University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Holmes, M;Kojadinovic, I

Title:
Open-end nonparametric sequential change-point detection based on the retrospective cusum statistic*

Date:
2021-01-01

Citation:
Holmes, M. & Kojadinovic, I. (2021). Open-end nonparametric sequential change-point detection based on the retrospective cusum statistic*. Electronic Journal of Statistics, 15 (1), pp.2288-2335. https://doi.org/10.1214/21-EJS1840.

Persistent Link:
https://hdl.handle.net/11343/288714

License:
CC-BY

Electronic Journal of Statistics
Vol. 15 (2021) 2288–2335
ISSN: 1935-7524
https://doi.org/10.1214/21-EJS1840

Open-end nonparametric sequential change-point detection based on the retrospective CUSUM statistic*

Mark Holmes[1] **and Ivan Kojadinovic**[2]

[1] *School of Mathematics & Statistics,*
The University of Melbourne,
Parkville, VIC 3010, Australia
e-mail: holmes.m@unimelb.edu.au

[2] *CNRS / Université de Pau et des Pays de l'Adour / E2S UPPA,*
Laboratoire de mathématiques et applications IPRA, UMR 5142,
B.P. 1155, 64013 Pau Cedex, France,
e-mail: ivan.kojadinovic@univ-pau.fr

Abstract: The aim of online monitoring is to issue an alarm as soon as there is significant evidence in the collected observations to suggest that the underlying data generating mechanism has changed. This work is concerned with open-end, nonparametric procedures that can be interpreted as statistical tests. The proposed monitoring schemes consist of computing the so-called retrospective CUSUM statistic (or minor variations thereof) after the arrival of each new observation. After proposing suitable threshold functions for the chosen detectors, the asymptotic validity of the procedures is investigated in the special case of monitoring for changes in the mean, both under the null hypothesis of stationarity and relevant alternatives. To carry out the sequential tests in practice, an approach based on an asymptotic regression model is used to estimate high quantiles of relevant limiting distributions. Monte Carlo experiments demonstrate the good finite-sample behavior of the proposed monitoring schemes and suggest that they are superior to existing competitors as long as changes do not occur at the very beginning of the monitoring. Extensions to statistics exhibiting an asymptotic mean-like behavior are briefly discussed. Finally, the application of the derived sequential change-point detection tests is succinctly illustrated on temperature anomaly data.

MSC 2010 subject classifications: Primary 62L99, 62E20; secondary 62G10.
Keywords and phrases: Change-point detection, online monitoring, open-end procedures, sequential testing.

Received July 2020.

Contents

1 Introduction 2289
2 The retrospective CUSUM for monitoring changes in the mean 2291

arXiv: 2007.08369
*Mark Holmes is supported by Future Fellowship FT160100166 from the Australian Research Council.

3 Asymptotics of the procedures 2296
4 Estimation of high quantiles of the limiting distributions 2301
5 Monte Carlo experiments 2305
6 Extensions to parameters whose estimators exhibit a mean-like behavior 2312
7 Data example 2314
8 Concluding remarks 2315
Acknowledgments 2316
A Proofs of the results under H_0 2316
B Proofs of the results under alternatives 2327
C Proofs of Propositions 4.1 and 6.2 2332
References 2333

1. Introduction

In situations in which observations are continuously collected over time, the aim of *sequential* or *online change-point detection* is to issue an alarm as soon as possible if it is thought that the probabilistic properties of the underlying unobservable data generating mechanism have changed. While this problem has a long history in *statistical process control* (see, e.g., Lai, 2001; Montgomery, 2007, for an overview), we adopt herein the alternative perspective proposed in the seminal work of Chu, Stinchcombe and White (1996) and treat the issue from the point of view of statistical testing. To fix ideas, assume that we have at hand an initial stretch $X_1, \ldots, X_m$ (frequently referred to as the *learning sample*) from a univariate stationary time series. As new observations arrive, the aim is to look for evidence against stationarity and issue an alarm if such evidence is deemed significant. Specifically, assuming that the kth observation has been collected, a positive statistic $D_m(k)$ measuring departure from stationarity is computed from the available observations $X_1, \ldots, X_k$ and compared to a suitably chosen threshold $w(k/m)$. If $D_m(k) > w(k/m)$, the hypothesis that $X_1, \ldots, X_k$ is a stretch from a stationary time series is rejected and the monitoring stops. Otherwise, a new observation X_{k+1} is collected and the previous steps are repeated using $X_1, \ldots, X_{k+1}$.

Among such procedures, one can distinguish between *closed-end* approaches for which the monitoring eventually stops if stationarity is not rejected after the arrival of observation X_n, $n > m$, and *open-end* approaches which, in principle, can continue indefinitely. The sequential testing procedures studied in this work pertain to this second category and, for that reason, their null hypothesis is

$$H_0 : X_1, \ldots, X_m, X_{m+1}, X_{m+2}, \ldots, \text{ is a stretch from a stationary time series.} \tag{1.1}$$

Their interpretation as statistical tests implies that, given a significance level $\alpha \in (0, 1/2)$, the thresholds $w(k/m)$, $k \geq m+1$, need to be chosen such that, ideally, under H_0,

$$\mathbb{P}\{D_m(k) \leq w(k/m) \text{ for all } k \geq m+1\} = \mathbb{P}\left\{\sup_{m+1 \leq k < \infty} \frac{D_m(k)}{w(k/m)} \leq 1\right\} \geq 1 - \alpha. \tag{1.2}$$

The previous display shows how closely the choice of the so-called *detector* D_m and the *threshold function* w are related. Intuitively, under H_0, the sample paths of the stochastic process $\{D_m(k)/w(k/m)\}_{k\geq m+1}$ should fluctuate but not exceed a constant threshold too often while, when H_0 is not true, they are expected to rapidly exceed it after a change in the data generating process has occurred.

The starting point of our investigations is the recent seminal work of Gösmann, Kley and Dette (2021) who study open-end monitoring schemes sensitive to potential changes in a parameter θ (such as the mean) of a time series. A first natural approach to tackle this problem, often referred to as the *ordinary CUSUM (cumulative sum)* in the sequential change-point detection literature, consists of comparing an estimator $\theta_{1:m}$ of θ computed from the learning sample $X_1, \ldots, X_m$ with an estimator $\theta_{m+1:k}$ of θ computed from the observations $X_{m+1}, \ldots, X_k$ collected after the monitoring has started; see, e.g., Horváth et al. (2004), Aue and Horváth (2004), Aue et al. (2006) as well as the references given in the recent review by Kirch and Weber (2018). The idea of Gösmann, Kley and Dette (2021) is to define detectors that take into account *all* of the differences $\theta_{1:j} - \theta_{j+1:k}$, $j \in \{m, \ldots, k-1\}$. This approach, that can be regarded as adapted from *retrospective* (or *offline*, *a posteriori*) change-point detection (which assumes that data collection has been completed *before* testing is carried out), treats each $j \in \{m, \ldots, k-1\}$ as a potential change-point. In their experiments, Gösmann, Kley and Dette (2021) found that their approach is not only more powerful than the ordinary CUSUM but also more powerful than the so-called *Page CUSUM* procedure which consists of defining a detector from the differences $\theta_{1:m} - \theta_{j+1:k}$, $j \in \{m, \ldots, k-1\}$; see, e.g, Fremdt (2015) and Kirch and Weber (2018).

When computed after the kth observation has been collected, the detector proposed by Gösmann, Kley and Dette (2021) (thus involving all the differences $\theta_{1:j} - \theta_{j+1:k}$, $j \in \{m, \ldots, k-1\}$) does not however coincide with the so-called *retrospective CUSUM statistic* that could be computed from $X_1, \ldots, X_k$ (also involving all the differences $\theta_{1:j} - \theta_{j+1:k}$, $j \in \{m, \ldots, k-1\}$; see, e.g., Csörgő and Horváth, 1997; Aue and Horváth, 2013 and the references therein). Reasons that led Gösmann, Kley and Dette (2021) not to consider such an approach are discussed in their Remark 2.1. As shall be explained in the next section, we believe that an open-end monitoring scheme using the retrospective CUSUM statistic as detector could be even more powerful than the procedure of Gösmann, Kley and Dette (2021) as long as changes do not occur at the very beginning of the monitoring.

The aim of this work is to address the theoretical and practical issues associated with defining a nonparametric detector for open-end monitoring such that it coincides at each k with the retrospective CUSUM statistic. The theoretical issues are mostly related to the choice of the threshold function, while the practical issues come from the fact that quantiles of the underlying limiting distribution required to carry out the sequential test are harder to estimate.

This paper is organized as follows. In the second section, we propose three open-end nonparametric monitoring schemes related to the retrospective CU-

SUM statistic designed to be sensitive to changes in the mean of univariate time series. Their asymptotic behavior as the size m of the learning sample tends to infinity is studied in the third section both under the null hypothesis of stationarity and relevant alternatives. Section 4 is concerned with the estimation of high quantiles of related limiting distributions necessary in practice to carry out the sequential tests. The fifth section presents a summary of extensive numerical experiments demonstrating the good finite-sample properties of the resulting sequential testing procedures. An extension to time series parameters whose estimators admit an asymptotic mean-like linearization as considered in Gösmann, Kley and Dette (2021) is briefly discussed in Section 6. A short illustration involving temperature anomaly data concludes the work.

All proofs are deferred to the appendices. Throughout the paper, all convergences are for $m \to \infty$ unless mentioned otherwise. A preliminary implementation of the studied tests is available in the package `npcp` (Kojadinovic, 2020) for the R statistical system (R Core Team, 2020).

2. The retrospective CUSUM for monitoring changes in the mean

Our aim is to derive open-end nonparametric sequential change-point detection procedures that are particularly sensitive to alternative hypotheses of the form

$$H_1 : \exists\, k^\star \geq m \text{ such that } \mathbb{E}(X_1) = \cdots = \mathbb{E}(X_{k^\star}) \neq \mathbb{E}(X_{k^\star+1}) = \mathbb{E}(X_{k^\star+2}) = \ldots \tag{2.1}$$

After the arrival of the kth observation with $k > m$, the data at hand consist of the stretch $X_1, \ldots, X_k$. If we were in the context of retrospective change-point detection, a natural test statistic would be the so-called *retrospective CUSUM statistic* (see, e.g., Csörgő and Horváth, 1997; Aue and Horváth, 2013, and the references therein) defined by

$$R_k = \max_{1 \leq j \leq k-1} \frac{j(k-j)}{k^{3/2}} |\bar{X}_{1:j} - \bar{X}_{j+1:k}|, \tag{2.2}$$

where

$$\bar{X}_{j:k} = \begin{cases} \dfrac{1}{k-j+1} \displaystyle\sum_{i=j}^{k} X_i, & \text{if } j \leq k, \\ 0, & \text{otherwise.} \end{cases} \tag{2.3}$$

In the definition of R_k, we see that every $j \in \{1, \ldots, k-1\}$ is treated as a potential change-point in the sequence $X_1, \ldots, X_k$. The maximum over j then implies that R_k will be large as soon as the difference between $\bar{X}_{1:j}$ and $\bar{X}_{j+1:k}$ is large for some j.

In the sequential context considered in this work, since $X_1, \ldots, X_m$ is the learning sample known to be a stretch from a stationary time series, a first natural modification of (2.2) is to restrict the maximum over j to $j \in \{m, \ldots, k-1\}$. This is the idea considered by Dette and Gösmann (2019) in a closed-end setting who, additionally, replaced the normalizing factor $k^{3/2}$ by $m^{3/2}$ so that

the asymptotics of the corresponding monitoring scheme could be studied as the size m of the learning sample tends to infinity. The resulting detector is

$$R_m(k) = \max_{m \leq j \leq k-1} \frac{j(k-j)}{m^{3/2}} |\bar{X}_{1:j} - \bar{X}_{j+1:k}|, \qquad k \geq m+1. \tag{2.4}$$

In an open-end setting, Gösmann, Kley and Dette (2021) choose however not to consider the detector R_m (see Remark 2.1 in the latter reference) but suggested instead the detector

$$E_m(k) = \max_{m \leq j \leq k-1} \frac{k-j}{m^{1/2}} |\bar{X}_{1:j} - \bar{X}_{j+1:k}|, \qquad k \geq m+1. \tag{2.5}$$

The difference between R_m and E_m evidently lies in the weighting of the absolute differences of means $|\bar{X}_{1:j} - \bar{X}_{j+1:k}|$, $j \in \{m, \ldots, k-1\}$. Instead of weighting $|\bar{X}_{1:j} - \bar{X}_{j+1:k}|$ by $j(k-j)/m^2$, (2.5) replaces j/m by 1. While this modification may be beneficial in terms of power when k is close to m, it could have a negative impact when k is substantially larger than m because, then, $j(k-j)/m^2$ can be substantially larger than $(k-j)/m$. In other words, we suspect that, for changes not occurring at the beginning of the monitoring, a suitable detection scheme based on R_m could be more powerful than the one proposed by Gösmann, Kley and Dette (2021) based on E_m.

Remark 2.1. As mentioned in the introduction, the simplest detector in an open-end setting is probably the so-called *ordinary CUSUM* initially considered by Horváth et al. (2004) for investigating changes in the parameters of linear models. With the aim of detecting changes in the mean, following Aue and Horváth (2004), it can be defined by

$$Q_m(k) = \frac{k-m}{m^{1/2}} |\bar{X}_{1:m} - \bar{X}_{m+1:k}|, \qquad k \geq m+1. \tag{2.6}$$

We will use it as a benchmark in our Monte Carlo experiments.

As explained in the introduction, the choice of a detector needs to be accompanied by the choice of a suitable threshold function. To heuristically justify our choice of a suitable threshold function for the detector R_m in (2.4), we momentarily consider the closed-end setting in which monitoring stops at the latest after observation X_n is collected. Following Dette and Gösmann (2019) among others, to be able to study the sequential testing scheme asymptotically, we set $n = \lfloor mT \rfloor$ for some real $T > 1$. Then, under H_0 in (1.1), assuming that the functional central limit theorem holds for the underlying stationary time series $(X_i)_{i \in \mathbb{Z}}$, it can be verified using relatively simple arguments (see, e.g., Dette and Gösmann, 2019, Section 3, or Kojadinovic and Verdier, 2021, Section 2.3) that $\{R_m(\lfloor mt \rfloor)\}_{t \in [1,T]}$ converges in distribution to $\{L(t)\}_{t \in [1,T]}$, where

$$L(t) = \sigma \sup_{1 \leq s \leq t} |tW(s) - sW(t)|, \qquad t \in [1, T],$$

W is a standard Brownian motion and $\sigma^2 = \sum_{i \in \mathbb{Z}} \mathrm{Cov}(X_0, X_i) > 0$ is the finite long-run variance of $(X_i)_{i \in \mathbb{Z}}$. For any fixed $t \in [1, T]$, by Brownian scaling and

the substitution $u = s/t$, $t^{-3/2}\sigma^{-1}L(t)$ is equal in distribution to

$$\sup_{1\le s\le t}\left|W\left(\frac{s}{t}\right)-\frac{s}{t}W(1)\right| = \sup_{1/t\le u\le 1}|W(u)-uW(1)|.$$

Hence, for large t, the distribution of $t^{-3/2}\sigma^{-1}L(t)$ is close to that of the supremum of a Brownian bridge on $[0,1]$. As a consequence, under H_0 in (1.1), the distribution of $t^{-3/2}\sigma^{-1}R_m(\lfloor mt\rfloor)$ stabilizes as m and t increase. The latter observation suggests the possibility of an open-end sequential testing scheme based on R_m in (2.4) with a threshold function that is not too different from $t\mapsto t^{3/2}$.

As shall become clear from Theorem 3.3 below, in order to ensure that (1.2) is fully meaningful when $D_m = R_m$, the corresponding threshold function actually needs to diverge to ∞ (as $t\to\infty$) slightly faster than $t^{3/2}$. We propose to use as threshold function for R_m

$$w_R(t) = t^{3/2+\eta}w_\gamma(t), \qquad t\in[1,\infty), \tag{2.7}$$

where $\eta > 0$ is a real parameter and

$$w_\gamma(t) = \max\left\{\left(\frac{t-1}{t}\right)^\gamma, \epsilon\right\}, \qquad t\in[1,\infty), \tag{2.8}$$

with $\gamma\ge 0$ another real parameter and $\epsilon > 0$ a technical constant that can be taken very small in practice (we used $\epsilon = 10^{-10}$ in our implementation).

Let us first explain the role of the parameter η. Following the perspective adopted in the discussion below (1.2), the resulting monitoring can be seen as consisting of computing $\{w_R(k/m)\}^{-1}R_m(k)$ for $k\ge m+1$. It is elementary that for any fixed $k\ge m+1$, as we increase η both the mean and the variance of $\{w_R(k/m)\}^{-1}R_m(k)$ decrease. The top-left plot of Figure 2.1 displays one sample path of $\{\{w_R(k/m)\}^{-1}R_m(k)\}_{m+1\le k\le m+5000}$ for $m = 100$, $\gamma = 0$ and $\eta = 0.1$ computed from a sequence of independent standard normals (solid line) and the sample path computed from the same sequence but with $\eta = 0.001$ instead (dotted line). Unsurprisingly, because of the factor $(k/m)^{-\eta}$ in $\{w_R(k/m)\}^{-1}R_m(k)$, the sample path with $\eta = 0.1$ is below the sample path with $\eta = 0.001$. This effect of η is confirmed by the top-right plot of Figure 2.1 which displays the corresponding empirical standard deviations at k against $k-m$ computed from 1000 sample paths. As expected, increasing the parameter η increases the rate of convergence (as $k\to\infty$) of $\{w_R(k/m)\}^{-1}R_m(k)$ (and its mean and variance) to zero. Intuitively, in the context of open-end monitoring, one would therefore want η to be very small so that there is little reduction in variability as time elapses. The practical choice of the parameter η will be discussed in detail in Section 4.

