Electronic Journal of Statistics
Vol. 11 (2017) 3633–3672
ISSN: 1935-7524
DOI: 10.1214/17-EJS1323

Change point estimation based on Wilcoxon tests in the presence of long-range dependence

Annika Betken*

Faculty of Mathematics
Ruhr-Universität Bochum
44780 Bochum, Germany
e-mail: annika.betken@rub.de

Abstract: We consider an estimator for the location of a shift in the mean of long-range dependent sequences. The estimation is based on the two-sample Wilcoxon statistic. Consistency and the rate of convergence for the estimated change point are established. In the case of a constant shift height, the $1/n$ convergence rate (with n denoting the number of observations), which is typical under the assumption of independent observations, is also achieved for long memory sequences. It is proved that if the change point height decreases to 0 with a certain rate, the suitably standardized estimator converges in distribution to a functional of a fractional Brownian motion. The estimator is tested on two well-known data sets. Finite sample behaviors are investigated in a Monte Carlo simulation study.

MSC 2010 subject classifications: Primary 62G05, 62M10; secondary 60G15, 60G22.
Keywords and phrases: Change point estimation, long-range dependence, Wilcoxon test, self-normalization.

Received January 2017.

Contents

1. Introduction

Suppose that the observations $X_1, \dots, X_n$ are generated by a stochastic process $(X_i)_{i\geq 1}$

*Research supported by the German National Academic Foundation and Collaborative Research Center SFB 823 *Statistical modelling of nonlinear dynamic processes*.

$$X_i = \mu_i + Y_i,$$

where $(\mu_i)_{i\geq 1}$ are unknown constants and where $(Y_i)_{i\geq 1}$ is a stationary, long-range dependent (LRD, in short) process with mean zero. A stationary process $(Y_i)_{i\geq 1}$ is called "long-range dependent" if its autocovariance function ρ, $\rho(k) := \operatorname{Cov}(Y_1, Y_{k+1})$, satisfies

$$\rho(k) \sim k^{-D}L(k), \text{ as } k \to \infty, \tag{1}$$

where $0 < D < 1$ (referred to as long-range dependence (LRD) parameter) and where L is a slowly varying function.

Furthermore, we assume that there is a change point in the mean of the observations, that is

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

where k_0 denotes the change point location and h_n is the height of the level-shift. Throughout the paper, we assume that $k_0 = \lfloor n\tau \rfloor$ with $0 < \tau < 1$ and with $\lfloor x \rfloor$ denoting the greatest integer less than or equal to x for any $x \in \mathbb{R}$.

In the following we differentiate between fixed and local changes. Under fixed changes we assume that $h_n = h$ for some $h \neq 0$. Local changes are characterized by a sequence h_n, $n \in \mathbb{N}$, with $h_n \longrightarrow 0$ as $n \longrightarrow \infty$; in other words, in a model where the height of the jump decreases with increasing sample size n.

In order to test the hypothesis

$$H : \mu_1 = \ldots = \mu_n$$

against the alternative

$$A : \mu_1 = \ldots = \mu_k \neq \mu_{k+1} = \ldots = \mu_n \text{ for some } k \in \{1, \ldots, n-1\}$$

the Wilcoxon change point test can be applied. It rejects the hypothesis for large values of the Wilcoxon test statistic defined by

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

(see Dehling, Rooch and Taqqu (2013a)). Under the assumption that there is a change point in the mean in k_0 we expect the absolute value of $W_{k_0,n}$ to exceed the absolute value of $W_{l,n}$ for any $l \neq k_0$. Therefore, it seems natural to define an estimator of k_0 by

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

Preceding papers that address the problem of estimating change point locations in dependent observations $X_1, \ldots, X_n$ with a shift in mean often refer to a family of estimators based on the CUSUM change point test statistics $C_n(\gamma) := \max_{1\leq k\leq n-1} |C_{k,n}(\gamma)|$, where

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

with parameter $0 \leq \gamma < 1$. The corresponding change point estimator is defined by

$$\hat{k}_{C,\gamma} = \hat{k}_{C,\gamma}(n) := \min\left\{k : |C_{k,n}(\gamma)| = \max_{1\leq i\leq n-1} |C_{i,n}(\gamma)|\right\}. \tag{2}$$

For long-range dependent Gaussian processes Horváth and Kokoszka (1997) derive the asymptotic distribution of the estimator $\hat{k}_{C,\gamma}$ under the assumption of a decreasing jump height h_n, i.e. under the assumption that h_n approaches 0 as the sample size n increases. Under non-restrictive constraints on the dependence structure of the data-generating process (including long-range dependent time series) Kokoszka and Leipus (1998) prove consistency of $\hat{k}_{C,\gamma}$ under the assumption of fixed as well as decreasing jump heights. Furthermore, they establish the convergence rate of the change point estimator as a function of the intensity of dependence in the data if the jump height is constant. Ben Hariz and Wylie (2005) show that under a similar assumption on the decay of the autocovariances the convergence rate that is achieved in the case of independent observations can be obtained for short- and long-range dependent data, as well. Furthermore, it is shown in their paper that for a decreasing jump height the convergence rate derived by Horváth and Kokoszka (1997) under the assumption of gaussianity can also be established under more general assumptions on the data-generating sequences.

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. 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, Leipus and Surgailis (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. A change point estimator based on a self-normalized CUSUM test statistic has been applied in Shao (2011) to real data sets. Although Shao assumes validity of using the estimator, the article does not cover a formal proof of consistency. Furthermore, it has been noted by Shao and Zhang (2010) that even under the assumption of short-range dependence it seems difficult to obtain the asymptotic distribution of the estimate.

In this paper we shortly address the issue of estimating the change point location on the basis of the self-normalized Wilcoxon test statistic proposed in Betken (2016).

In order to construct the self-normalized Wilcoxon test statistic, we have to consider the ranks R_i, $i = 1, \ldots, n$, of the observations $X_1, \ldots, X_n$. These are defined by $R_i := \operatorname{rank}(X_i) = \sum_{j=1}^{n} 1_{\{X_j \leq X_i\}}$ for $i = 1, \ldots, n$. The self-normalized two-sample test statistic is defined by

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

where

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

The self-normalized Wilcoxon change point test for the test problem (H, A) rejects the hypothesis for large values of $T_n(\tau_1, \tau_2) = \max_{k \in \{\lfloor n\tau_1 \rfloor, \ldots, \lfloor n\tau_2 \rfloor\}} |SW_{k,n}|$, where $0 < \tau_1 < \tau_2 < 1$. Note that the proportion of the data that is included in the calculation of the supremum is restricted by τ_1 and τ_2. A common choice for these parameters is $\tau_1 = 1 - \tau_2 = 0.15$; see Andrews (1993).

A natural change point estimator that results from the self-normalized Wilcoxon test statistic is

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

We will prove consistency of the estimator $\hat{k}_{SW}$ under fixed changes and under local changes whose height converges to 0 with a rate depending on the intensity of dependence in the data. Nonetheless, the main aim of this paper is to characterize the asymptotic behavior of the change point estimator $\hat{k}_W$. In Section 2 we establish consistency of $\hat{k}_W$ and $\hat{k}_{SW}$, derive the optimal convergence rate of $\hat{k}_W$ and finally consider its asymptotic distribution. Applications to two well-known data sets can be found in Section 3. The finite sample properties of the estimators are investigated by simulations in Section 4. Proofs of the theoretical results are given in Section 5.

