e. we multiply $X_{\lfloor n\tau \rfloor+1}, \ldots, X_n$ by h such that $\mathrm{Var}\, X_j = \omega^2$ for $j = 1, \ldots, \lfloor n\tau \rfloor$ while $\mathrm{Var}\, X_j = h^2\omega^2$ for $j = \lfloor n\tau \rfloor + 1, \ldots, n$. The rejection rates of the testing procedures were computed on the basis of 5000 realizations of time series with sample sizes 500, 1000 and 2000. The simulation results are reported in Tables 4.3 and 4.4. The frequency of a type 1 error, i.e. the rejection rate under the hypothesis, corresponds to the values in the columns that are superscribed by $h = 1$.

We make the following observations with respect to the behavior of the different testing procedures under the hypothesis:

- The empirical size of the non-self-normalized tests tends to increase when the dependence in the simulated time series decreases, i.e. when H decreases, while the empirical size of the self-normalized tests seems to be independent of H.
- For none of the four change-point tests, the empirical size seems to be significantly affected by the value of the tail parameter α.
- The CUSUM test suffers from severe size distortions. It rejects the hypothesis too frequently for every combination of the parameters α and H. In particular, the number of rejections increases as the number of observations increases. Apparently, unrealistically large sample sizes are required for the asymptotics to apply. The same observation has been made in De Pooter and Van Dijk (2004) where the CUSUM statistic is used to test for breaks in the variance of GARCH(1, 1) time series.
- The Wilcoxon test rejects the hypothesis too often. Yet, the rejection rates seem to approach the level of significance as the length of time series increases.

- The self-normalized CUSUM test tends to be undersized. Moreover, the rejection rates do not seem to approach the level of significance when the sample size increases.
- The rejection rates of the self-normalized Wilcoxon test are, even for relatively small sample sizes, close to the level of significance.

Under the alternative we make the following observations:

- The empirical power of all four change-point tests indicates consistency, i.e. it increases when the number of observations increases.
- For all four change-point tests, an increase in correlation, i.e. an increase of H, leads to a decrease in the empirical power.
- For all four change-point tests, the empirical power does not seem to be significantly affected by a change in the value of the tail parameter α.
- For a change-point height $h = 0.5$, the power of the non-self-normalized CUSUM test decreases when τ changes from 0.5 to 0.25; the empirical size exceeds the empirical power in both cases. Nonetheless, the empirical power tends to increase when τ changes from 0.5 to 0.25 in the presence of a change-point with height $h = 2$. Independent of the change-point location, an increase of the variance, characterized by $h = 2$, is better detected by the CUSUM test than a decrease of the variance, characterized by $h = 0.5$.
- For a change-point height $h = 0.5$, the power of the self-normalized CUSUM test increases when τ changes from 0.5 to 0.25. However, it decreases when τ changes from 0.5 to 0.25 for a change-point height $h = 2$. In this case, the empirical size of the self-normalized CUSUM test does not differ much from its empirical power. For a change-point in $\tau = 0.5$, an increase and a decrease of the variance are almost equally well detected by the self-normalized CUSUM test, while for a change-point in $\tau = 0.25$ a decrease in the variance, characterized by $h = 0.25$, is better detected than an increase of the variance.
- Independent of the change-point location, an increase and a decrease of the variance are almost equally well detected by both Wilcoxon tests. Independent of the change-point height, a change from $\tau = 0.5$ to $\tau = 0.25$ leads to a decrease in the empirical power of both Wilcoxon-based testing procedures.

A comparison of the finite sample performance of the testing procedures shows that:

- Again, the so-called *better size but less power* phenomenon for self-normalized tests can be observed.
- The empirical power of the self-normalized CUSUM test cannot compete with the empirical power of any other test.

- Even though the CUSUM test has high size distortions, its power can only compete with the power of the Wilcoxon-based testing procedures if the variance increases, i.e. if $h = 2$. For a decrease of variance, i.e. $h = 0.5$, the empirical power of the CUSUM test is smaller than the empirical power of the Wilcoxon-based testing procedures .
- The power of the Wilcoxon test exceeds the power of the self-normalized Wilcoxon test for every parameter combination that has been considered.

All in all, the simulation results show that both CUSUM tests are outperformed by the Wilcoxon-based testing procedures. Obviously, CUSUM-based tests are highly unreliable when testing for a change in the variance.

Change in the tail index

We will now investigate the finite sample performance of the CUSUM and Wilcoxon change-point tests for detecting changes in the tail parameter α of LMSV time series $X_n, n \in \mathbb{N}$, i.e. we choose $\psi(x) = \log|x|$ for the test statistics described in Section 4.3. For this purpose, we simulate observations $X_1, \ldots, X_n$ which satisfy (4.12) for random variables ε_j, $j = 1, \ldots, n$, that follow a non-centered Pareto$(\alpha, 1)$ distribution. The choice of ψ is justified since $P(|X_1| > x) = x^{-\alpha}$, $x > 1$, so that $\mathrm{E}\log|X_1| = \alpha^{-1}$.

CUSUM: Given the above conditions, $\mathrm{E}(\psi(X_1) \mid \mathcal{F}_0) = \log|\sigma(\xi_1)| + \mathrm{E}\log|\varepsilon_1| \neq 0$. If $rD < 1$ (with r denoting the Hermite rank of Ψ), it follows from Corollary 4 that the CUSUM statistic converges to

$$\frac{1}{r!}|J_r(\Psi)| \sup_{t\in[0,1]} |Z_{r,H}(t) - tZ_{r,H}(1)|$$

with $J_r(\Psi) = J_r(\log\circ|\sigma|)$, while the self-normalized CUSUM statistic converges to the limit in formula (4.14), where, in both cases, $H = 1 - \frac{rD}{2}$. In particular, the Hermite rank of Ψ equals the Hermite rank of $\log\circ\sigma$, so that for $\sigma(z) = \exp(z)$ and ε_1 centered Pareto distributed, $J_1(\Psi) = \mathrm{E}\,\xi_1^2 = 1$.

Remark 15. For $\psi(x) = \log|x|$, the asymptotic distributions of the CUSUM statistics can also be derived from an application of Donsker's theorem and a non-central limit theorem for the partial sum process of subordinated Gaussian sequences. To see this, note that

$$\begin{aligned}&\frac{1}{d_{n,r}}\sum_{j=1}^{\lfloor nt\rfloor}(\log|X_j| - \mathrm{E}\log|X_j|)\\ =&\frac{1}{d_{n,r}}\sum_{j=1}^{\lfloor nt\rfloor}(\log|\sigma(\xi_j)| - \mathrm{E}\log|\sigma(\xi_j)|) + \frac{\sqrt{n}}{d_{n,r}}\frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor nt\rfloor}(\log|\varepsilon_j| - \mathrm{E}\log|\varepsilon_j|).\end{aligned}$$

As $\sqrt{n} = o(d_{n,r})$, the second summand on the right-hand side of the above equality converges to 0 in probability, uniformly in t, according to Donsker's theorem.

As a consequence of Theorem 4.1 in Taqqu (1975), the partial sum process non-central limit theorem for subordinated Gaussian sequences,

$$\frac{1}{d_{n,r}}\sum_{j=1}^{\lfloor nt\rfloor}\left(\log|\sigma(\xi_j)| - \mathrm{E}\log|\sigma(\xi_j)|\right) \xrightarrow{\mathcal{D}} \frac{1}{r!}J_r\left(\log\circ|\sigma|\right)Z_{r,H}(t), \quad t\in[0,1].$$

Wilcoxon: According to Corollary 5, the limit of the Wilcoxon statistic corresponds to the expression in formula (4.13) while the limit of the self-normalized Wilcoxon statistic is given by (4.14). In order to apply the non-self-normalized Wilcoxon test, the value of the multiplicative factor $\left|\int_{\mathbb{R}} J_1(\Psi_x\circ\sigma)dF_{\psi(X_1)}(x)\right|$ has to be computed under the given assumptions. For $x\geq 0$, integration by parts yields

$$J_1(\Psi_x\circ\sigma) = \alpha x^{-\alpha}\int_0^\infty z^{\alpha-1}\varphi\left(\log z\right)1_{\{x\geq z\}}dz.$$

Moreover, the probability density function $f_{\psi(X_1)}$ of $\psi(X_1)$ corresponds to

$$f_{\psi(X_1)}(x) = \alpha x^{-\alpha-1}\int_0^\infty z^{\alpha-1}\varphi\left(\log z\right)1_{\{x\geq z\}}dz.$$

As a result,

$$\left|\int_{\mathbb{R}} J_1(\Psi_x\circ\sigma)dF_{\psi(X_1)}(x)\right| = \int_0^\infty \alpha^2 x^{-2\alpha-1}\left(\int_0^\infty z^{\alpha-1}\varphi\left(\log z\right)1_{\{x\geq z\}}dz\right)^2 dx.$$

Since the above integral cannot be computed analytically, critical values for the Wilcoxon test are based on an approximation by numerical integration.

In order to compare the finite sample behavior of the change-point tests, the empirical size and the empirical power of the testing procedures is computed. For the computation of the empirical power, we consider LMSV time series with a change-point of height h after a proportion τ of the data, i.e. we consider non-centered Pareto distributed random variables ε_j, $j=1,\ldots,n$ with shape shape parameters α_j, $j=1,\ldots,n$ such that $\alpha_j=\alpha$ for $j=1,\ldots,\lfloor n\tau\rfloor$ while $\alpha_j=\alpha+h$ for $j=\lfloor n\tau\rfloor+1,\ldots,n$. The rejection rates of the testing procedures were computed on the basis of 5000 realizations of time series with sample sizes 500, 1000 and 2000. The simulation results are reported in Tables 4.5 and 4.6. The frequency of a type 1 error, i.e. the rejection rate under the hypothesis, corresponds to the values in the columns that are superscribed by $h=0$.

We make the following observations with respect to the behavior of the different testing procedures under the hypothesis:

- The empirical size of the non-self-normalized tests increases when the dependence in the simulated time series decreases, i.e. when H decreases. The empirical size of the self-normalized tests does not seem to be significantly affected by a change of H.

- Lighter tails, i.e. higher values of α, yield an empirical size that is closer the level of significance than heavier tails.
- The non-self-normalized tests tend to be oversized, while the self-normalized tests tend to be undersized.

Under the alternative we make the following observations:

- The empirical power of the change-point tests increases when the number of observations increases or the height of the level shift increases. It is higher for breakpoints located in the middle of the sample than for change-point locations that lie close to the boundary of the testing region.
- For all four change-point tests, an increase in correlation, i.e. an increase of H, leads to a decrease in the empirical power.
- For all four change-point tests, the empirical power decreases as heavy-tailedness decreases, i.e. as α increases.

A comparison of the finite sample performance of the testing procedures shows that:

- Once more, the *better size but less power* phenomenon for self-normalized tests can be observed.
- A comparison of the finite sample performance of the non-self-normalized tests does not give rise to choosing one of both tests over the other: While the empirical size of the Wilcoxon test tends to be closer to the significance level, the empirical power of the CUSUM test tends to be higher.
- A comparison of the finite sample performance of the self-normalized tests shows that although both tests have similar empirical size, the empirical power of the self-normalized CUSUM test exceeds the empirical power of the self-normalized Wilcoxon test for almost every combination of parameters.

All in all, the simulation results indicate that the self-normalized CUSUM test outperforms the self-normalized Wilcoxon test when testing for a change in the tail index. However, a comparison of the rejection rates of the self-normalized CUSUM test with those of the non-self-normalized testing procedures does not give rise to choosing one of these testing procedures over the others.
For a comparison of self-normalized and non-self-normalized change-point tests, it is important to note that the considered finite sample results are based on simulations which were executed under the assumption that the normalization of the non-self-normalized tests and the multiplicative quantities that appear in the limits of the corresponding test statistics are known. In particular, normalization and limit of the non-self-normalized statistics usually depend on the parameters H, r, the slowly-varying function L_γ that characterizes the autocovariances of the Gaussian random variables ξ_n, $n \in \mathbb{N}$, the distribution of ε_n, $n \in \mathbb{N}$, (or at least the tail parameter α), and the function σ.