Let us now explain the role of the parameter γ. The multiplication of a candidate threshold function by w_γ in (2.7) was initially considered in Horváth et al. (2004). From a practical perspective, as we shall see below, this may improve the finite-sample performance of the sequential testing scheme at the beginning of the monitoring (and has a negligible effect later). From a theoretical perspective,

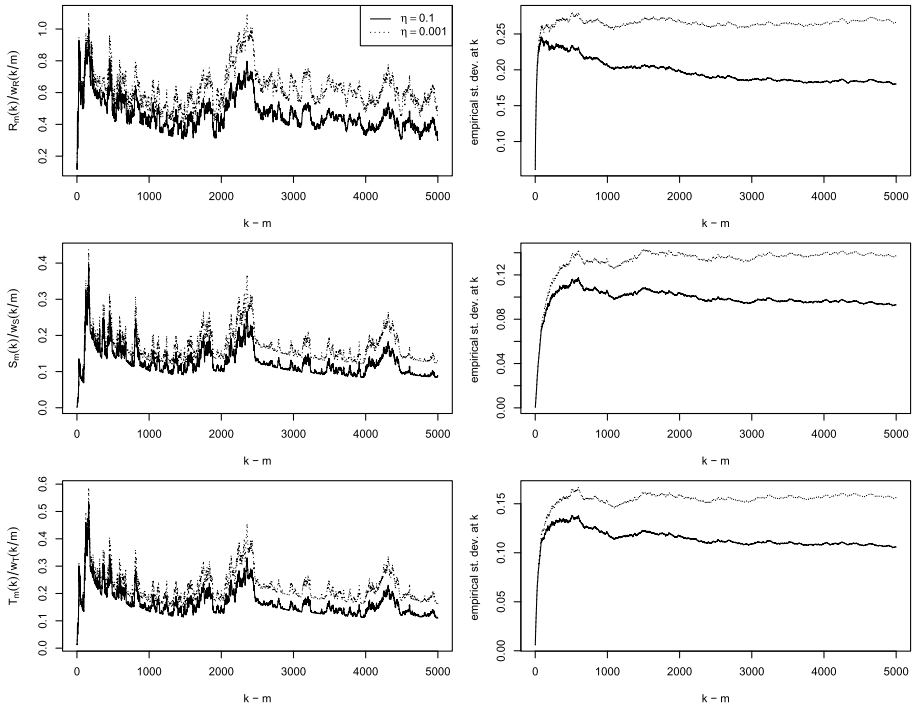


FIG 2.1. *Left column: for* $m = 100$, $\gamma = 0$ *and* $\epsilon = 10^{-10}$, *the solid lines represent one sample path of* $\{\{w_R(k/m)\}^{-1}R_m(k)\}_{m+1\le k\le m+5000}$ *(first row),* $\{\{w_S(k/m)\}^{-1}S_m(k)\}_{m+1\le k\le m+5000}$ *(second row) and* $\{\{w_T(k/m)\}^{-1}T_m(k)\}_{m+1\le k\le m+5000}$ *(third row) for* $\eta = 0.1$ *computed from an independent sequence of standard normal random variables. The dotted lines represent the sample paths computed from the same sequence but with* $\eta = 0.001$ *instead. Right column: corresponding empirical standard deviations at* k *against* $k - m$ *computed from 1000 sample paths.*

by doing so for the ordinary CUSUM detector Q_m in (2.6), Aue and Horváth (2004) were able, under alternatives related to (2.1), to derive the limiting distribution of the detection delay (the time delay after which the detector exceeds the threshold function). The use of such a modification of a threshold function turns out to be common in the literature and can also be found for instance in Fremdt (2015), Kirch and Weber (2018) and Gösmann, Kley and Dette (2021), among many others. Notice that, unlike what is frequently done in the literature, we do not impose that γ be strictly smaller than $1/2$. To provide some further insight, consider the top-left plot of Figure 2.2 which displays the empirical 95% quantile (solid line), empirical standard deviation (dashed line) and sample mean (dotted line) of $\{w_R(k/m)\}^{-1}R_m(k)$ for $m = 100$, $\eta = 0.001$ and $\gamma = 0$ against $k - m$ computed from 1000 sample paths computed from independent standard normal sequences. As one can see, because of the small value of η, the distribution of $\{w_R(k/m)\}^{-1}R_m(k)$ appears to stabilize rather quickly

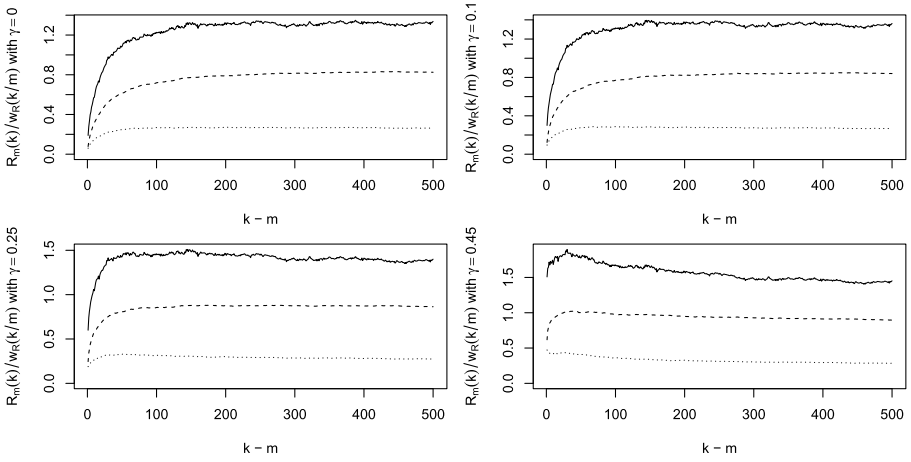


FIG 2.2. *The solid (resp. dashed, dotted) line represents the empirical 95% quantile (resp. empirical standard deviation, sample mean) of* $\{w_R(k/m)\}^{-1}R_m(k)$ *with* $m = 100$, $\eta = 0.001$ *and* $\gamma \in \{0, 0.1, 0.25, 0.45\}$ *against* $k - m$ *computed from 1000 sample paths.*

as k increases. The speed at which this occurs can be increased by increasing γ. For instance, by comparing the top-left and bottom-left plots of Figure 2.2, one can clearly see that the distribution of $\{w_R(k/m)\}^{-1}R_m(k)$ stabilizes quicker as k increases for $\gamma = 0.25$ than for $\gamma = 0$. However, as one can see from the plot for $\gamma = 0.45$, the value of γ should not be taken too large.

In addition to the detector R_m in (2.4), we shall also consider the detectors

$$S_m(k) = \frac{1}{m}\sum_{j=m}^{k-1} \frac{j(k-j)}{m^{3/2}} |\bar{X}_{1:j} - \bar{X}_{j+1:k}|, \qquad k \geq m+1, \tag{2.9}$$

$$T_m(k) = \sqrt{\frac{1}{m}\sum_{j=m}^{k-1} \left\{ \frac{j(k-j)}{m^{3/2}} (\bar{X}_{1:j} - \bar{X}_{j+1:k}) \right\}^2}, \qquad k \geq m+1, \tag{2.10}$$

with corresponding threshold functions

$$w_S(t) = t^{5/2+\eta} w_\gamma(t), \tag{2.11}$$

$$w_T(t) = t^{2+\eta} w_\gamma(t). \tag{2.12}$$

As one can see, S_m and T_m could be regarded as the L_1 and L_2 versions, respectively, of R_m in (2.4).

The parameters η and γ in (2.11) and (2.12) play the same role as in (2.7). For η, this can be empirically verified from the second and third rows of graphs in Figure 2.1. For γ, plots similar to Figure 2.2 for S_m and T_m reveal that values of γ larger than 0.25 seem meaningful for these two detectors. Specifically, it seems possible to improve the finite-sample performance of the corresponding schemes at the beginning of the monitoring by taking γ as large as 0.85 for S_m and as large as 0.45 for T_m.

3. Asymptotics of the procedures

To study the asymptotic behavior of the three considered monitoring schemes under H_0 in (1.1), we follow Horváth et al. (2004), Aue et al. (2006), Fremdt (2015) and Gösmann, Kley and Dette (2021), among others, and assume that the observations satisfy the following condition.

Condition 3.1. *The data are a stretch from a stationary time series $(X_i)_{i\in\mathbb{Z}}$ such that $\sigma^2 = \sum_{i\in\mathbb{Z}} \mathrm{Cov}(X_0, X_i)$, the long-run variance of $(X_i)_{i\in\mathbb{Z}}$, is strictly positive and finite. Furthermore, for all $m \in \mathbb{N}$, there exists two independent standard Brownian motions $W_{m,1}$ and $W_{m,2}$ such that, for some $0 < \xi < 1/2$,*

$$\sup_{m+1\le k\le\infty} \frac{1}{(k-m)^\xi} \left| \sum_{i=m+1}^{k} \{X_i - \mathbb{E}(X_1)\} - \sigma W_{m,1}(k-m) \right| = O_{\mathbb{P}}(1) \tag{3.1}$$

and

$$\frac{1}{m^\xi} \left| \sum_{i=1}^{m} \{X_i - \mathbb{E}(X_1)\} - \sigma W_{m,2}(m) \right| = O_{\mathbb{P}}(1). \tag{3.2}$$

As mentioned in Remark 2.6 (i) of Gösmann, Kley and Dette (2021), the validity of Condition 3.1 can be verified by proceeding along the lines of Aue et al. (2006) using the strong approximation results of Eberlein (1986). The latter is actually already carried out in Lemma A.1 of Aue et al. (2006) for augmented GARCH processes. Note that the verification of (3.1) is also discussed in Section 2 of Aue and Horváth (2004) for various types of stochastic processes.

Remark 3.2. In the prototypical situation in which $(X_i)_{i\in\mathbb{Z}}$ is a sequence of independent normal random variables with variance σ^2, there exists a probability space on which the above conditions are trivially satisfied with $O_{\mathbb{P}}(1)$ replaced with 0. This is essentially just the statement that the increments of a Brownian motion are standard normal random variables. To be more precise, let $(B_j)_{j\ge0}$ be independent standard Brownian bridges that are independent of $(X_i)_{i\in\mathbb{Z}}$. Without loss of generality, we may assume that the X_i are standard normal, and we can define $W_{1,m}$ by first specifying its values at integer times by $W_{1,m}(k-m) = \sum_{i=m+1}^{k} X_i$, $k \ge m$, and then interpolating between these values using the bridges $(B_{m+j})_{j\ge0}$: for $t \in (k-m, k-m+1)$, set $W_{1,m}(t) = W_{1,m}(k-m) + B_k\{t-(k-m)\} + \{t-(k-m)\}\{W_{1,m}(k-m+1) - W_{1,m}(k-m)\}$ (it is an exercise to check that the resulting process $W_{1,m}$ is a standard Brownian motion). Then, the term in absolute values in (3.1) is exactly 0 for each k and m. Similarly, setting $W_{2,m}(j) = \sum_{i=1}^{j} X_i$ for $j \le m$ and interpolating with the bridges $(B_j)_{0\le j\le m-1}$ makes (3.2) hold with 0 in the absolute value for each m. Furthermore, by construction, $W_{1,m}$ and $W_{2,m}$ are independent.

The following result is proven in Appendix A.

Theorem 3.3. *Under H_0 in (1.1) and Condition 3.1, for any fixed $\eta > 0$, $\epsilon > 0$ and $\gamma \ge 0$,*

$$\sigma^{-1} \sup_{m+1\le k<\infty} \{w_R(k/m)\}^{-1} R_m(k) \rightsquigarrow \sup_{1\le s\le t<\infty} \{w_R(t)\}^{-1} |tW(s) - sW(t)|,$$

$$\sigma^{-1} \sup_{m+1\le k<\infty} \{w_S(k/m)\}^{-1} S_m(k) \rightsquigarrow \sup_{t\in[1,\infty)} \{w_S(t)\}^{-1} \int_1^t |tW(s) - sW(t)| \mathrm{d}s$$

and

$$\sigma^{-1} \sup_{m+1\le k<\infty} \{w_T(k/m)\}^{-1} T_m(k) \rightsquigarrow \sup_{t\in[1,\infty)} \{w_T(t)\}^{-1} \sqrt{\int_1^t \{tW(s) - sW(t)\}^2 \mathrm{d}s},$$

where the arrow '$\rightsquigarrow$' denotes convergence in distribution, the detectors R_m, S_m and T_m are defined in (2.4), (2.9) and (2.10), respectively, the threshold functions w_R, w_S and w_T are defined in (2.7), (2.11) and (2.12), respectively, and W is a standard Brownian motion. In addition, all the limiting random variables are almost surely finite.

Imposing that η is strictly positive in the previous theorem is necessary for ensuring that the limiting random variables are almost surely finite. Recall the definition of the function w_γ in (2.8). Since the function $t \mapsto 1/w_\gamma(t)$ is bounded from below by 1, the following result, proven in Appendix A, implies that for the monitoring scheme based on R_m and w_R, this condition is necessary and sufficient.

Proposition 3.4. *For any fixed $M > 0$,*

$$\mathbb{P}\left\{ \sup_{1\le s\le t<\infty} \frac{1}{t^{3/2}} |tW(s) - sW(t)| \ge M \right\} = 1,$$

where W is a standard Brownian motion.

Remark 3.5. We thus have that $\sup_{1\le s\le t<\infty} \{w_R(t)\}^{-1} |tW(s) - sW(t)|$ is almost surely finite for $\eta > 0$ and infinite for $\eta = 0$. By the law of the iterated logarithm for Brownian motion, this supremum remains finite if t^η in w_R in (2.7) is replaced by $h(t)$, where $h(t) = \sqrt{\log\log t}$ when $t > e^e$ and $h(t) = 1$ when $t \le e^e$. We expect that Theorem 3.3 remains valid with such a modification which could be considered optimal in the sense that, as $t \to \infty$, h diverges slower to infinity than $t \mapsto t^\eta$ for any $\eta > 0$. The latter implies that the use of $h(t)$ instead of t^η entails the lowest possible variability reduction for the monitoring scheme (in the sense of the discussion on the role of η in the previous section) in the limit as $k \to \infty$.

The next corollary provides an operational version of Theorem 3.3 to carry out the three sequential change-point detection tests in practice.

Corollary 3.6. *The statement of Theorem 3.3 remains true if the long-run variance σ^2 is replaced by an estimator σ_m^2 of σ^2 computed from the learning sample $X_1, \ldots, X_m$ such that $\sigma_m^2 - \sigma^2 = o_{\mathbb{P}}(1)$.*

To fix ideas, let us briefly explain how Corollary 3.6 can be used to carry out the sequential test based on R_m in (2.4). Given a significance level $\alpha \in (0, 1/2)$,

it is first necessary to accurately estimate $q_{R,1-\alpha}$, the $(1-\alpha)$-quantile of the continuous random variable $\sup_{1\leq s\leq t<\infty}\{w_R(t)\}^{-1}|tW(s)-sW(t)|$ (this aspect will be discussed in detail in Section 4). Then, under H_0 in (1.1), from the Portmanteau theorem,

$$\mathbb{P}\left\{\sigma_m^{-1}\sup_{m+1\leq k<\infty}\{w_R(k/m)\}^{-1}R_m(k)\leq q_{R,1-\alpha}\right\} \to \mathbb{P}\left\{\sup_{1\leq s\leq t<\infty}\{w_R(t)\}^{-1}|tW(s)-sW(t)|\leq q_{R,1-\alpha}\right\}=1-\alpha. \quad (3.3)$$

Hence, for large m, we can expect that, under H_0,

$$\mathbb{P}\{R_m(k)\leq \sigma_m q_{R,1-\alpha}w_R(k/m) \text{ for all } k\geq m+1\}\approx 1-\alpha.$$

In practice, after the arrival of observation X_k, $k>m$, $R_m(k)$ is computed from $X_1,\ldots,X_k$ and compared to the threshold $\sigma_m q_{R,1-\alpha}w_R(k/m)$ (or, equivalently, $\sigma_m^{-1}\{w_R(k/m)\}^{-1}R_m(k)$ is computed and compared to the threshold $q_{R,1-\alpha}$). If $R_m(k)>\sigma_m q_{R,1-\alpha}w_R(k/m)$, the null hypothesis is rejected and the monitoring stops. Otherwise, the next observation is collected and the previous iteration is carried out with the $k+1$ available data points. Corollary 3.6 and (3.3) in particular guarantee that such a sequential testing procedure will have asymptotic level α. Steps to carry out the tests based on S_m in (2.9) and T_m in (2.10) can be obtained *mutatis mutandis*.

We now turn to the asymptotic behavior of the monitoring schemes under sequences of alternatives related to H_1 in (2.1). We start with the procedure based on R_m in (2.4) which we study under a condition similar to the one used by Gösmann, Kley and Dette (2021, Theorem 2.13).

Condition 3.7. *The data are a stretch, for some $m\in\mathbb{N}$, from the sequence of random variables $(X_i^{(m)})_{i\in\mathbb{N}}$ defined by*

$$X_i^{(m)}=\begin{cases}Y_i^{(0)}, & \text{if } i\leq k_m^\star,\\ Y_i^{(m)}, & \text{otherwise,}\end{cases}$$

where $(k_m^\star)_{m\in\mathbb{N}}$ is a sequence of integers such that $k_m^\star\geq m$ and $(Y_i^{(0)})_{i\in\mathbb{Z}}$, $(Y_i^{(1)})_{i\in\mathbb{Z}}$, …, $(Y_i^{(m)})_{i\in\mathbb{Z}}$, … are sequences of random variables defined on the same probability space and satisfying

(i) for each $m\geq 0$, $\big(\mathbb{E}(Y_i^{(m)})\big)_{i\in\mathbb{Z}}$ is a constant sequence,
(ii) $\sqrt{m}|\mathbb{E}(Y_1^{(0)})-\mathbb{E}(Y_1^{(m)})|\to\infty$,
(iii)

$$\frac{1}{\sqrt{k_m^\star}}\sum_{i=1}^{k_m^\star}\{Y_i^{(0)}-\mathbb{E}(Y_1^{(0)})\}=O_{\mathbb{P}}(1), \quad (3.4)$$

and either

(iv) there exist constants $c > 0$ *and* $C_1 > c_1 > 0$ *such that* $c_1 \leq m^{-1} k_m^\star \leq C_1$ *for all* $m \in \mathbb{N}$ *and*

$$\frac{1}{\sqrt{m}} \sum_{i=k_m^\star+1}^{k_m^\star+\lfloor cm \rfloor} \{Y_i^{(m)} - \mathbb{E}(Y_1^{(m)})\} = O_{\mathbb{P}}(1), \tag{3.5}$$

or

(v) $k_m^\star/m \to \infty$*, and there exists a constant* $c > 0$ *such that*

$$\frac{1}{\sqrt{k_m^\star}} \sum_{i=k_m^\star+1}^{k_m^\star+\lfloor ck_m^\star \rfloor} \{Y_i^{(m)} - \mathbb{E}(Y_1^{(m)})\} = O_{\mathbb{P}}(1). \tag{3.6}$$

Remark 3.8. Statement (iv) (resp. (v)) in Condition 3.7 is related to what were called *early* (resp. *late*) changes in Gösmann, Kley and Dette (2021) because $k_m^\star/m = O(1)$ (resp. $k_m^\star/m \to \infty$).

Remark 3.9. Suppose that $(Y_i^{(0)})_{i\in\mathbb{Z}}$ is a stationary centered sequence for which the central limit theorem holds, $(a_m)_{m\in\mathbb{N}}$ is a sequence such that $\sqrt{m}a_m \to \infty$ and, for any $m \in \mathbb{N}$, $Y_i^{(m)} = Y_i^{(0)} + a_m$, $i \in \mathbb{Z}$. It is then easy to verify that the sequences $(Y_i^{(m)})_{i\in\mathbb{Z}}$, $m \geq 0$, satisfy (i) and (ii) in Condition 3.7. Let us verify that they also satisfy (iii), (iv) and (v). Let $\varepsilon > 0$. Since the central limit theorem holds for $(Y_i^{(0)})_{i\in\mathbb{Z}}$, there exists $M > 0$ such that

$$\sup_{n\in\mathbb{N}} \mathbb{P}\Big(n^{-1/2} \sum_{i=1}^{n} Y_i^{(0)} > M\Big) < \varepsilon. \tag{3.7}$$

Hence, (3.4) holds for any sequence $(k_m^\star)_{m\in\mathbb{N}}$. For $c = 1$, the quantity on the left-hand side of (3.5) is equal to

$$\frac{1}{\sqrt{m}} \sum_{i=k_m^\star+1}^{k_m^\star+m} Y_i^{(0)}.$$

Since $\mathbb{P}(m^{-1/2}\sum_{i=1}^{m} Y_i^{(0)} > M) = \mathbb{P}(m^{-1/2}\sum_{i=n+1}^{n+m} Y_i^{(0)} > M)$ for all $m, n \in \mathbb{N}$, (3.7) implies that $\sup_{n\in\mathbb{N}} \sup_{m\in\mathbb{N}} \mathbb{P}(m^{-1/2}\sum_{i=n+1}^{n+m} Y_i^{(0)} > M) < \varepsilon$ which in turn implies that (3.5) holds with $c = 1$ for any sequence $(k_m^\star)_{m\in\mathbb{N}}$. Similarly, for $c = 1$, the quantity on the left hand side of (3.6) is equal to

$$\frac{1}{\sqrt{k_m^\star}} \sum_{i=k_m^\star+1}^{2k_m^\star} Y_i^{(0)},$$

and since $\sup_{n\in\mathbb{N}} \mathbb{P}(n^{-1/2}\sum_{i=n+1}^{2n} Y_i^{(0)} > M) < \varepsilon$, (3.6) holds with $c = 1$ for any sequence $(k_m^\star)_{m\in\mathbb{N}}$.

The following result proven in Appendix B establishes the consistency of the procedure based on R_m under Condition 3.7.

Theorem 3.10. *Under Condition 3.7, for any fixed $\eta \geq 0$, $\epsilon \geq 0$ and $\gamma \geq 0$,*

$$\sup_{m+1 \leq k < \infty} \{w_R(k/m)\}^{-1} R_m(k) \xrightarrow{\mathbb{P}} \infty,$$

where the arrow '$\xrightarrow{\mathbb{P}}$' denotes convergence in probability.

Assume that, under Condition 3.7, the estimator σ_m^2 appearing in Corollary 3.6 converges in probability to a strictly positive constant. An immediate corollary of the previous theorem is then that, under Condition 3.7, for any fixed $\eta \geq 0$, $\epsilon \geq 0$ and $\gamma \geq 0$, $\sigma_m^{-1} \sup_{m+1 \leq k < \infty} \{w_R(k/m)\}^{-1} R_m(k)$ diverges in probability to infinity. In other words, under the considered conditions, for any fixed $\eta \geq 0$, $\epsilon \geq 0$ and $\gamma \geq 0$, and any threshold $\kappa > 0$,

$$\mathbb{P}\left[\sigma_m^{-1} \sup_{m+1 \leq k < \infty} \{w_R(k/m)\}^{-1} R_m(k) > \kappa\right] \to 1.$$

Proving the consistency of the procedures based on S_m in (2.9) and T_m in (2.10) for late changes turns out to be more difficult. The following result, proven in Appendix B, shows that they are consistent for early changes. Note that the considered condition is very similar to those considered for instance in Dette and Gösmann (2019, Theorem 3.8) or Kojadinovic and Verdier (2021, Proposition 2.7) in the context of closed-end monitoring.