2. Main results

Recall that for fixed x, $x \in \mathbb{R}$, the Hermite expansion of $1_{\{G(\xi_i)\leq x\}} - F(x)$ is given by

$$1_{\{G(\xi_i)\leq x\}} - F(x) = \sum_{q=1}^{\infty} \frac{J_q(x)}{q!} H_q(\xi_i),$$

where H_q denotes the q-th order Hermite polynomial and where

$$J_q(x) = \mathrm{E}\left(1_{\{G(\xi_i)\leq x\}} H_q(\xi_i)\right).$$

Assumption 1. Let $Y_i = G(\xi_i)$, where $(\xi_i)_{i\geq 1}$ is a stationary, long-range dependent Gaussian process with mean 0, variance 1 and LRD parameter D. We assume that $0 < D < \frac{1}{r}$, where r denotes the Hermite rank of the class of functions $1_{\{G(\xi_i)\leq x\}} - F(x)$, $x \in \mathbb{R}$, defined by

$$r := \min\left\{q \geq 1 : J_q(x) \neq 0 \text{ for some } x \in \mathbb{R}\right\}.$$

Moreover, we assume that $G : \mathbb{R} \longrightarrow \mathbb{R}$ is a measurable function and that $(Y_i)_{i\geq 1}$ has a continuous distribution function F.

Let

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

and define

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

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

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

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

The following result states that $\frac{\hat{k}_W}{n}$ and $\frac{\hat{k}_{SW}}{n}$ are consistent estimators for the change point location under fixed as well as certain local changes.

Proposition 1. *Suppose that Assumption 1 holds. Under fixed changes, $\frac{\hat{k}_W}{n}$ and $\frac{\hat{k}_{SW}}{n}$ are consistent estimators for the change point location. The estimators are also consistent under local changes if $h_n^{-1} = o\left(\frac{n}{d_{n,r}}\right)$ and if F has a bounded density f. In other words, we have*

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

in both situations. Furthermore, it follows that the Wilcoxon test is consistent under these assumptions (in the sense that $\frac{1}{nd_{n,r}} \max_{1\leq k\leq n-1} |W_{k,n}| \xrightarrow{P} \infty$).

The following theorem establishes a convergence rate for the change point estimator $\hat{k}_W$. Note that only under local changes the convergence rate depends on the intensity of dependence in the data.

Theorem 1. *Suppose that Assumption 1 holds and let* $m_n := g^{-}_{D,r}(h_n^{-1})$*. Then, we have*

$$\left|\hat{k}_W - k_0\right| = \mathcal{O}_P(m_n)$$

if either

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

or

- $\lim_{n\to\infty} h_n = 0$ *with* $h_n^{-1} = o\left(\frac{n}{d_{n,r}}\right)$ *and* F *has a bounded density* f*.*

Remark 1.

1. Under fixed changes m_n is constant. As a consequence, $|\hat{k}_W - k_0| = \mathcal{O}_P(1)$. This result corresponds to the convergence rates obtained by Ben Hariz and Wylie (2005) for the CUSUM-test based change point estimator and by Lavielle and Moulines (2000) for the least-squares estimate of the change point location. Surprisingly, in this case the rate of convergence is independent of the intensity of dependence in the data characterized by the value of the LRD parameter D. An explanation for this phenomenon might be the occurrence of two opposing effects: increasing values of the LRD parameter D go along with a slower convergence of the test statistic $W_{k,n}$ (making estimation more difficult), but a more regular behavior of the random component (making estimation easier) (see Ben Hariz and Wylie (2005)).
2. Note that if $h_n^{-1} = o\left(\frac{n}{d_{n,r}}\right)$ and $m_n = g^{-}_{D,r}(h_n^{-1})$, it holds that
   - $m_n \longrightarrow \infty$,
   - $\frac{m_n}{n} \longrightarrow 0$,
   - $\frac{d_{m_n,r}}{m_n} \sim h_n$,

   as $n \longrightarrow \infty$.

Based on the previous results it is possible to derive the asymptotic distribution of the change point estimator $\hat{k}_W$:

Theorem 2. *Let* $(B_H(t))_{t\in\mathbb{R}}$ *be a (standard) fractional Brownian motion process, i.e.* B_H *is a Gaussian process with almost surely continuous sample paths,* $\mathrm{E}\, B_H(t) = 0$ *for all* $t \in \mathbb{R}$ *and* $\mathrm{Cov}(B_H(t)B_H(s)) = \frac{1}{2}\left\{|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right\}$ *for all* $s, t \in \mathbb{R}$ *(see Definition 3.23 in Beran et al. (2013)). Suppose that Assumption 1 holds with* $r = 1$ *and assume that* F *has a bounded density* f*. Let*

$m_n := g_{D,1}^{-}(h_n^{-1})$ and define $h(s;\tau)$ by

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

If $h_n^{-1} = o\left(\frac{n}{d_{n,1}}\right)$, then, for all $M > 0$,

$$\frac{1}{e_n}\left(W^2_{k_0+\lfloor m_n s\rfloor,n} - W^2_{k_0,n}\right), \ -M \leq s \leq M,$$

with $e_n = n^3 h_n d_{m_n,1}$, converges in distribution to

$$2\tau(1-\tau)\int_{\mathbb{R}} f^2(x)dx\left(B_H(s)\int_{\mathbb{R}} J_1(x)dF(x) + h(s;\tau)\right), \ -M \leq s \leq M, \tag{3}$$

in the Skorohod space $D[-M,M]$. Furthermore, it follows that $m_n^{-1}(\hat{k}_W - k_0)$ converges in distribution to

$$\underset{-\infty<s<\infty}{\operatorname{argmax}}\left(B_H(s)\int_{\mathbb{R}} J_1(x)dF(x) + h(s;\tau)\right). \tag{4}$$

Remark 2.

1. Under local changes the assumption on h_n is equivalent to Assumption C.5 (i) in Horváth and Kokoszka (1997). Moreover, the limit distribution (4) closely resembles the limit distribution of the CUSUM-based change point estimator considered in that paper.
2. The proof of Theorem 2 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 many 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, Hörmann and Schauer (2009) for S-mixing processes, Giraitis and Surgailis (1999) for long memory linear (or moving average) processes, Dehling, Durieu and Tusche (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 proof of Theorem 2 for subordinated Gaussian processes. 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), should converge in distribution to a limit that corresponds to (4) (up to multiplicative constants).

3. In the proof of Theorem 2, convergence of $m_n^{-1}(\hat{k}_W - k_0)$ is derived from a continuous mapping theorem for the argmax functional which presumes unimodality of the considered limit process. The limit process in formula (3) attains its maximum at a unique point according to Lifshits' criterion for unimodality of Gaussian processes. For this reason, the argument relies on the assumption that the Hermite rank r of the class of functions $1_{\{G(\xi_i)\leq x\}} - F(x)$, $x \in \mathbb{R}$, equals 1, guaranteeing a Gaussian limit process. If $r > 1$, the limit process in formula (3) is non-Gaussian. Since Lifshits' criterion applies to Gaussian processes exclusively, an alternative argument is needed for non-Gaussian limit processes. Moreover, an application of Lemma 4 yields convergence of the sargmax computed with respect to compact intervals $[-M, M]$ only. An extension of convergence to the sargmax computed with respect to the whole real line is based on the observation that the limit in (3) is subjected to a negative drift, meaning that it diverges to $-\infty$ as $|s|$ tends to ∞. For the proof of Theorem 2, this behavior is deduced from the law of the iterated logarithm for fractional Brownian motion processes. In order to generalize Theorem 2 to limits determined by a Hermite rank $r > 1$, a corresponding result for a more general class of processes is required, e.g. a law of the iterated logarithm for general Hermite processes; see Mori and Oodaira (1986).

3. Applications

We consider two well-known data sets which have been analyzed before. We compute the estimator $\hat{k}_W$ based on the given observations and put our results into context with the findings and conclusions of other authors.