For all practical purposes, these quantities are unknown, and for this reason have to be estimated. In contrast, the self-normalized test statistic can be computed from the given data while its limit depends on the parameters r and H only. For an adequate comparison of the testing procedures, this has to be taken into consideration.
All in all, the simulation studies show that the choice of the change-point test should depend on the particular test situation that is considered. In general, an application of Wilcoxon-type tests reduces the influence of heavy tails in data-generating processes on test decisions. As a result, Wilcoxon-based testing procedures yield better results when testing for changes in the mean and the variance of LMSV time series, while it might be advisable to choose CUSUM-based testing procedures when testing for a change in the tail parameter.

			CUSUM						Wilcoxon					
			$\alpha = 2.5$			$\alpha = 4$			$\alpha = 2.5$			$\alpha = 4$		
	H	n	$h = 0$	$h = 0.5$	$h = 1$	$h = 0$	$h = 0.5$	$h = 1$	$h = 0$	$h = 0.5$	$h = 1$	$h = 0$	$h = 0.5$	$h = 1$
$\tau = 0.25$	0.6	500	0.035	0.048	0.213	0.047	0.866	1.000	0.628	1.000	1.000	0.802	1.000	1.000
		1000	0.034	0.085	0.745	0.051	0.993	1.000	0.578	1.000	1.000	0.751	1.000	1.000
		2000	0.034	0.288	0.986	0.048	1.000	1.000	0.524	1.000	1.000	0.713	1.000	1.000
	0.7	500	0.035	0.050	0.209	0.053	0.864	1.000	0.331	1.000	1.000	0.475	1.000	1.000
		1000	0.037	0.087	0.752	0.058	0.994	1.000	0.270	1.000	1.000	0.384	1.000	1.000
		2000	0.039	0.285	0.983	0.051	1.000	1.000	0.207	1.000	1.000	0.300	1.000	1.000
	0.8	500	0.045	0.055	0.207	0.073	0.879	1.000	0.191	0.974	1.000	0.273	1.000	1.000
		1000	0.041	0.108	0.757	0.066	0.994	1.000	0.144	0.994	1.000	0.187	1.000	1.000
		2000	0.042	0.280	0.983	0.066	1.000	1.000	0.108	1.000	1.000	0.132	1.000	1.000
	0.9	500	0.057	0.069	0.191	0.080	0.901	1.000	0.188	0.863	0.984	0.232	0.995	1.000
		1000	0.059	0.111	0.783	0.090	0.992	1.000	0.139	0.888	0.990	0.165	0.996	1.000
		2000	0.064	0.238	0.984	0.092	1.000	1.000	0.108	0.917	0.996	0.121	0.998	1.000
$\tau = 0.5$	0.6	500		0.081	0.625		0.986	1.000		1.000	1.000		1.000	1.000
		1000		0.239	0.969		1.000	1.000		1.000	1.000		1.000	1.000
		2000		0.638	1.000		1.000	1.000		1.000	1.000		1.000	1.000
	0.7	500		0.092	0.621		0.986	1.000		1.000	1.000		1.000	1.000
		1000		0.239	0.961		1.000	1.000		1.000	1.000		1.000	1.000
		2000		0.648	0.999		1.000	1.000		1.000	1.000		1.000	1.000
	0.8	500		0.096	0.622		0.984	1.000		0.987	1.000		1.000	1.000
		1000		0.224	0.966		1.000	1.000		0.997	1.000		1.000	1.000
		2000		0.637	0.999		1.000	1.000		1.000	1.000		1.000	1.000
	0.9	500		0.100	0.627		0.986	1.000		0.919	0.993		0.994	1.000
		1000		0.207	0.973		1.000	1.000		0.929	0.995		0.999	1.000
		2000		0.635	1.000		1.000	1.000		0.954	0.997		0.999	1.000

Table 4.1.: Rejection rates of the CUSUM and Wilcoxon test for LMSV time series of length n with Hurst parameter H, tail index α and a shift in the mean of height h after a proportion τ. The calculations are based on 5000 simulation runs.

			self-norm. CUSUM						self-norm. Wilcoxon					
			$\alpha = 2.5$			$\alpha = 4$			$\alpha = 2.5$			$\alpha = 4$		
	H	n	$h = 0$	$h = 0.5$	$h = 1$	$h = 0$	$h = 0.5$	$h = 1$	$h = 0$	$h = 0.5$	$h = 1$	$h = 0$	$h = 0.5$	$h = 1$
$\tau = 0.25$	0.6	500	0.046	0.181	0.539	0.042	0.688	0.958	0.032	0.879	0.991	0.030	0.995	1.000
		1000	0.049	0.290	0.722	0.044	0.862	0.990	0.032	0.973	1.000	0.030	1.000	1.000
		2000	0.053	0.458	0.875	0.026	0.967	0.999	0.034	0.999	1.000	0.028	1.000	1.000
	0.7	500	0.051	0.204	0.552	0.042	0.697	0.954	0.029	0.680	0.938	0.021	0.960	0.997
		1000	0.050	0.295	0.727	0.046	0.866	0.990	0.032	0.856	0.988	0.027	0.993	1.000
		2000	0.049	0.455	0.868	0.042	0.966	0.998	0.037	0.948	0.999	0.030	0.999	1.000
	0.8	500	0.045	0.226	0.580	0.044	0.720	0.951	0.031	0.424	0.772	0.021	0.815	0.964
		1000	0.042	0.338	0.736	0.040	0.870	0.989	0.033	0.559	0.862	0.024	0.915	0.984
		2000	0.050	0.498	0.881	0.052	0.960	0.998	0.034	0.673	0.938	0.023	0.958	0.998
	0.9	500	0.044	0.329	0.645	0.041	0.760	0.947	0.031	0.309	0.582	0.020	0.640	0.861
		1000	0.051	0.446	0.761	0.042	0.871	0.980	0.039	0.369	0.650	0.034	0.734	0.912
		2000	0.041	0.585	0.869	0.048	0.949	0.996	0.049	0.422	0.719	0.039	0.791	0.947
$\tau = 0.5$	0.6	500		0.384	0.801		0.904	0.990		0.994	1.000		1.000	1.000
		1000		0.564	0.909		0.973	0.998		1.000	1.000		1.000	1.000
		2000		0.744	0.962		0.993	1.000		1.000	1.000		1.000	1.000
	0.7	500		0.401	0.801		0.902	0.989		0.950	1.000		1.000	1.000
		1000		0.565	0.904		0.972	0.998		0.993	1.000		1.000	1.000
		2000		0.744	0.966		0.994	0.999		1.000	1.000		1.000	1.000
	0.8	500		0.424	0.804		0.899	0.990		0.776	0.977		0.987	0.999
		1000		0.589	0.905		0.966	0.997		0.896	0.995		0.998	1.000
		2000		0.761	0.959		0.994	0.999		0.963	0.999		1.000	1.000
	0.9	500		0.527	0.815		0.893	0.982		0.622	0.890		0.912	0.990
		1000		0.650	0.898		0.959	0.997		0.708	0.936		0.956	0.996
		2000		0.781	0.954		0.989	0.999		0.779	0.960		0.976	0.998

Table 4.2.: Rejection rates of the self-normalized CUSUM and the self-normalized Wilcoxon test for LMSV time series of length n with Hurst parameter H, tail index α and a shift in the mean of height h after a proportion τ. The calculations are based on 5000 simulation runs.

			CUSUM						**Wilcoxon**					
				$\alpha = 4.5$			$\alpha = 6$			$\alpha = 4.5$			$\alpha = 6$	
	H	n	$h = 1$	$h = 0.5$	$h = 2$	$h = 1$	$h = 0.5$	$h = 2$	$h = 1$	$h = 0.5$	$h = 2$	$h = 1$	$h = 0.5$	$h = 2$
$\tau = 0.25$	0.6	500	0.464	0.252	0.963	0.467	0.270	0.978	0.119	0.925	0.937	0.130	0.929	0.929
		1000	0.589	0.383	0.995	0.596	0.385	0.997	0.112	0.994	0.995	0.118	0.995	0.996
		2000	0.708	0.529	1.000	0.694	0.574	1.000	0.108	1.000	1.000	0.104	1.000	1.000
	0.7	500	0.330	0.164	0.852	0.330	0.174	0.882	0.078	0.584	0.587	0.078	0.606	0.591
		1000	0.374	0.197	0.937	0.404	0.207	0.961	0.066	0.781	0.784	0.068	0.781	0.780
		2000	0.431	0.263	0.983	0.443	0.273	0.991	0.060	0.932	0.929	0.061	0.934	0.936
	0.8	500	0.235	0.116	0.670	0.244	0.111	0.686	0.074	0.314	0.328	0.074	0.332	0.319
		1000	0.258	0.131	0.770	0.256	0.132	0.786	0.067	0.398	0.386	0.067	0.398	0.382
		2000	0.275	0.139	0.837	0.271	0.137	0.861	0.060	0.502	0.499	0.059	0.505	0.501
	0.9	500	0.179	0.088	0.470	0.170	0.084	0.486	0.088	0.251	0.254	0.088	0.252	0.262
		1000	0.184	0.100	0.513	0.177	0.097	0.523	0.078	0.267	0.277	0.079	0.272	0.272
		2000	0.191	0.101	0.564	0.191	0.099	0.566	0.072	0.283	0.283	0.065	0.279	0.277
$\tau = 0.5$	0.6	500		0.377	0.954		0.417	0.967		0.988	0.990		0.990	0.990
		1000		0.565	0.992		0.594	0.996		1.000	1.000		1.000	1.000
		2000		0.774	0.999		0.814	0.999		1.000	1.000		1.000	1.000
	0.7	500		0.252	0.821		0.252	0.839		0.808	0.808		0.806	0.809
		1000		0.313	0.932		0.333	0.946		0.934	0.937		0.936	0.935
		2000		0.416	0.984		0.436	0.987		0.992	0.992		0.990	0.991
	0.8	500		0.170	0.623		0.166	0.651		0.529	0.525		0.515	0.528
		1000		0.196	0.733		0.199	0.757		0.605	0.615		0.624	0.631
		2000		0.226	0.838		0.230	0.848		0.741	0.738		0.746	0.734
	0.9	500		0.127	0.440		0.132	0.445		0.410	0.419		0.404	0.406
		1000		0.137	0.487		0.140	0.506		0.442	0.440		0.452	0.424
		2000		0.135	0.542		0.148	0.540		0.466	0.467		0.468	0.473

Table 4.3.: Rejection rates of the CUSUM and Wilcoxon test for LMSV time series of length n with Hurst parameter H, tail index α and a shift in the variance of height h^2 after a proportion τ. The calculations are based on 5000 simulation runs.

			self-norm. CUSUM						self-norm. Wilcoxon					
			$\alpha = 4.5$			$\alpha = 6$			$\alpha = 4.5$			$\alpha = 6$		
	H	n	$h = 1$	$h = 0.5$	$h = 2$	$h = 1$	$h = 0.5$	$h = 2$	$h = 1$	$h = 0.5$	$h = 2$	$h = 1$	$h = 0.5$	$h = 2$
$\tau = 0.25$	0.6	500	0.035	0.229	0.040	0.038	0.237	0.043	0.040	0.518	0.495	0.041	0.511	0.494
		1000	0.033	0.291	0.043	0.035	0.317	0.050	0.043	0.736	0.739	0.047	0.731	0.745
		2000	0.033	0.383	0.048	0.034	0.423	0.060	0.044	0.901	0.897	0.042	0.908	0.905
	0.7	500	0.022	0.146	0.022	0.024	0.161	0.025	0.046	0.252	0.248	0.048	0.263	0.259
		1000	0.022	0.192	0.023	0.025	0.225	0.027	0.044	0.380	0.381	0.046	0.380	0.378
		2000	0.020	0.263	0.030	0.021	0.296	0.034	0.049	0.526	0.528	0.050	0.531	0.536
	0.8	500	0.015	0.086	0.013	0.019	0.098	0.015	0.038	0.120	0.119	0.042	0.136	0.127
		1000	0.015	0.107	0.017	0.016	0.128	0.017	0.049	0.159	0.155	0.045	0.165	0.156
		2000	0.015	0.140	0.014	0.017	0.171	0.016	0.047	0.198	0.192	0.046	0.197	0.198
	0.9	500	0.019	0.079	0.018	0.023	0.088	0.021	0.048	0.100	0.101	0.049	0.097	0.096
		1000	0.019	0.099	0.015	0.021	0.112	0.020	0.053	0.109	0.107	0.050	0.110	0.105
		2000	0.020	0.123	0.020	0.023	0.133	0.023	0.056	0.118	0.114	0.047	0.104	0.123
$\tau = 0.5$	0.6	500		0.145	0.152		0.158	0.164		0.816	0.831		0.823	0.816
		1000		0.206	0.193		0.227	0.221		0.965	0.963		0.962	0.960
		2000		0.277	0.270		0.324	0.318		0.998	0.998		0.996	0.996
	0.7	500		0.090	0.090		0.100	0.093		0.524	0.530		0.534	0.521
		1000		0.125	0.121		0.130	0.135		0.701	0.695		0.698	0.699
		2000		0.165	0.168		0.196	0.193		0.859	0.863		0.854	0.863
	0.8	500		0.045	0.046		0.053	0.054		0.270	0.274		0.270	0.278
		1000		0.063	0.064		0.062	0.070		0.357	0.361		0.350	0.361
		2000		0.073	0.086		0.099	0.083		0.439	0.443		0.454	0.454
	0.9	500		0.044	0.037		0.047	0.040		0.200	0.205		0.195	0.198
		1000		0.053	0.052		0.064	0.058		0.228	0.224		0.246	0.228
		2000		0.066	0.065		0.076	0.074		0.263	0.248		0.248	0.264

Table 4.4.: Rejection rates of the self-normalized CUSUM and the self-normalized Wilcoxon test for LMSV time series of length n with Hurst parameter H, tail index α and a shift in the variance of height h^2 after a proportion τ. The calculations are based on 5000 simulation runs.