Condition 3.11. *The data are a stretch, for some $m \in \mathbb{N}$, from the sequence of random variables $(X_i^{(m)})_{i \in \mathbb{N}}$ defined by*

$$X_i^{(m)} = \begin{cases} Y_i, & \text{if } i \leq k_m^\star, \\ Z_i, & \text{otherwise,} \end{cases}$$

where $k_m^\star = \lfloor mc \rfloor$ for some constant $c \geq 1$, and $(Y_i)_{i \in \mathbb{Z}}$ and $(Z_i)_{i \in \mathbb{Z}}$ are stationary sequences defined on the same probability space such that $\mathbb{E}(Y_1) \neq \mathbb{E}(Z_1)$ and for which the functional central limit theorem holds.

Theorem 3.12. *Under Condition 3.11, for any fixed $T > c$, the empirical process $\{m^{-1/2} H_m(s,t)\}_{1 \leq s \leq t \leq T}$ converges in probability to $\{K_c(s,t)\}_{1 \leq s \leq t \leq T}$, where, for any $1 \leq s \leq t$,*

$$H_m(s,t) = m^{-3/2} \lfloor ms \rfloor (\lfloor mt \rfloor - \lfloor ms \rfloor)(\bar{X}^{(m)}_{1:\lfloor ms \rfloor} - \bar{X}^{(m)}_{\lfloor ms \rfloor + 1:\lfloor mt \rfloor}),$$
$$K_c(s,t) = (s \wedge c)\{(t \vee c) - (s \vee c)\}\{\mathbb{E}(X^{(m)}_{k_m^\star}) - \mathbb{E}(X^{(m)}_{k_m^\star + 1})\},$$

and $\vee$ and $\wedge$ are the maximum and minimum operators, respectively. Consequently, for any fixed $\eta \geq 0$, $\epsilon \geq 0$ and $\gamma \geq 0$,

$$\sup_{m+1\leq k<\infty}\{w_S(k/m)\}^{-1}S_m(k)\xrightarrow{\mathbb{P}}\infty \quad \textit{and} \quad \sup_{m+1\leq k<\infty}\{w_T(k/m)\}^{-1}T_m(k)\xrightarrow{\mathbb{P}}\infty. \tag{3.8}$$

Since the estimator σ_m^2 appearing in Corollary 3.6 converges in probability to a strictly positive constant under Condition 3.11, an immediate corollary of Theorem 3.12 is that, under the same conditions,

$$\sigma_m^{-1}\sup_{m+1\leq k<\infty}\{w_S(k/m)\}^{-1}S_m(k)\xrightarrow{\mathbb{P}}\infty$$

and

$$\sigma_m^{-1}\sup_{m+1\leq k<\infty}\{w_T(k/m)\}^{-1}T_m(k)\xrightarrow{\mathbb{P}}\infty.$$

4. Estimation of high quantiles of the limiting distributions

From the discussion following Corollary 3.6, we know that accurate estimations of high quantiles of the limiting distributions appearing in Theorem 3.3 are necessary to carry out the sequential tests based on the detectors R_m, S_m and T_m defined in (2.4), (2.9) and (2.10), respectively. Before we present the approach considered in this work, let us briefly explain how Horváth et al. (2004) and Gösmann, Kley and Dette (2021) proceeded for the procedure based on Q_m in (2.6) and the one based on E_m in (2.5), respectively. From the latter references, suitable threshold functions for these two detectors are $w_Q(t)=w_E(t)=tw_\gamma(t)$, $t\in[1,\infty)$, where the function w_γ is defined in (2.7) but with γ restricted to the interval $[0,1/2)$. Furthermore, from Gösmann, Kley and Dette (2021), under H_0 in (1.1) and Condition 3.1, one has

$$\sigma_m^{-1}\sup_{m+1\leq k<\infty}\{w_Q(k/m)\}^{-1}Q_m(k)\rightsquigarrow\sup_{t\in[0,1]}\frac{W(t)}{\max(t^\gamma,\epsilon)},$$
$$\sigma_m^{-1}\sup_{m+1\leq k<\infty}\{w_E(k/m)\}^{-1}E_m(k)\rightsquigarrow\sup_{0\leq s\leq t\leq 1}\frac{1}{\max(t^\gamma,\epsilon)}|W(t)-W(s)|,$$

where W is a standard Brownian motion. Notice that, using the law of the iterated logarithm/local modulus of continuity for Brownian motion, it can be verified, since $\gamma<1/2$, that the limiting random variables are almost surely finite even if ϵ is taken equal to zero. When $\gamma=0$, it seems particularly natural to estimate high quantiles of these limiting distributions by Monte Carlo simulation using an approximation of W on a fine grid of $[0,1]$ (one should keep in mind that although simulating such a process at discrete times always underestimates the true supremum, one can get arbitrarily close with probability as close to 1 as one wishes by taking a sufficiently fine grid). Note that in practice Horváth et al. (2004) and Gösmann, Kley and Dette (2021) use $\epsilon=0$ even when $\gamma>0$.

The following result, proven in Appendix C, suggests that the latter approach could be attempted for the first limiting distribution appearing in Theorem 3.3 related to the procedure based on R_m in (2.4).

Proposition 4.1. *For any fixed $\eta > 0$, $\epsilon > 0$ and $\gamma \geq 0$, the random variable*

$$\sup_{1 \leq s \leq t < \infty} \{w_R(t)\}^{-1} |tW(s) - sW(t)| = \sup_{1 \leq s \leq t < \infty} \frac{1}{t^{3/2+\eta} \max[\{(t-1)/t\}^{\gamma}, \epsilon]} |tW(s) - sW(t)| \quad (4.1)$$

and the random variable

$$\sup_{0 < u \leq v \leq 1} \frac{u^{1/2+\eta}}{v \max\{(1-u)^{\gamma}, \epsilon\}} |W(v) - W(u)| \quad (4.2)$$

are equal in distribution, where W is a standard Brownian motion.

Preliminary numerical experiments revealed however that there do not seem to be advantages to work with expressions of the form (4.2) as these seem to transpose practical issues due to the presence of ∞ in (4.1) into practical issues near zero. For that reason, we opted for a heuristic estimation approach which we detail in the rest of this section in the case of the procedure based on R_m in (2.4) (and which we used *mutatis mutandis* for the procedures based on S_m in (2.9) and T_m in (2.10) as well).

Fix $\eta > 0$, $\gamma \geq 0$ and $\alpha \in (0, 1/2)$, and let $q_{R,1-\alpha}$ denote the $(1-\alpha)$-quantile of $\sup_{1 \leq s \leq t < \infty} \{w_R(t)\}^{-1} |tW(s) - sW(t)|$. To estimate $q_{R,1-\alpha}$, we propose to take m relatively large and simulate a large number N of sample paths of $\{\{w_R(k/m)\}^{-1} R_m(k)\}_{m+1 \leq k \leq m+2^p}$ for p large from sequences of independent standard normal random variables. At this point, it may be tempting to use the $(1-\alpha)$-empirical quantile of $\sup_{m+1 \leq k \leq m+2^p} \{w_R(k/m)\}^{-1} R_m(k)$ as an estimate of $q_{R,1-\alpha}$. Although the latter is expected to be a good estimate of the $(1-\alpha)$-quantile of $\sup_{1 \leq s \leq t \leq T} \{w_R(t)\}^{-1} |tW(s) - sW(t)|$ for $T = 1 + 2^p/m$, it may underestimate $q_{R,1-\alpha}$ since the supremum for $1 \leq s \leq t < \infty$ obviously dominates the supremum for $1 \leq s \leq t \leq T$ for any $T > 1$. The latter will actually depend on the value η. Indeed, recall from our discussion in Section 2 that, because of the factor $(k/m)^{-\eta}$, increasing η decreases the variance of $\{w_R(k/m)\}^{-1} R_m(k)$ as k increases. This has two consequences as far as the above mentioned empirical estimation of $q_{R,1-\alpha}$ is concerned: (i) for any fixed $\eta > 0$, since the variability of $\{w_R(k/m)\}^{-1} R_m(k)$ decreases as k increases, if we could set p large enough, we would be able to obtain a good estimate of $q_{R,1-\alpha}$; unfortunately, our margin of action in that respect is limited by computational resources; (ii) for any fixed p, as η decreases to 0, the probability that our estimate of the $(1-\alpha)$-quantile of $\sup_{1 \leq s \leq t \leq 1+2^p/m} \{w_R(t)\}^{-1} |tW(s) - sW(t)|$ is also a good estimate of the $(1-\alpha)$-quantile of $\sup_{1 \leq s \leq t < \infty} \{w_R(t)\}^{-1} |tW(s) - sW(t)|$ decreases to zero; in other words, for any fixed p, as η decreases, it is less and less likely that the distribution of $\sup_{m+1 \leq k \leq m+2^p} \{w_R(k/m)\}^{-1} R_m(k)$ is a good approximation of the distribution of $\sup_{1 \leq s \leq t < \infty} \{w_R(t)\}^{-1} |tW(s) - sW(t)|$.

To try to empirically solve this quantile estimation problem, we propose, for any fixed $\eta > 0$, to attempt to model the relationship between p and the $(1-\alpha)$-empirical quantile of $\sup_{m+1 \leq k \leq m+2^p} \{w_R(k/m)\}^{-1} R_m(k)$, and to use the fitted

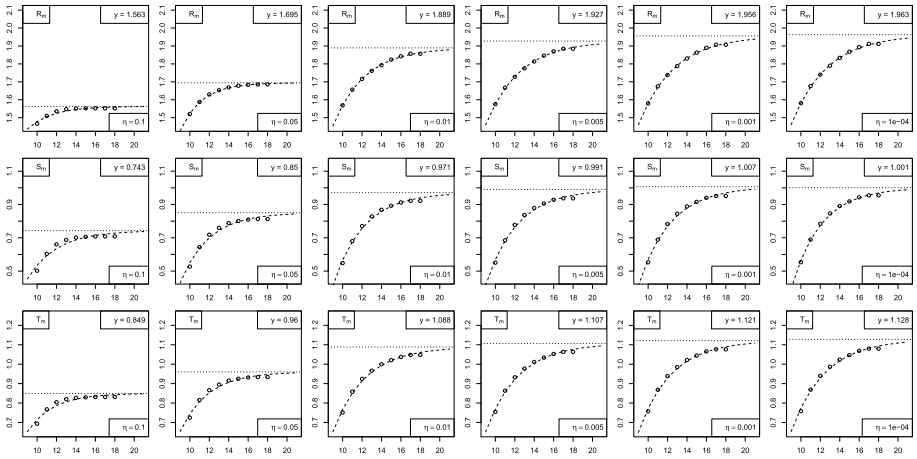


FIG 4.1. *For $\alpha = 0.05$ and $\gamma = 0$, scatter plots of $\{(p, q^{(p)}_{R,1-\alpha})\}_{p\in\{10,\ldots,18\}}$ (first row), $\{(p, q^{(p)}_{S,1-\alpha})\}_{p\in\{10,\ldots,18\}}$ (second row) and $\{(p, q^{(p)}_{T,1-\alpha})\}_{p\in\{10,\ldots,18\}}$ (third row), corresponding fitted asymptotic regression models (dashed curves) and estimates of the upper horizontal asymptotes (dotted lines) for $\eta \in \{0.1, 0.05, 0.01, 0.005, 0.001, 0.0001\}$.*

model to extrapolate the value of the quantile for larger p. To do so, we set m to 500, generated $N = 15000$ sample paths of $\{\{w_R(k/m)\}^{-1}R_m(k)\}_{m+1\le k\le m+2^{18}}$ using a computer cluster and, for $p \in \{10,\ldots,18\}$, estimated the $(1-\alpha)$-quantiles of $\sup_{m+1\le k\le m+2^p}\{w_R(k/m)\}^{-1}R_m(k)$. Let $q^{(p)}_{R,1-\alpha}$, $p \in \{10,\ldots,18\}$, denote the resulting estimates. Then, we fitted a so-called *asymptotic regression model* to the points $(p, q^{(p)}_{R,1-\alpha})$, $p \in \{10,\ldots,18\}$. The considered model, often used for the analysis of dose–response curves, is a three-parameter model with mean function $f(x) = c + (d-c)\{1-\exp(-x/e)\}$, where $y = d$ is the equation of the upper horizontal asymptote of f. Its fitting was carried out using the R package `drc` (Ritz et al., 2015). A candidate estimate of $q_{R,1-\alpha}$, the $(1-\alpha)$-quantile of $\sup_{1\le s\le t<\infty}\{w_R(t)\}^{-1}|tW(s)-sW(t)|$, is then the resulting estimate of the parameter d.

The first row of graphs in Figure 4.1 shows the scatter plots of the points $\{(p, q^{(p)}_{R,1-\alpha})\}_{p\in\{10,\ldots,18\}}$ for $\eta \in \{0.1, 0.05, 0.01, 0.005, 0.001, 0.0001\}$, $\alpha = 0.05$ and $\gamma = 0$. The corresponding fitted asymptotic regression models are represented by dashed curves. The estimated upper horizontal asymptotes are represented by dotted horizontal lines whose equations are given in the upper right corners of the plots.

As one can see from the first two graphs in the first row of Figure 4.1, the scatter plots for $\eta = 0.1$ and $\eta = 0.05$ reveal the presence of a plateau for larger values of p. The latter is an empirical indication of the fact that the distribution of the random variable $\sup_{m+1\le k\le m+2^p}\{w_R(k/m)\}^{-1}R_m(k)$, say for $p = 18$, seems to be a good approximation of the distribution of the random variable $\sup_{m+1\le k<\infty}\{w_R(k/m)\}^{-1}R_m(k)$ and thus of $\sup_{1\le s\le t<\infty}\{w_R(t)\}^{-1}|tW(s) -$

TABLE 4.1

For $\alpha \in \{0.01, 0.05, 0.1\}$, estimated $(1-\alpha)$-quantiles of the limiting distributions appearing in Theorem 3.3 related to the monitoring schemes based on R_m and S_m for $\eta = 0.001$ and different values of γ. The quantiles were estimated using asymptotic regression models. Standard errors of the estimates are provided between parentheses. Estimated quantiles for larger values of η are available in the R package `npcp`.

	R_m		S_m	
$1-\alpha$	$\gamma = 0$	$\gamma = 0.25$	$\gamma = 0$	$\gamma = 0.85$
0.99	2.157(0.006)	2.278(0.002)	1.145(0.019)	1.199(0.008)
0.95	1.956(0.009)	2.054(0.003)	1.007(0.015)	1.058(0.007)
0.90	1.837(0.008)	1.952(0.007)	0.939(0.014)	0.987(0.006)

$sW(t)|$. In other words, because of the relatively large values of η leading to a relatively quick reduction of the variability of $\{w_R(k/m)\}^{-1}R_m(k)$ as k increases, the supremum of $\{\{w_R(k/m)\}^{-1}R_m(k)\}_{m+1\leq k<\infty}$ tends to occur for k relatively close to m (and smaller than $m + 2^{18}$). Specifically, under H_0 in (1.1) and with $\eta = 0.1$ (resp. 0.05), the standard deviation of $\{w_R(k/m)\}^{-1}R_m(k)$ at $k = 2^{18}$ could be very roughly approximated to be $(2^{18}/500)^{-0.1} \approx 53\%$ (resp. $(2^{18}/500)^{-0.05} \approx 73\%$) of the standard deviation of $\{w_R(k/m)\}^{-1}R_m(k)$ in the early stages of the monitoring. The corresponding standard deviation reductions for $\eta = 0.01$, 0.005, 0.001 and 0.0001 are very roughly 7%, 3%, 1% and less than 0.1%, respectively.

Because $\eta \mapsto \{w_R(k/m)\}^{-1}R_m(k)$ is decreasing for any fixed $k \geq m + 1$, the estimated quantiles $q^{(p)}_{R,1-\alpha}$, $p \in \{10, \ldots, 18\}$, and their corresponding estimated asymptotes (which are candidate estimates of $q_{R,1-\alpha}$) are increasing as η decreases in the first row of plots in Figure 4.1. As one can see, the rate of increase of the estimated quantiles decreases as η decreases. For instance, there are hardly any differences between the plot for $\eta = 0.001$ and the plot for $\eta = 0.0001$. Unsurprisingly, this last plot is actually hardly indistinguishable from the plot for $\eta = 0$. The latter empirical observation practically implies that the proposed estimation technique will provide a finite estimate of $q_{R,1-\alpha}$ even when, according to Proposition 3.4, $q_{R,1-\alpha}$ is known to be infinite. It follows that it is not meaningful in practice to consider values of η that are "very small". Based on the previous approximate variance reduction calculations and the plots given in Figure 4.1, our intuitive recommendation is not to take η smaller than 0.001. Notice, that with $m = 100$, this value of η induces an approximate standard deviation reduction under H_0 for $\{w_R(k/m)\}^{-1}R_m(k)$ after 10^{10} monitoring steps of less than 2%.

The estimated quantiles for $\eta = 0.001$, $\alpha \in \{0.01, 0.05, 0.1\}$, $\gamma = 0$ as well for the largest meaningful value of γ (among those that were considered) for the three monitoring schemes are reported in Tables 4.1 and 4.2 along with standard errors. Estimated quantiles for larger values of η are available in the R package `npcp` (Kojadinovic, 2020).

Remark 4.2. It is a research project of its own to investigate more thoroughly the estimation of the quantiles both empirically and theoretically. For instance,

TABLE 4.2
For $\alpha \in \{0.01, 0.05, 0.1\}$, estimated $(1-\alpha)$-quantiles of the limiting distributions appearing in Theorem 3.3 related to the monitoring scheme based on T_m for $\eta = 0.001$ and $\gamma \in \{0, 0.45\}$. The quantiles were estimated using asymptotic regression models. Standard errors of the estimates are provided between parentheses. Estimated quantiles for larger values of η are available in the R package `npcp`.

	T_m	
$1-\alpha$	$\gamma = 0$	$\gamma = 0.45$
0.99	1.246(0.006)	1.324(0.007)
0.95	1.121(0.011)	1.164(0.005)
0.90	1.046(0.010)	1.087(0.004)

one may wish to investigate bounds on the probability for a given $\eta > 0$ and $T > 1$ that $\sup_{1 \le s \le t \le T} \{w_R(t)\}^{-1} |tW(s) - sW(t)|$ occurs at some (s, t) with $t > T$, and whether such occurrences tend to be associated with small suprema or large suprema.

5. Monte Carlo experiments

To investigate the finite-sample properties of the studied open-end sequential change-point detection procedures, we carried out extensive Monte Carlo experiments. One should however keep in mind that numerical experiments cannot provide a full insight into the finite-sample behavior of open-end approaches as finite computing resources impose that monitoring has to be stopped eventually.

To estimate the long-run variance σ^2 related to the learning sample, we used the approach of Andrews (1991) based on the quadratic spectral kernel with automatic bandwidth selection as implemented in the function `lrvar()` of the R package `sandwich` (Zeileis, 2004). We first considered 10 data generating models, denoted M1, ..., M10. Models M1, ..., M5 are simple AR(1) models with normal innovations whose autoregressive parameter is equal to 0, 0.1, 0.3, 0.5 and 0.7, respectively. These models were chosen among others to allow a comparison of our results with those of Gösmann, Kley and Dette (2021, Section 4.1). Model M6 generates independent observations from the Student t distribution with 5 degrees of freedom. Model M7 is a GARCH(1,1) model with normal innovations and parameters $\omega = 0.012$, $\beta = 0.919$ and $\alpha = 0.072$ to mimic SP500 daily log-returns following Jondeau, Poon and Rockinger (2007). Models M8 and M9 are the nonlinear autoregressive model used in Paparoditis and Politis (2001, Section 3.3) and the exponential autoregressive model considered in Auestad and Tjøstheim (1990) and Paparoditis and Politis (2001, Section 3.3). The underlying generating equations are

$$X_i = 0.6 \sin(X_{i-1}) + \epsilon_i$$

and

$$X_i = \{0.8 - 1.1 \exp(-50 X_{i-1}^2)\} X_{i-1} + 0.1 \epsilon_i,$$

respectively, where the ϵ_i are independent standard normal innovations. Note that, for all time series models, a burn-out sample of 100 observations was used. Finally, in order to mimic count data, model M10 generates independent observations from a Poisson distribution with parameter 3.