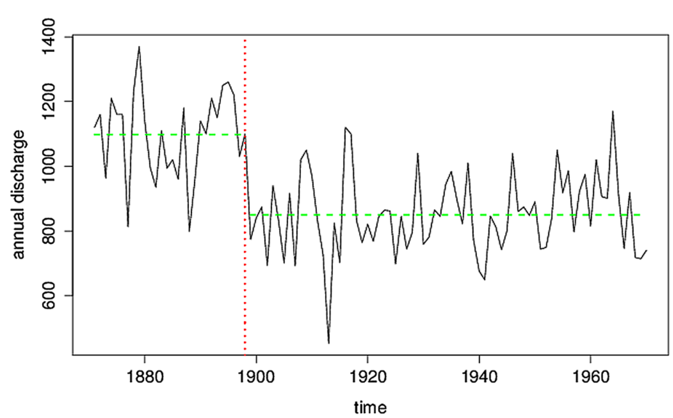

FIG 1. *Measurements of the annual discharge of the river Nile at Aswan in* $10^8 m^3$ *for the years 1871-1970. The dotted line indicates the potential change point estimated by* $\hat{k}_W$*; the dashed lines designate the sample means for the pre-break and post-break samples.*

The plot in Figure 1 depicts the annual volume of discharge from the Nile river at Aswan in $10^8 m^3$ for the years 1871 to 1970. The data set is included in any standard distribution of R. Amongst others, Cobb (1978), Macneill, Tang and Jandhyala (1991), Wu and Zhao (2007), Shao (2011) and Betken and Wendler (2015) provide 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 around the turn of the century. Yet, Cobb gave another explanation for the decrease in water volume by citing rainfall records which suggest a decline of tropical rainfall at that time. In fact, an application of the change point estimator $\hat{k}_W$ 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) indicates a change in 1898. Balke (1993) and Wu and Zhao (2007) suggest that the change occurred in 1899.

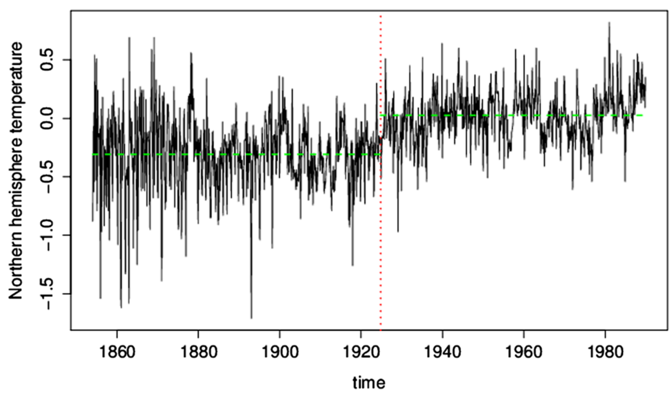

FIG 2. *Monthly temperature of the Northern hemisphere for the years 1854-1989 from the data base held at the Climate Research Unit of the University of East Anglia, Norwich, England. The temperature anomalies (in degrees C) are calculated with respect to the reference period 1950-1979. The dotted line indicates the location of the potential change point; the dashed lines designate the sample means for the pre-break and post-break samples.*

The second data set consists of the seasonally adjusted monthly deviations of the temperature (degrees C) for the Northern hemisphere during the years 1854 to 1989 from the monthly averages over the period 1950 to 1979. The data has been taken from the `longmemo` package in R. It results from spatial averaging of temperatures measured over land and sea. In view of the plot in Figure 2 it seems natural to assume that the data generating process is non-stationary. 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. However, the consideration of a more general stochastic model by the assumption of so-called semiparametric fractional autoregressive (SEMIFAR) processes in Beran and Feng (2002) does not confirm the conjecture of a trend-like behavior.

Neither does the investigation of the global temperature data in Wang (2007) support the hypothesis of an increasing trend. It is pointed out by Wang that the trend-like behavior of the Northern hemisphere temperature data may have been generated by stationary long-range dependent processes. Yet, it is shown in Shao (2011) and also in Betken and Wendler (2015) that under model assumptions that include long-range dependence an application of change point tests leads to a rejection of the hypothesis that the time series is stationary. According to Shao (2011) an estimation based on a self-normalized CUSUM test statistic suggests a 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 previously mentioned estimator $\hat{k}_{C,0}$ considered in Horváth and Kokoszka (1997). In this regard estimation by $\hat{k}_W$ seems to be in good accordance with the results of alternative change point estimators.

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 $\hat{k}_{\mathrm{C},0}$ (based on the CUSUM test statistic with parameter $\gamma = 0$). For this purpose, we consider two different scenarios:

1. Normal margins: We generate fractional Gaussian noise time series $(\xi_i)_{i\geq 1}$ and choose $G(t) = t$ in Assumption 1. As a result, the simulated observations $(Y_i)_{i\geq 1}$ are Gaussian with autocovariance function ρ satisfying

$$\rho(k) \sim \left(1 - \frac{D}{2}\right)(1 - D)\, k^{-D}.$$

Note that in this case the Hermite coefficient $J_1(x)$ is not equal to 0 for all $x \in \mathbb{R}$ (see Dehling, Rooch and Taqqu (2013a)) so that $m = 1$, where m denotes the Hermite rank of $1_{\{G(\xi_i)\leq x\}} - F(x), x \in \mathbb{R}$. Therefore, Assumption 1 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

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

with parameters $k, \beta > 0$ and with Φ denoting the standard normal distribution function. Since G is a strictly decreasing function, it follows by Theorem 2 in Dehling, Rooch and Taqqu (2013a) that the Hermite rank of $1_{\{G(\xi_i)\leq x\}} - F(x), x \in \mathbb{R}$, is $m = 1$ so that Assumption 1 holds for all values of $D \in (0, 1)$.

To analyze the behavior of the estimators we simulated 500 time series of length 600 and added a level shift of height h after a proportion τ of the data. We have done so for several choices of h and τ. The descriptive statistics, i.e. mean, sample standard deviation (S.D.) and quartiles, are reported in Tables 1, 2, and 3 for the three change point estimators $\hat{k}_W$, $\hat{k}_{SW}$ and $\hat{k}_{C,0}$.

The following observations, made on the basis of Tables 1, 2, and 3, correspond to the expected behavior of consistent change point estimators:

- 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 when there is very strong dependence, i.e. H is large, the variance of the series increases, so that it becomes harder to accurately estimate the location of a level shift.

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

- In most cases the estimator $\hat{k}_{SW}$ has a smaller bias, especially for an early change point location. Nevertheless, the difference between the biases of $\hat{k}_{SW}$ and $\hat{k}_W$ is not big.
- In general the sample standard deviation of $\hat{k}_W$ is smaller than that of $\hat{k}_{SW}$. Indeed, it is only slightly better for $\tau = 0.25$, but there is a clear difference for $\tau = 0.5$.

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 $\hat{k}_{SW}$.

Comparing the finite sample performance 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 slightly better, at least when $\tau = 0.25$ and especially for relatively high level shifts. Nonetheless, the deviations are in most cases negligible.
- If the change happens in the middle of a sample 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(3, 1) time series $\hat{k}_W$ clearly outperforms $\hat{k}_{C,0}$ by yielding smaller biases and decisively smaller variances 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 the Wilcoxon change point test is more robust against outliers in data sets than the CUSUM-like change point tests, i.e. the Wilcoxon test outperforms 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.

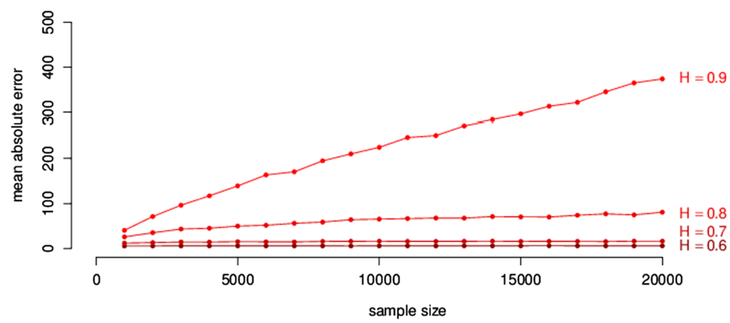

Fig 3. *The MAE of $\hat{k}_W$ for different values of H.*

As noted in Remark 1, $\hat{k}_W - k_0 = \mathcal{O}_P(1)$ under the assumption of a constant change point height h. This observation is illustrated by simulations of the mean absolute error

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

where $\hat{k}_{W,i}$, $i = 1, \ldots, m$, denote the estimates for k_0, computed on the basis of $m = 5000$ different sequences of fractional Gaussian noise time series.

