Nonparametric sequential change-point detection for multivariate time series based on empirical distribution functions

Ivan Kojadinovic[1] and Ghislain Verdier[1]

[1] *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; ghislain.verdier@univ-pau.fr*

Abstract: The aim of sequential change-point detection is to issue an alarm when it is thought that certain probabilistic properties of the monitored observations have changed. This work is concerned with nonparametric, closed-end testing procedures based on differences of empirical distribution functions that are designed to be particularly sensitive to changes in the contemporary distribution of multivariate time series. The proposed detectors are adaptations of statistics used in a posteriori (offline) change-point testing and involve a weighting allowing to give more importance to recent observations. The resulting sequential change-point detection procedures are carried out by comparing the detectors to threshold functions estimated through resampling such that the probability of false alarm remains approximately constant over the monitoring period. A generic result on the asymptotic validity of such a way of estimating a threshold function is stated. As a corollary, the asymptotic validity of the studied sequential tests based on empirical distribution functions is proven when these are carried out using a dependent multiplier bootstrap for multivariate time series. Large-scale Monte Carlo experiments demonstrate the good finite-sample properties of the resulting procedures. The application of the derived sequential tests is illustrated on financial data.

MSC 2010 subject classifications: Primary 62E20, 62H15; secondary 62G09.
Keywords and phrases: asymptotic validity results, dependent multiplier bootstrap, online monitoring, resampling, threshold function estimation.

Contents

1. Introduction

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$, $m \geq 1$, be a stretch from a d-dimensional stationary times series of continuous random vectors with unknown contemporary distribution function (d.f.) F given by $F(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{X}_1 \leq \boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^d$. These available observations will be referred to as the *learning sample* as we continue. The context of this work is that of *sequential change-point detection*: new observations $\boldsymbol{X}_{m+1}, \boldsymbol{X}_{m+2}, \ldots$ arrive sequentially and we wish to issue an alarm as soon as possible if the contemporary distribution of the most recent observations is not equal to F anymore. If there is no evidence of a change in the contemporary distribution, the monitoring stops after the arrival of observation $\boldsymbol{X}_n$ for some $n > m$.

The theoretical framework of our investigations is that adopted in the seminal paper of Chu, Stinchcombe and White (1996). Unlike classical approaches in statistical process control (SPC) usually calibrated in terms of *average run length* (ARL) (see, e.g., Lai, 2001; Montgomery, 2007, for an overview) and leading in general to the rejection of the underlying null hypothesis of stationarity with probability one, the approach of Chu, Stinchcombe and White (1996) guarantees that, asymptotically (that is, as the size m of the learning sample tends to infinity), stationarity, if it holds, will only be rejected with a small probability α to be interpreted as a type I error and thus called the *probability of false alarm*. The latter paradigm is increasingly considered in the literature; see, e.g., Horváth et al. (2004), Aue et al. (2012a), Aue et al. (2012b), Fremdt (2015), Kirch and Weber (2018), Dette and Gösmann (2019) or Kirch and Stoehr (2019).

Among the approaches *à la* Chu, Stinchcombe and White (1996), one can distinguish between *closed-end* and *open-end* procedures. The latter can in principle continue indefinitely if no evidence against the null is observed. Our approach is of the former type in the sense that at most $n-m$ new observations will be considered before the monitoring stops.

As already mentioned, the null hypothesis of the procedure that we shall investigate is that of stationarity and can be more formally stated as follows:

$$H_0 : \boldsymbol{X}_1, \ldots, \boldsymbol{X}_m, \boldsymbol{X}_{m+1}, \ldots, \boldsymbol{X}_n \text{ is a stretch from a stationary time series} \tag{1.1}$$
$$\text{with contemporary d.f. } F.$$

Notice that, if one is additionally ready to assume the independence of the observations, H_0 simplifies to

$$H_0^{\text{ind}} : \ \boldsymbol{X}_1, \ldots, \boldsymbol{X}_m, \boldsymbol{X}_{m+1}, \ldots, \boldsymbol{X}_n \text{ are independent random vectors with d.f. } F. \tag{1.2}$$

The aim of this work is to derive nonparametric sequential change-point detection procedures particularly sensitive to the alternative hypothesis

$$H_1 : \exists \, k^\star \in \{m, \ldots, n-1\} \text{ such that } \boldsymbol{X}_1, \ldots, \boldsymbol{X}_{k^\star} \text{ is a stretch from a stationary}$$
$$\text{time series with contemporary d.f. } F \text{ and } \boldsymbol{X}_{k^\star+1}, \ldots, \boldsymbol{X}_n \text{ is a stretch} \tag{1.3}$$
$$\text{from a stationary time series with contemporary d.f. } G \neq F.$$

In other words, unlike some of the approaches reported for instance in Kirch and Weber (2018), Dette and Gösmann (2019) or Kirch and Stoehr (2019), we are not solely interested in being sensitive to a change in a given parameter of the d-dimensional time series such as the mean vector or the covariance matrix. We aim at deriving nonparametric monitoring procedures that, in principle, provided m and n are large enough, can detect all types of changes in the contemporary d.f.