In a first series of experiments, we attempted to assess how well the studied tests hold their level. To do so, we generated 5000 samples from models M1–M10 and used the first m observations of each sample as learning sample. All the sequential tests were carried out at the 5% nominal level using the estimated quantiles available in the R package `npcp` (see also Tables 4.1 and 4.2). Because monitoring cannot go on indefinitely, we stopped the sequential testing procedures after the arrival of observation $n = m + 10000$. The percentages of rejection of H_0 in (1.1) for the procedures based on R_m in (2.4), S_m in (2.9), T_m in (2.10) with $\gamma = 0$ and $\eta \in \{0.05, 0.01, 0.005, 0.001\}$, as well as for the procedures based on E_m in (2.5) and Q_m in (2.6) with $\gamma = 0$, are reported in Tables 5.1 and 5.2 (to carry out the procedures based on Q_m and E_m, we used the quantiles reported in Table 1 of Horváth et al. (2004) and Gösmann, Kley and Dette (2021), respectively). Given the closed-end nature of the experiments, one should keep in mind that the empirical levels would have been higher had larger values of n been considered. As one can see from Tables 5.1 and 5.2, for most models, the empirical levels drop below the 5% threshold rather quickly as m increases. When the tests are too liberal, it is probably mostly a consequence of the difficulty of the estimation of the long-run variance σ^2. It is for instance unsurprising that a large value of m is necessary to obtain a reasonably accurate estimate of σ^2 for model M5 (the AR(1) model with the strongest serial dependence) or model M9 (an exponential autoregressive model). Notice that, overall, the tests based on E_m and Q_m are less liberal then the tests based on R_m, S_m and T_m when the serial dependence is strong and m is small. The opposite tends to occur as m increases. Among the three proposed procedures, the one based on S_m is the most conservative, followed by the procedure based on T_m. The fact that the empirical levels are higher for large values of η is due to the factor $(k/m)^{-\eta}$ which favors the occurrence of *false alarms* (threshold exceedences) at the beginning of the monitoring.

In a second series of experiments, we studied the finite-sample behavior of the tests under H_1 in (2.1) using simulation scenarios similar to those considered in Gösmann, Kley and Dette (2021, Section 4.1). We generated 1000 samples of size $n = m + 7000$ from models M1 with $m = 100$ and M4 with $m = 200$ and, for each sample, added a positive offset of δ to all observations after position $k^\star \in \{m, m + 500, m + 1000, m + 5000\}$. We started by investigating the influence of η on the most conservative procedure according to Tables 5.1 and 5.2. The rejection percentages for the test based on S_m with $\gamma = 0$ and $\eta \in \{0.1, 0.01, 0.001\}$, as well as for the procedures based on E_m and Q_m with $\gamma = 0$ are represented in Figure 5.1. Note that only threshold exceedences after position $k^\star$ were taken into account in the computation of rejection percentages. Those occurring before (*false alarms*) were ignored. As one can see, when the change occurs right at the beginning of the monitoring, the larger η, the more powerful the procedure based on S_m. As $k^\star$ increases, the influence of the fac-

TABLE 5.1

Percentages of rejection of H_0 in (1.1) for the procedures based on R_m and S_m with $\gamma = 0$ and $\eta \in \{0.1, 0.05, 0.01, 0.001\}$. The rejection percentages are computed from 5000 samples of size $n = m + 10000$ generated from the time series models M1 – M10.

		R_m with $\eta =$				S_m with $\eta =$			
Model	m	0.05	0.01	0.005	0.001	0.05	0.01	0.005	0.001
M1	100	8.7	7.8	7.4	7.1	5.5	4.8	4.6	4.5
	200	6.7	5.2	4.7	4.5	3.7	2.8	2.8	2.8
	400	5.4	3.6	3.1	2.8	2.4	1.6	1.4	1.2
	800	4.3	2.5	2.1	1.7	1.7	0.9	0.7	0.7
M2	100	9.0	8.2	7.8	7.6	6.0	5.1	4.9	4.8
	200	6.9	5.3	5.0	4.7	3.8	3.0	2.9	2.8
	400	5.4	3.6	3.2	2.7	2.4	1.6	1.4	1.3
	800	4.2	2.5	2.1	1.9	1.7	0.9	0.7	0.7
M3	100	10.4	9.8	9.5	9.3	7.1	6.2	6.0	5.9
	200	7.5	5.9	5.5	5.3	4.1	3.4	3.2	3.1
	400	5.5	3.6	3.2	2.9	2.6	1.6	1.5	1.4
	800	4.2	2.5	2.1	1.9	1.8	0.9	0.9	0.7
M4	100	13.1	12.3	11.9	11.6	9.0	8.2	8.0	7.9
	200	8.9	7.3	6.8	6.4	4.8	4.0	3.8	3.6
	400	6.1	4.0	3.6	3.3	2.8	1.8	1.7	1.6
	800	4.4	2.4	2.1	1.9	1.9	1.0	0.9	0.8
M5	100	18.4	18.3	17.7	17.4	13.9	13.1	12.8	12.6
	200	11.5	10.1	9.6	9.4	7.3	6.3	5.9	5.6
	400	7.6	5.2	4.7	4.4	3.8	2.6	2.5	2.2
	800	4.9	2.9	2.5	2.2	2.1	1.3	1.1	1.1
M6	100	10.8	10.0	9.4	8.9	6.8	5.9	5.7	5.6
	200	8.3	6.4	5.8	5.4	4.3	3.3	3.1	2.9
	400	5.9	3.8	3.4	3.1	2.9	1.8	1.7	1.6
	800	4.5	2.4	2.1	1.9	1.9	1.0	1.0	0.9
M7	100	12.8	11.4	10.8	10.5	7.7	6.9	6.8	6.5
	200	9.6	7.6	7.1	6.7	4.8	4.1	3.9	3.7
	400	6.4	4.2	3.6	3.4	2.9	1.8	1.5	1.5
	800	5.0	3.0	2.4	2.1	1.8	1.2	1.1	1.0
M8	100	9.3	8.5	8.0	7.8	6.7	5.9	5.7	5.6
	200	6.9	5.6	5.2	4.9	4.1	3.4	3.2	3.1
	400	5.4	3.5	3.1	2.8	2.5	1.5	1.4	1.3
	800	4.2	2.4	2.2	1.9	1.8	1.0	0.9	0.8
M9	100	36.0	35.7	35.1	34.8	28.0	26.9	26.6	26.1
	200	28.8	26.3	25.3	24.7	18.6	15.7	15.2	15.0
	400	20.5	17.0	16.0	15.3	11.0	8.6	8.2	7.8
	800	14.8	10.3	9.4	8.8	5.5	3.7	3.4	3.1
M10	100	8.9	7.9	7.4	7.1	5.5	5.0	4.8	4.8
	200	7.8	5.6	5.1	4.9	4.2	3.2	3.1	3.0
	400	5.0	3.3	2.9	2.6	2.4	1.5	1.4	1.3
	800	4.1	2.3	2.1	1.9	1.5	0.9	0.8	0.7

tor $(k/m)^{-\eta}$ comes into effect and the opposite tends to occurs (although the difference in power does not seem of practical importance for the monitoring

TABLE 5.2

Percentages of rejection of H_0 in (1.1) for the procedure based on T_m with $\gamma = 0$ and $\eta \in \{0.1, 0.05, 0.01, 0.001\}$, as well as for the procedures based on E_m and Q_m with $\gamma = 0$. The rejection percentages are computed from 5000 samples of size $n = m + 10000$ generated from the time series models M1 – M10.

		T_m with $\eta =$					
Model	m	0.05	0.01	0.005	0.001	E_m	Q_m
M1	100	6.9	5.4	5.4	5.4	6.1	6.7
	200	4.6	3.7	3.6	3.5	4.9	5.1
	400	3.3	2.3	2.1	2.0	4.7	5.0
	800	2.6	1.4	1.2	1.1	3.1	3.9
M2	100	7.2	6.1	5.9	6.0	6.3	6.6
	200	4.9	3.8	3.7	3.6	4.8	5.2
	400	3.5	2.3	2.2	2.1	4.8	5.1
	800	2.6	1.4	1.2	1.1	3.2	3.9
M3	100	8.4	7.4	7.4	7.3	6.7	7.2
	200	5.4	4.2	4.1	4.0	4.9	5.6
	400	3.7	2.5	2.4	2.3	4.9	5.2
	800	2.7	1.5	1.2	1.2	3.3	4.0
M4	100	10.7	9.7	9.6	9.6	7.5	7.8
	200	6.3	5.0	4.8	4.7	5.3	6.0
	400	4.0	2.7	2.5	2.5	5.0	5.4
	800	2.8	1.7	1.5	1.4	3.3	4.0
M5	100	15.5	14.8	14.5	14.6	9.6	9.7
	200	8.9	7.2	7.0	6.9	6.6	7.2
	400	5.0	3.5	3.4	3.3	5.3	6.0
	800	3.1	2.0	1.8	1.7	3.7	4.4
M6	100	8.2	6.9	6.8	6.8	6.8	6.9
	200	5.7	4.4	4.2	4.1	5.2	5.6
	400	3.7	2.7	2.4	2.3	4.7	5.2
	800	2.7	1.6	1.4	1.4	3.4	3.8
M7	100	9.3	8.3	8.1	7.9	6.7	6.8
	200	6.1	5.0	4.8	4.8	5.3	5.5
	400	3.8	2.6	2.5	2.5	4.4	4.8
	800	2.9	1.6	1.5	1.4	3.4	3.9
M8	100	7.9	6.8	6.7	6.6	6.6	7.1
	200	5.1	3.9	3.8	3.7	5.1	5.5
	400	3.9	2.4	2.1	2.1	4.8	5.1
	800	2.8	1.5	1.4	1.3	3.7	4.1
M9	100	31.1	29.9	29.6	29.6	15.2	13.0
	200	22.6	19.6	19.1	18.8	11.7	10.4
	400	14.4	11.5	11.0	10.7	9.6	8.7
	800	9.0	5.8	5.4	5.1	7.5	7.2
M10	100	6.6	5.6	5.6	5.6	6.4	6.5
	200	5.1	4.3	4.2	4.0	5.0	5.4
	400	3.5	2.2	2.0	2.0	5.0	5.2
	800	2.4	1.3	1.2	1.1	3.5	3.6

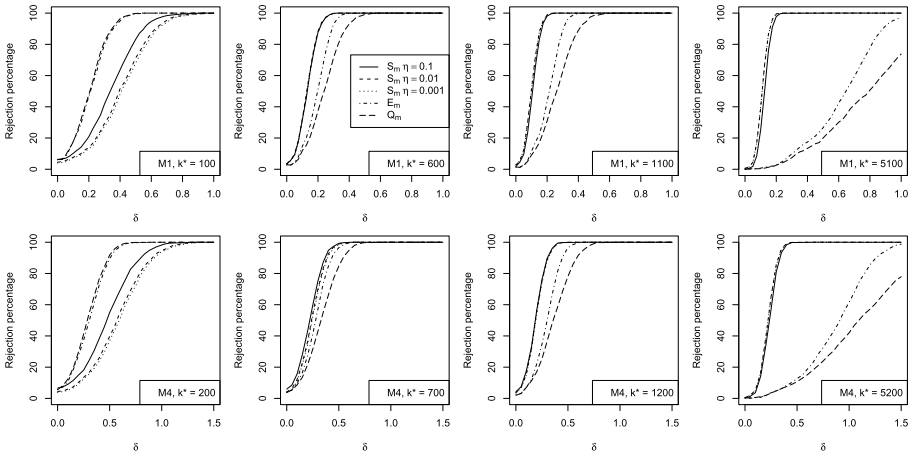


FIG 5.1. *Rejection percentages of H_0 in (1.1) for the procedure based on S_m with $\gamma = 0$ and $\eta \in \{0.1, 0.05, 0.01, 0.001\}$, as well as for the procedures based on E_m and Q_m with $\gamma = 0$, estimated from 1000 samples of size $n = m + 7000$ from model M1 with $m = 100$ or M4 with $m = 200$ such that, for each sample, a positive offset of δ was added to all observations after position $k^\star$.*

periods under consideration). As far as the procedures based on E_m and Q_m are concerned, they are more powerful than the procedure based on S_m when the change occurs at the beginning of the monitoring but, as expected, become less powerful as $k^\star$ increases. The fourth column of plots in Figure 5.1 shows that the difference in terms of power can be substantial. One should however keep in mind that, because of the factor $(k/m)^{-\eta}$, the test based on E_m for instance might again become more powerful than the procedure based on S_m for $k^\star$ extremely large. Nevertheless, even if we consider the least favorable setting ($\eta = 0.1$), we believe that such a scenario is extremely unlikely to occur in practice given the (relatively) slow decrease of the function $t \mapsto t^{-\eta}$ and the fact that the weighting used in the definition of E_m penalizes in some sense late changes as explained in the discussion below (2.5).

From the second row of plots of Figure 5.1, we see that the previous conclusions seem to remain qualitatively the same when model M4 is used although, unsurprisingly, the stronger serial dependence gives the impression that the values of $k^\star$ are smaller.

Figure 5.2 reports the rejection percentages of H_0 in (1.1) for the same experiments but for the procedures based on R_m, S_m, T_m with $\gamma = 0$ and $\eta = 0.005$, as well as for the procedures based on E_m and Q_m with $\gamma = 0$. As one can see, among the three studied procedures, those based on R_m and T_m seem more powerful than the one based S_m when the change occurs at the beginning of the monitoring (which is in accordance with the fact that the procedure based on S_m is the most conservative as can be seen from Tables 5.1 and 5.2), while there seems to be very little difference between the three tests when the change

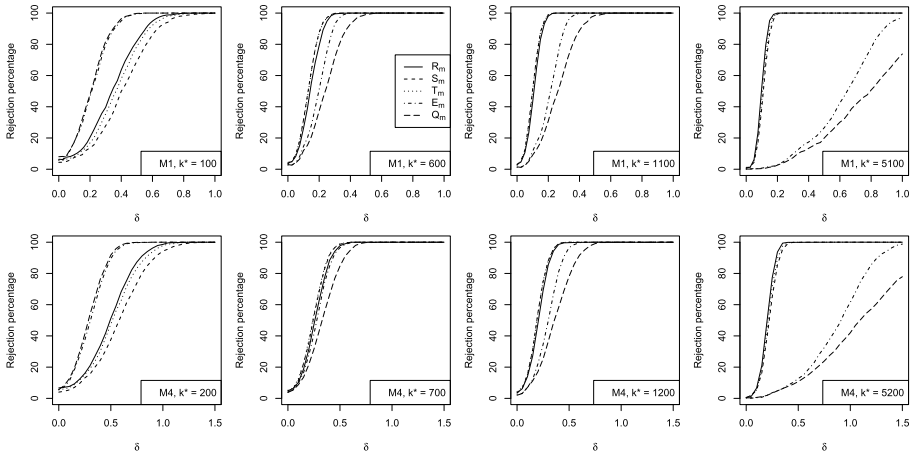


FIG 5.2. *Rejection percentages of H_0 in (1.1) for the procedures based on R_m, S_m, T_m with $\gamma = 0$ and $\eta = 0.005$, as well as for the procedures based on E_m and Q_m with $\gamma = 0$, estimated from 1000 samples of size $n = m + 7000$ from model M1 with $m = 100$ or M4 with $m = 200$ such that, for each sample, a positive offset of δ was added to all observations after position $k^\star$.*

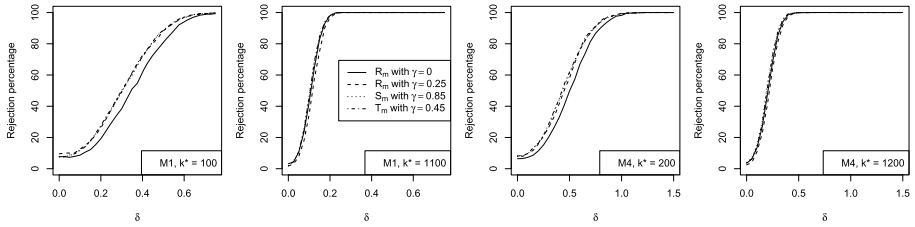


FIG 5.3. *For $\eta = 0.001$, rejection percentages of H_0 in (1.1) for the procedures based on R_m with $\gamma = 0$, R_m with $\gamma = 0.25$, S_m with $\gamma = 0.85$ and T_m with $\gamma = 0.45$ estimated from 1000 samples of size $n = m + 7000$ from model M1 with $m = 100$ or M4 with $m = 200$ such that, for each sample, a positive offset of δ was added to all observations after position $k^\star$.*

occurs later.

The increase in power resulting from taking γ equal to its largest meaningful value is illustrated in Figure 5.3. As expected, the improvement is visible only when changes occur at the beginning of the monitoring. In practice, we suggest to increase the value of γ only when it is believed that the size m of the learning sample permits a reasonably accurate estimation of the long-run variance σ^2.

As a second to last experiment, we considered a longer monitoring period with late changes. Table 5.3 provides the percentages of rejection of H_0 in (1.1) for the procedures based on the detectors R_m, S_m, T_m with $\gamma = 0$ and $\eta \in \{0.1, 0.05, 0.01, 0.005, 0.001\}$ estimated from 2000 samples of size $n = 20000$ from model M1 with $m = 100$ such that, for each sample, a positive offset of

TABLE 5.3

Percentages of rejection of H_0 in (1.1) for the procedures based on R_m, S_m, T_m with $\gamma = 0$ and $\eta \in \{0.1, 0.05, 0.01, 0.005, 0.001\}$ estimated from 2000 samples of size $n = 20000$ from model M1 with $m = 100$ such that, for each sample, a positive offset of $\delta = 0.1$ was added to all observations after position $k^\star = 15000$. The corresponding rejection percentages of the procedures based on E_m and Q_m are 0.7 and 0.7, respectively.

	$\eta = 0.1$	$\eta = 0.05$	$\eta = 0.01$	$\eta = 0.005$	$\eta = 0.001$
R_m	62.3	88.9	95.5	95.6	95.7
S_m	64.4	79.7	87.1	87.6	87.7
T_m	66.3	84.9	92.1	92.5	92.8

$\delta = 0.1$ was added to all observations after position $k^\star = 15000$. These results highlight again the role of η and its influence on power through the factor $(k/m)^{-\eta}$. Notice that the corresponding rejection percentages of the procedures based E_m and Q_m are less than 1%.

Finally, we report the results of a similar experiment illustrating the influence of a change in the serial dependence in the data generating mechanism on the power of the tests. Specifically, we considered two additional models: for Model M11 (resp. M12), observations consist of a realization of a sequence of m independent standard normals followed by observations generated from an AR(1) model with standard normal innovations and autoregressive parameter evolving linearly from 0 to 0.9 (resp. 0 to -0.9) over the $n - m$ remaining observations. Table 5.4 provides the percentages of rejection of H_0 in (1.1) for the procedures based on R_m, S_m, T_m with $\gamma = 0$ and $\eta = 0.001$, as well as for the procedures E_m and Q_m with $\gamma = 0$, estimated from 2000 samples of size $n = 20000$ from model M1, M11 or M12 with $m = 100$ such that, for each sample, a positive offset of $\delta \in \{0, 0.5, 0.1\}$ was added to all observations after position $k^\star = 15000$. As one can see from the first horizontal block of the table (corresponding to $\delta = 0$), the tests based on R_m, S_m and T_m reject very frequently H_0 in (1.1) when data are generated from Model M11 although the underlying (non-stationary) time series has mean zero. The latter is actually not surprising. On one hand, the variance estimate σ_m involved in the procedure (see the practical description of the tests below Corollary 3.6) was computed from the initial i.i.d. learning sample of size m and should therefore not be too different from 1. On the other hand, the progressive increase of the autoregressive parameter in Model M11 entails an increase of the variance of the observations as time elapses and makes it more and more probable that large observations in absolute value are followed by large observations in absolute values of the same sign. The two previous facts lead to an increase, as time elapses, of the probability that the detectors will exceed their corresponding thresholds. The high rejection percentages observed when data are generated from Model M11 confirm that the null hypothesis of the studied procedures is not that the underlying time series has mean zero but that it is stationary as stated in (1.1). The closeness of the rejection percentages to the 5% nominal level when data are generated from Model M12 is a consequence of the fact that, as time elapses, large observations in absolute value tend to be followed by large observations in absolute values but of the opposite

TABLE 5.4

Percentages of rejection of H_0 in (1.1) for the procedures based on R_m, S_m, T_m with $\gamma = 0$ and $\eta = 0.001$, as well as for the procedures E_m and Q_m with $\gamma = 0$, estimated from 2000 samples of size $n = 20000$ from model M1, M11 or M12 with $m = 100$ such that, for each sample, a positive offset of $\delta \in \{0, 0.05, 0.1\}$ was added to all observations after position $k^\star = 15000$.

δ	Model	R_m	S_m	T_m	E_m	Q_m
0	M1	8.8	5.5	7.0	6.3	6.7
	M11	99.1	88.9	92.5	8.2	8.1
	M12	5.3	3.1	4.2	6.7	6.8
0.05	M1	23.3	17.5	19.6	0.3	0.2
	M11	98.2	86.3	90.5	1.1	1.0
	M12	8.2	10.2	10.9	0.3	0.2
0.1	M1	95.7	87.7	92.8	0.7	0.7
	M11	99.5	92.2	94.4	1.6	1.2
	M12	98.0	92.2	96.3	1.0	0.4

sign. As already explained when describing the first series of experiments, the fact that the rejection percentages are slightly above the 5% nominal level when data are generated from Model M1 is due to the relatively small learning sample ($m = 100$). We expect the empirical levels to improve for larger values of m as in Tables 5.1 and 5.2. The last two horizontal blocks of Table 5.4 (corresponding to $\delta \in \{0.05, 0.1\}$) confirm that, even in the presence of a change in the serial dependence, the studied procedures have some power when a change in mean occurs.