Figure 3 depicts a plot of MAE against the sample size n with n varying between 1000 and 20000.

Since $\hat{k}_W - k_0 = \mathcal{O}_P(1)$ due to Theorem 1, we expect MAE to approach a constant as n tends to infinity. This can be clearly seen in Figure 3 for $H \in \{0.6, 0.7, 0.8\}$. For a high intensity of dependence in the data (characterized by $H = 0.9$) convergence becomes slower. This is due to a slower convergence of the test statistic $W_n(k)$ which, in finite samples, is not canceled out by the effect of a more regular behavior of the sample paths of the limit process.

5. Proofs

In the following 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.

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

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

Table 1. *Descriptive statistics of the sampling distribution of* $\hat{k}_W$ *for a change in the mean based on* 500 *fractional Gaussian noise and Pareto time series of length* 600 *with Hurst parameter* H *and a change in mean in* τ *of height* h*.*

margins	τ	h		$H = 0.6$	$H = 0.7$	$H = 0.8$	$H = 0.9$
normal	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$(3, 1)$	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. *Descriptive statistics of the sampling distribution of $\hat{k}_{SW}$ for a change in the mean based on 500 replications of fractional Gaussian noise and Pareto time series of length 600 with Hurst parameter H and a change in mean in τ of height h.*

margins	τ	h		$H = 0.6$	$H = 0.7$	$H = 0.8$	$H = 0.9$
normal	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$(3, 1)$	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 3. *Descriptive statistics of the sampling distribution of $\hat{k}_{C,0}$ for a change in the mean based on* 500 *replications of fractional Gaussian noise and Pareto time series of length* 600 *with Hurst parameter H and a change in mean in τ of height h.*

margins	τ	h		$H = 0.6$	$H = 0.7$	$H = 0.8$	$H = 0.9$
normal	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)	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)

For notational convenience we write $W_n(k)$ instead of $W_{k,n}$, $SW_n(k)$ instead of $SW_{k,n}$, and $\int$ instead of $\int_{\mathbb{R}}$. The proofs in this section as well as the proofs in the appendix are partially influenced by arguments that have been established in Horváth and Kokoszka (1997), Bai (1994) and Dehling, Rooch and Taqqu (2013a). In particular, some arguments are based on the empirical process non-central limit theorem of Dehling and Taqqu (1989) which states that

$$d_{n,r}^{-1}\lfloor n\lambda\rfloor(F_{\lfloor n\lambda\rfloor}(x)-F(x)) \xrightarrow{\mathcal{D}} \frac{1}{r!}J_r(x)Z_H^{(r)}(\lambda),$$

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

We write

$$\begin{aligned} X_n(\lambda,x) &:= d_{n,r}^{-1}\lfloor n\lambda\rfloor(F_{\lfloor n\lambda\rfloor}(x)-F(x)),\\ X(\lambda,x) &:= \frac{1}{r!}J_r(x)Z_H^{(r)}(\lambda), \end{aligned}$$

so that X_n, $n\in\mathbb{N}$, can be considered as a sequence of random variables with values in $D\left([-\infty,\infty]\times[0,1]\right)$ converging in distribution to X. Note that J_r is bounded and continuous. Moreover, the Hermite process $Z_H^{(r)}$ is almost surely continuous. With $C\left([-\infty,\infty]\times[0,1]\right)$ denoting the set of all continuous, real-valued functions with domain $[-\infty,\infty]\times[0,1]$, it follows that $X\in C\left([-\infty,\infty]\times[0,1]\right)$ with probability 1. Since $C\left([-\infty,\infty]\times[0,1]\right)$ is a separable subset of $D\left([-\infty,\infty]\times[0,1]\right)$, the Dudley-Wichura version of Skorohod's representation theorem (see Shorack and Wellner (1986), Theorem 2.3.4) implies that there exists another probability space $(\Omega',\mathcal{F}',P')$ and random variables X_n', $n\in\mathbb{N}$, and X' defined on it with $X_n' \overset{\mathcal{D}}{=} X_n$, $n\in\mathbb{N}$, and $X'\overset{\mathcal{D}}{=} X$ such that

$$\sup_{\lambda\in[0,1],x\in\mathbb{R}} |X_n'(\lambda,x)-X'(\lambda,x)| \longrightarrow 0 \tag{5}$$

almost surely. The line of argument in the proofs of Theorem 1 and Theorem 2 is partly based on this inference. In this context, it is important to note that, for notational convenience, we write

$$\sup_{\lambda\in[0,1],x\in\mathbb{R}}\left|d_{n,r}^{-1}\lfloor n\lambda\rfloor\left(F_{\lfloor n\lambda\rfloor}(x)-F(x)\right)-\frac{1}{r!}J_r(x)Z_H^{(r)}(\lambda)\right| \longrightarrow 0 \ \text{ a. s.} \tag{6}$$

only to indicate the convergence in (5). Generally speaking, it is not possible to infer that $\sup_{\lambda\in[0,1],x\in\mathbb{R}}|X_n(\lambda,x)-X(\lambda,x)|$ converges to 0 a.s. Since, whenever the argument in the proofs is based on the almost sure convergence in (6), we are only interested in distributional properties, this notation is always justified (although, in general, it is not possible to conclude that (6) holds).

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

Proof of Proposition 1. The proof of Proposition 1 is based on an application of Lemma 1 in the appendix. According to Lemma 1 it holds that, under the assumptions of Proposition 1,

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

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

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

and C denotes some non-zero constant.

It directly follows that $\frac{1}{nd_{n,r}}\max_{1\leq k\leq n-1}|W_n(k)|\xrightarrow{P}\infty$.

Furthermore,

$$\frac{1}{n^2 h_n}\max_{1\leq k\leq\lfloor n(\tau-\varepsilon)\rfloor}\left|\sum_{i=1}^{k}\sum_{j=k+1}^{n}\left(1_{\{X_i\leq X_j\}}-\frac{1}{2}\right)\right|$$

converges in probability to

$$C\sup_{0\leq\lambda\leq\tau-\varepsilon}\delta_\tau(\lambda)=C(\tau-\varepsilon)(1-\tau)$$

for any $0\leq\varepsilon<\tau$.

For $\varepsilon>0$ define

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

As $Z_{n,\varepsilon}\xrightarrow{P}C(1-\tau)\varepsilon$, it follows that $P(\hat{k}_W<\lfloor n(\tau-\varepsilon)\rfloor)=P(Z_{n,\varepsilon}=0)\longrightarrow 0$.

An analogous line of argument yields

$$P(\hat{k}_W>\lfloor n(\tau+\varepsilon)\rfloor)\longrightarrow 0.$$

All in all, it follows that for any $\varepsilon>0$

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

This proves consistency of the change point estimator which is based on the Wilcoxon test statistic.