			CUSUM						**Wilcoxon**					
			$\alpha = 0.5$			$\alpha = 1$			$\alpha = 0.5$			$\alpha = 1$		
	H	n	$h = 0$	$h = 0.25$	$h = 0.5$	$h = 0$	$h = 0.25$	$h = 0.5$	$h = 0$	$h = 0.25$	$h = 0.5$	$h = 0$	$h = 0.25$	$h = 0.5$
$\tau = 0.25$	0.6	500	0.457	0.884	0.994	0.139	0.215	0.393	0.303	0.807	0.985	0.123	0.213	0.388
		1000	0.419	0.982	1.000	0.121	0.298	0.588	0.277	0.931	1.000	0.113	0.269	0.538
		2000	0.388	0.999	1.000	0.113	0.436	0.825	0.259	0.991	1.000	0.100	0.374	0.747
	0.7	500	0.213	0.601	0.879	0.084	0.108	0.184	0.148	0.538	0.828	0.078	0.132	0.220
		1000	0.177	0.761	0.977	0.071	0.134	0.249	0.123	0.635	0.930	0.067	0.144	0.272
		2000	0.141	0.907	0.998	0.071	0.151	0.350	0.105	0.790	0.987	0.066	0.159	0.347
	0.8	500	0.131	0.329	0.590	0.064	0.081	0.105	0.104	0.340	0.587	0.065	0.113	0.153
		1000	0.107	0.379	0.716	0.064	0.081	0.115	0.090	0.357	0.665	0.064	0.108	0.155
		2000	0.087	0.491	0.836	0.054	0.083	0.138	0.078	0.428	0.756	0.058	0.102	0.170
	0.9	500	0.087	0.201	0.376	0.056	0.058	0.079	0.092	0.275	0.474	0.086	0.129	0.171
		1000	0.075	0.216	0.425	0.054	0.068	0.080	0.082	0.279	0.493	0.075	0.126	0.163
		2000	0.061	0.221	0.480	0.051	0.061	0.074	0.067	0.273	0.520	0.072	0.110	0.147
$\tau = 0.5$	0.6	500		0.971	1.000		0.312	0.597		0.915	0.998		0.280	0.540
		1000		0.997	1.000		0.446	0.812		0.985	1.000		0.381	0.739
		2000		1.000	1.000		0.628	0.953		0.999	1.000		0.526	0.901
	0.7	500		0.794	0.976		0.161	0.296		0.674	0.932		0.160	0.291
		1000		0.916	0.998		0.188	0.402		0.795	0.988		0.180	0.381
		2000		0.982	1.000		0.241	0.531		0.915	0.999		0.221	0.483
	0.8	500		0.509	0.812		0.107	0.164		0.434	0.735		0.119	0.188
		1000		0.611	0.902		0.109	0.192		0.501	0.817		0.123	0.208
		2000		0.709	0.957		0.123	0.227		0.567	0.888		0.134	0.234
	0.9	500		0.339	0.611		0.070	0.106		0.342	0.588		0.120	0.179
		1000		0.360	0.656		0.068	0.121		0.345	0.606		0.116	0.183
		2000		0.391	0.716		0.071	0.127		0.356	0.645		0.113	0.183

Table 4.5.: Rejection rates of the CUSUM and Wilcoxon test for LMSV time series of length n with Hurst parameter H and a change in the tail index α of height h after a proportion τ. The calculations are based on 5000 simulation runs.

			self-norm. CUSUM						self-norm. Wilcoxon					
			$\alpha = 0.5$			$\alpha = 1$			$\alpha = 0.5$			$\alpha = 1$		
	H	n	$h = 0$	$h = 0.25$	$h = 0.5$	$h = 0$	$h = 0.25$	$h = 0.5$	$h = 0$	$h = 0.25$	$h = 0.5$	$h = 0$	$h = 0.25$	$h = 0.5$
$\tau = 0.25$	0.6	500	0.035	0.373	0.725	0.040	0.087	0.181	0.037	0.221	0.461	0.043	0.072	0.145
		1000	0.033	0.596	0.912	0.045	0.124	0.277	0.041	0.383	0.707	0.049	0.100	0.211
		2000	0.038	0.819	0.985	0.045	0.206	0.453	0.037	0.584	0.888	0.045	0.159	0.334
	0.7	500	0.026	0.189	0.435	0.041	0.065	0.100	0.032	0.120	0.242	0.042	0.059	0.082
		1000	0.038	0.294	0.596	0.043	0.079	0.137	0.046	0.167	0.347	0.046	0.069	0.115
		2000	0.040	0.452	0.774	0.047	0.098	0.191	0.042	0.263	0.502	0.049	0.081	0.151
	0.8	500	0.028	0.092	0.194	0.041	0.046	0.056	0.034	0.062	0.111	0.043	0.044	0.053
		1000	0.034	0.130	0.276	0.047	0.053	0.064	0.043	0.090	0.146	0.049	0.054	0.056
		2000	0.040	0.177	0.350	0.046	0.050	0.078	0.043	0.107	0.188	0.047	0.049	0.068
	0.9	500	0.030	0.073	0.133	0.047	0.044	0.058	0.039	0.059	0.081	0.049	0.048	0.060
		1000	0.038	0.083	0.157	0.050	0.055	0.060	0.047	0.066	0.095	0.053	0.058	0.061
		2000	0.046	0.101	0.189	0.052	0.055	0.058	0.054	0.077	0.099	0.055	0.057	0.057
$\tau = 0.5$	0.6	500		0.603	0.907		0.137	0.318		0.456	0.813		0.112	0.259
		1000		0.843	0.989		0.227	0.517		0.703	0.960		0.179	0.420
		2000		0.974	1.000		0.363	0.732		0.901	0.996		0.288	0.612
	0.7	500		0.364	0.692		0.085	0.167		0.262	0.535		0.078	0.136
		1000		0.551	0.869		0.120	0.248		0.388	0.717		0.099	0.197
		2000		0.750	0.969		0.157	0.344		0.562	0.882		0.124	0.275
	0.8	500		0.195	0.414		0.057	0.091		0.136	0.290		0.057	0.083
		1000		0.296	0.539		0.065	0.106		0.193	0.370		0.063	0.094
		2000		0.383	0.671		0.082	0.135		0.234	0.463		0.076	0.115
	0.9	500		0.135	0.288		0.059	0.083		0.102	0.192		0.057	0.080
		1000		0.181	0.350		0.059	0.092		0.120	0.220		0.060	0.088
		2000		0.216	0.410		0.060	0.085		0.143	0.263		0.060	0.080

Table 4.6.: Rejection rates of the self-normalized CUSUM and the self-normalized Wilcoxon test for LMSV time series of length n with Hurst parameter H and a change in the tail index α of height h after a proportion τ. The calculations are based on 5000 simulation runs.

5. Data examples

On the basis of corresponding evaluations in Betken and Wendler (2015), in the following, three different data sets are analyzed with regard to changes in the mean by applications of the sampling-window method considered in Chapter 3 and the change-point estimator $\hat{k}_W$ established in Chapter 2. All three data sets are included in any standard distribution of `R`.

5.1. Nile river discharge

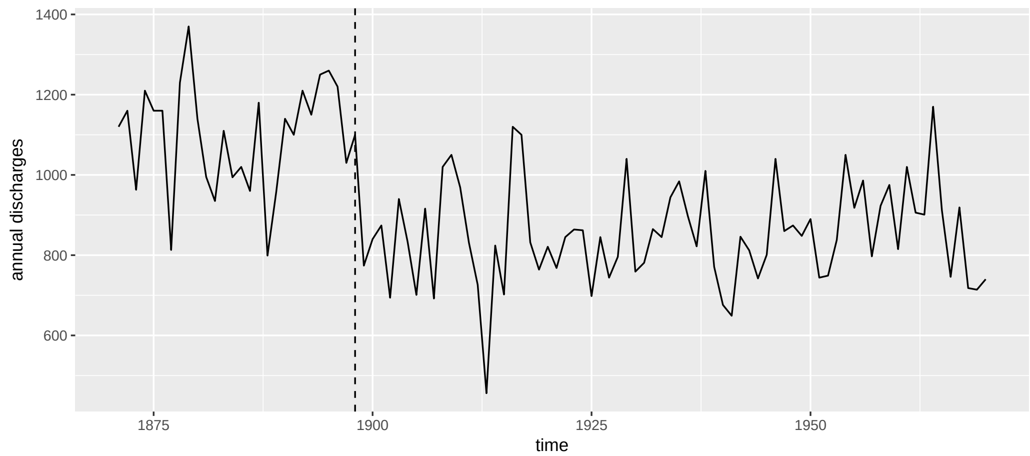

Figure 5.1.: *Measurements of the annual discharge volume of the river Nile at Aswan in* $10^8 m^3$ *for the years 1871-1970. The dashed line indicates the potential change-point location estimated by* $\hat{k}_W$*.*

The first data set consists of annual measurements of the discharge volume from the Nile river at Aswan in $10^8 m^3$ for the years 1871 to 1970; see Figure 5.1. The data has been taken from the `datasets` package in `R`. It has been analyzed for the detection of a change-point by numerous authors under differing assumptions concerning the data generating random process and by usage of diverse methods. Amongst others, Cobb (1978), MacNeill et al. (1991), Wu and Zhao (2007), and Shao (2011) provided statistically significant evidence for a decrease of the Nile's annual discharge towards the end of the 19th century. The construction of the Aswan Low Dam between 1898 and 1902 serves as a popular explanation for an abrupt change in the data.

Yet, Cobb (1978) gave another explanation for the decrease in water volume by citing rainfall records which suggest a decline of tropical rainfall at that time.
Computed with respect to this data set, the value of the self-normalized Wilcoxon test statistic, defined by (1.9) in Chapter 1, corresponds to $SW_n(\tau_1, \tau_2) = 13.48729$ when $\tau_1 = 1 - \tau_2 = 0.15$. A comparison of this value with the 95%-quantile of the statistic's asymptotic distribution, reported in Betken (2016), leads to a rejection of the hypothesis for every possible choice of H. Based on an approximation of the finite sample distribution of the self-normalized Wilcoxon test statistic by the sampling-window method with block size $l_n \in \{\lfloor n^{\gamma} \rfloor | \ \gamma = 0.5, 0.6, 0.7\} = \{10, 15, 25\}$, subsampling indicates the existence of a change-point in the mean of the data, even if the 99%-quantile of the empirical distribution function $\hat{F}_{m_n, l_n}$, $m_n = n - l_n + 1$, defined by (3.1) in Chapter 3, is considered.
Previous analysis of the Nile data, done by Wu and Zhao (2007) and Balke (1993), suggests that the change in the discharge volume occurred in 1899. An application of the change-point estimator $\hat{k}_W$, defined by (2.1) in Chapter 2, identifies a change in 1898. This result seems to be in good accordance with the estimated change-point locations suggested by other authors: Cobb's analysis of the Nile data leads to the conjecture of a significant decrease in discharge volume in 1898. Moreover, computation of the CUSUM-based change-point estimator $\hat{k}_{C,0}$, considered in Horváth and Kokoszka (1997) and defined by (2.2) in Chapter 2, indicates a change in 1898.
An application of the self-normalized Wilcoxon test statistic to the corresponding pre-break and post-break samples provides evidence for stationarity of the time series before and after the estimated change-point: For both subsamples, neither a comparison of $SW_n(\tau_1, \tau_2)$, $\tau_1 = 1 - \tau_2 = 0.15$, with the 90%-quantile of the sampling distribution $\hat{F}_{m_n, l_n}$ nor a comparison with the 90%-quantile of its limit distribution, leads to a rejection of the hypothesis (for any possible choice of the block length and any value of H). Based on the whole sample, local Whittle estimation with bandwidth parameter $b_n = \lfloor n^{2/3} \rfloor$, as previously considered in Section 3.2, suggests the existence of long-range dependence characterized by a Hurst parameter $\hat{H} = 0.962$, whereas the estimates for the pre-break and post-break samples, given by $\hat{H}_1 = 0.517$ and $\hat{H}_2 = 0.5$, respectively, should be considered as indication of short-range dependent data. These findings support the conjecture of spurious long-range dependence caused by a change-point and therefore agree with the results of Shao (2011).