Taking into account all the empirical results summarized in this section, we recommend to set η to 0.005 or 0.001 and use either the procedure based on T_m or the procedure based on S_m as they are more conservative than the monitoring scheme based on R_m while almost as powerful (except when changes occur at the beginning of the monitoring). If it is believed that the size m of the learning sample permits a reasonably accurate estimation of the long-run variance σ^2, to improve the behavior of the procedures at the beginning of the monitoring, one can additionally set γ to 0.45 for T_m and 0.85 for S_m.

6. Extensions to parameters whose estimators exhibit a mean-like behavior

In their seminal work, Gösmann, Kley and Dette (2021) actually considered monitoring schemes sensitive to changes in time series parameters whose estimators exhibit an asymptotic mean-like behavior. The aim of this section is to briefly demonstrate that the same type of extension is possible for the sequential procedures studied in this work. For the sake of keeping the notation simple, we restrict our discussion to univariate time series and univariate parameters.

Let F be a univariate distribution function (d.f.) and let $\theta = \theta(F)$ be a univariate parameter of F (such as the expectation, the variance, etc). Let $F_{j:k}$ be the empirical d.f. computed from the stretch $X_j, \ldots, X_k$ of available observa-

tions. More formally, for any integers $j, k \geq 1$ and $x \in \mathbb{R}$, let

$$F_{j:k}(x) = \begin{cases} \dfrac{1}{k-j+1} \displaystyle\sum_{i=j}^{k} \mathbf{1}(X_i \leq x), & \text{if } j \leq k, \\ 0, & \text{otherwise,} \end{cases}$$

and let $\theta_{j:k} = \theta(F_{j:k})$ be the corresponding plug-in estimator of θ computed from the stretch $X_j, \ldots, X_k$. Natural extensions of the detectors R_m, S_m and T_m defined in (2.4), (2.9) and (2.10), respectively, for monitoring changes in the parameter θ are then given, for any $k \geq m+1$, by

$$R_m^{\theta}(k) = \max_{m \leq j \leq k-1} \frac{j(k-j)}{m^{3/2}} |\theta_{1:j} - \theta_{j+1:k}|, \tag{6.1}$$

$$S_m^{\theta}(k) = \frac{1}{m} \sum_{j=m}^{k-1} \frac{j(k-j)}{m^{3/2}} |\theta_{1:j} - \theta_{j+1:k}|, \tag{6.2}$$

$$T_m^{\theta}(k) = \sqrt{\frac{1}{m} \sum_{j=m}^{k-1} \left\{ \frac{j(k-j)}{m^{3/2}} (\theta_{1:j} - \theta_{j+1:k}) \right\}^2}. \tag{6.3}$$

Furthermore, assuming it exists, let

$$I\!F(x, F, \theta) = \lim_{\varepsilon \downarrow 0} \frac{\theta\{(1-\varepsilon)F + \varepsilon \delta_x\} - \theta(F)}{\varepsilon}$$

denote the influence function related to θ and F at $x \in \mathbb{R}$, where $\delta_x(\cdot) = \mathbf{1}(x \leq \cdot)$ is the d.f. of the Dirac measure at x. To be able to study the asymptotic validity of monitoring schemes based on R_m^{θ} in (6.1), S_m^{θ} in (6.2) and T_m^{θ} in (6.3), we follow Gösmann, Kley and Dette (2021) and focus on parameters θ that admit an asymptotic linearization in terms of the influence function, that is, such that

$$\theta_{j:k} - \theta = \theta(F_{j:k}) - \theta(F) = \frac{1}{k-j+1} \sum_{i=j}^{k} I\!F(X_i, F, \theta) + R_{j,k}, \tag{6.4}$$

where the remainders $R_{j,k}$ are asymptotically negligible in the sense of the following condition.

Condition 6.1. *The remainders in* (6.4) *satisfy*

$$k^{-1/2} \max_{1 \leq i < j \leq k} (j-i+1) R_{i,j} \xrightarrow{as} 0 \qquad \text{as } k \to \infty,$$

where the arrow ' $\xrightarrow{as}$ *' denotes almost sure convergence.*

In the rest of this section, we assume that H_0 in (1.1) holds. Moreover, for any integers $j, k \geq 1$, let

$$\bar{I\!F}_{j:k} = \begin{cases} \dfrac{1}{k-j+1} \displaystyle\sum_{i=j}^{k} I\!F(X_i, F, \theta), & \text{if } j \leq k, \\ 0, & \text{otherwise.} \end{cases}$$

If the random variables $I\!F(X_1, F, \theta), \ldots, I\!F(X_m, F, \theta), I\!F(X_{m+1}, F, \theta), \ldots$ were observable, one could naturally consider analogues of the detectors R_m, S_m and T_m in (2.4), (2.9) and (2.10), respectively, defined, for $k \geq m+1$, by

$$R_m^{I\!F}(k) = \max_{m \leq j \leq k-1} \frac{j(k-j)}{m^{3/2}} |\bar{I\!F}_{1:j} - \bar{I\!F}_{j+1:k}|,$$

$$S_m^{I\!F}(k) = \frac{1}{m} \sum_{j=m}^{k-1} \frac{j(k-j)}{m^{3/2}} |\bar{I\!F}_{1:j} - \bar{I\!F}_{j+1:k}|,$$

$$T_m^{I\!F}(k) = \sqrt{\frac{1}{m} \sum_{j=m}^{k-1} \left\{ \frac{j(k-j)}{m^{3/2}} (\bar{I\!F}_{1:j} - \bar{I\!F}_{j+1:k}) \right\}^2}.$$

Upon assuming that Condition 3.1 holds for the sequence $\big(I\!F(X_i, F, \theta)\big)_{i \in \mathbb{Z}}$, one immediately obtains an analogue of Theorem 3.3 for the detectors $R_m^{I\!F}$, $S_m^{I\!F}$ and $T_m^{I\!F}$. The next result, proven in Appendix C, shows that, if Condition 6.1 is additionally assumed, an analogue of Theorem 3.3 also holds for the computable detectors R_m^θ in (6.1), S_m^θ in (6.2) and T_m^θ in (6.3).

Proposition 6.2. *Under H_0 in (1.1) and Condition 6.1, for any $\eta > 0$, $\epsilon > 0$ and $\gamma \geq 0$,*

$$\sup_{m+1 \leq k \leq \infty} \{w_R(k/m)\}^{-1} |R_m^\theta(k) - R_m^{I\!F}(k)| = o_{\mathbb{P}}(1),$$

$$\sup_{m+1 \leq k \leq \infty} \{w_S(k/m)\}^{-1} |S_m^\theta(k) - S_m^{I\!F}(k)| = o_{\mathbb{P}}(1),$$

$$\sup_{m+1 \leq k \leq \infty} \{w_T(k/m)\}^{-1} |T_m^\theta(k) - T_m^{I\!F}(k)| = o_{\mathbb{P}}(1),$$

where the threshold functions w_R, w_S and w_T are defined in (2.7), (2.11) and (2.12), respectively.

Remark 6.3. As mentioned in Gösmann, Kley and Dette (2021), the verification of Condition 6.1 is highly non-trivial. When θ is the variance or a quantile of F, it was shown to hold in probability in Section 4 of Dette and Gösmann (2019). In the multivariate parameter and time series case, it was verified in Gösmann, Kley and Dette (2021, Section 3.2) for a time-dependent linear model. In a related way, note that an inspection of the proof of Proposition 6.2 reveals that Condition 6.1 could actually be replaced by the requirement that the remainders in (6.4) satisfy

$$\sup_{m+1 \leq k < \infty} k^{-1/2} \max_{1 \leq i < j \leq k} (j - i + 1) |R_{i,j}| = o_{\mathbb{P}}(1).$$

7. Data example

As a small data example, we consider a fictitious scenario consisting of monitoring global temperature anomalies for changes in the mean. Specifically, we use the time series of monthly global (land and ocean) temperature anomalies

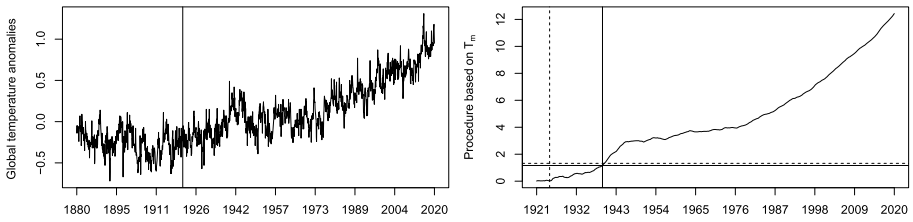


FIG 7.1. *Left: monthly global (land and ocean) temperature anomalies in degrees Celsius for the period January 1880 – May 2020. The solid vertical line corresponds to September 1921 and marks the beginning of the fictitious monitoring. Right: the solid curve displays the normalized detector* $\sigma_m^{-1}\{w_T(k/m)\}^{-1}T_m(k)$ *with* $\eta = 0.001$ *and* $\gamma = 0.45$ *against* $k \geq m+1$. *The solid (resp. dashed) horizontal line represents the estimated 0.95-quantile (resp. 0.99-quantile) of the corresponding limiting distribution in Theorem 3.3. The solid (resp. dashed) vertical line corresponds to November 1939, the date of exceedence (resp. April 1925, the estimated date of change).*

available at http://www.climate.gov/ which covers the period January 1880 – May 2020. The time series in degrees Celsius is represented in the left panel of Figure 7.1. The solid vertical line marks the beginning of the fictitious monitoring and corresponds to September 1921 (and thus to a learning sample of size $m = 500$). Note that this monitoring scenario is indeed fully fictitious, among other things, because temperature anomalies are computed with respect to the 20th century average (see, e.g., Smith et al., 2008) and, therefore, the corresponding time series would not have been available until the beginning of the current century.

The solid curve in the right panel of Figure 7.1 displays the evolution of the normalized detector $\sigma_m^{-1}\{w_T(k/m)\}^{-1}T_m(k)$ with $\eta = 0.001$ and $\gamma = 0.45$ against $k \geq m + 1$. The solid (resp. dashed) horizontal line represents the estimated 0.95-quantile (resp. 0.99-quantile) of the corresponding limiting distribution in Theorem 3.3. The solid vertical line represents the date at which the normalized detector exceeded the aforementioned 0.95-quantile and corresponds to November 1939. Note that to estimate the point of change corresponding to an exceedence at position k, one could use

$$\text{argmax}_{m\leq j\leq k-1}\frac{j(k-j)}{m^{3/2}}|\bar{X}_{1:j} - \bar{X}_{j+1:k}| + 1.$$

The date of change corresponding to an exceedence in November 1939 is thus estimated to be April 1925 and is marked by a dashed vertical line in the right panel of Figure 7.1.

8. Concluding remarks

This work has demonstrated that it is relevant to define open-end sequential change-point tests such that the underlying detectors coincide with (or are related to) the retrospective CUSUM statistic at each monitoring step. From a practical perspective, when focusing on changes in the mean, such an approach

was observed to lead to an increase in power with respect to existing competitors except when changes occur at the very beginning of the monitoring. Given the potentially very long term nature of open-end monitoring (by definition), it can be argued that having extra power for all but very early changes is strongly desirable. The price to pay for the additional power is a more complicated theoretical setting and the fact that quantiles of the underlying limiting distributions required to carry out the sequential tests in practice are harder to estimate. As far as monitoring for changes in other parameters than the mean is concerned, extensions are possible as long as the underlying estimators exhibit a mean-like asymptotic behavior as considered in Gösmann, Kley and Dette (2021).

Acknowledgments

The authors would like to thank Ed Perkins for helpful discussions as well as Josua Gösmann and an anonymous Referee for their constructive comments on an earlier version of this manuscript.

Appendix A: Proofs of the results under H_0

The three following lemmas are used in the proof of Theorem 3.3.

Lemma A.1. *Assume that Condition 3.1 holds and, for any $k \geq m+1$, let*

$$\tilde{R}_m(k) = \sigma m^{-1/2} \max_{m \leq j \leq k-1} \left| \frac{k}{m}\{W_{m,2}(m) + W_{m,1}(j-m)\} - \frac{j}{m}\{W_{m,2}(m) + W_{m,1}(k-m)\} \right|, \tag{A.1}$$

$$\tilde{S}_m(k) = \sigma m^{-1/2} \frac{1}{m} \sum_{j=m}^{k-1} \left| \frac{k}{m}\{W_{m,2}(m) + W_{m,1}(j-m)\} - \frac{j}{m}\{W_{m,2}(m) + W_{m,1}(k-m)\} \right|, \tag{A.2}$$

$$\tilde{T}_m(k) = \sigma m^{-1/2} \left(\frac{1}{m} \sum_{j=m}^{k-1} \left[\frac{k}{m}\{W_{m,2}(m) + W_{m,1}(j-m)\} - \frac{j}{m}\{W_{m,2}(m) + W_{m,1}(k-m)\} \right]^2 \right)^{1/2}. \tag{A.3}$$

Then, for any fixed $\eta > 0$, $\epsilon > 0$ and $\gamma \geq 0$,

$$\sup_{m+1 \leq k \leq \infty} \{w_R(k/m)\}^{-1} |R_m(k) - \tilde{R}_m(k)| = o_{\mathbb{P}}(1), \tag{A.4}$$

$$\sup_{m+1 \leq k \leq \infty} \{w_S(k/m)\}^{-1} |S_m(k) - \tilde{S}_m(k)| = o_{\mathbb{P}}(1), \tag{A.5}$$

$$\sup_{m+1 \leq k \leq \infty} \{w_T(k/m)\}^{-1} |T_m(k) - \tilde{T}_m(k)| = o_{\mathbb{P}}(1), \tag{A.6}$$

where w_R, w_S and w_T are defined in (2.7), (2.11) and (2.12), respectively, and R_m, S_m and T_m are defined in (2.4), (2.9) and (2.10), respectively.

Proof. Let us first show (A.4). From (2.4) and (A.1), using the reverse triangle inequality for the maximum norm, we have that, for any $k \geq m+1$, $\{w_R(k/m)\}^{-1}|R_m(k) - \tilde{R}_m(k)| \leq U_m(k)$, where

$$U_m(k) = \epsilon^{-1} m^{-1/2} \left(\frac{k}{m}\right)^{-3/2-\eta} \max_{m \leq j \leq k-1} \left| \frac{j(k-j)}{m} \{\bar{X}_{1:j} - \bar{X}_{j+1:k}\} \right.$$
$$\left. - \frac{k}{m}\sigma\{W_{m,2}(m) + W_{m,1}(j-m)\} + \frac{j}{m}\sigma\{W_{m,2}(m) + W_{m,1}(k-m)\} \right|, \quad \text{(A.7)}$$

using the fact that $t \mapsto 1/w_\gamma(t)$ is bounded by ϵ^{-1}. From (2.3) and under Condition 3.1, we obtain that, for any $k \geq m+1$ and $j \in \{m, \ldots, k-1\}$,

$$\begin{aligned} j(k-j)\{\bar{X}_{1:j} - \bar{X}_{j+1:k}\} =& (k-j)\sum_{i=1}^{j}\{X_i - \mathbb{E}(X_1)\} - j\sum_{i=j+1}^{k}\{X_i - \mathbb{E}(X_1)\} \\ =& k\sum_{i=1}^{j}\{X_i - \mathbb{E}(X_1)\} - j\sum_{i=1}^{k}\{X_i - \mathbb{E}(X_1)\} \\ =& k\sum_{i=1}^{m}\{X_i - \mathbb{E}(X_1)\} + k\sum_{i=m+1}^{j}\{X_i - \mathbb{E}(X_1)\} \\ & - j\sum_{i=1}^{m}\{X_i - \mathbb{E}(X_1)\} - j\sum_{i=m+1}^{k}\{X_i - \mathbb{E}(X_1)\}. \end{aligned}$$

Hence, by the triangle inequality, we have that

$$\sup_{m+1 \leq k \leq \infty} \{w_R(k/m)\}^{-1}|R_m(k) - \tilde{R}_m(k)|$$
$$\leq \sup_{m+1 \leq k \leq \infty} U_m(k) \leq \epsilon^{-1}(I_m + I'_m + J_m + J'_m), \quad \text{(A.8)}$$

where

$$I_m = m^{-1/2} \sup_{m+1 \leq k < \infty} \left(\frac{k}{m}\right)^{-1/2-\eta} \max_{m \leq j \leq k-1} \left| \sum_{i=1}^{m}\{X_i - \mathbb{E}(X_1)\} - \sigma W_{m,2}(m) \right|,$$
$$I'_m = m^{-1/2} \sup_{m+1 \leq k < \infty} \left(\frac{k}{m}\right)^{-1/2-\eta}$$
$$\times \max_{m \leq j \leq k-1} \left| \sum_{i=m+1}^{j}\{X_i - \mathbb{E}(X_1)\} - \sigma W_{m,1}(j-m) \right|,$$
$$J_m = m^{-1/2} \sup_{m+1 \leq k < \infty} \left(\frac{k}{m}\right)^{-3/2-\eta}$$

$$\times \max_{m \leq j \leq k-1} \frac{j}{m} \left| \sum_{i=1}^{m} \{X_i - \mathbb{E}(X_1)\} - \sigma W_{m,2}(m) \right|, \tag{A.9}$$

$$J'_m = m^{-1/2} \sup_{m+1 \leq k < \infty} \left(\frac{k}{m}\right)^{-3/2-\eta} \times \max_{m \leq j \leq k-1} \frac{j}{m} \left| \sum_{i=m+1}^{k} \{X_i - \mathbb{E}(X_1)\} - \sigma W_{m,1}(k-m) \right|. \tag{A.10}$$

To prove (A.4), it remains to show that I_m, I'_m, J_m and J'_m converge to zero in probability. For any $\xi \in (0, 1/2)$,

$$I_m = \frac{1}{m^{\xi}} \left| \sum_{i=1}^{m} \{X_i - \mathbb{E}(X_1)\} - W_{m,2}(m) \right| m^{\xi+\eta} \sup_{m+1 \leq k < \infty} k^{-1/2-\eta},$$

which, under Condition 3.1, is smaller than a term bounded in probability times $m^{\xi-1/2}$, which converges to zero because $\xi < 1/2$. Similarly,

$$I'_m \leq m^{\eta} \sup_{m+1 \leq k < \infty} k^{-1/2-\eta} \max_{m \leq j \leq k-1} (j-m)^{\xi} \times \sup_{m+1 \leq \ell < \infty} \frac{1}{(\ell-m)^{\xi}} \left| \sum_{i=m+1}^{\ell} \{X_i - \mathbb{E}(X_1)\} - \sigma W_{m,1}(\ell-m) \right|.$$

Using again Condition 3.1, the latter converges to zero in probability since

$$m^{\eta} \sup_{m+1 \leq k < \infty} k^{-1/2-\eta} \max_{m \leq j \leq k-1} (j-m)^{\xi} \leq m^{\eta} \sup_{m+1 \leq k < \infty} k^{-1/2-\eta+\xi} \leq m^{\xi-1/2} \to 0.$$

The fact that J_m in (A.9) and J'_m in (A.10) converge to zero in probability can be checked by proceeding similarly. Hence, we have shown (A.4). It remains to prove (A.5) and (A.6).