In the following it is shown that $\frac{1}{n}\hat{k}_{SW}$ is a consistent estimator, too. For this purpose, we consider the process $SW_n(\lfloor n\lambda\rfloor)$, $0\leq\lambda\leq 1$. 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 process

$$\frac{1}{a_n}\sum_{i=1}^{\lfloor n\lambda\rfloor}\sum_{j=\lfloor n\lambda\rfloor+1}^{n}\left(1_{\{X_i\leq X_j\}}-\frac{1}{2}\right),\ 0\leq\lambda\leq 1,$$

where a_n denotes an appropriate normalization. Therefore, it follows by the corresponding argument in Betken (2016) that

$$SW_n(\lfloor n\lambda\rfloor)\xrightarrow{P}\frac{|\delta_\tau(\lambda)|}{\left\{\int_0^\lambda\left(\delta_\tau(t)-\frac{t}{\lambda}\delta_\tau(\lambda)\right)^2dt+\int_\lambda^1\left(\delta_\tau(t)-\frac{1-t}{1-\lambda}\delta_\tau(\lambda)\right)^2dt\right\}^{\frac{1}{2}}}$$

uniformly in $\lambda\in[0,1]$. Elementary calculations yield

$$\sup_{\lfloor n\tau_1\rfloor\leq k\leq k_0-n\varepsilon}SW_n(k)\xrightarrow{P}\sup_{\tau_1\leq\lambda\leq\tau-\varepsilon}\frac{\sqrt{3}\lambda\sqrt{1-\lambda}}{(\tau-\lambda)},$$

$$\sup_{k_0+n\varepsilon\leq k\leq\lfloor n\tau_2\rfloor}SW_n(k)\xrightarrow{P}\sup_{\tau+\varepsilon\leq\lambda\leq\tau_2}\frac{\sqrt{3}\sqrt{\lambda}(1-\lambda)}{(\tau-\lambda)}.$$

As $SW_n(k_0)\xrightarrow{P}\infty$ due to Theorem 2 in Betken (2016), we 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 $\frac{1}{n}\hat{k}_{SW}\xrightarrow{P}\tau$. $\square$

Proof of Theorem 1. In the following we write $\hat{k}$ instead of $\hat{k}_W$. For convenience, 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, respectively. Furthermore, we subsume both 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 1, we need to show that for all $\varepsilon>0$ there exists an $n(\varepsilon)\in\mathbb{N}$ and an $M>0$ such that

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

for all $n\geq n(\varepsilon)$.

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

We have

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

with

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

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

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

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

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

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

In the following we will consider the first summand only. (For the second summand analogous implications result from the same argument.)

For this, we define

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

where

$$\delta_n(k) := \begin{cases} k(n-k_0), & k \leq k_0 \\ k_0(n-k), & k > k_0 \end{cases}$$

and

$$\Delta(h_n) := \int \left(F(x+h_n) - F(x)\right) dF(x).$$

Note that

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

We have

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

Due to Lemma 2 in the appendix and Theorem 1.1 in Dehling, Rooch and Taqqu (2013a)

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

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

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

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 above inequality diverges if $h_n = h$ is fixed or if $h_n^{-1} = o\left(\frac{n}{d_{n,r}}\right)$. Therefore, it is possible to find an $n(\varepsilon) \in \mathbb{N}$ such that

$$\begin{aligned}
P_{2,1} &\leq P\left(2 \sup_{\lambda\in[0,\tau]} \left|W_n(\lfloor n\lambda\rfloor) - \widehat{W}_n(\lfloor n\lambda\rfloor)\right| \geq k_0(n-k_0)\Delta(h_n)\right) \\
&\leq P\left(2 \sup_{\lambda\in[0,\tau]} \left|W_n(\lfloor n\lambda\rfloor) - \widehat{W}_n(\lfloor n\lambda\rfloor)\right| \geq Knd_{n,r}\right) \\
&< \varepsilon
\end{aligned}$$

for all $n \geq n(\varepsilon)$.

We will now turn to the summand P_1. We have $P_1 \leq P_{1,1} + P_{1,2}$, where

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

In the following we will consider the first summand only. (For the second summand analogous implications result from the same argument.)

We define a random sequence k_n, $n \in \mathbb{N}$, by choosing $k_n \in D_{n,M}(1)$ such that

$$\begin{aligned}
&\sup_{k\in D_{n,M}(1)} \left(W_n(k) - \widehat{W}_n(k) + \widehat{W}_n(k_0) - W_n(k_0)\right) \\
&= W_n(k_n) - \widehat{W}_n(k_n) + \widehat{W}_n(k_0) - W_n(k_0).
\end{aligned}$$

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 $k_n \in D_{n,M}(1)$ and $m_n \longrightarrow \infty$ we have

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

for n sufficiently large. Thus, we have

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

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 1.

In any case, for $\delta > 0$ it is possible to find an $n_0 \in \mathbb{N}$ 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 $n \geq n_0$.

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

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

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

Some 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

$$A_{n,1}(k) := -(n-k_0)(k_0-k) \int \left(F_{k+1,k_0}(x+h_n) - F(x+h_n) \right) dF_{k_0+1,n}(x),$$

$$A_{n,2}(k) := -(n-k_0)(k_0-k) \int \left(F_{k_0+1,n}(x) - F(x) \right) dF(x+h_n),$$

$$A_{n,3}(k) := (k_0-k)k \int \left(F_k(x) - F(x) \right) dF_{k+1,k_0}(x),$$

$$A_{n,4}(k) := -k(k_0-k) \int \left(F_{k+1,k_0}(x) - F(x) \right) dF(x).$$

Thus, for $n \geq n_0$

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

For each $i \in \{1, \dots, 4\}$ it will be shown that

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

for n and M sufficiently large.

1. Note that

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

Due to stationarity

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

Note that

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

Since

$$\sup_{x \in \mathbb{R}} \left| d_{n,r}^{-1} n\left(F_n(x) - F(x)\right) - \frac{1}{r!} Z_H^{(r)}(1) J_r(x)\right| \longrightarrow 0 \ a.s.$$

if $n \longrightarrow \infty$, and as $k_0 - k \geq Mm_n$ with $m_n \longrightarrow \infty$, it follows that

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

converges to 0 almost surely. Therefore,

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

for n sufficiently large. Note that $\sup_{x\in\mathbb{R}} |J_r(x)| < \infty$. Furthermore, it is well-known that all moments of Hermite processes are finite. As a result, it follows by Markov's inequality that for some $M_0 \in \mathbb{R}$

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

for all $M \geq M_0$.

2. We have

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

for n sufficiently large. As a result,

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

Due to the empirical process non-central limit theorem of Dehling and Taqqu (1989) we have

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

Moreover,

$$\frac{1}{r!}\left|Z_H^{(r)}(1)-Z_H^{(r)}(\tau)\right|\sup_{x\in\mathbb{R}}|J_r(x)| \overset{\mathcal{D}}{=} \frac{1}{r!}(1-\tau)^H\left|Z_H^{(r)}(1)\right|\sup_{x\in\mathbb{R}}|J_r(x)|$$

since $Z_H^{(r)}$ is a H-self-similar process with stationary increments. Thus, we have

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

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

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

for M sufficiently large.

3. Note that

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

for n sufficiently large. Therefore,

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

The expression on the right hand side of the inequality converges in distribution to

$$\frac{1}{r!} \sup_{0\leq\lambda\leq 1} \left| Z_H^{(r)}(\lambda) \right| \sup_{x\in\mathbb{R}} |J_r(x)|$$

due to the empirical process non-central limit theorem. Since

$$\left\{ Z_H^{(r)}(\lambda),\ 0 \leq \lambda \leq 1 \right\} \overset{\mathcal{D}}{=} \left\{ \lambda^H Z_H^{(r)}(1),\ 0 \leq \lambda \leq 1 \right\},$$

we have

$$\sup_{0\leq\lambda\leq 1} \left| Z_H^{(r)}(\lambda) \right| \overset{\mathcal{D}}{=} |Z_H^{(r)}(1)|.$$

As a result, the aforementioned argument yields

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

for n and M sufficiently large.

4. We have

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

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

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

for n and M sufficiently large.

All in all, it follows that for all $\varepsilon > 0$ there exists an $n(\varepsilon) \in \mathbb{N}$ and an $M > 0$ such that

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

for all $n \geq n(\varepsilon)$. This proves Theorem 1. □

Proof of Theorem 2. 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 show 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 $l_n(s):=k_0+\lfloor m_n s\rfloor$. We have

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

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

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

We will show 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 $D\left[-M,M\right]$.