5.2. Northern hemisphere temperature

The second data set consists of the seasonally adjusted monthly deviations of the temperature (degrees Celsius) for the Northern hemisphere during the years 1854 to 1989 from the monthly averages over the period 1950 to 1979; see Figure 5.2. The data has been taken from the `longmemo` package in `R`. It results from spatial averaging of temperatures measured over land and sea.
At first sight, the plot in Figure 5.2 suggests non-stationarity of the data-generating process, possibly caused by an increasing trend and an abrupt change of the temperature

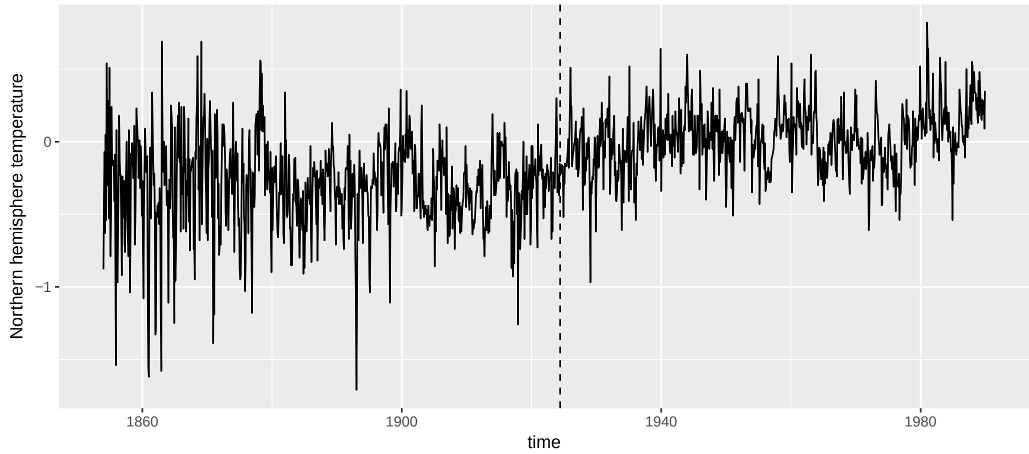

Figure 5.2.: *Monthly temperature of the Northern hemisphere for the years 1854 – 1989 from the data base held at the Climate Research Unit of the University of East Anglia, Norwich, England. The temperature anomalies (in degrees C) are calculated with respect to the reference period 1950 – 1979. The dashed line indicates the potential change-point location estimated by* $\hat{k}_W$.

deviations. Statistical evidence for a positive deterministic trend can be interpreted as affirmation of the conjecture that there has been global warming during the last decades. The question of whether the Northern hemisphere temperature data acts as an indicator for global warming of the atmosphere is a controversial issue. Previous analysis of this data offers different explanations for the irregular behavior of the time series. Deo and Hurvich (1998) fitted a linear trend to the data, thereby providing statistical evidence for global warming during the last decades. Beran and Feng (2002) considered a more general stochastic model by the assumption of so-called semiparametric fractional autoregressive (SEMIFAR) processes. Their method did not deliver sufficient statistical evidence for a deterministic trend. Neither does the investigation of the global temperature data in Wang (2007) support the hypothesis of an increasing trend. In fact, Wang (2007) offers an alternative explanation for the occurrence of a trend-like behavior by pointing out that it may have been generated by stationary long-range dependent processes. In contrast, it is shown in Shao (2011) that under model assumptions that include long-range dependence, the existence of a change-point in the mean yields yet another explanation for the performance of the data.

The value of the self-normalized Wilcoxon test statistic for the temperature data is $SW_n(\tau_1, \tau_2) = 18.98636$, when $\tau_1 = 1 - \tau_2 = 0.15$. By comparison of this value with the asymptotic critical values for the corresponding hypothesis test, the hypothesis of stationarity is rejected for every value of H at a level of significance of 1%; see Betken (2016). An application of the sampling-window method with respect to the self-normalized Wilcoxon test statistic based on comparison of $SW_n(\tau_1, \tau_2)$ with the 99%-quantile of the sampling distribution $\hat{F}_{m_n,l_n}$, $m_n = n - l_n + 1$, yields a test decision in favor of a change-point in the mean for any choice of the block length

$l_n \in \{\lfloor n^\gamma \rfloor |\ \gamma = 0.3, 0.4, \ldots, 0.9\} = \{9, 19, 40, 84, 177, 371, 778\}$. All in all, both testing procedures provide strong evidence for the existence of a change in the mean.

According to Shao (2011), estimation of the change-point location based on a self-normalized CUSUM test statistic suggests a structural change around October 1924. Computation of the change-point estimator $\hat{k}_W$ corresponds to a change-point located around June 1924. The same change-point location results from an application of the estimator $\hat{k}_{C,0}$. In this regard, estimation by $\hat{k}_W$ seems to be in good accordance with the results of alternative change-point estimators.

Based on the whole sample, local Whittle estimation with bandwidth $b_n = \lfloor n^{2/3} \rfloor$ provides an estimator $\hat{H} = 0.811$. The estimated Hurst parameters for the pre-break and post-break sample are $\hat{H}_1 = 0.597$ and $\hat{H}_2 = 0.88$, respectively. On the basis of a significance level of 10%, neither of both testing procedures, i.e. subsampling with respect to the self-normalized Wilcoxon test statistic and comparison of the value of $SW_n(\tau_1, \tau_2)$, $\tau_1 = 1 - \tau_2 = 0.15$, with the corresponding critical values of its limit distribution, provides evidence for another change-point in the pre-break or post-break sample (for any possible choice of block length and any value of H).

In Appendix B.2 testing procedures that allow for more than one change-point are considered. These are based on test statistics resulting from corresponding modifications of the Wilcoxon statistics defined in Sections 1.2.2 and 1.2.3. Computation of the self-normalized Wilcoxon statistic that allows for two breakpoints, denoted by $SW_n(\tau_1, \tau_2, \varepsilon)$ and defined by formula (B.2) in the appendix, yields $SW_n(\tau_1, \tau_2, \varepsilon) = 17.88404$ for $\tau_1 = 1 - \tau_2 = \varepsilon = 0.15$. This value only surpasses the critical value corresponding to $H = 0.501$ at a significance level of 10% (see Table B.1), but does not exceed any of the other quantiles. Subsampling with respect to the test statistic $SW_n(\tau_1, \tau_2, \varepsilon)$ does not support the conjecture of two changes, either. In fact, subsampling leads to a rejection of the hypothesis when the block length equals $l_n = \lfloor n^{0.7} \rfloor = 177$ (based on a comparison of $SW_n(\tau_1, \tau_2, \varepsilon)$ with the 95%-quantile of the corresponding sampling distribution $\hat{F}_{m_n, l_n}$, but yields a test decision in favor of the hypothesis for block lengths $l_n \in \{\lfloor n^\gamma \rfloor |\ \gamma = 0.5, 0.6, 0.8, 0.9\} = \{40, 84, 371, 778\}$ and for comparison with the 90%-quantile of $\hat{F}_{m_n, l_n}$. Therefore, it seems safe to conclude that the appearance of long-range dependence in the post-break sample is not caused by another change-point in the mean.

The pronounced difference between the local Whittle estimators $\hat{H}_1$ and $\hat{H}_2$ can be interpreted as indication of a change in the dependence structure of the time series. Another explanation could be a gradual change of the temperature in the post-break period.

5.3. Ethernet traffic

The third data set consists of the arrival rate of Ethernet data (bytes per 10 milliseconds) from a local area network (LAN) measured at Bellcore Research and Engineering Center in 1989. The data has been taken from the `longmemo` package in `R`. For more information on the LAN traffic monitoring, see Leland and Wilson (1991) and Beran (1994).

Figure 5.3 reveals that the observations are strongly right-skewed. As Wilcoxon-like statistics can be computed from ranks, this is not expected to affect tests and estimators

that are based on these statistics.
Coulon et al. (2009) examined this data set in view of change-points under the assumption that a FARIMA model holds for segments of the data. The number of different sections and the location of potential change-points are chosen by a model selection criterion. The algorithm proposed by Coulon et al. (2009) detects multiple changes in the parameters of the corresponding FARIMA time series. In contrast, an application of the self-normalized Wilcoxon change-point test does not provide evidence for a change-point in the mean: the value of the test statistic is given by $SW_n(\tau_1, \tau_2) = 3.270726$ when $\tau_1 = 1 - \tau_2 = 0.15$. Even for a level of significance of 10%, the self-normalized Wilcoxon change-point test does not reject the hypothesis for any value of H. Subsampling with respect to the self-normalized Wilcoxon test statistic does not lead to a rejection of the hypothesis, either (for any choice of block length $l_n \in \{\lfloor n^\gamma \rfloor \mid \gamma = 0.3, 0.4, \ldots, 0.9\} = \{12, 27, 63, 144, 332, 761, 1745\}$ and for comparison with the 90%-quantile of the sampling distribution $\hat{F}_{m_n,l_n}$, $m_n = n - l_n + 1$).

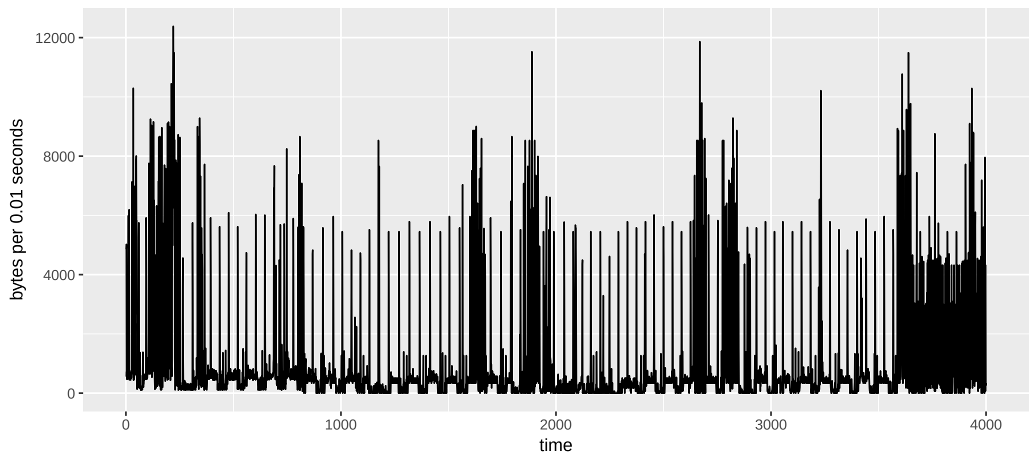

Figure 5.3.: *Ethernet traffic in bytes per 10 milliseconds from a LAN measured at Bellcore Research Engineering Center.*

Since the value 0 appears several times in the Ethernet traffic data, it seems reasonable to also consider change-point tests that allow for ties in the data. In Appendix B.1, corresponding testing procedures, which are based on test statistics resulting from modifications of the Wilcoxon statistics defined in Sections 1.2.2 and 1.2.3, are considered. As noted in Appendix B.1, test decisions on the basis of the asymptotic distributions of the modified statistics are not feasible without specific knowledge of the data-generating process. Yet, it is possible to apply subsampling with respect to the modified self-normalized Wilcoxon test statistic $SW_n^*(\tau_1, \tau_2)$ defined by (B.1) in the appendix. A comparison of $SW_n^*(\tau_1, \tau_2)$, $\tau_1 = 1 - \tau_2 = 0.15$, with the 90%-quantile of the corresponding sampling distribution $\hat{F}_{m_n,l_n}$ does not lead to a rejection of the hypothesis (for any possible choice of the block length).