From (2.9) and (A.2), using the reverse triangle inequality for the L_1 norm, it can be verified that, for any $k \geq m+1$,

$$\{w_S(k/m)\}^{-1} |S_m(k) - \tilde{S}_m(k)| \leq \epsilon^{-1} m^{-1/2} \left(\frac{k}{m}\right)^{-5/2-\eta} \frac{1}{m} \sum_{j=m}^{k-1} \left| \frac{j(k-j)}{m} \{\bar{X}_{1:j} - \bar{X}_{j+1:k}\} - \frac{k}{m} \sigma \{W_{m,2}(m) + W_{m,1}(j-m)\} + \frac{j}{m} \sigma \{W_{m,2}(m) + W_{m,1}(k-m)\} \right| \leq \frac{k-m}{k} U_m(k),$$

where $U_m(k)$ is defined in (A.7). Similarly, from (2.10) and (A.3), using the reverse triangle inequality for the Euclidean norm,

$$\{w_T(k/m)\}^{-1}|T_m(k)-\tilde{T}_m(k)|$$
$$\leq \epsilon^{-1}m^{-1/2}\left(\frac{k}{m}\right)^{-2-\eta}\left(\frac{1}{m}\sum_{j=m}^{k-1}\left[\frac{j(k-j)}{m}\{\bar{X}_{1:j}-\bar{X}_{j+1:k}\}\right.\right.$$
$$\left.\left.-\frac{k}{m}\sigma\{W_{m,2}(m)+W_{m,1}(j-m)\}+\frac{j}{m}\sigma\{W_{m,2}(m)+W_{m,1}(k-m)\}\right]^2\right)^{1/2}$$
$$\leq \sqrt{\frac{k-m}{k}}U_m(k).$$

The fact that (A.5) and (A.6) hold then immediately follows from the two previous displays, (A.8) and the fact that I_m, I'_m, J_m and J'_m converge to zero in probability. □

Lemma A.2. *For any fixed $\eta>0$, the random function R_η defined by*

$$R_\eta(s,t)=\frac{1}{t^{3/2+\eta}}|(t-s)W_2(1)+tW_1(s-1)-sW_1(t-1)|,\qquad 1\leq s\leq t<\infty, \tag{A.11}$$

where W_1 and W_2 are independent standard Brownian motions, is almost surely bounded and uniformly continuous.

Proof. Fix $\eta>0$ and let us first verify that R_η is almost surely bounded on $A=\{(s,t)\in[1,\infty)^2 : s\leq t\}$. By the triangle inequality, $\sup_{1\leq s\leq t<\infty}R_\eta(s,t)$ is smaller than

$$\begin{aligned}
&\sup_{1\leq s\leq t<\infty}t^{-3/2-\eta}(t-s)|W_2(1)|+\sup_{1\leq s\leq t<\infty}t^{-1/2-\eta}|W_1(s-1)|\\
&\quad+\sup_{1\leq s\leq t<\infty}t^{-3/2-\eta}s|W_1(t-1)|\\
&\leq|W_2(1)|\sup_{1\leq s\leq t<\infty}t^{-1/2-\eta}+\sup_{1\leq s<\infty}s^{-1/2-\eta}|W_1(s-1)|\\
&\quad+\sup_{1\leq t<\infty}t^{-1/2-\eta}|W_1(t-1)|\\
&\leq|W_2(1)|+2\sup_{1\leq t<\infty}t^{-1/2-\eta}|W_1(t-1)|.
\end{aligned}$$

It thus remains to verify that $\sup_{1\leq t<\infty}t^{-1/2-\eta}|W_1(t-1)|$ is almost surely bounded. From the law of the iterated logarithm for Brownian motion, we have that, almost surely,

$$\begin{aligned}
\limsup_{t\to\infty}t^{-1/2-\eta}|W_1(t-1)|&=\lim_{T\to\infty}\sup_{T\leq t<\infty}t^{-1/2-\eta}|W_1(t-1)|\\
&=\lim_{T\to\infty}\sup_{T\leq t<\infty}(t+1)^{-1/2-\eta}|W_1(t)|\\
&\leq\lim_{T\to\infty}\sup_{T\leq t<\infty}(2t\log\log t)^{-1/2}|W_1(t)|\\
&\quad\times\lim_{T\to\infty}\sup_{T\leq t<\infty}(2t\log\log t)^{1/2}(t+1)^{-1/2-\eta}
\end{aligned}$$

$$=1 \times \lim_{t\to\infty} (2t \log\log t)^{1/2}(t+1)^{-1/2-\eta} = 0.$$

Hence, on one hand, there exists $T \in (1,\infty)$ such that $\sup_{T\le t<\infty} t^{-1/2-\eta}|W_1(t-1)| \le 1$ almost surely. On the other hand, since $t \mapsto t^{-1/2-\eta}|W_1(t-1)|$ is a process whose sample paths are almost surely continuous, $\sup_{1\le t\le T} < t^{-1/2-\eta}|W_1(t-1)| < \infty$ with probability one.

It remains to prove that R_η is almost surely uniformly continuous on $A = \{(s,t) \in [1,\infty)^2 : s \le t\}$. Let $\varepsilon > 0$. We must show that there exists some (random) $\delta > 0$ such that, for all $(s,t),(s',t') \in A$ such that $d((s,t),(s',t')) < \delta$, $|R_\eta(s,t) - R_\eta(s',t')| \le \varepsilon$.

By the law of the iterated logarithm, there exists $T_\varepsilon \ge 2$ random such that

$$\sup_{u\ge T_\varepsilon} \frac{|W_1(u)|}{u^{1/2+\eta}} \le \frac{\varepsilon}{8}, \tag{A.12}$$

and since $W_2(1)$ is almost surely finite we can choose T_ε to also satisfy

$$\frac{|W_2(1)|}{(T_\varepsilon)^{1/2+\eta}} \le \frac{\varepsilon}{8}. \tag{A.13}$$

Since $[1, T_\varepsilon + 3]$ is compact, there exists $T'_\varepsilon > T_\varepsilon + 5$ (also random) such that

$$\sup_{u\in[1,T_\varepsilon+3]} \frac{|W_1(u)|}{(T'_\varepsilon)^{1/2+\eta}} < \frac{\varepsilon}{8}. \tag{A.14}$$

We will consider the following three subsets of A whose union is A:

$$\begin{aligned} A_0 &= \{(s,t) \in A : s \ge T_\varepsilon + 2\},\\ A_1 &= \{(s,t) \in A : t \le T'_\varepsilon + 1\},\\ A_2 &= \{(s,t) \in A : s \le T_\varepsilon + 2, t \ge T'_\varepsilon + 1\}, \end{aligned}$$

and find a single $\delta \le 1$ that works no matter which of the above (s,t) is in.

Let $(s,t) \in A_0$, and let (s',t') be such that $d((s,t),(s',t')) \le 1$ (so $s' \ge T_\varepsilon + 1$). Then,

$$|R_\eta(s,t) - R_\eta(s',t')| \le \frac{1}{t^{3/2+\eta}} |(t-s)W_2(1) + tW_1(s-1) - sW_1(t-1)| \tag{A.15}$$

$$+ \frac{1}{t'^{3/2+\eta}} |(t'-s')W_2(1) + t'W_1(s'-1) - s'W_1(t'-1)| \tag{A.16}$$

$$\le \frac{2|W_2(1)|}{(T_\varepsilon)^{1/2+\eta}} + \frac{|W_1(s-1)|}{(s-1)^{1/2+\eta}} + \frac{|W_1(s'-1)|}{(s'-1)^{1/2+\eta}} + \frac{|W_1(t-1)|}{(t-1)^{1/2+\eta}} + \frac{|W_1(t'-1)|}{(t'-1)^{1/2+\eta}}.$$

By (A.12) and (A.13), this is at most $6\varepsilon/8$.

Let $(s,t) \in A_2$ and (s',t') be such that $d((s,t),(s',t')) \leq 1$ (so $s' \leq T_\varepsilon + 3$). Then, using (A.15) and (A.16) again, we get that

$$|R_\eta(s,t) - R_\eta(s',t')| \leq \frac{2|W_2(1)|}{(T'_\varepsilon)^{1/2+\eta}} + \frac{|W_1(s-1)|}{(T'_\varepsilon)^{1/2+\eta}} + \frac{|W_1(s'-1)|}{(T'_\varepsilon)^{1/2+\eta}} + \frac{|W_1(t-1)|}{(t-1)^{1/2+\eta}} + \frac{|W_1(t'-1)|}{(t'-1)^{1/2+\eta}}.$$

Using (A.12), (A.14) and (A.13), we get that this is at most $6\varepsilon/8$.

Finally, suppose that $(s,t) \in A_1$. Since the set $A'_1 = \{(s,t) \in A : t \leq T'_\varepsilon + 2\}$ (note that $A'_1 \supset A_1$) is compact and R_η is continuous, it follows that R_η is uniformly continuous on A'_1, so there exists some (random) $\delta_0 > 0$ such that whenever $(s,t),(s',t') \in A'_1$ and $d((s,t),(s',t')) \leq \delta_0$, then we have $|R_\eta(s,t) - R_\eta(s',t')| \leq \varepsilon/2$. Note that if $(s,t) \in A_1$ and $d((s,t),(s',t')) \leq 1$, then (s,t) and (s',t') are both in A'_1.

Let $\delta = \min(\delta_0, 1)$. Then, whenever $d((s,t),(s',t')) \leq \delta$, we have that (s,t) is in (at least) one of A_0, A_1, A_2, so by the above $|R_\eta(s,t) - R_\eta(s',t')| < \varepsilon$. □

Lemma A.3. *For any fixed $\eta > 0$, $\epsilon > 0$ and $\gamma \geq 0$,*

$$\sup_{m+1 \leq k < \infty} \{w_R(k/m)\}^{-1} \tilde{R}_m(k) \rightsquigarrow \sigma \sup_{1 \leq s \leq t < \infty} \{w_\gamma(t)\}^{-1} R_\eta(s,t), \tag{A.17}$$

$$\sup_{m+1 \leq k < \infty} \{w_S(k/m)\}^{-1} \tilde{S}_m(k) \rightsquigarrow \sigma \sup_{1 \leq t < \infty} \{w_\gamma(t)\}^{-1} t^{-1} \int_1^t |R_\eta(s,t)| \mathrm{d}s, \tag{A.18}$$

$$\sup_{m+1 \leq k < \infty} \{w_T(k/m)\}^{-1} \tilde{T}_m(k) \rightsquigarrow \sigma \sup_{1 \leq t < \infty} \{w_\gamma(t)\}^{-1} t^{-1/2} \sqrt{\int_1^t \{R_\eta(s,t)\}^2 \mathrm{d}s}, \tag{A.19}$$

where w_R, w_S and w_T are defined in (2.7), (2.11) and (2.12), respectively, $\tilde{R}_m$, $\tilde{S}_m$ and $\tilde{T}_m$ are defined in (A.1), (A.2) and (A.3), respectively, w_γ is defined in (2.8) and the random function R_η is defined in (A.11). In addition, all the limiting random variables are almost surely finite.

Proof. Fix $\eta > 0$, $\epsilon > 0$ and $\gamma \geq 0$ and let W_1 and W_2 be independent standard Brownian motions. Then, from (A.1), for any $m \in \mathbb{N}$, the random variable $\sup_{m+1 \leq k < \infty} \{w_R(k/m)\}^{-1} \tilde{R}_m(k)$ is equal in distribution to

$$\sigma m^{-1/2} \sup_{m+1 \leq k < \infty} \{w_R(k/m)\}^{-1} \max_{m \leq j \leq k-1} \left| \frac{k}{m} \{W_2(m) + W_1(j-m)\} - \frac{j}{m} \{W_2(m) + W_1(k-m)\} \right|.$$

Next, notice that, for any $k \geq m+1$ and any $j \in \{m, \ldots, k-1\}$, there exists $1 \leq s \leq t$ such that $k = \lfloor mt \rfloor$ and $j = \lfloor ms \rfloor$. Hence, the previous expression can be rewritten as

$$\sigma m^{-1/2} \sup_{t\in[1,\infty)} \{w_R(\lfloor mt\rfloor/m)\}^{-1} \sup_{s\in[1,t]} \left| \frac{\lfloor mt\rfloor}{m}\{W_2(m) + W_1(\lfloor ms\rfloor - m)\} - \frac{\lfloor ms\rfloor}{m}\{W_2(m) + W_1(\lfloor mt\rfloor - m)\}\right|.$$

By Brownian scaling, the latter is equal in distribution to

$$\sigma \sup_{1\le s\le t<\infty} \{w_R(\lfloor mt\rfloor/m)\}^{-1} \left| \frac{\lfloor mt\rfloor}{m}\left\{W_2(1) + W_1\left(\frac{\lfloor ms\rfloor}{m} - 1\right)\right\} - \frac{\lfloor ms\rfloor}{m}\left\{W_2(1) + W_1\left(\frac{\lfloor mt\rfloor}{m} - 1\right)\right\}\right|,$$

which, using (2.7), (2.8) and the function R_η defined in (A.11), can be expressed as

$$\sigma \sup_{1\le s\le t<\infty} \{w_\gamma(\lfloor mt\rfloor/m)\}^{-1} R_\eta(\lfloor ms\rfloor/m, \lfloor mt\rfloor/m).$$

Using additionally the fact that, for any functions f, g,

$$\left|\sup_x |f(x)| - \sup_x |g(x)|\right| \le \sup_x |f(x) - g(x)|, \tag{A.20}$$

we obtain that

$$\begin{aligned}
&\Big| \sup_{1\le s\le t<\infty} \{w_\gamma(\lfloor mt\rfloor/m)\}^{-1} R_\eta(\lfloor ms\rfloor/m, \lfloor mt\rfloor/m) \\
&\qquad\qquad - \sup_{1\le s\le t<\infty} \{w_\gamma(t)\}^{-1} R_\eta(s,t)\Big| \\
&\le \sup_{1\le s\le t<\infty} \left|\{w_\gamma(\lfloor mt\rfloor/m)\}^{-1} R_\eta(\lfloor ms\rfloor/m, \lfloor mt\rfloor/m) - \{w_\gamma(t)\}^{-1} R_\eta(s,t)\right| \\
&\le \sup_{\substack{1\le s\le t<\infty, 1\le s'\le t'<\infty \\ |s-s'|\le 1/m, |t-t'|\le 1/m}} \left|\{w_\gamma(t')\}^{-1} R_\eta(s',t') - \{w_\gamma(t)\}^{-1} R_\eta(s,t)\right|, \qquad \text{(A.21)}
\end{aligned}$$

since $\sup_{t\in[1,\infty)} |\lfloor mt\rfloor/m - t| \le 1/m$.

From Lemma A.2, we know that R_η is almost surely bounded and uniformly continuous on $\{(s,t) \in [1,\infty)^2 : s \le t\}$. Furthermore, the function $t \mapsto 1/w_\gamma(t)$ being bounded, continuous and converging to zero as $t \to \infty$, it is also uniformly continuous on $[1,\infty)$. The latter facts imply that the function $(s,t) \mapsto \{w_\gamma(t)\}^{-1} R_\eta(s,t)$ is almost surely bounded and uniformly continuous on $\{(s,t) \in [1,\infty)^2 : s \le t\}$ and, therefore, that (A.21) converges almost surely to zero and, finally, that (A.17) holds with the limit being almost surely finite.

Let us now show (A.18). First, notice that the limiting random variable is almost surely finite as an immediate consequence of the inequality

$$\sup_{1\le t<\infty} t^{-1} \int_1^t |R_\eta(s,t)| \mathrm{d}s \le \sup_{1\le s\le t<\infty} R_\eta(s,t).$$

Then, from (A.2), for any $m \in \mathbb{N}$, $\sup_{m+1\le k<\infty} \{w_S(k/m)\}^{-1} \tilde{S}_m(k)$ is equal in distribution to

$$\sigma m^{-1/2} \sup_{m+1\le k<\infty} \{w_S(k/m)\}^{-1} \times \frac{1}{m}\sum_{j=m}^{k-1}\left|\frac{k}{m}\{W_2(m)+W_1(j-m)\}-\frac{j}{m}\{W_2(m)+W_1(k-m)\}\right|.$$

Using Brownian scaling, the latter is equal in distribution to

$$\begin{aligned}
&\sigma \sup_{m+1\le k<\infty}\{w_S(k/m)\}^{-1}\frac{1}{m}\sum_{j=m}^{k-1}\left|\frac{k}{m}\{W_2(1)+W_1(j/m-1)\}\right.\\
&\qquad\left.-\frac{j}{m}\{W_2(1)+W_1(k/m-1)\}\right|\\
&=\sigma \sup_{m+1\le\lfloor mt\rfloor<\infty}\left\{w_S\left(\frac{\lfloor mt\rfloor}{m}\right)\right\}^{-1}\\
&\qquad\times\frac{1}{m}\sum_{j=m}^{\lfloor mt\rfloor-1}\left|\frac{\lfloor mt\rfloor}{m}\left\{W_2(1)+W_1\left(\frac{j}{m}-1\right)\right\}\right.\\
&\qquad\left.-\frac{j}{m}\left\{W_2(1)+W_1\left(\frac{\lfloor mt\rfloor}{m}-1\right)\right\}\right|\\
&=\sigma \sup_{1\le t<\infty}\left\{w_\gamma\left(\frac{\lfloor mt\rfloor}{m}\right)\right\}^{-1}\left(\frac{\lfloor mt\rfloor}{m}\right)^{-5/2-\eta}\\
&\qquad\times\int_1^t\left|\frac{\lfloor mt\rfloor}{m}\left\{W_2(1)+W_1\left(\frac{\lfloor ms\rfloor}{m}-1\right)\right\}\right.\\
&\qquad\left.-\frac{\lfloor ms\rfloor}{m}\left\{W_2(1)+W_1\left(\frac{\lfloor mt\rfloor}{m}-1\right)\right\}\right|\mathrm{d}s\\
&=\sigma \sup_{1\le t<\infty}\left\{w_\gamma\left(\frac{\lfloor mt\rfloor}{m}\right)\right\}^{-1}\left(\frac{\lfloor mt\rfloor}{m}\right)^{-1}\int_1^t R_\eta\left(\frac{\lfloor ms\rfloor}{m},\frac{\lfloor mt\rfloor}{m}\right)\mathrm{d}s,
\end{aligned}$$

where the second equality follows from the fact that the integrand is zero on the interval $[\lfloor mt\rfloor/m, t]$ since $\lfloor ms\rfloor = \lfloor mt\rfloor$ for $s \in [\lfloor mt\rfloor/m, t]$. Then, from (A.20), we obtain that

$$\begin{aligned}
&\left|\sup_{1\le t<\infty}\left\{w_\gamma\left(\frac{\lfloor mt\rfloor}{m}\right)\right\}^{-1}\left(\frac{\lfloor mt\rfloor}{m}\right)^{-1}\int_1^t R_\eta\left(\frac{\lfloor ms\rfloor}{m},\frac{\lfloor mt\rfloor}{m}\right)\mathrm{d}s\right.\\
&\qquad\left.-\sup_{1\le t<\infty}\{w_\gamma(t)\}^{-1}t^{-1}\int_1^t R_\eta(s,t)\mathrm{d}s\right|\\
&\le\sup_{1\le t<\infty}\left|\left\{w_\gamma\left(\frac{\lfloor mt\rfloor}{m}\right)\right\}^{-1}\left(\frac{\lfloor mt\rfloor}{m}\right)^{-1}\int_1^t R_\eta\left(\frac{\lfloor ms\rfloor}{m},\frac{\lfloor mt\rfloor}{m}\right)\mathrm{d}s\right.\\
&\qquad\left.-\{w_\gamma(t)\}^{-1}t^{-1}\int_1^t R_\eta(s,t)\mathrm{d}s\right|,
\end{aligned}$$

which is smaller than $I_m + J_m$, where

$$I_m = \sup_{1\leq t<\infty} \left\{ w_\gamma\left(\frac{\lfloor mt\rfloor}{m}\right)\right\}^{-1} \left(\frac{\lfloor mt\rfloor}{m}\right)^{-1} \left| \int_1^t R_\eta\left(\frac{\lfloor ms\rfloor}{m}, \frac{\lfloor mt\rfloor}{m}\right) \mathrm{d}s - \int_1^t R_\eta(s,t)\mathrm{d}s \right|,$$

$$J_m = \sup_{1\leq t<\infty} \left| \left\{ w_\gamma\left(\frac{\lfloor mt\rfloor}{m}\right)\right\}^{-1} \left(\frac{\lfloor mt\rfloor}{m}\right)^{-1} - \{w_\gamma(t)\}^{-1} t^{-1} \right| \int_1^t R_\eta(s,t)\mathrm{d}s.$$

To show (A.18), we shall verify that both I_m and J_m converge to zero almost surely. For I_m, we have

$$I_m \leq \sup_{1\leq t<\infty} \left\{ w_\gamma\left(\frac{\lfloor mt\rfloor}{m}\right)\right\}^{-1} \left(\frac{\lfloor mt\rfloor}{m}\right)^{-1} t \times \sup_{1\leq s\leq t<\infty} \left| R_\eta\left(\frac{\lfloor ms\rfloor}{m}, \frac{\lfloor mt\rfloor}{m}\right) - R_\eta(s,t)\right|.$$

Using (2.8) and the fact that $\sup_{1\leq t<\infty} mt/\lfloor mt\rfloor \leq 2$ as soon as $m \geq 2$, the first supremum on the right is bounded by $2\epsilon^{-1}$. The second supremum converges to zero almost surely from the almost sure uniform continuity of R_η on $\{(s,t) \in [1,\infty)^2 : s \leq t\}$ shown in Lemma A.2. For J_m, we have

$$J_m \leq \sup_{1\leq t<\infty} \left| \left\{ w_\gamma\left(\frac{\lfloor mt\rfloor}{m}\right)\right\}^{-1} \frac{mt}{\lfloor mt\rfloor} - \{w_\gamma(t)\}^{-1} \right| \times \sup_{1\leq s\leq t<\infty} R_\eta(s,t).$$

Since the second supremum on the right is almost surely bounded by Lemma A.2, it suffices to verify that the first supremum converges to zero. The latter is smaller than

$$\begin{aligned} &\sup_{1\leq t<\infty} \left| \left\{ w_\gamma\left(\frac{\lfloor mt\rfloor}{m}\right)\right\}^{-1} \frac{mt}{\lfloor mt\rfloor} \right. \\ &\qquad \left. - \{w_\gamma(t)\}^{-1} \frac{mt}{\lfloor mt\rfloor} + \{w_\gamma(t)\}^{-1} \frac{mt}{\lfloor mt\rfloor} - \{w_\gamma(t)\}^{-1} \right| \\ &\leq 2 \sup_{1\leq t<\infty} \left| \left\{ w_\gamma\left(\frac{\lfloor mt\rfloor}{m}\right)\right\}^{-1} - \{w_\gamma(t)\}^{-1} \right| + \epsilon^{-1} \sup_{1\leq t<\infty} \frac{mt - \lfloor mt\rfloor}{\lfloor mt\rfloor}. \end{aligned}$$

The first term on the right-hand side converges to zero by the uniform continuity of the function $t \mapsto 1/w_\gamma(t)$ on $[1,\infty)$. The second term converges to zero since it is smaller than $\epsilon^{-1}m^{-1}$.