We rewrite $\tilde{V}_n(l_n(s))$ in the following way:

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

if $s<0$,

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

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

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

due to Lemma 3 and stationarity of the random sequence Y_i, $i \geq 1$. Note that

$$\begin{aligned} &d_{m_n}^{-1}(k_0 - l_n(s)) \int (F(x+h_n) - F(x))\, dF(x) \\ &= -d_{m_n}^{-1} \lfloor m_n s \rfloor h_n \int \frac{1}{h_n} (F(x+h_n) - F(x))\, dF(x). \end{aligned}$$

The above expression 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 3 and stationarity of the random sequence Y_i, $i \geq 1$. Note that

$$\begin{aligned} &d_{m_n}^{-1}(l_n(s) - k_0) \int (F(x+h_n) - F(x))\, dF(x) \\ &= d_{m_n}^{-1} \lfloor m_n s \rfloor h_n \int \frac{1}{h_n} (F(x+h_n) - F(x))\, dF(x) \end{aligned}$$

The above expression 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

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

Note that if $s < 0$,

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

If $s > 0$, we have

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

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

The arguments that appear in the proof of Lemma 3 can also be applied to show 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

$$
A_{1,n}(s) := (-l_n(s) - n + k_0)(k_0 - l_n(s)) \int \left(F_{l_n(s)+1,k_0}(x) - F(x)\right) dF(x)
$$

if $s < 0$,

$$
A_{1,n}(s) := (n - l_n(s) + k_0)(l_n(s) - k_0) \int \left(F_{k_0+1,l_n(s)}(x) - F(x)\right) dF(x)
$$

if $s > 0$,

$$
A_{2,n}(s) := \begin{cases} (k_0 - l_n(s))l_n(s) \int (F_{l_n(s)}(x) - F(x))dF(x) & \text{if } s < 0 \\ -(l_n(s) - k_0)(n - l_n(s)) \int \left(F_{l_n(s)+1,n}(x) - F(x)\right) dF(x) & \text{if } s > 0 \end{cases},
$$

$$
A_{3,n}(s) := \begin{cases} (k_0 - l_n(s))(n - k_0) \int (F_{k_0+1,n}(x) - F(x))\, dF(x) & \text{if } s < 0 \\ -(l_n(s) - k_0)k_0 \int (F_{k_0}(x) - F(x))\, dF(x) & \text{if } s > 0 \end{cases}.
$$

Note that for $s < 0$

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

The above expression converges to 0 uniformly in $s \in [-M, 0]$, since $\frac{m_n}{d_{m_n}} = o(\frac{n}{d_n})$ and since

$$
\begin{aligned}
&\sup_{-M \le s \le 0} \left| d_n^{-1} l_n(s) \int (F_{l_n(s)}(x) - F(x))dF(x) \right| \\
&\le \sup_{x,\lambda} \left| d_n^{-1} \lfloor n\lambda \rfloor (F_{\lfloor n\lambda \rfloor}(x) - F(x)) - B_H(\lambda) J(x) \right| \\
&\quad + \sup_{0 \le \lambda \le 1} |B_H(\lambda)| \left| \int J(x) dF(x) \right|,
\end{aligned}
$$

i.e. $\sup_{-M \le s \le 0} \left| d_n^{-1} l_n(s) \int (F_{l_n(s)}(x) - F(x))dF(x) \right|$ 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 show 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, $\frac{1}{nd_{m_n}}A_{1,n}(s)$, $s\in[0,M]$, converges in distribution to $B_H(s)\times\int J(x)dF(x)$, $s\in[0,M]$, in $D\,[0,M]$. Furthermore, we have

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

Note that

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

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

Considering $\frac{1}{nd_{m_n}}A_{1,n}(s)$, $s\in[-M,M]$, as a stochastic process with path space $D\,[-M,M]$, we note that for $s\in[0,M]$ and $t\in[-M,0]$

$$\left(\frac{1}{nd_{m_n}}A_{1,n}(s),\frac{1}{nd_{m_n}}A_{1,n}(t)\right)^\top \overset{\mathcal{D}}{=} \begin{pmatrix} e_n(s-t)-e_n(-t)\\ -e_n(-t)\end{pmatrix}+o_P(1),$$

where

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

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

$$\left(\frac{1}{nd_{m_n}}A_{1,n}(s),\frac{1}{nd_{m_n}}A_{1,n}(t)\right)^\top \xrightarrow{\mathcal{D}} \left(B_H(s-t)-B_H(-t),-B_H(-t)\right)^\top.$$

The limit is Gaussian with mean 0 and covariances $\mathrm{Cov}(B_H(s-t)-B_H(-t), -B_H(-t)) = \frac{1}{2}\left(|s|^{2H}+|t|^{2H}-|s-t|^{2H}\right)$, 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 2. By an extension of the argument to

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

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 $D[-M, M]$.

Furthermore, it follows that 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 $\frac{1}{h_n n^2}W_n(k_0)$.

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 vanishes as n tends to ∞, since $h_n^{-1} = o(n/d_n)$. Due to Lemma 3 the limit of $d_n^{-1}k_0\int(F_{k_0}(x+h_n) - F_{k_0}(x))dF_{k_0+1,n}(x)$ corresponds to the limit of $d_n^{-1}k_0\int(F(x+h_n) - F(x))dF(x)$. Therefore,

$$h_n^{-1}\int \left(F_{k_0}(x+h_n) - F_{k_0}(x)\right)dF_{k_0+1,n}(x) \longrightarrow \int f^2(x)dx \quad a.s.$$

In addition, $\frac{k_0}{n}\frac{(n-k_0)}{n} \longrightarrow \tau(1-\tau)$.

From this we can conclude that

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

in $D[-M, M]$. This completes the proof of the first assertion in Theorem 2.

In order to show that

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

we make use of Lemma 4.

For this purpose, we note that according to Lifshits' criterion for unimodality of Gaussian processes (see Theorem 1.1 in Ferger (1999)) the random function

$G_{H,\tau}(s) = B_H(s) \int J(x)dF(x) + h(s;\tau)$ attains its maximal value in $[-M, M]$ at a unique point with probability 1 for every $M > 0$. Hence, an application of Lemma 4 in the appendix yields

$$\operatorname*{sargmax}_{s \in [-M,M]} \frac{1}{e_n} \left(W_n^2(k_0 + \lfloor m_n s \rfloor) - W_n^2(k_0) \right) \xrightarrow{\mathcal{D}} \operatorname*{argmax}_{s \in [-M,M]} G_{H,\tau}(s).$$

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 for fractional Brownian motions we have $\lim_{|s|\to\infty} \frac{B_H(s)}{s} = 0$ a.s. so that $B_H(s) \int J(x)dF(x) + h(s;\tau) \longrightarrow -\infty$ a.s. if $|s| \to \infty$. Therefore, the limit corresponds to $\operatorname{argmax}_{s\in(-\infty,\infty)} G_{H,\tau}(s)$ if M is sufficiently large.

For $M > 0$ define

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

Note that

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

Therefore, 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 infinity. Note that

$$\begin{aligned}
P\left(\hat{k} = \hat{\hat{k}}(M) \right) &= P\left(\left| \hat{k} - k_0 \right| \le M m_n \right) \\
&= 1 - P\left(\left| \hat{k} - k_0 \right| > M m_n \right).
\end{aligned}$$

Furthermore, we have

$$\begin{aligned}
&\lim_{M\to\infty} \liminf_{n\to\infty} \left(1 - P\left(|\hat{k} - k_0| > M m_n \right) \right) \\
&= 1 - \lim_{M\to\infty} \limsup_{n\to\infty} P\left(|\hat{k} - k_0| > M m_n \right) \\
&= 1
\end{aligned}$$

because $|\hat{k} - k_0| = O_P(m_n)$ by Theorem 1. As a result, we have

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

Hence, for all $\varepsilon > 0$ there is an $M_0 \in \mathbb{R}$ and an $n_0 \in \mathbb{N}$ such that

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

for all $n \ge n_0$ and all $M \ge M_0$. This concludes the proof of Theorem 2. $\square$

Appendix A: Auxiliary results

In the following we prove some Lemmas that are needed for the proofs of our main results. Lemma 1 characterizes the asymptotic behavior of the Wilcoxon process under the assumption of a change-point in the mean. It is used to prove consistency of the change-point estimators $\hat{k}_{\mathrm{W}}$ and $\hat{k}_{SW}$.