An application of the test statistic constructed for the detection of two changes yields $SW_n(\tau_1, \tau_2, \varepsilon) = 15.24527$ when $\varepsilon = \tau_1 = 1 - \tau_2 = 0.15$. Comparing this value to the 90%-quantiles of the asymptotic distribution of $SW_n(\tau_1, \tau_2, \varepsilon)$, reported in Table B.1 in Section B.1 of the appendix, does not lead to a rejection of the hypothesis for any value of the parameter H. Subsampling based on a comparison of $SW_n(\tau_1, \tau_2, \varepsilon)$ with the 90%-quantile of the corresponding sampling-window estimate $\hat{F}_{m_n,l_n}$ does not provide evidence for more than one change-point in the data for any choice of the block length $l_n \in \{\lfloor n^\gamma \rfloor | \gamma = 0.5, 0.6, 0.7, 0.8\} = \{63, 144, 332, 761\}$, either. These results do not coincide with the analysis of Coulon et al. (2009). This may be due to the fact that the considered methods differ considerably from the testing procedures applied before. The change-point estimation algorithm proposed in Coulon et al. (2009) is not robust to skewness or heavy-tailed distributions and decisively relies on the assumption of FARIMA time series. This seems to contradict observations made by Bhansali and Kokoszka (2001) as well as Taqqu and Teverovsky (1997) who stress that the Ethernet traffic data is very unlikely to be generated by FARIMA processes.
Estimation of the Hurst parameter by the local Whittle procedure with bandwidth parameter $b_n = \lfloor n^{2/3} \rfloor$ yields an estimate $\hat{H} = 0.845$ indicating long-range dependence. This is consistent with the results of Leland et al. (1994) and Taqqu and Teverovsky (1997).

A. Skorohod spaces

Given a stationary sequence of random variables X_n, $n \in \mathbb{N}$, with marginal distribution function F, the determination of the asymptotic distribution of many non-parametric statistics, such as U-statistics or von Mises statistics, can be derived from the asymptotic behavior of the two-parameter empirical process

$$e_n(x,t) := \sum_{i=1}^{\lfloor nt \rfloor} \left(1_{\{X_i \leq x\}} - F(x)\right), \; x \in [-\infty, \infty], \; t \in [0,1].$$

In this thesis, limit theorems for the empirical process are considered in Chapters 1, 2, and 4 with the objective of deriving the asymptotic distribution of Wilcoxon-type test statistics and change-point estimators.
Since the sample paths of the two-parameter process $e_n(x,t)$, $x \in [-\infty, \infty]$, $t \in [0,1]$, contain jumps, the space $C([-\infty, \infty] \times [0,1])$ of continuous, real-valued functions on $[-\infty, \infty] \times [0,1]$ cannot be considered as a suitable path space. For this reason, weak convergence is studied in the Skorohod space $D([-\infty, \infty] \times [0,1])$ which allows for sample paths with certain discontinuities.

A.1. Topologies on Skorohod spaces

Given a compact interval $I \subset \bar{\mathbb{R}}$, the space $D(I)$ is defined as the set of all functions on I which are right-continuous and have left limits. An element of $D(I)$ is also referred to as càdlàg function (French: ***c**ontinue **à** **d**roite, **l**imite **à** **g**auche*).
For the definition of $D := D(K)$, where $K = [a_1, b_1] \times [a_2, b_2]$ with $a_1, a_2, b_1, b_2 \in \bar{\mathbb{R}}$, $a_1 < b_1$, $a_2 < b_2$, the terms *right-continuity* and *one-sided limits* are generalized by the expressions *continuity from above* and *quadrant limits*. For a suitable definition of continuity from above, which applies to functions with bidimensional domain K, we consider the intervals $I_k(s,t)$, $t \in [a_k, b_k]$, $s \in \{a_k, b_k\}$, defined by

$$I_k(s,t) := \begin{cases} [a_k, t) & \text{if } s = a_k, \\ (t, b_k] & \text{if } s = b_k \end{cases}$$

for $k = 1, 2$. With $\mathcal{V}$ denoting the set of all vertices of K, i.e.

$$\mathcal{V} := \left\{(a_1, a_2)^t, (a_1, b_2)^t, (b_1, a_2)^t, (b_1, b_2)^t\right\},$$

we write

$$Q(v,x) := I_1(v_1, x_1) \times I_2(v_2, x_2)$$

for any $v \in \mathcal{V}$ and $x = (x_1, x_2)^t \in [a_1, b_1] \times [a_2, b_2]$.

Definition 13 (Seijo and Sen (2011))**.** Given $v \in \mathcal{V}$, $x \in K$, with K and $\mathcal{V}$ defined as above, and a function $f : K \longrightarrow \mathbb{R}$, we say that l is the *v-limit* of f at x if for every sequence x_n, $n \in \mathbb{N}$, in $Q(v, x)$ that converges to x, we have $\lim_{n\to\infty} f(x_n) = l$. In this case, we define $f(x + 0_v) := l$. When $v = (b_1, b_2)^t$ we write $f(x + 0_+) := f(x + 0_b)$. The function f is said to be *continuous from above* at x if $f(x + 0_+) = f(x)$.

Based on the notion of continuity from above, it is possible to define the Skorohod space on rectangles K.

Definition 14 (Seijo and Sen (2011))**.** The *Skorohod space* D is defined as the collection of all functions $f : K \longrightarrow \mathbb{R}$ for which all v-limits exist and which are continuous from above for every $x \in K$.

Equipped with the uniform metric $d_u : D \times D \to \mathbb{R}$ defined by

$$d_u(x, y) := \sup_{t\in K} |x(t) - y(t)|$$

for $x, y \in D$, the space (D, d_u) is a complete metric space. However, it is not separable; Billingsley (1968). For this reason, we consider the Skorohod metric as an alternative distance function on D.

Definition 15 (Seijo and Sen (2011))**.** Let Λ_K denote the class of all (with respect to the coordinatewise order on $\mathbb{R}^2$) strictly increasing, continuous bijections $\lambda : K \longrightarrow K$ and let $\mathrm{id} : K \longrightarrow K$ denote the identity function. The *Skorokhod metric* d_s on D is defined by

$$d_s(x, y) := \inf_{\lambda\in\Lambda_K} \max \{d_u(\lambda, \mathrm{id}), d_u(y, x \circ \lambda)\}$$

for $x, y \in D$.

Every $\lambda \in \Lambda_K$ in the definition of the Skorohod metric can be interpreted as the representation of a small perturbation in time whose size is measured by $d_u(\lambda, \mathrm{id})$ while, at the same time, $d_u(y, x \circ \lambda)$ measures the size of perturbations in space. Therefore, d_s allows for small deformations of functions in space and time whereas d_u only allows for perturbations in space.
A sequence of càdlàg functions x_n, $n \in \mathbb{N}$, in D converges to a limit x in the Skorohod topology if and only if there exist functions λ_n, $n \in \mathbb{N}$, in Λ_K such that $d_u(\lambda_n, \mathrm{id})$ and $d_u(x_n, x \circ \lambda_n)$ converge to 0. As a result, convergence with respect to the metric d_u implies convergence with respect to the metric d_s.
With $\mathcal{U}$ and $\mathcal{S}$ denoting the Borel σ-algebras with respect to d_u and d_s, respectively, it follows that $\mathcal{S} \subset \mathcal{U}$, i.e. the uniform topology is finer than the Skorohod topology. On the other hand, a sequence of càdlàg functions x_n, $n \in \mathbb{N}$, in D that converges with respect to the Skorohod metric to a (uniformly) continuous function x, also converges to x with respect to the uniform metric since

$$d_u(x_n, x) \leq 2 \max \{d_u(x_n, x \circ \lambda_n), d_u(x \circ \lambda_n, x)\}.$$

Therefore, the Skorohod topology and the uniform topology coincide if relativized to the space $C := C(K)$ of continuous, real-valued functions on K. Nonetheless, the incongruity of the topologies implies subtle difficulties when considering weak convergence of random elements with values in D.

A.2. Weak convergence in Skorohod spaces

Given probability measures P_n, $n \in \mathbb{N}$, and P defined on $\mathcal{U}$ (and therefore also defined on $\mathcal{S}$), weak convergence of P_n, $n \in \mathbb{N}$, to P, denoted by $P_n \Rightarrow P$, is defined by requiring

$$\lim_{n\to\infty} \int h \, dP_n = \int h \, dP \quad \text{for all } h \in C_b(D),$$

where $C_b(D)$ denotes the class of all bounded, continuous real-valued functions on D. Based on this definition of weak convergence of probability measures, a sequence of random variables X_n, $n \in \mathbb{N}$, is said to converge in distribution to some random variable X, if and only if the induced laws of X_n, $n \in \mathbb{N}$, converge weakly to the induced law of X, i.e. if and only if

$$\lim_{n\to\infty} \mathrm{E}\, h(X_n) = \mathrm{E}\, h(X) \quad \text{for all } h \in C_b(D). \tag{A.1}$$

We indicate convergence in distribution by the notation $X_n \xrightarrow{\mathcal{D}} X$.
Since continuity depends on the topology that is considered, we write $P_n \Rightarrow_u P$ for convergence with respect to the uniform topology and $P_n \Rightarrow_s P$ for convergence with respect to the Skorohod topology. Weak convergence with respect to the uniform topology requires the probability measures P_n, $n \in \mathbb{N}$, and P to be defined on $\mathcal{U}$, while probability measures which converge weakly with respect to the Skorohod topology have to be defined on the smaller σ-field $\mathcal{S}$ only. For this reason, the possibility of an extension of probability measures defined on $\mathcal{S}$ to the larger σ-field $\mathcal{U}$ is crucial to relating weak convergence with respect to d_s and weak convergence with respect to d_u.

Theorem 13 (Billingsley (1968))**.** *Suppose P_n, $n \in \mathbb{N}$, and P are probability measures defined on $\mathcal{U}$.*

a) If $P_n \Rightarrow_u P$, then $P_n \Rightarrow_s P$.

b) If $P_n \Rightarrow_s P$ and $P(C) = 1$, then $P_n \Rightarrow_u P$.

For a probability measure P on $\mathcal{S}$ with $P(C) = 1$, an extension of P from $\mathcal{S}$ to $\mathcal{U}$ can be based on the fact that the uniform topology and the Skorohod topology coincide if relativized to C. However, in general, an extension is not possible. As shown in Billingsley, the empirical process of independent, uniformly distributed random variables is an $\mathcal{S}$-measurable, $D[0,1]$-valued random variable, which is not measurable with respect to $\mathcal{U}$, so that the distribution of the empirical process cannot be extended from $\mathcal{S}$ to $\mathcal{U}$. To overcome the problems resulting from non-separability of the metric space (D, d_u), Hoffmann-Jørgensen (1991) suggests to drop the requirement of Borel measurability by replacing the expected values in the convergence condition (A.1) by outer expectations, thereby extending the classical notion of weak convergence to a theory of *weak convergence of laws without laws being defined*:

Definition 16 (Van Der Vaart and Wellner (1996))**.** Let (M, d) be a metric space. Given probability spaces $(\Omega, \mathcal{F}, P)$, $(\Omega_n, \mathcal{F}_n, P_n)$, $n \in \mathbb{N}$, and arbitrary (possibly non-measurable) maps $X_n : \Omega_n \longrightarrow M$, $n \in \mathbb{N}$, and $X : \Omega \longrightarrow M$, X_n, $n \in \mathbb{N}$, is said to *converge in distribution in Hoffmann – Jørgensen's sense* to X, denoted by $X_n \rightsquigarrow X$, if $P(X \in M_0) = 1$ for some separable Borel set $M_0 \subset M$ and if

$$\lim_{n\to\infty} \mathrm{E}^* h(X_n) = \mathrm{E}\, h(X) \quad \text{for all } h \in C_b(M),$$

where $C_b(M)$ denotes the class of all bounded, continuous real-valued functions defined on M, and where the outer expectation E^* is defined by

$$\mathrm{E}^* f := \inf \{\mathrm{E}\, h \mid f \leq h \text{ and } h \text{ is measurable}\}.$$

Hoffmann – Jørgensen's definition of convergence in distribution corresponds to the classical definition of weak convergence for Borel measurable random variables X_n, $n \in \mathbb{N}$, and, even in the case of non-measurability, parallels the classical theory to a large extent: Van Der Vaart and Wellner (1989) establish fundamental results of the classical weak convergence theory on the basis of the above definition. These include a Portmanteau theorem, continuous mapping theorems, Prohorov's theorem, tightness and tools for establishing tightness.
An alternative theory of weak convergence has been established in Dudley (1966, 1967). Instead of allowing for non-measurability by operating with outer expectations, Dudley suggests to replace the Borel σ-algebra, i.e. the σ-field generated by the open sets, in the definition of weak convergence by the σ-field generated by the open balls (the so-called ball σ-field).