It remains to prove (A.19). Notice first that the limit in (A.19) is almost surely finite since

$$\sup_{1\leq t<\infty} t^{-1/2} \sqrt{\int_1^t \{R_\eta(s,t)\}^2 \mathrm{d}s} \leq \sup_{1\leq s\leq t<\infty} R_\eta(s,t).$$

Then, starting from (A.3) and proceeding as previously, it can be verified that the random variable $\sup_{m+1\leq k<\infty}\{w_T(k/m)\}^{-1}\tilde{T}_m(k)$ is equal in distribution to

$$\sigma \sup_{1\leq t<\infty}\left\{w_T\left(\frac{\lfloor mt\rfloor}{m}\right)\right\}^{-1}\left(\int_1^t\left[\frac{\lfloor mt\rfloor}{m}\left\{W_2(1)+W_1\left(\frac{\lfloor ms\rfloor}{m}-1\right)\right\}\right.\right.$$
$$\left.\left.\frac{\lfloor ms\rfloor}{m}\left\{W_2(1)+W_1\left(\frac{\lfloor mt\rfloor}{m}-1\right)\right\}\right]^2\mathrm{d}s\right)^{1/2}$$
$$=\sigma\sup_{1\leq t<\infty}\left\{w_\gamma\left(\frac{\lfloor mt\rfloor}{m}\right)\right\}^{-1}\left(\frac{\lfloor mt\rfloor}{m}\right)^{-1/2}\sqrt{\int_1^t\left\{R_\eta\left(\frac{\lfloor ms\rfloor}{m},\frac{\lfloor mt\rfloor}{m}\right)\right\}^2\mathrm{d}s}.$$

Then, from (A.20),

$$\left|\sup_{1\leq t<\infty}\left\{w_\gamma\left(\frac{\lfloor mt\rfloor}{m}\right)\right\}^{-1}\left(\frac{\lfloor mt\rfloor}{m}\right)^{-1/2}\sqrt{\int_1^t\left\{R_\eta\left(\frac{\lfloor ms\rfloor}{m},\frac{\lfloor mt\rfloor}{m}\right)\right\}^2\mathrm{d}s}\right.$$
$$\left.-\sup_{1\leq t<\infty}\{w_\gamma(t)\}^{-1}t^{-1/2}\sqrt{\int_1^t\{R_\eta(s,t)\}^2\mathrm{d}s}\right|$$
$$\leq\sup_{1\leq t<\infty}\left|\left\{w_\gamma\left(\frac{\lfloor mt\rfloor}{m}\right)\right\}^{-1}\left(\frac{\lfloor mt\rfloor}{m}\right)^{-1/2}\sqrt{\int_1^t\left\{R_\eta\left(\frac{\lfloor ms\rfloor}{m},\frac{\lfloor mt\rfloor}{m}\right)\right\}^2\mathrm{d}s}\right.$$
$$\left.-\{w_\gamma(t)\}^{-1}t^{-1/2}\sqrt{\int_1^t\{R_\eta(s,t)\}^2\mathrm{d}s}\right|,$$

which is smaller than $I'_m+J'_m$, where

$$I'_m=\sup_{1\leq t<\infty}\left\{w_\gamma\left(\frac{\lfloor mt\rfloor}{m}\right)\right\}^{-1}\left(\frac{\lfloor mt\rfloor}{m}\right)^{-1/2}$$
$$\times\left|\sqrt{\int_1^t\left\{R_\eta\left(\frac{\lfloor ms\rfloor}{m},\frac{\lfloor mt\rfloor}{m}\right)\right\}^2\mathrm{d}s}-\sqrt{\int_1^t\{R_\eta(s,t)\}^2\mathrm{d}s}\right|,$$
$$J'_m=\sup_{1\leq t<\infty}\left|\left\{w_\gamma\left(\frac{\lfloor mt\rfloor}{m}\right)\right\}^{-1}\left(\frac{\lfloor mt\rfloor}{m}\right)^{-1/2}-\{w_\gamma(t)\}^{-1}t^{-1/2}\right|$$
$$\times\sqrt{\int_1^t\{R_\eta(s,t)\}^2\mathrm{d}s}.$$

Using Minkwoski's inequality in the form $|\|f\|_2-\|g\|_2|\leq\|f-g\|_2$, and using similar arguments as previously,

$$I'_m\leq\sup_{1\leq t<\infty}\left\{w_\gamma\left(\frac{\lfloor mt\rfloor}{m}\right)\right\}^{-1}\left(\frac{\lfloor mt\rfloor}{m}\right)^{-1/2}$$

$$\times \sqrt{\int_1^t \left\{ R_\eta\left(\frac{\lfloor ms \rfloor}{m}, \frac{\lfloor mt \rfloor}{m}\right) - R_\eta(s,t) \right\}^2 \mathrm{d}s}$$

$$\leq \epsilon^{-1} \sup_{1 \leq t < \infty} \left(\frac{mt}{\lfloor mt \rfloor} \right)^{1/2} \times \sup_{1 \leq s \leq t < \infty} \left| R_\eta\left(\frac{\lfloor ms \rfloor}{m}, \frac{\lfloor mt \rfloor}{m}\right) - R_\eta(s,t) \right| \overset{as}{\to} 0.$$

Proceeding as for J_m, $J'_m \overset{as}{\to} 0$ since

$$\sup_{1 \leq t < \infty} \left| \left\{ w_\gamma\left(\frac{\lfloor mt \rfloor}{m}\right) \right\}^{-1} \left(\frac{mt}{\lfloor mt \rfloor} \right)^{1/2} - \{w_\gamma(t)\}^{-1} \right|$$

$$\leq \sqrt{2} \sup_{1 \leq t < \infty} \left| \left\{ w_\gamma\left(\frac{\lfloor mt \rfloor}{m}\right) \right\}^{-1} - \{w_\gamma(t)\}^{-1} \right| + \epsilon^{-1} \sup_{1 \leq t < \infty} \frac{\sqrt{mt} - \sqrt{\lfloor mt \rfloor}}{\sqrt{\lfloor mt \rfloor}} \overset{as}{\to} 0$$

by the uniform continuity of the functions $t \mapsto 1/w_\gamma(t)$ and the fact that the argument of the second supremum on the right hand side is at most $\lfloor mt \rfloor^{-1}$. □

Proof of Theorem 3.3. From Lemmas A.1 and A.3, we have that, for any fixed $\eta > 0$, $\epsilon > 0$ and $\gamma \geq 0$,

$$\sup_{m+1 \leq k < \infty} \{w_R(k/m)\}^{-1} R_m(k) \rightsquigarrow \sigma \sup_{1 \leq s \leq t < \infty} \{w_\gamma(t)\}^{-1} R_\eta(s,t),$$

$$\sup_{m+1 \leq k < \infty} \{w_S(k/m)\}^{-1} S_m(k) \rightsquigarrow \sigma \sup_{1 \leq t < \infty} \{w_\gamma(t)\}^{-1} t^{-1} \int_1^t |R_\eta(s,t)| \mathrm{d}s,$$

$$\sup_{m+1 \leq k < \infty} \{w_T(k/m)\}^{-1} T_m(k) \rightsquigarrow \sigma \sup_{1 \leq t < \infty} \{w_\gamma(t)\}^{-1} t^{-1/2} \sqrt{\int_1^t \{R_\eta(s,t)\}^2 \mathrm{d}s},$$

where w_γ is defined in (2.8) and the random function R_η is defined in (A.11).

It thus remains to verify that the expressions of the limiting random variables can be simplified to coincide in distribution with those given in the statement of the theorem. The latter is merely a consequence of the fact that the random functions

$$U(s,t) = (t-s)W_2(1) + tW_1(s-1) - sW_1(t-1), \qquad 1 \leq s \leq t,$$

and

$$V(s,t) = tW(s) - sW(t), \qquad 1 \leq s \leq t,$$

are equal in distribution. Since U and V are centered Gaussian processes whose sample paths are continuous almost surely, the equality in distribution is a direct consequence of the equality of their covariance functions. Indeed, for any $1 \leq s \leq t$ and $1 \leq s' \leq t'$, it is an exercise to verify by direct computation that

$$\mathrm{Cov}\{U(s,t), U(s',t')\} = \mathrm{Cov}\{V(s,t), V(s',t')\}.$$ □

Proof of Proposition 3.4. We have for all $k \geq 1$

$$\sup_{1 \leq s \leq t < \infty} \frac{1}{t^{3/2}} |tW(s) - sW(t)| \geq \frac{1}{(2^k)^{3/2}} |2^k W(2^{k-1}) - 2^{k-1} W(2^k)|$$

$$= \frac{2^{k-1}}{2^{3k/2}} |W(2^{k-1}) - \{W(2^k) - W(2^{k-1})\}|$$
$$= \frac{1}{2^{k/2+1}} |W(2^{k-1}) - \{W(2^k) - W(2^{k-1})\}|.$$

Consider an arbitrary fixed $M > 0$ and the events $D_k = \{|W(2^{k-1}) - \{W(2^k) - W(2^{k-1})\}| \geq M 2^{k/2+1}\}$. It is sufficient to show that $\mathbb{P}\left(\bigcup_{k=1}^{\infty} D_k\right) = 1$ (since this shows that the supremum is at least M with probability 1), or equivalently, $\mathbb{P}\left(\bigcap_{k=1}^{\infty} D_k^{\mathrm{C}}\right) = 0$. Now,

$$\mathbb{P}\left(\bigcap_{k=1}^{\infty} D_k^{\mathrm{C}}\right) = \lim_{r\to\infty} \mathbb{P}\left(\bigcap_{k=1}^{r} D_k^{\mathrm{C}}\right) = \lim_{r\to\infty} \mathbb{P}(D_1^{\mathrm{C}}) \prod_{k=2}^{r} \mathbb{P}\left(D_k^{\mathrm{C}} \,\middle|\, \bigcap_{j=1}^{k-1} D_j^{\mathrm{C}}\right)$$
$$= \mathbb{P}(D_1^{\mathrm{C}}) \lim_{r\to\infty} \prod_{k=2}^{r} \left\{1 - \mathbb{P}\left(D_k \,\middle|\, \bigcap_{j=1}^{k-1} D_j^{\mathrm{C}}\right)\right\},$$

so that it is enough to show that there exists $\delta_M > 0$ such that

$$\mathbb{P}\left(D_k \mid \bigcap_{j=1}^{k-1} D_j^{\mathrm{C}}\right) \geq \delta_M$$

for all $k \geq 2$. The latter holds with $\delta_M = \mathbb{P}(|Z| \geq 2^{3/2} M)/2 > 0$, where Z is a standard normal random variable, since for all $k \geq 2$, with $A_k = \{W(2^k) - W(2^{k-1})$ and $W(2^{k-1})$ have opposite sign$\}$,

$$\mathbb{P}\left(D_k \,\middle|\, \bigcap_{j=1}^{k-1} D_j^{\mathrm{C}}\right) \geq \mathbb{P}\Big(A_k, |W(2^k) - W(2^{k-1})| \geq M 2^{k/2+1}\Big)$$
$$= \mathbb{P}\Big(|W(2^k) - W(2^{k-1})| \geq M 2^{k/2+1}\Big)\mathbb{P}(A_k)$$
$$= \mathbb{P}(|Z| \geq 2^{3/2} M)/2.$$

In the above, we have used the fact that the sign of the increment and its magnitude are independent of the past and each other, and that the increment is Gaussian with mean 0 and variance 2^{k-1}. □

Appendix B: Proofs of the results under alternatives

Proof of Theorem 3.10. Let us first prove the claim when (iv) in Condition 3.7 holds. Recall the definition of the function w_γ in (2.8) and notice that $w_\gamma(t) \leq 1$ for all $t \in [1, \infty)$, $\epsilon \geq 0$ and $\gamma \geq 0$. Thus, for all $\eta \geq 0$, $\epsilon \geq 0$ and $\gamma \geq 0$,

$$\sup_{m+1\leq k<\infty} \{w_R(k/m)\}^{-1} R_m(k) \geq \sup_{m+1\leq k<\infty} (k/m)^{-3/2-\eta} R_m(k)$$

$$
\begin{aligned}
k=k_m^\star+\lfloor cm\rfloor \geq & \left(\frac{k_m^\star+\lfloor cm\rfloor}{m}\right)^{-3/2-\eta} \\
& \times \max_{m\leq j\leq k_m^\star+\lfloor cm\rfloor-1} \frac{j(k_m^\star+\lfloor cm\rfloor-j)}{m^{3/2}}\big|\bar{X}_{1:j}^{(m)}-\bar{X}_{j+1:k_m^\star+\lfloor cm\rfloor}^{(m)}\big| \\
j=k_m^\star \geq & \left(\frac{k_m^\star}{m}+c\right)^{-3/2-\eta}\frac{k_m^\star\lfloor cm\rfloor}{m^{3/2}}\big|\bar{X}_{1:k_m^\star}^{(m)}-\bar{X}_{k_m^\star+1:k_m^\star+\lfloor cm\rfloor}^{(m)}\big| \\
\geq & \left(\frac{k_m^\star}{m}+c\right)^{-3/2-\eta}\frac{k_m^\star(cm-1)}{m^{3/2}}\big|\bar{X}_{1:k_m^\star}^{(m)}-\mathbb{E}(X_{k_m^\star}^{(m)})+\mathbb{E}(X_{k_m^\star}^{(m)}) \\
& -\mathbb{E}(X_{k_m^\star+1}^{(m)})+\mathbb{E}(X_{k_m^\star+1}^{(m)})-\bar{X}_{k_m^\star+1:k_m^\star+\lfloor cm\rfloor}^{(m)}\big| \\
\geq & \, m^{1/2}\left(\frac{k_m^\star}{m}+c\right)^{-3/2-\eta}\left(\frac{k_m^\star}{m}\right)\left(c-\frac{1}{m}\right)\Big\{|\mathbb{E}(X_{k_m^\star+1}^{(m)})-\mathbb{E}(X_{k_m^\star}^{(m)})| \\
& -|\bar{X}_{1:k_m^\star}^{(m)}-\mathbb{E}(X_{k_m^\star}^{(m)})-\bar{X}_{k_m^\star+1:k_m^\star+\lfloor cm\rfloor}^{(m)}+\mathbb{E}(X_{k_m^\star+1}^{(m)})|\Big\} \qquad \text{(B.1)}
\end{aligned}
$$

by the triangle inequality. Since (iv) in Condition 3.7 holds, for m large enough, (B.1) is at least $(C_1+c)^{-3/2-\eta}\,c_1\,(c/2)$ times

$$
m^{1/2}\big\{|\mathbb{E}(Y_1^{(m)})-\mathbb{E}(Y_1^{(0)})|-|\bar{Y}_{1:k_m^\star}^{(0)}-\mathbb{E}(Y_1^{(0)})-\bar{Y}_{k_m^\star+1:k_m^\star+\lfloor cm\rfloor}^{(m)}+\mathbb{E}(Y_1^{(m)})|\big\}. \qquad \text{(B.2)}
$$

Notice that (3.4) implies that $\sqrt{m}\{\bar{Y}_{1:k_m^\star}^{(0)}-\mathbb{E}(Y_1^{(0)})\}=(m/k_m^\star)^{1/2}\times\sqrt{k_m^\star}\{\bar{Y}_{1:k_m^\star}^{(0)}-\mathbb{E}(Y_1^{(0)})\}=O_{\mathbb{P}}(1)$ and that (3.5) implies that $\sqrt{m}\{\bar{Y}_{k_m^\star+1:k_m^\star+\lfloor cm\rfloor}^{(m)}-\mathbb{E}(Y_1^{(m)})\}=O_{\mathbb{P}}(1)$. Hence, (B.2) diverges in probability to infinity as a consequence of (ii) in Condition 3.7.

Assume now that (v) in Condition 3.7 holds. Then, we can use the fact that

$$
\begin{aligned}
& \sup_{m+1\leq k<\infty}\{w_R(k/m)\}^{-1}R_m(k) \\
k=k_m^\star+\lfloor ck_m^\star\rfloor \geq & \left(\frac{k_m^\star+\lfloor ck_m^\star\rfloor}{m}\right)^{-3/2-\eta} R_m(k_m^\star+\lfloor ck_m^\star\rfloor) \\
\geq & \frac{m^{3/2+\eta}}{(k_m^\star+ck_m^\star)^{3/2+\eta}} \\
& \times \max_{m\leq j\leq k_m^\star+\lfloor ck_m^\star\rfloor-1}\frac{j(k_m^\star+\lfloor ck_m^\star\rfloor-j)}{m^{3/2}}\big|\bar{X}_{1:j}^{(m)}-\bar{X}_{j+1:k_m^\star+\lfloor ck_m^\star\rfloor}^{(m)}\big| \\
j=k_m^\star \geq & \frac{m^\eta k_m^\star(ck_m^\star-1)}{(k_m^\star+ck_m^\star)^{3/2+\eta}}\big|\bar{X}_{1:k_m^\star}^{(m)}-\bar{X}_{k_m^\star+1:k_m^\star+\lfloor ck_m^\star\rfloor}^{(m)}\big| \\
\geq & \frac{m^\eta(k_m^\star)^{1/2-\eta}(c-1/k_m^\star)}{(1+c)^{3/2+\eta}}\big|\bar{X}_{1:k_m^\star}^{(m)}-\mathbb{E}(X_{k_m^\star}^{(m)})+\mathbb{E}(X_{k_m^\star}^{(m)}) \\
& -\mathbb{E}(X_{k_m^\star+1}^{(m)})+\mathbb{E}(X_{k_m^\star+1}^{(m)})-\bar{X}_{k_m^\star+1:k_m^\star+\lfloor ck_m^\star\rfloor}^{(m)}\big|,
\end{aligned}
$$

which, for m large enough, is larger than

$$
\frac{m^{1/2}c}{2(1+c)^{3/2+\eta}}\big\{|\mathbb{E}(Y_1^{(m)})-\mathbb{E}(Y_1^{(0)})|-|\bar{Y}_{1:k_m^\star}^{(0)}-\mathbb{E}(Y_1^{(0)})-\bar{Y}_{k_m^\star+1:k_m^\star+\lfloor ck_m^\star\rfloor}^{(m)}+\mathbb{E}(Y_1^{(m)})|\big\}. \qquad \text{(B.3)}
$$

This time, since $k^\star/m \to \infty$, (3.4) implies that

$$\sqrt{m}\{\bar{Y}^{(0)}_{1:k^\star_m} - \mathbb{E}(Y^{(0)}_1)\} = (m/k^\star_m)^{1/2} \times \sqrt{k^\star_m}\{\bar{Y}^{(0)}_{1:k^\star_m} - \mathbb{E}(Y^{(0)}_1)\} = o_{\mathbb{P}}(1)$$

while (3.6) implies that

$$(m/k^\star_m)^{1/2}\sqrt{k^\star_m}\{\bar{Y}^{(m)}_{k^\star_m+1:k^\star_m+\lfloor ck^\star_m \rfloor} - \mathbb{E}(Y^{(m)}_1)\} = o_{\mathbb{P}}(1).$$

Therefore, (B.3) diverges in probability to infinity as a consequence of (*ii*) in Condition 3.7. □

Proof of Theorem 3.12. We adapt the proof of Proposition 2.7 of Kojadinovic and Verdier (2021) to the current setting. Given a set $\mathcal{S}$, let $\ell^\infty(\mathcal{S})$ denote the space of all bounded real-valued functions on $\mathcal{S}$ equipped with the uniform metric. Fix $T > c$. For any $s \in [1, T]$, let

$$W_{m,Y}(s) = m^{-1/2} \sum_{i=1}^{\lfloor ms \rfloor} \{Y_i - \mathbb{E}(Y_1)\}$$

and

$$W_{m,Z}(s) = m^{-1/2} \sum_{i=1}^{\lfloor ms \rfloor} \{Z_i - \mathbb{E}(Z_1)\}.$$

From Condition 3.11, we have that $W_{m,Y}$ converges weakly to a standard Brownian motion W_Y in $\ell^\infty([1, T])$ and $W_{m,Z}$ converges weakly to a standard Brownian motion W_Z in $\ell^\infty([1, T])$.