Lemma 1. *Define* $\delta_\tau : [0,1] \longrightarrow \mathbb{R}$ *by*

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

Assume that Assumption 1 holds and that either

a) $h_n = h$ *with* $h \neq 0$,

or

b) $\lim_{n\to\infty} h_n = 0$ *with* $h_n^{-1} = o\left(\frac{n}{d_{n,r}}\right)$ *and* F *has a bounded density* f.

Then, we have

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

where

$$C := \begin{cases} \frac{1}{h}\int (F(x+h) - F(x))\, dF(x) & \text{if } h_n = h, \ h \neq 0, \\ \int f^2(x)dx & \text{if } \lim_{n\to\infty} h_n = 0 \text{ and } h_n^{-1} = o\left(\frac{n}{d_{n,r}}\right) \end{cases}.$$

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

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

By Lemma 1 in Betken (2016) the first summand on the right hand side of the equation converges in probability to $\lambda(1-\tau)\int (F(x+h) - F(x))\, dF(x)$ uniformly in $\lambda \leq \tau$. The second summand vanishes as n tends to ∞.

If $\lfloor n\lambda \rfloor > \lfloor n\tau \rfloor$,

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

In this case, the first summand on the right hand side of the equation converges in probability to $(1-\lambda)\tau \int (F(x+h)-F(x))\, dF(x)$ uniformly in $\lambda \geq \tau$ due to Lemma 1 in Betken (2016) while the second summand converges in probability to zero. All in all, it follows that

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

uniformly in $\lambda \in [0,1]$.

If $\lim_{n\to\infty} h_n = 0$, the process

$$\frac{1}{nd_{n,r}}\sum_{i=1}^{\lfloor n\lambda\rfloor}\sum_{j=\lfloor n\lambda\rfloor+1}^{n}\left(1_{\{X_i\leq X_j\}}-\frac{1}{2}\right)$$
$$-\frac{n}{d_{n,r}}\delta_\tau(\lambda)\int (F(x+h_n)-F(x))\, dF(x),\ 0\leq\lambda\leq 1,$$

converges in distribution to

$$\frac{1}{r!}\int J_r(x)dF(x)\left(Z_H^{(r)}(\lambda)-\lambda Z_H^{(r)}(1)\right),\ 0\leq\lambda\leq 1,$$

due to Theorem 3.1 in Dehling, Rooch and Taqqu (2013b). By assumption $h_n^{-1} = o\left(\frac{n}{d_{n,r}}\right)$, so that

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

The proof of Theorem 1, which establishes a convergence rate for the estimator $\hat{k}_{\mathrm{W}}$, requires the following result:

Lemma 2. *Suppose that Assumption 1 holds and let h_n, $n\in\mathbb{N}$, be a sequence of real numbers with $\lim_{n\to\infty} h_n = h$.*

1. *The process*

$$\frac{1}{nd_{n,r}}\sum_{i=1}^{\lfloor n\lambda\rfloor}\sum_{j=\lfloor n\tau\rfloor+1}^{n}\left(1_{\{Y_i\leq Y_j+h_n\}}-\int F(x+h_n)dF(x)\right),\ 0\leq\lambda\leq\tau,$$

converges in distribution to

$$(1-\tau)\frac{1}{r!}Z_H^{(r)}(\lambda)\int J_r(x+h)dF(x)$$
$$-\lambda\frac{1}{r!}\left(Z_H^{(r)}(1)-Z_H^{(r)}(\tau)\right)\int J_r(x)dF(x+h)$$

uniformly in $\lambda\leq\tau$.

2. *The process*

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

converges in distribution to

$$\begin{aligned}&(1-\lambda)\frac{1}{r!}Z_H^{(r)}(\tau)\int J_r(x+h)dF(x)\\&-\tau\frac{1}{r!}\left(Z_H^{(r)}(1)-Z_H^{(r)}(\lambda)\right)\int J_r(x)dF(x+h)\end{aligned}$$

uniformly in $\lambda\geq\tau$.

Proof. We give a proof for the first assertion only as the convergence of the second term follows by an analogous argument. The steps in this proof correspond to the argument that proves Theorem 1.1 in Dehling, Rooch and Taqqu (2013a).

For $\lambda\leq\tau$ it follows that

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

This yields the following decomposition:

$$\begin{aligned}&\frac{1}{nd_{n,r}}\sum_{i=1}^{\lfloor n\lambda\rfloor}\sum_{j=\lfloor n\tau\rfloor+1}^{n}\left(1_{\{Y_i\leq Y_j+h_n\}}-\int F(x+h_n)dF(x)\right)\\&=\frac{n-\lfloor n\tau\rfloor}{n}d_{n,r}^{-1}\lfloor n\lambda\rfloor\int\left(F_{\lfloor n\lambda\rfloor}(x+h_n)-F(x+h_n)\right)dF_{\lfloor n\tau\rfloor+1,n}(x)\\&\quad+\frac{n-\lfloor n\tau\rfloor}{n}d_{n,r}^{-1}\lfloor n\lambda\rfloor\int F(x+h_n)d\left(F_{\lfloor n\tau\rfloor+1,n}-F\right)(x).\end{aligned}\tag{7}$$

For the first summand we have

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

We will show that each of the summands on the right hand side converges to 0. The first summand converges to 0 because of the empirical non-central limit theorem of Dehling and Taqqu (1989). In order to show convergence of the second and third summand, note that $\sup_{0\leq\lambda\leq\tau}|Z_H^{(r)}(\lambda)| < \infty$ a.s. since the sample paths of the Hermite processes are almost surely continuous.

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

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

The first expression on the right hand side converges to 0 by the Glivenko-Cantelli theorem and the fact that $\int|H_r(y)|\varphi(y)dy<\infty$; 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

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

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 in (7) we have

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

Since $\frac{\lfloor n\lambda\rfloor}{n} \longrightarrow \lambda$ uniformly in λ, consider

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

The first and second summand on the right hand side 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. □

Lemma 3 and Lemma 4 are needed for the proof of Theorem 2.

Lemma 3. *Suppose that Assumption 1 holds and let l_n, $n \in \mathbb{N}$, and h_n, $n \in \mathbb{N}$, be two sequences with $l_n \to \infty$, $\lim_{n\to\infty} h_n = h$ and $l_n = \mathcal{O}(n)$. Then, it holds that*

$$\begin{aligned}
\sup_{0\leq s\leq 1}\Bigg| & d_{l_n,r}^{-1}\lfloor l_n s\rfloor \int \left(F_{\lfloor l_n s\rfloor}(x+h_n)-F_{\lfloor l_n s\rfloor}(x+h)\right) dF_n(x) \\
&- d_{l_n,r}^{-1}\lfloor l_n s\rfloor \int \left(F(x+h_n)-F(x+h)\right) dF(x)\Bigg| \qquad (8)
\end{aligned}$$

and

$$\sup_{0\leq s\leq 1}\Bigg| d_{l_n,r}^{-1}\lfloor l_n s\rfloor \int \left(F_n(x+h_n)-F_n(x+h)\right) dF_{\lfloor l_n s\rfloor}(x)$$

$$- d_{l_n,r}^{-1} \lfloor l_n s \rfloor \int (F(x+h_n) - F(x+h))\, dF(x) \Bigg| \tag{9}$$

converge to 0 *almost surely.*

Proof. For the expression (8) the triangle inequality yields

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

The first summand converges to 0 because of the empirical non-central limit theorem. Moreover, $\sup_{0\le s\le 1} \left| Z_H^{(r)}(s) \right| < \infty$ a.s. due to the fact that $Z_H^{(r)}$ is continuous with probability 1. It is shown in the proof of Lemma 2 that $\left| \int (J_r(x+h_n) - J_r(x+h)) dF_n(x) \right| \longrightarrow 0$. As a result, the second summand vanishes as n tends to ∞.