Definition 17 (Dudley (1966))**.** Let (M, d) be a metric space. Given probability spaces $(\Omega, \mathcal{F}, P)$, $(\Omega_n, \mathcal{F}_n, P_n)$, $n \in \mathbb{N}$, and maps $X_n : \Omega_n \longrightarrow M$, $n \in \mathbb{N}$, and $X : \Omega \longrightarrow M$, which are measurable with respect to the ball σ-field $\mathcal{M}_b$ on M, X_n is said to *converge in distribution* to X, denoted by $X_n \rightharpoonup X$, if

$$\lim_{n\to\infty} \mathrm{E}\, h(X_n) = \mathrm{E}\, h(X) \quad \text{for all } h \in C_b(M, \mathcal{M}_b),$$

where $C_b(M, \mathcal{M}_b)$ denotes the class of all bounded, continuous real-valued functions defined on M which are $\mathcal{M}_b$-measurable.

In general, the Borel σ-algebra is finer than the ball σ-field. Especially, with $\mathcal{U}_b$ denoting the ball σ-field on D equipped with the uniform metric, $\mathcal{U}_b \subset \mathcal{U}$. Since (D, d_u) is non-separable, the ball σ-field $\mathcal{U}_b$ is a proper subset of the Borel σ-algebra $\mathcal{U}$ in this case. However, on separable metric spaces ball and Borel σ-algebras coincide. Specifically, $\mathcal{S}_b = \mathcal{S}$, where $\mathcal{S}_b$ denotes the σ-field generated by the d_s-open balls.

Moreover, it can be shown that the σ-field on D generated by the d_u-open balls equals the σ-field generated by the d_s-open sets, i.e. all in all, the following relations hold:

$$\mathcal{S}_b = \mathcal{S} = \mathcal{U}_b \subset \mathcal{U}.$$

In particular, Dudley's concept of convergence in distribution of random variables with values in D corresponds to the classical definition of weak convergence if the Skorohod metric is considered.
If the limit variable X in Definition 17 is concentrated on a separable subset, it can be shown that Dudley's notion of weak convergence and the (in this case more general) concept of weak convergence in the sense of Hoffmann – Jørgensen are equivalent.

Theorem 14 (Van Der Vaart and Wellner (1989))**.** *Let (M, d) be a metric space. Moreover, let $(\Omega, \mathcal{F}, P)$, $(\Omega_n, \mathcal{F}_n, P_n)$, $n \in \mathbb{N}$, be probability spaces and $X_n : \Omega_n \longrightarrow M$, $n \in \mathbb{N}$, and $X : \Omega \longrightarrow M$ maps which are measurable with respect to the ball σ-field $\mathcal{M}_b$ on M. If $P(X \in M_0) = 1$ for some separable Borel set $M_0 \subset M$, then $X_n \rightharpoonup X$ if and only if $X_n \rightsquigarrow X$.*

Due to the fact that within this thesis all random variables with values in some Skorohod space $D(K)$ are at least Borel measurable with respect to the Skorohod topology and since the corresponding limits are always concentrated on a separable subset of $D(K)$, namely on $C(K)$, the difficulties that arise with the definition of weak convergence of measures on $D(K)$ can be avoided by resorting to Dudley's or Hoffmann –Jørgensen's concept of weak convergence. Because in this case the two definitions are equivalent, we refer to weak convergence of probability measures by $\Rightarrow$ and to convergence in distribution of random variables by $\stackrel{\mathcal{D}}{\longrightarrow}$ without specifying the metric on $D(K)$ or the convergence concept.
A useful technique for proving weak convergence of stochastic processes with values in function spaces such as $C[0, 1]$ (equipped with the uniform metric) and $D[0, 1]$ (equipped with the Skorohod metric), consists in establishing tightness and weak convergence of the finite-dimensional distributions; see Billingsley (1968). In fact, this result can be easily generalized to processes with values in $D(E)$, where $E := [0, 1] \times [0, 1]$ denotes the two-dimensional unit cube in $\mathbb{R}^2$. For this purpose, we define the modulus of continuity of an element x of $D(E)$ by

$$\omega_x(\delta) := \sup_{\|s-t\|<\delta} |x(s) - x(t)|,$$

where $\|\cdot\|$ denotes the Euclidean norm, and the natural projections from $D(E)$ to $\mathbb{R}^m$ by

$$\pi_{t_1,\ldots,t_m}(x) = (x(t_1), \ldots, x(t_m))$$

for $t_j \in E$, $j = 1, \ldots, m$, and $m \in \mathbb{N}$.
Analogous to the corresponding results for stochastic processes with one-dimensional time parameter, the following theorem establishes tightness and convergence of the finite-dimensional distributions as sufficient conditions for proving weak convergence.

Theorem 15 (Neuhaus (1971))**.** *Let* P_n*,* $n \in \mathbb{N}$*, be a sequence of probability measures on* $(D(E), \mathcal{B})$*, where* $\mathcal{B}$ *denotes the Borel* σ*-algebra on* $(D(E), d_s)$*. Suppose that*

$$P_n \circ \pi^{-1}_{t_1,\dots,t_m} \xrightarrow{\mathcal{D}} P \circ \pi^{-1}_{t_1,\dots,t_m}$$

for some probability measure P *on* $(D(E), \mathcal{B})$ *with* $P(C(E)) = 1$ *and for all* $t_1, \dots t_m \in E$*,* $m \in \mathbb{N}$*. Suppose further that*

$$\lim_{\delta \to 0} \limsup_{n \to \infty} P_n\left(\{x \in D(E) : \omega_x(\delta) \geq \varepsilon\}\right) = 0$$

for all $\varepsilon > 0$*. Then* $P_n \Rightarrow P$*.*

B. Modified Wilcoxon-type change-point tests

The Wilcoxon-based statistics considered in this thesis are designed for the identification of a single change-point in time series with continuous marginal distribution. However, as seen in Chapter 5, in practice it may be necessary to allow for data with ties or multiple change-points. Corresponding modifications of the Wilcoxon statistics, originally proposed in Betken and Wendler (2015), are introduced in the following sections.

B.1. Wilcoxon-based change-point tests for data with ties

Given time series data $X_1, \ldots, X_n$, generated by random variables with non-continuous marginal distribution, there is a positive probability that $X_i = X_j$ for $i \neq j$, i.e. there may be ties in the data.
By replacing the ranks R_i, $i = 1, \ldots, n$, in the Wilcoxon statistics by the modified ranks

$$R_i^{\star} = \sum_{j=1}^{n} \left(1_{\{X_j < X_i\}} + \frac{1}{2} 1_{\{X_j = X_i\}} \right), \ i = 1, \ldots, n,$$

change-point tests taking the possibility of ties into consideration are obtained. More precisely, we arrive at the following definitions of test statistics:

$$W_n^{\star} := \max_{1 \leq k \leq n-1} \left| \sum_{i=1}^{k} R_i^{\star} - \frac{k}{n} \sum_{i=1}^{n} R_i^{\star} \right|,$$

$$SW_n^{\star}(\tau_1, \tau_2) := \max_{\{\lfloor n\tau_1 \rfloor \leq k \leq \lfloor n\tau_2 \rfloor\}} \frac{\left| \sum_{i=1}^{k} R_i^{\star} - \frac{k}{n} \sum_{i=1}^{n} R_i^{\star} \right|}{\left\{ \frac{1}{n} \sum_{t=1}^{k} S_t^2(1, k) + \frac{1}{n} \sum_{t=k+1}^{n} S_t^2(k+1, n) \right\}^{1/2}}, \tag{B.1}$$

where

$$S_t(j, k) = \sum_{i=j}^{t} \left(R_i^{\star} - \frac{1}{k - j + 1} \sum_{i=j}^{k} R_i^{\star} \right).$$

Suppose the considered data corresponds to a mean-zero subordinated Gaussian time series $X_n = G(\xi_n)$, $n \in \mathbb{N}$, according to Model 1 introduced in Section 1.1.2, where ξ_n, $n \in \mathbb{N}$, is a long-range dependent Gaussian sequence with LRD parameter D.

Under the additional assumption that G is piecewise monotone on finitely many pieces, the asymptotic distribution of these statistics can be derived in the same way as the limits of the Wilcoxon and self-normalized Wilcoxon statistics considered in Sections 1.2.2 and 1.2.3. However, instead of considering the Wilcoxon process defined by (1.7) in Section 1.2.2, we consider the modified Wilcoxon process

$$W_n^\star(t) := \sum_{i=1}^{[nt]} \sum_{j=[nt]+1}^{n} h(\xi_i, \xi_j), \ t \in [0,1]\,,$$

with $h(x,y) = h_G(x,y) := 1_{\{G(x)<G(y)\}} + \frac{1}{2}1_{\{G(x)=G(y)\}} - \frac{1}{2}$. Basic transformations yield

$$W_n^\star = \sup_{t\in[0,1]} |W_n^\star(t)|\,.$$

Following the proof of Theorem 1 in Betken (2016), the self-normalized Wilcoxon statistic can be approximated by a function of $W_n^\star(t)$, $t \in [0,1]$, as follows:

$$SW_n^\star(\tau_1,\tau_2) = \sup_{t\in[\tau_1,\tau_2]} \left|G_{W_n^\star}(t)\right| + \mathcal{O}_P(1),$$

where for $f \in D\,[0,1]$ the function $G_f \in D\,[0,1]$ is defined by

$$G_f(t) := \frac{f(t)}{V_f(t)}, \ V_f(t) := \left\{\int_0^t \left(f(s) - \frac{s}{t}f(t)\right)^2 ds + \int_t^1 \left(f(s) - \frac{1-s}{1-t}f(t)\right)^2 ds\right\}^{\frac{1}{2}}.$$

Let $d_n := d_{n,1}$ denote the normalizing sequence defined by (1.5) in Section 1.2.2. Then, according to Theorem 2.2 in Dehling et al. (2017b), the standardized process

$$\frac{1}{nd_n} W_n^\star(t), \ t \in [0,1]\,,$$

converges in distribution to the limit process $W^\star(t)$, $t \in [0,1]$, defined by

$$\begin{aligned} W^\star(t) :=& -(1-t)B_H(t)\int_{\mathbb{R}} \varphi(x)\,dh_1(x) \\ &- t(B_H(1) - B_H(t))\int_{\mathbb{R}}\left(\int_{\mathbb{R}} \varphi(y)dh(x,y)(y)\right)\varphi(x)\,dx, \end{aligned}$$

where B_H denotes a fractional Brownian motion, φ the density function of the standard normal distribution and $h_1(x) := E\left(h(x,\xi_i)\right)$. As a result,

$$\frac{1}{nd_n} W_n^\star \xrightarrow{\mathcal{D}} \sup_{t\in[\tau_1,\tau_2]} |W^\star(t)|\,,$$

$$SW_n^\star(\tau_1,\tau_2) \xrightarrow{\mathcal{D}} \sup_{t\in[\tau_1,\tau_2]} |G_{W^\star}(t)|\,.$$

Since the limit distributions depend on the transformation G, critical values can only be computed under additional assumptions on the data-generating process. Nevertheless, convergence in distribution to non-degenerate limits justifies an application of the sampling-window method considered in Chapter 3.