Let $J_m(s, t) = m^{-1/2}H_m(s, t) - K_c(s, t)$, $(s, t) \in \Delta_T = \{(s, t) \in [1, T]^2 : s \le t\}$. The fact that $m^{-1/2}H_m \xrightarrow{\mathbb{P}} K_c$ in $\ell^\infty(\Delta_T)$ is proven if we show that

$$\sup_{(s,t)\in\Delta_T} |J_m(s, t)| \xrightarrow{\mathbb{P}} 0. \tag{B.4}$$

The supremum on the left-hand side of (B.4) is equal to

$$\max\left\{ \sup_{1\le s\le t\le c} |J_m(s, t)|, \sup_{1\le s\le c\le t\le T} |J_m(s, t)|, \sup_{c\le s\le t\le T} |J_m(s, t)| \right\}. \tag{B.5}$$

Notice first that

$$K_c(s, t) = \begin{cases} 0, & \text{if } 1 \le s \le t \le c, \\ s(t - c)\{\mathbb{E}(Y_1) - \mathbb{E}(Z_1)\}, & \text{if } 1 \le s \le c \le t \le T, \\ c(t - s)\{\mathbb{E}(Y_1) - \mathbb{E}(Z_1)\}, & \text{if } c \le s \le t \le T. \end{cases} \tag{B.6}$$

Furthermore, for any $(s, t) \in \Delta_T \cap [1, c]^2$, let

$$D_{m,Y}(s, t) = \sqrt{m}\lambda_m(s, t)\{\bar{X}^{(m)}_{\lfloor ms \rfloor+1:\lfloor mt \rfloor} - \mathbb{E}(X^{(m)}_{k^\star_m})\} = W_{m,Y}(t) - W_{m,Y}(s),$$

where $\lambda_m(s,t) = (\lfloor mt \rfloor - \lfloor ms \rfloor)/m$, $(s,t) \in \Delta_T$, and, for any $(s,t) \in \Delta_T \cap [c,T]^2$, let

$$D_{m,Z}(s,t) = \sqrt{m}\lambda_m(s,t)\{\bar{X}^{(m)}_{\lfloor ms \rfloor + 1:\lfloor mt \rfloor} - \mathbb{E}(X^{(m)}_{k^\star_m+1})\} = W_{m,Z}(t) - W_{m,Z}(s).$$

Under the assumptions of the theorem, from the continuous mapping theorem, $D_{m,Y} \rightsquigarrow D_Y$ in $\ell^\infty(\Delta_T \cap [1,c]^2)$ and $D_{m,Z} \rightsquigarrow D_Z$ in $\ell^\infty(\Delta_T \cap [c,T]^2)$, where $D_Y(s,t) = W_Y(t) - W_Y(s)$ and $D_Z(s,t) = W_Z(t) - W_Z(s)$.

From the expression of K_c given in (B.6), for the first supremum in (B.5), we obtain that

$$\sup_{1 \le s \le t \le c} |J_m(s,t)| = m^{-1/2} \sup_{1 \le s \le t \le c} |H_m(s,t)| = o(1) \times O_{\mathbb{P}}(1) \xrightarrow{\mathbb{P}} 0,$$

since H_m converges weakly to $(s,t) \mapsto (t-s)D_Y(0,s) - sD_Y(s,t)$ in $\ell^\infty(\Delta_T \cap [1,c]^2)$ as a consequence of the fact that, for any $1 \le s \le t \le c$, $H_m(s,t) = \lambda_m(s,t)D_{m,Y}(0,s) - \lambda_m(0,s)D_{m,Y}(s,t)$ and $\sup_{(s,t)\in\Delta_T} |\lambda_m(s,t) - (t-s)| \le 2/m$, and from the continuous mapping theorem.

Regarding the second supremum, for any $1 \le s \le c \le t \le T$, we have that

$$\lambda_m(s,t)\bar{X}_{\lfloor ms \rfloor + 1:\lfloor mt \rfloor} = \lambda_m(s,c)\bar{X}_{\lfloor ms \rfloor + 1:\lfloor mc \rfloor} + \lambda_m(c,t)\bar{X}_{\lfloor mc \rfloor + 1:\lfloor mt \rfloor}.$$

Thus, on one hand,

$$m^{-1/2}H_m(s,t) = \lambda_m(0,s)\{\lambda_m(s,t)\bar{X}_{1:\lfloor ms \rfloor} - \lambda_m(s,c)\bar{X}_{\lfloor ms \rfloor + 1:\lfloor mc \rfloor} - \lambda_m(c,t)\bar{X}_{\lfloor mc \rfloor + 1:\lfloor mt \rfloor}\}.$$

On the other hand, from (B.6) and the fact that $\sup_{(s,t)\in\Delta_T} |\lambda_m(s,t) - (t-s)| \le 2/m$,

$$K_c(s,t) = \lambda_m(0,s)\{\lambda_m(s,t)\mathbb{E}(Y_1) - \lambda_m(s,c)\mathbb{E}(Y_1) - \lambda_m(c,t)\mathbb{E}(Z_1)\} + O(1/m),$$

where the term $O(1/m)$ is uniform in s,t,c. By the triangle inequality and using the fact that $\sup_{(s,t)\in\Delta_T} |\lambda_m(s,t)| \le T$, it then follows that

$$\sup_{1 \le s \le c \le t \le T} |J_m(s,t)| \le m^{-1/2} T \left[\sup_{1 \le s \le c} |D_{m,Y}(0,s)| + \sup_{1 \le s \le c} |D_{m,Y}(s,c)| + \sup_{c \le t \le T} |D_{m,Z}(c,t)| \right] = o(1) \times O_{\mathbb{P}}(1).$$

Similarly, for the third supremum, for any $c \le s \le T$,

$$\lambda_m(0,s)\bar{X}_{1:\lfloor ms \rfloor} = \lambda_m(0,c)\bar{X}_{1:\lfloor mc \rfloor} + \lambda_m(c,s)\bar{X}_{\lfloor mc \rfloor + 1:\lfloor ms \rfloor},$$

and, hence, on one hand, for any $c \le s \le t \le T$,

$$m^{-1/2}H_m(s,t) = \lambda_m(s,t)\{\lambda_m(0,c)\bar{X}_{1:\lfloor mc \rfloor} + \lambda_m(c,s)\bar{X}_{\lfloor mc \rfloor + 1:\lfloor ms \rfloor} - \lambda_m(0,s)\bar{X}_{\lfloor ms \rfloor + 1:\lfloor mt \rfloor}\},$$

while, on the other hand,

$$K_c(s,t) = \lambda_m(s,t)\{\lambda_m(0,c)\mathbb{E}(Y_1) + \lambda_m(c,s)\mathbb{E}(Z_1) - \lambda_m(0,s)\mathbb{E}(Z_1)\} + O(1/m),$$

with the term $O(1/m)$ again uniform in s, t, c. Finally, by the triangle inequality,

$$\sup_{c\le s\le t\le T} |J_m(s,t)| \le m^{-1/2}T\,[|D_{m,Y}(0,c)| + \sup_{c\le s\le T}|D_{m,Z}(c,s)| + \sup_{c\le s\le t\le T}|D_{m,Z}(s,t)|\Big] = o(1)\times O_{\mathbb{P}}(1),$$

which completes the proof of (B.4).

It remains to prove (3.8). Recall that from the definition of the function w_γ in (2.8), $w_\gamma(t) \le 1$ for all $t \ge 1$, $\epsilon > 0$ and $\gamma \ge 0$. Hence,

$$\begin{aligned}
&\sup_{m+1\le k<\infty}\{w_S(k/m)\}^{-1}S_m(k)\\
&\ge \sup_{m+1\le k\le \lfloor mT\rfloor}(k/m)^{-5/2-\eta}\frac{1}{m}\sum_{j=m}^{k-1}\frac{j(k-j)}{m^{3/2}}|\bar X^{(m)}_{1:j} - \bar X^{(m)}_{j+1:k}|\\
&= \sup_{t\in[1,T]}\left(\frac{\lfloor mt\rfloor}{m}\right)^{-5/2-\eta}\int_1^t |H_m(s,t)|\mathrm{d}s \ge \sup_{t\in[1,T]} t^{-5/2-\eta}\int_1^t |H_m(s,t)|\mathrm{d}s\\
&\ge T^{-5/2-\eta}\,m^{1/2}\sup_{t\in[1,T]}\int_1^t |m^{-1/2}H_m(s,t)|\mathrm{d}s \xrightarrow{\mathbb{P}} \infty
\end{aligned}$$

since, by the continuous mapping theorem,

$$\sup_{t\in[1,T]}\int_1^t |m^{-1/2}H_m(s,t)|\mathrm{d}s \xrightarrow{\mathbb{P}} \sup_{t\in[1,T]}\int_1^t |K_c(s,t)|\mathrm{d}s > 0.$$

Similarly, $\sup_{m+1\le k<\infty}\{w_T(k/m)\}^{-1}T_m(k)$ is larger than

$$\begin{aligned}
&\sup_{m+1\le k\le\lfloor mT\rfloor}(k/m)^{-2-\eta}\sqrt{\frac{1}{m}\sum_{j=m}^{k-1}\left\{\frac{j(k-j)}{m^{3/2}}(\bar X_{1:j} - \bar X_{j+1:k})\right\}^2}\\
&\qquad\ge \sup_{t\in[1,T]} t^{-2-\eta}\sqrt{\int_1^t\{H_m(s,t)\}^2\mathrm{d}s}\\
&\qquad\ge T^{-2-\eta}m^{1/2}\sup_{t\in[1,T]}\sqrt{\int_1^t\{m^{-1/2}H_m(s,t)\}^2\mathrm{d}s} \xrightarrow{\mathbb{P}} \infty
\end{aligned}$$

since, by the continuous mapping theorem,

$$\sup_{t\in[1,T]}\sqrt{\int_1^t\{m^{-1/2}H_m(s,t)\}^2\mathrm{d}s} \xrightarrow{\mathbb{P}} \sup_{t\in[1,T]}\sqrt{\int_1^t\{K_c(s,t)\}^2\mathrm{d}s} > 0. \qquad \square$$

Appendix C: Proofs of Propositions 4.1 and 6.2

Proof of Proposition 4.1. For any fixed $\eta > 0$, $\epsilon > 0$ and $\gamma \geq 0$, we have

$$\begin{aligned}
&\sup_{1 \leq s \leq t < \infty} \frac{1}{t^{3/2+\eta} \max[\{(t-1)/t\}^{\gamma}, \epsilon]} |tW(s) - sW(t)| \\
&\quad = \sup_{1 \leq s \leq t < \infty} \frac{1}{t^{3/2+\eta} \max\{(1-1/t)^{\gamma}, \epsilon\}} \left| \frac{ts}{s} W(s) - \frac{st}{t} W(t) \right| \\
&\quad = \sup_{1 \leq s \leq t < \infty} \frac{ts}{t^{3/2+\eta} \max\{(1-1/t)^{\gamma}, \epsilon\}} \left| \frac{1}{s} W(s) - \frac{1}{t} W(t) \right| \\
&\quad = \sup_{1 \leq s \leq t < \infty} \frac{s}{t^{1/2+\eta} \max\{(1-1/t)^{\gamma}, \epsilon\}} \left| \frac{1}{s} W(s) - \frac{1}{t} W(t) \right|.
\end{aligned}$$

Let $u = 1/t$ and $v = 1/s$. Then, the last expression on the right is equal to

$$\begin{aligned}
&\sup_{1 \leq 1/v \leq 1/u < \infty} \frac{1/v}{(1/u)^{1/2+\eta} \max\{(1-u)^{\gamma}, \epsilon\}} |vW(1/v) - uW(1/u)| \\
&\qquad = \sup_{0 < u \leq v \leq 1} \frac{u^{1/2+\eta}}{v \max\{(1-u)^{\gamma}, \epsilon\}} |vW(1/v) - uW(1/u)|.
\end{aligned}$$

The claim finally follows from the Brownian inversion property stating that $\{tW(1/t)\}_{t \geq 0}$ is also a standard Brownian motion. □

Proof of Proposition 6.2. Fix $\eta > 0$, $\epsilon > 0$ and $\gamma \geq 0$, and, for any $k \geq m+1$, let

$$\begin{aligned}
V_m(k) =& \epsilon^{-1} \sup_{m+1 \leq k \leq \infty} (k/m)^{-3/2-\eta} \max_{m \leq j \leq k-1} \frac{j(k-j)}{m^{3/2}} |\theta_{1:j} - \theta_{j+1:k} \\
&- \bar{IF}_{1:j} + \bar{IF}_{j+1:k}| \\
=& \epsilon^{-1} \sup_{m+1 \leq k \leq \infty} (k/m)^{-3/2-\eta} \max_{m \leq j \leq k-1} \frac{j(k-j)}{m^{3/2}} |R_{1,j} - R_{j+1:k}|. \qquad \text{(C.1)}
\end{aligned}$$

Then, using the reverse triangle inequality for the maximum norm,

$$\begin{aligned}
&\sup_{m+1 \leq k \leq \infty} \{w_R(k/m)\}^{-1} |R_m^{\theta}(k) - R_m^{I\!F}(k)| \leq V_m(k) \\
&\qquad \leq \epsilon^{-1} m^{\eta} \sup_{m+1 \leq k \leq \infty} k^{-3/2-\eta} \left\{ \max_{m \leq j \leq k-1} j(k-j)|R_{1,j}| \right. \\
&\qquad\qquad \left. + \max_{m \leq j \leq k-1} j(k-j)|R_{j+1,k}| \right\} = o_{\mathbb{P}}(1)
\end{aligned}$$

since

$$m^{\eta} \sup_{m+1 \leq k \leq \infty} k^{-3/2-\eta} \max_{m \leq j \leq k-1} j(k-j)|R_{1,j}|$$

$$\leq m^{\eta} \sup_{m+1\leq k\leq\infty} k^{-1/2-\eta} \max_{m\leq j\leq k-1} j|R_{1,j}|$$
$$\leq m^{\eta} \sup_{m+1\leq k\leq\infty} k^{-\eta} \times \sup_{m+1\leq k\leq\infty} k^{-1/2} \max_{m\leq j\leq k-1} j|R_{1,j}|$$
$$\leq \sup_{m+1\leq k\leq\infty} k^{-1/2} \max_{1\leq i<j\leq k} (j-i+1)|R_{i,j}| = o_{\mathbb{P}}(1)$$

as a consequence of Condition 6.1 and, similarly,

$$m^{\eta} \sup_{m+1\leq k\leq\infty} k^{-3/2-\eta} \max_{m\leq j\leq k-1} j(k-j)|R_{j+1,k}|$$
$$\leq m^{\eta} \sup_{m+1\leq k\leq\infty} k^{-1/2-\eta} \max_{m\leq j\leq k-1} (k-j)|R_{j+1,k}|$$
$$\leq m^{\eta} \sup_{m+1\leq k\leq\infty} k^{-\eta} \times \sup_{m+1\leq k\leq\infty} k^{-1/2} \max_{m\leq j\leq k-1} (k-j)|R_{j+1,k}|$$
$$\leq \sup_{m+1\leq k\leq\infty} k^{-1/2} \max_{1\leq i<j\leq k} (j-i+1)|R_{i,j}| = o_{\mathbb{P}}(1).$$

Using the reverse triangle inequality for the L_1 norm, the claim for S_m^{θ} and S_m^{IF} follows from the fact that

$$\sup_{m+1\leq k\leq\infty} \{w_S(k/m)\}^{-1} |S_m^{\theta}(k) - S_m^{\mathit{IF}}(k)|$$
$$\leq \epsilon^{-1} \sup_{m+1\leq k\leq\infty} (k/m)^{-5/2-\eta} \frac{1}{m} \sum_{j=m}^{k-1} \frac{j(k-j)}{m^{3/2}} |R_{1,j} - R_{j+1:k}| \leq \frac{k-m}{k} V_m(k),$$

where V_m is defined in (C.1). Similarly, the claim for T_m^{θ} and T_m^{IF} follows from the fact that

$$\sup_{m+1\leq k\leq\infty} \{w_T(k/m)\}^{-1} |T_m^{\theta}(k) - T_m^{\mathit{IF}}(k)|$$
$$\leq \epsilon^{-1} \sup_{m+1\leq k\leq\infty} (k/m)^{-2-\eta} \sqrt{\frac{1}{m} \sum_{j=m}^{k-1} \left\{ \frac{j(k-j)}{m^{3/2}} (R_{1,j} - R_{j+1:k}) \right\}^2}$$
$$\leq \sqrt{\frac{k-m}{k}} V_m(k). \qquad \square$$

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59** 817–858. MR1106513

Aue, A. and Horváth, L. (2013). Structural breaks in time series. *J. Time Series Anal.* **34** 1–16. MR3008012 MR3008012

Aue, A. and Horváth, L. (2004). Delay time in sequential detection of change. *Statistics and Probability Letters* **67** 221 – 231. MR2053524

Aue, A., Horváth, L., Hušková, M. and Kokoszka, P. (2006). Change-point monitoring in linear models. *The Econometrics Journal* **9** 373-403. MR2328569

Auestad, B. and Tjøstheim, D. (1990). Identification of nonlinear time series: First order characterization and order determination. *Biometrika* **77** 669–687. MR1086681

Chu, C.-S. J., Stinchcombe, M. and White, H. (1996). Monitoring Structural Change. *Econometrica* **64** 1045–1065.

Csörgő, M. and Horváth, L. (1997). *Limit theorems in change-point analysis. Wiley Series in Probability and Statistics*. John Wiley and Sons, Chichester, UK. MR2743035

Dette, H. and Gösmann, J. (2019). A Likelihood Ratio Approach to Sequential Change Point Detection for a General Class of Parameters. *Journal of the American Statistical Association* **0** 1-17. MR4143471

Eberlein, E. (1986). On Strong Invariance Principles Under Dependence Assumptions. *Ann. Probab.* **14** 260–270. MR0815969

Fremdt, S. (2015). Page's sequential procedure for change-point detection in time series regression. *Statistics* **49** 128-155. MR3304372

Gösmann, J., Kley, T. and Dette, H. (2021). A new approach for open-end sequential change point monitoring. *Journal of the Time Series Analysis* **42** 63–84.

Horváth, L., Hušková, M., Kokoszka, P. and Steinebach, J. (2004). Monitoring changes in linear models. *Journal of Statistical Planning and Inference* **126** 225–251. MR2090695

Jondeau, E., Poon, S. H. and Rockinger, M. (2007). *Financial modeling under non-Gaussian distributions*. Springer, London. MR2260802

Kirch, C. and Weber, S. (2018). Modified sequential change point procedures based on estimating functions. *Electron. J. Statist.* **12** 1579–1613.

Kojadinovic, I. (2020). npcp: Some Nonparametric Tests for Change-Point Detection in Possibly Multivariate Observations R package version 0.2-2.

Kojadinovic, I. and Verdier, G. (2021). Nonparametric sequential change-point detection for multivariate time series based on empirical distribution functions. *Electron. J. Statist.* **15** 773–829. MR4203345

Lai, T. L. (2001). Sequential analysis: some classical problems and new challenges. *Statistica Sinica* **11** 303–351. MR1844531

Montgomery, D. C. (2007). *Introduction to statistical quality control*. John Wiley & Sons.

Paparoditis, E. and Politis, D. N. (2001). Tapered block bootstrap. *Biometrika* **88** 1105–1119. MR1872222

Ritz, C., Baty, F., Streibig, J. C. and Gerhard, D. (2015). Dose-Response Analysis Using R. *PLOS ONE* **10**.

Smith, T. M., Reynolds, R. W., Peterson, T. C. and Lawrimore, J. (2008). Improvements to NOAA's Historical Merged Land–Ocean Surface Temperature Analysis (1880–2006). *Journal of Climate* **21** 2283-2296.

R Core Team (2020). R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria.

Zeileis, A. (2004). Econometric Computing with HC and HAC Covariance Matrix Estimators. *Journal of Statistical Software* **11** 1–17.