Furthermore, note that

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

for some constant K and n sufficiently large, since $l_n = \mathcal{O}(n)$. 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 (9), note that

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

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. It holds that

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

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 converges to 0, as well. $\square$

Lemma 4 establishes a condition under which convergence in distribution of a sequence of random variables entails convergence of the smallest argmax of the sequence.

Lemma 4. *Let K be a compact interval and denote by $D(K)$ the corresponding Skorohod space, i.e. the collection of all functions $f : K \longrightarrow \mathbb{R}$ which are right-continuous with left limits. Assume that Z_n, $n\in\mathbb{N}$, are random variables taking values in $D(K)$ and that $Z_n \xrightarrow{\mathcal{D}} Z$, where (with probability 1) Z is continuous and Z has a unique maximizer. Then* $\text{sargmax}(Z_n) \xrightarrow{\mathcal{D}} \text{argmax}(Z)$.

Proof. Due to Skorohod's representation theorem there exist random variables $\tilde{Z}_n$ and $\tilde{Z}$ defined on a common probability space $(\tilde{\Omega}, \tilde{F}, \tilde{P})$, such that $\tilde{Z}_n \overset{\mathcal{D}}{=} Z_n$, $\tilde{Z} \overset{\mathcal{D}}{=} Z$ and $\tilde{Z}_n \xrightarrow{a.s.} \tilde{Z}$. Due to Lemma 2.9 in Seijo et al. (2011) the smallest argmax functional is continuous at W (with respect to the Skorohod-metric and the sup-norm metric) if $W \in D(K)$ is a continuous function which has a unique maximizer. Since (with probability 1) Z is continuous with unique maximizer, $\text{sargmax}(\tilde{Z}_n) \xrightarrow{a.s.} \text{argmax}(\tilde{Z})$. As almost sure convergence implies convergence in distribution, we have $\text{sargmax}(\tilde{Z}_n) \xrightarrow{\mathcal{D}} \text{argmax}(\tilde{Z})$ and therefore $\text{sargmax}(Z_n) \xrightarrow{\mathcal{D}} \text{argmax}(Z)$. □

References

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* **61** 821–856. MR1231678

Bai, J. (1994). Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis* **15** 453–472. MR1292161

Bai, J. (1997). Estimation of a change point in multiple regression models. *Review of Economics and Statistics* **79** 551–563.

Balke, N. S. (1993). Detecting level shifts in time series. *Journal of Business & Economic Statistics* **11** 81–92.

Ben Hariz, S. and Wylie, J. J. (2005). 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. MR2188874

Beran, J. and Feng, Y. (2002). SEMIFAR models - a semiparametric framework for modelling trends, long-range dependence and nonstationarity. *Computational Statistics & Data Analysis* **40** 393–419. MR1924017

Beran, J., Feng, Y., Ghosh, S. and Kulik, R. (2013). *Long-memory processes*. Springer-Verlag Berlin Heidelberg. MR3075595

Berkes, I., Hörmann, S. and Schauer, J. (2009). Asymptotic results for the empirical process of stationary sequences. *Stochastic Processes and their Applications* **119** 1298–1324. MR2508575

Berkes, I. and Philipp, W. (1977). An almost sure invariance principle for the empirical distribution function of mixing random variables. *Probability Theory and Related Fields* **41** 115–137. MR0464344

Betken, A. (2016). Testing for change-points in long-range dependent time series by means of a self-normalized Wilcoxon test. *Journal of Time Series Analysis* **20** 785–809. MR3564308

Betken, A. and Wendler, M. (2015). Subsampling for General Statistics under Long Range Dependence. *arXiv:1509.05720*.

Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular variation*. Cambridge University Press. MR0898871

Cobb, G. W. (1978). The problem of the Nile: Conditional solution to a change-point problem. *Biometrika* **65** 243–251. MR0513930

Darkhovskh, B. S. (1976). 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.

Dehling, H., Durieu, O. and Tusche, M. (2014). Approximating class approach for empirical processes of dependent sequences indexed by functions. *Bernoulli* **20** 1372–1403. MR3217447

Dehling, H., Rooch, A. and Taqqu, M. S. (2013a). Non-parametric change-point tests for long-range dependent data. *Scandinavian Journal of Statistics* **40** 153–173.

Dehling, H., Rooch, A. and Taqqu, M. S. (2013b). Power of change-point tests for long-range dependent data. *arXiv:1303.4917*. MR3024037

Dehling, H. and Taqqu, M. S. (1989). The empirical process of some long-range dependent sequences with an application to U-statistics. *The Annals of Statistics* **17** 1767–1783. MR1026312

Deo, R. S. and Hurvich, C. M. (1998). Linear trend with fractionally integrated errors. *Journal of Time Series Analysis* **19** 379–397. MR1650546

Ferger, D. (1999). On the uniqueness of maximizers of Markov–Gaussian processes. *Statistics & Probability Letters* **45** 71–77. MR1718353

Giraitis, L., Leipus, R. and Surgailis, D. (1996). The change-point problem for dependent observations. *Journal of Statistical Planning and Inference* **53** 297–310. MR1407644

Giraitis, L. and Surgailis, D. (1999). Central limit theorem for the empirical process of a linear sequence with long memory. *Journal of Statistical Planning and Inference* **80** 81–93. MR1713796

Horváth, L. and Kokoszka, P. (1997). The effect of long-range dependence on change-point estimators. *Journal of Statistical Planning and Inference* **64** 57–81. MR1492361

Kiefer, J. (1972). Skorohod embedding of multivariate RV's, and the sample DF. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **24** 1–35. MR1554013

Kokoszka, P. and Leipus, R. (1998). Change-point in the mean of dependent observations. *Statistics & Probability Letters* **40** 385–393. MR1664564

Kuan, C. M. and Hsu, C. C. (1998). Change-point estimation of fractionally integrated processes. *Journal of Time Series Analysis* **19** 693–708. MR1665953

Lavielle, M. and Moulines, E. (2000). Least-squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis* **21** 33–59. MR1766173

MacNeill, I. B., Tang, S. M. and Jandhyala, V. K. (1991). A Search for the Source of the Nile's Change-Points. *Environmetrics* **2** 341–375.

Mori, T. and Oodaira, H. (1986). The law of the iterated logarithm for self-similar processes represented by multiple Wiener integrals. *Probability theory and related fields* **71** 367–391. MR0824710

Müller, D. W. (1970). On Glivenko-Cantelli convergence. *Probability Theory and Related Fields* **16** 195–210. MR0388616

Philipp, W. and Pinzur, L. (1980). Almost sure approximation theorems for the multivariate empirical process. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **54** 1–13. MR0595473

Seijo, E., Sen, B. et al. (2011). A continuous mapping theorem for the smallest argmax functional. *Electronic Journal of Statistics* **5** 421–439. MR2802050

Shao, X. (2011). A simple test of changes in mean in the possible presence of long-range dependence. *Journal of Time Series Analysis* **32** 598–606. MR2846559

Shao, X. and Zhang, X. (2010). Testing for change points in time series. *Journal of the American Statistical Association* **105** 1228–1240. MR2752617

Shorack, G. R. and Wellner, J. A. (1986). *Empirical processes with applications to statistics*. John Wiley & Sons, New York. MR0838963

Taqqu, M. S. (1979). Convergence of integrated processes of arbitrary Hermite rank. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **50** 53–83. MR0550123

Wang, L. (2007). Gradual changes in long memory processes with applications. *Statistics* **41** 221–240. MR2328228

Wu, W. B. and Zhao, Z. (2007). Inference of trends in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69** 391–410. MR2323759