B.2. Wilcoxon-based change-point tests for multiple breakpoints

An extension of the Wilcoxon-based tests to testing procedures that allow for two or more change-points can be based on a modification of the test statistics according to an approach proposed in Shao (2011) for the self-normalized CUSUM statistic. To illustrate the general idea that underlies the construction of the modified test statistics, we focus on alternative hypotheses that assume two level-shifts. By dividing the observations $X_1, \ldots, X_n$ according to a pair (k_1, k_2), $1 \leq k_1 < k_2 \leq n$, of potential change-point locations, the modified test statistics stem from an application of the original test statistic to the subsamples $X_1, \ldots, X_{k_2}$ and $X_{k_1+1}, \ldots, X_n$.
The rank of an observation X_i with respect to the subsample $X_{j+1}, \ldots, X_l$ is given by

$$R_{i;j,l} := \sum_{k=j+1}^{l} 1_{\{X_k \leq X_i\}}.$$

Evaluations of the two-sample Wilcoxon statistics with respect to this subsample lead to the following definitions:

$$W_{k;j,l} := \sum_{i=j+1}^{k} R_{i;j,l} - \frac{k-j}{l-j} \sum_{i=j+1}^{l} R_{i;j,l},$$

$$SW_{k;j,l} := \frac{W_{k;j,l}}{\left\{ \frac{1}{n} \sum_{t=j+1}^{k} S_{t;j,l}^2(j+1,k) + \frac{1}{n} \sum_{t=k+1}^{l} S_{t;j,l}^2(k+1,l) \right\}^{1/2}},$$

where

$$S_{t;j,l}(k,m) := \sum_{i=k}^{t} \left(R_{i;j,l} - \bar{R}_{k,m} \right) \quad \text{and} \quad \bar{R}_{k,m} := \frac{1}{m-k+1} \sum_{i=k}^{m} R_{i;j,l}.$$

Under the assumption that in the presence of two breakpoints the change-point locations are unknown and separated by at least $\lfloor n\varepsilon \rfloor$ observations, the set of potential change-point locations is defined by

$$\Omega_n(\tau_1, \tau_2, \varepsilon) := \{(k_1, k_2) \mid \lfloor n\tau_1 \rfloor \leq k_1 < k_2 \leq \lfloor n\tau_2 \rfloor, \ k_2 - k_1 \geq \lfloor n\varepsilon \rfloor\}$$

for $0 < \tau_1 < \tau_2 < 1$ and $\varepsilon \in (0, \tau_2 - \tau_1)$.
As a result, the Wilcoxon statistic $W_n(\varepsilon)$ and the self-normalized Wilcoxon statistic $SW_n(\varepsilon)$ are defined by

$$W_n(\varepsilon) := \max_{(k_1,k_2) \in \Omega_n(0,1,\varepsilon)} \left\{ |W_{k_1;0,k_2}| + |W_{k_2;k_1,n}| \right\},$$

and

$$SW_n(\tau_1, \tau_2, \varepsilon) := \max_{(k_1,k_2)\in\Omega_n(\tau_1,\tau_2,\varepsilon)} \{|SW_{k_1;0,k_2}| + |SW_{k_2;k_1,n}|\}. \tag{B.2}$$

Basic transformations yield

$$W_{\lfloor ns\rfloor,\lfloor nt\rfloor,\lfloor nu\rfloor} = W_n(s,t,u) := W_n(t,t) - W_n(s,t) - W_n(t,u),$$

and

$$SW_{\lfloor ns\rfloor,\lfloor nt\rfloor,\lfloor nu\rfloor} = \frac{W_n(s,t,u)}{V_{W_n}(s,t,u)} + \mathcal{O}_P(1),$$

where

$$W_n(s,t) := \sum_{i=1}^{\lfloor ns\rfloor} \sum_{j=\lfloor nt\rfloor+1}^{n} \left(1_{\{X_i\leq X_j\}} - \frac{1}{2}\right)$$

and

$$V_f(s,t,u)$$
$$:= \left\{\int_s^t \left(f(s,v,u) - \frac{v-s}{u-s} f(s,t,u)\right)^2 dv + \int_t^u \left(f(s,v,u) - \frac{u-v}{u-t} f(s,t,u)\right)^2 dv\right\}^{\frac{1}{2}}.$$

According to the above representations, both test statistics, $W_n(\varepsilon)$ and $SW_n(\tau_1, \tau_2, \varepsilon)$, can be considered as functions of the process $W_n(s,t)$, $0 \leq s \leq t \leq 1$.

Given a mean-zero subordinated Gaussian time series $Y_n = G(\xi_n)$, $n \in \mathbb{N}$, according to Model 1 introduced in Section 1.1.2, where ξ_n, $n \in \mathbb{N}$, is a long-range dependent Gaussian sequence with LRD parameter D, the standardized process

$$\frac{1}{nd_{n,r}} W_n(s,t), \ 0 \leq s \leq t \leq 1,$$

with r denoting the Hermite rank of the class of functions $1_{\{G(\xi_1)\leq x\}} - F(x)$, $x \in \mathbb{R}$, and with $d_{n,r}$ denoting the corresponding normalizing sequence defined by (1.5) in Section 1.2.2, converges in distribution to

$$\{(1-t)Z_{r,H}(s) - s(Z_{r,H}(1) - Z_{r,H}(t))\} \frac{1}{r!} \int_{\mathbb{R}} J_r(x)\, dF(x), \ 0 \leq s \leq t \leq 1,$$

where $Z_{r,H}$ is an r-th order Hermite process with Hurst parameter $H = 1 - \frac{rD}{2}$ and where $J_r(x) = \mathrm{E}\left(H_r(\xi_1) 1_{\{G(\xi_1)\leq x\}}\right)$; see Lemma 2 in Chapter 2.
It follows by applications of the continuous mapping theorem that the statistics

$$\frac{1}{nd_{n,r}} W_n(\varepsilon) \ \text{ and } \ SW_n(\tau_1, \tau_2, \varepsilon)$$

converge in distribution to $W(\varepsilon)$ and $SW(\tau_1, \tau_2, \varepsilon)$ defined by

$$W(\varepsilon) := \sup_{(t_1,t_2)\in\Omega(0,1,\varepsilon)} \{|W(t_1; 0, t_2)| + |W(t_2; t_1, 1)|\} \frac{1}{r!} \int_{\mathbb{R}} J_r(x)\, dF(x)$$

$$SW(\tau_1, \tau_2, \varepsilon) := \sup_{(t_1,t_2)\in\Omega(\tau_1,\tau_2,\varepsilon)} \{|SW(t_1; 0, t_2)| + |SW(t_2; t_1, 1)|\}$$

with

$$W(t; s, u) := Z_{r,H}(t) - \frac{t-s}{u-s} Z_{r,H}(u) - \frac{u-t}{u-s} Z_{r,H}(s),$$

$$SW(t; s, u) := \frac{W(t; s, u)}{\left\{ \int_s^t W^2(u; s, t)\, du + \int_t^u W^2(u; t, 1)\, du \right\}^{\frac{1}{2}}},$$

and

$$\Omega(\tau_1, \tau_2, \varepsilon) := \{(t_1, t_2) \mid \tau_1 \le t_1 < t_2 \le \tau_2,\ t_2 - t_1 \ge \varepsilon\}.$$

The critical values corresponding to the asymptotic distributions of the test statistics are reported in Table B.1.

	$W(\varepsilon)$			$W(\tau_1, \tau_2, \varepsilon)$		
H	10%	5%	1%	10%	5%	1%
0.501	0.6198999	0.6704254	0.7713137	17.79236	19.76166	24.12842
0.6	0.4716089	0.5109281	0.5972226	19.79540	22.38011	27.67941
0.7	0.3513135	0.3822559	0.4461184	22.07942	24.94961	30.46419
0.8	0.2477967	0.2718783	0.3205061	24.23847	27.61217	34.03470
0.9	0.1552892	0.1713176	0.2024479	26.49583	30.11468	37.77919
0.999	0.0138900	0.0153919	0.0182338	28.27782	32.32295	41.23974

Table B.1.: *Simulated critical values for the distributions of* $W(\varepsilon)$ *and* $SW(\tau_1, \tau_2, \varepsilon)$ *when* $r = 1$, $[\tau_1, \tau_2] = [0.15, 0.85]$, *and* $\varepsilon = 0.15$. *The sample size is* 1000, *the number of replications is* 10,000.

List of Symbols

P	probability measure
$\mathrm{E}\,X$	expected value of a random variable X
$\mathrm{Var}\,X$	variance of a random variable X
$\mathrm{Cov}(X,Y)$	covariance of two random variables X and Y
$\mathrm{Corr}(X,Y)$	correlation of two random variables X and Y
$\sigma\,(X_1,\ldots,X_n)$	σ-field generated by random variables $X_1,\ldots,X_n$
$\mathrm{E}\,(X \mid \mathcal{F})$	conditional expectation of a random variable X given a σ-field $\mathcal{F}$
1_A	indicator function of a set A
$\lfloor x \rfloor$	(floor) integer part of a real number x
$\lceil x \rceil$	(ceiling) integer part of a real number x
$L^2(\Omega,\mathcal{F},P)$	the space of all $\mathcal{F}$-measurable, real-valued functions on Ω which are square-integrable with respect to the measure P
M'	transpose of a matrix M
o	little-o notation
$\mathcal{O}$	big-O notation
$\mathcal{O}_P$	stochastic boundedness
$\xrightarrow{a.s.}$	almost sure convergence
$\xrightarrow{P}$	convergence in probability
$\xrightarrow{\mathcal{D}}$	convergence in distribution
$\overset{\mathcal{D}}{=}$	equality in distribution
$f \sim g$	asymptotic equivalence of two functions f and g
$f \lesssim g$	existence of a function h with $f \leq h$ and $h \sim g$
α	tail parameter
D	long-range dependence parameter
H	Hurst parameter
r	Hermite rank
B	Brownian motion
B_H	fractional Brownian motion with Hurst parameter H
$Z_{r,H}$	r-th order Hermite process with self-similarity parameter H

Bibliography

Rolf K. Adenstedt. On Large-Sample Estimation for the Mean of a Stationary Random Sequence. *The Annals of Statistics*, 2:1095 – 1107, 1974.

Donald W. K. Andrews. Tests for Parameter Instability and Structural Change With Unknown Change Point. *Econometrica*, 61:821 – 856, 1993.

Miguel A. Arcones. On the law of the iterated logarithm for Gaussian processes. *Journal of Theoretical Probability*, 8:877 – 903, 1995.

Jushan Bai. Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis*, 15:453 – 472, 1994.

Jushan Bai. Estimation of a Change Point in Multiple Regression Models. *Review of Economics and Statistics*, 79:551 – 563, 1997.

Shuyang Bai and Murad S. Taqqu. On the validity of resampling methods under long memory. *arXiv:1512.00819*, 2015.

Shuyang Bai, Murad S. Taqqu, and Ting Zhang. A unified approach to self-normalized block sampling. *Stochastic Processes and their Applications*, 126:2465 – 2493, 2016.

Nathan S. Balke. Detecting Level Shifts in Time Series. *Journal of Business & Economic Statistics*, 11:81 – 92, 1993.

Samir Ben Hariz and Jonathan J. Wylie. Convergence rates for estimating a change-point with long-range dependent sequences. *Comptes Rendus de l'Académie des Sciences Paris - Series I*, 341:765 – 768, 2005.

Jan Beran. *Statistics for Long-Memory Processes.* Chapman & Hall, 1994.

Jan Beran and Yuanhua Feng. SEMIFAR Models - A Semiparametric Framework for Modelling Trends, Long Range Dependence and Nonstationarity. *Computational Statistics & Data Analysis*, 40:393 – 419, 2002.

Jan Beran, Yuanhua Feng, Sucharita Ghosh, and Rafał Kulik. *Long-Memory Processes.* Springer-Verlag Berlin Heidelberg, 2013.

István Berkes and Walter Philipp. An Almost Sure Invariance Principle for the Empirical Distribution Function of Mixing Random Variables. *Probability Theory and Related Fields*, 41:115 – 137, 1977.

István Berkes, Siegfried Hörmann, and Johannes Schauer. Asymptotic results for the empirical process of stationary sequences. *Stochastic Processes and their Applications*, 119:1298 – 1324, 2009.

Annika Betken. Strukturbruchtests mit Selbstnormierung. Master's thesis, Ruhr-Universität Bochum, 2013.

Annika Betken. Testing for Change-Points in Long-Range Dependent Time Series by Means of a Self-Normalized Wilcoxon Test. *Journal of Time Series Analysis*, 37:785 – 809, 2016.

Annika Betken. Change point estimation based on Wilcoxon tests in the presence of long-range dependence. *Electronic Journal of Statistics*, 11:3633 – 3672, 2017.

Annika Betken and Rafał Kulik. Testing for Change in Stochastic Volatility with Long Range Dependence. *arXiv:1706.06351*, 2017.

Annika Betken and Martin Wendler. Subsampling for General Statistics under Long Range Dependence. *Statistica Sinica*, 2015. doi: 10.5705/ss.202015.0435.

Rajendra J. Bhansali and Piota S. Kokoszka. Estimation of the Long-Memory Parameter: A Review of Recent Developments and an Extension. *Lecture Notes-Monograph Series*, 37:125 – 150, 2001.

Patrick Billingsley. *Convergence of Probability Measures.* John Wiley & Sons, Inc., 1968.

Patrick Billingsley. *Probability and Measure.* John Wiley & Sons, Inc., 1995.

Patrick Billingsley. *Convergence of Probability Measures.* John Wiley & Sons, Inc., 1999.

Nicholas H. Bingham, Charles M. Goldie, and Jozef L. Teugels. *Regular Variation.* Cambridge University Press, 1987.

Richard C. Bradley. *Introduction to Strong Mixing Conditions.* Kendrick press, 2007.

F. Jay Breidt, Nuno Crato, and Pedro De Lima. The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics*, 83:325 – 348, 1998.

Leo Breiman. On Some Limit Theorems Similar to the Arc-Sin Law. *Theory of Probability & Its Applications*, 10, 1965.

Peter J. Brockwell and Richard A. Davis. *Introduction to Time Series and Forecasting.* Springer-Verlag New York, Inc., 2002.

Peter J. Brockwell and Richard A. Davis. *Time Series: Theory and Methods.* Springer Science & Business Media, 2013.

George W. Cobb. The Problem of the Nile: Conditional Solution to a Changepoint Problem. *Biometrika*, 65:243 – 251, 1978.

Martial Coulon, Marie Chabert, and Ananthram Swami. Detection of Multiple Changes in Fractional Integrated ARMA Processes. *IEEE Transactions on Signal Processing*, 57:48 – 61, 2009.

B. S. Darkhovskh. A Nonparametric Method for the a Posteriori Detection of the "Disorder" Time of a Sequence of Independent Random Variables. *Theory of Probability & Its Applications*, 21:178 – 183, 1976.

Michiel De Pooter and Dick Van Dijk. Testing for changes in volatility in heteroskedastic time series – a further examination. Technical report, 2004.

Herold Dehling and Murad S. Taqqu. The Empirical Process of some Long-Range Dependent Sequences with an Application to U -Statistics. *The Annals of Statistics*, 17: 1767 – 1783, 1989.

Herold Dehling, Aeneas Rooch, and Murad S. Taqqu. Non-Parametric Change-Point Tests for Long-Range Dependent Data. *Scandinavian Journal of Statistics*, 40:153 – 173, 2013.

Herold Dehling, Olivier Durieu, and Marco Tusche. Approximating class approach for empirical processes of dependent sequences indexed by functions. *Bernoulli*, 20:1372 – 1403, 2014.

Herold Dehling, Aeneas Rooch, and Murad S. Taqqu. Power of change-point tests for long-range dependent data. *Electronic Journal of Statistics*, 11:2168 – 2198, 2017a.

Herold Dehling, Aeneas Rooch, and Martin Wendler. Two-sample U-statistic processes for long- range dependent data. *Statistics*, 51:84 – 104, 2017b.

Rohit Deo, Mengchen Hsieh, Clifford M. Hurvich, and Philippe Soulier. Long Memory in Nonlinear Processes. In *Dependence in probability and statistics*, volume 187, pages 221 – 244. Springer, 2006.

Rohit S. Deo and Clifford M. Hurvich. Linear Trend with Fractionally Integrated Errors. *Journal of Time Series Analysis*, 19:379 – 397, 1998.

Richard M. Dudley. Weak convergence of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces. *Illinois Journal of Mathematics*, 10:109 – 126, 1966.

Richard M. Dudley. Measures on non-separable metric spaces. *Illinois Journal of Mathematics*, 11:449 – 453, 1967.

Aryeh Dvoretzky. Asymptotic normality for sums of dependent random variables. volume 2, pages 513 – 535. The Regents of the University of California, University of California Press, 1972.

Zhenwen Fan. *Statistical issues and developments in time series analysis and educational measurement*. BiblioBazaar, 2012.

Dietmar Ferger. On the uniqueness of maximizers of Markov–Gaussian processes. *Statistics & Probability Letters*, 45:71 – 77, 1999.

Liudas Giraitis and Donatas Surgailis. Central limit theorem for the empirical process of a linear sequence with long memory. *Journal of Statistical Planning and Inference*, 80:81 – 93, 1999.

Liudas Giraitis, Remigijus Leipus, and Donatas Surgailis. The change-point problem for dependent observations. *Journal of Statistical Planning and Inference*, 53:297 – 310, 1996.

Peter Hall and Bingyi Jing. On Sample Reuse Methods for Dependent Data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 58:727 – 737, 1996.

Peter Hall, Bing-Yi Jing, and Soumendra N. Lahiri. On the sampling window method for long-range dependent data. *Statistica Sinica*, 8:1189 – 1204, 1998.

Andrew C. Harvey. Long memory in stochastic volatility. In *Forecasting Volatility in the Financial Markets*, pages 307 – 320. Butterwoth-Heinemann Finance, 2002.

Jørgen Hoffmann-Jørgensen. *Stochastic processes on Polish spaces*. Aarhus Universitet. Matematisk Institut, 1991.

Lajos Horváth and Piotr Kokoszka. The effect of long-range dependence on change-point estimators. *Journal of Statistical Planning and Inference*, 64:57 – 81, 1997.

Harold E. Hurst. The problem of long-term storage in reservoirs. *Hydrological Sciences Journal*, 1:13 – 27, 1956.

Clifford M. Hurvich and Philippe Soulier. *Stochastic Volatility Models with Long Memory*, pages 345 – 354. Springer, 2009.

Ildar A. Ibragimov and Yurii A. Rozanov. *Gaussian random processes*. Springer, New York, 1978.

Gail Ivanoff. Stopping times and tightness in two dimensions. Technical report, University of Ottawa, 1983.

Agnieszka Jach, Tucker McElroy, and Dimitris N. Politis. Subsampling inference for the mean of heavy-tailed long-memory time series. *Journal of Time Series Analysis*, 33: 96 – 111, 2012.

Jack Kiefer. Skorohod embedding of multivariate RV's, and the sample DF. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 24:1 – 35, 1972.

Nicholas M. Kiefer and Timothy J. Vogelsang. Heteroskedasticity-Autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation. *Econometrica*, 70:2093 – 2095, 2002.

Piotr Kokoszka and Remigijus Leipus. Change-point in the mean of dependent observations. *Statistics & Probability Letters*, 40:385 – 393, 1998.

Andrei N. Kolmogorov and Yurii A. Rozanov. On Strong Mixing Conditions for Stationary Gaussian Processes. *Theory of Probability & Its Applications*, 5:204 – 208, 1960.

Chung-Ming Kuan and Chih-Chiang Hsu. Change-Point Estimation of Fractionally Integrated Processes. *Journal of Time Series Analysis*, 19:693 – 708, 1998.

Rafał Kulik and Philippe Soulier. The tail empirical process for long memory stochastic volatility sequences. *Stochastic Processes and their Applications*, 121:109 – 134, 2011.

Hans R. Künsch. Statistical aspects of self-similar processes. In *Proceedings of the first World Congress of the Bernoulli Society*, volume 1, pages 67 – 74. VNU Science Press Utrecht, The Netherlands, 1987.

Marc Lavielle and Eric Moulines. Least-squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis*, 21:33 – 59, 2000.

Will E. Leland and Daniel V. Wilson. High time-resolution measurement and analysis of LAN traffic: Implications for LAN interconnection. In *INFOCOM'91. Proceedings. Tenth Annual Joint Conference of the IEEE Computer and Communications Societies. Networking in the 90s.*, pages 1360 – 1366. IEEE, 1991.

Will E. Leland, Murad S. Taqqu, Walter Willinger, and Daniel V. Wilson. On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, 2:1 – 15, 1994.

Mikhail A. Lifshits. On the Absolute Continuity of Distributions of Functionals of Random Processes. *Teoriya Veroyatnostei i ee Primeneniya*, 27:559 – 566, 1982.

Ignacio N. Lobato. Testing That a Dependent Process Is Uncorrelated. *Journal of the American Statistical Association*, 96:1066 – 1076, 2001.

I. B. MacNeill, S. M. Tang, and V. K. Jandhyala. A Search for the Source of the Nile's Change-Points. *Environmetrics*, 2:341 – 375, 1991.

Tucker McElroy and Dimitris Politis. Self-Normalization for Heavy-Tailed Time Series with Long Memory. *Statistica Sinica*, 17:199 – 220, 2007.

Thomas Mikosch. *ELEMENTARY STOCHASTIC CALCULUS with Finance in View.* World Scientific Publishing Co. Pte. Ltd., 1998.

Dietrich W. Müller. On Glivenko-Cantelli Convergence. *Probability Theory and Related Fields*, 16:195 – 210, 1970.

Georg Neuhaus. On Weak Convergence of Stochastic Processes with Multidimensional Time Parameter. *The Annals of Mathematical Statistics*, 42:1285 – 1295, 1971.

Daniel J. Nordman and Soumendra N. Lahiri. Validity of the Sampling Window Method for Long-Range Dependent Linear Processes. *Econometric Theory*, 21:1087 – 1111, 2005.

Walter Philipp and Laurence Pinzur. Almost Sure Approximation Theorems for the Multivariate Empirical Process. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 54:1 – 13, 1980.

Vladas Pipiras and Murad S. Taqqu. *Long-Range Dependence and Self-Similarity*, volume 45. Cambridge University Press, 2017.

Dimitris N. Politis and Joseph P. Romano. Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions. *The Annals of Statistics*, 22:2031 – 2050, 1994.

Peter M. Robinson. Gaussian Semiparametric Estimation of Long Range Dependence. *The Annals of Statistics*, 23:1630 – 1661, 1995.

Emilio Seijo and Bodhisattva Sen. A continuous mapping theorem for the smallest argmax functional. *Electronic Journal of Statistics*, 5:421 – 439, 2011.

Xiaofeng Shao. A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72: 343 – 366, 2010.

Xiaofeng Shao. A simple test of changes in mean in the possible presence of long-range dependence. *Journal of Time Series Analysis*, 32:598 – 606, 2011.

Xiaofeng Shao and Xianyang Zhang. Testing for change points in time series. *Journal of the American Statistical Association*, 105:1228 – 1240, 2010.

Michael Sherman and Edward Carlstein. Replicate Histograms. *Journal of the American Statistical Association*, 91:566 – 576, 1996.

Galen R. Shorack and Jon A. Wellner. *Empirical Processes with Applications to Statistics*. John Wiley & Sons, New York, 1986.

Yu G. Sinai. Self-Similar Probability Distributions. *Theory of Probability & Its Applications*, 21:64 – 80, 1976.

Murad S. Taqqu. Weak Convergence to Fractional Brownian Motion and to the Rosenblatt Process. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 31: 287 – 302, 1975.

Murad S. Taqqu. Convergence of Integrated Processes of Arbitrary Hermite Rank. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 50:53 – 83, 1979.

Murad S. Taqqu. Fractional Brownian Motion and Long-Range. Dependence. In *Theory and Applications of Long-Range Dependence*. Birkhäuser Boston Inc., 2003.

Murad S. Taqqu and Vadim Teverovsky. Robustness of Whittle-type Estimators for Time Series with Long-Range Dependence. *Communications in Statistics. Stochastic Models*, 13:723 – 757, 1997.

Stephen J. Taylor. *Modelling Financial Time Series.* Wiley, New York, 1986.

Omar Toussoun. *Mémoire sur l'histoire du Nil.* Memoires de l'Institut d'Egypte. l'Institut français d'archéologie orientale, 1925.

Aad W. Van Der Vaart and Jon A. Wellner. Convolution and Asymptotic Minimax Theorems for Nonmeasurable "Estimators". 1989.

Aad W. Van Der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes.* Springer, 1996.

Lihong Wang. Gradual changes in long memory processes with applications. *Statistics*, 41:221 – 240, 2007.

Wei Biao Wu and Zhibiao Zhao. Inference of trends in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69:391 – 410, 2007.

Ting Zhang, Hwai-Chung Ho, Martin Wendler, and Wei Biao Wu. Block sampling under strong dependence. *Stochastic Processes and their Applications*, 123:2323 – 2339, 2013.