In the considered context, the two main ingredients of a sequential change-point detection procedure are a sequence of positive statistics $D_m(k)$, $k \in \{m+1, \ldots, n\}$, and a sequence of suitably chosen strictly positive thresholds $w_m(k)$, $k \in \{m+1, \ldots, n\}$. For $k \in \{m+1, \ldots, n\}$, the statistic $D_m(k)$ (called a *detector* in the literature) is used to assess a possible departure from H_0 in (1.1) using only observations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k$ and is such that the larger $D_m(k)$, the more evidence against H_0. After the arrival of observation $\boldsymbol{X}_k$, $k \in \{m+1, \ldots, n\}$, the detector $D_m(k)$ is computed and compared to the threshold $w_m(k)$. If $D_m(k) > w_m(k)$, the available evidence against stationarity is considered to be large enough and an alarm is issued resulting in the monitoring to stop. If $D_m(k) \leq w_m(k)$ and $k < n$, a new observation $\boldsymbol{X}_{k+1}$ is collected and the previous iteration is repeated. This monitoring process can be naturally represented by a graph illustrating the evolution of the sequence of detectors against the sequence of thresholds; see, e.g., Figures 6.1 and 6.2. To ensure that the monitoring can be interpreted as a testing procedure, the sequence of thresholds $w_m(k)$, $k \in \{m+1, \ldots, n\}$, needs to be chosen such that, under stationarity,

$$\mathbb{P}\{D_m(m+1) \leq w_m(m+1), \ldots, D_m(n) \leq w_m(n)\} \geq 1 - \alpha, \tag{1.4}$$

for some small significance level α.

The detectors $D_m(k)$, $k \in \{m+1, \ldots, n\}$, proposed in this work are defined from differences of empirical distribution functions. Specifically, as shall be explained in detail in Section 2, particularly powerful monitoring procedures can be obtained by defining $D_m(k)$ as a maximized, suitably normalized difference of empirical distribution functions computed from $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_j$ and $\boldsymbol{X}_{j+1}, \ldots, \boldsymbol{X}_k$, respectively, where the maximum is taken over all $j \in \{m, \ldots, k-1\}$. When sequentially investigating changes in real-valued parameters of multivariate time series, such an approach (related to what was called a Page-CUSUM procedure in Kirch and Weber (2018) as a consequence of the work of Fremdt (2015)) was recently suggested in Dette and Gösmann (2019) and justified through a likelihood ratio approach. The latter principle had actually already been considered in the SPC literature (to the best of our knowledge, without any asymptotic theory, however) since at least the work of Hawkins, Qiu and Kang (2003); see also Hawkins and Zamba (2005), Ross, Tasoulis and Adams (2011) and Ross and Adams (2012). As we shall see, our detectors can be cast into the framework of Dette and Gösmann (2019) with the difference that they involve a weighting allowing to give more or less importance to recent observations. In that sense, they could also be regarded as an adaptation of the statistics considered in Csörgő and Szyszkowicz (1994a) for a posteriori (offline) change-point detection to the sequential setting considered in this work.

As far as the thresholds $w_m(k)$, $k \in \{m+1, \ldots, n\}$, are concerned, unlike in the recent literature in which these are calibrated through parametric functions defined up to a positive multiplicative constant such that (1.4) holds (see, e.g., Kirch and Weber, 2018; Dette and Gösmann, 2019; Kirch and Stoehr, 2019), we define them through simulations or resampling with the aim that, under H_0 in (1.1), the probability of rejection of H_0 be roughly the same at every step $k \in \{m+1, \ldots, n\}$ of the procedure. In the considered sequential testing framework *à la* Chu, Stinchcombe and White (1996), the latter requirement had actually already been proposed by Anatolyev and Kosenok (2018). Interestingly enough, as we shall see, it is strongly related to an approach appearing in SPC in Margavio et al. (1995); see also, for instance, Hawkins and Zamba (2005), Ross (2014), Ross and Adams (2012) and the R package cpm described in Ross (2015).

From the point of view of their main ingredients, the sequential change-point tests studied in this work are related to the approach put forward in SPC in Ross and Adams (2012) (without any asymptotic theory) and implemented in the R package cpm (Ross, 2015). However, unlike the latter, they are not restricted to univariate observations and they can deal with time series, among several other differences. To achieve this, as soon as the learning data set $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ is multivariate $(d > 1)$ or serially dependent, resampling, under the form of a *dependent multiplier bootstrap à la* Bücher and Kojadinovic (2016), is used to carry out the monitoring procedures. By adapting the theoretical framework considered in Dette and Gösmann (2019), the asymptotic validity of the investigated procedures is proven under strong mixing both under H_0 in (1.1) and under sequences of alternatives related to H_1 in (1.3).

This paper is organized as follows. In the second section, we propose three classes of detectors based on differences of empirical d.f.s and study their asymptotics, both

under H_0 and sequences of alternatives related to H_1. A more general perspective is adopted in the third section: for an arbitrary detector in the considered closed-end setting, a procedure for estimating the threshold function such that the probability of false alarm remains approximately constant over the monitoring period is investigated and its asymptotic validity is proven under both H_0 and sequences of alternatives related to H_1 when the estimation is based on an asymptotically valid resampling scheme. The results of the third section are applied in the fourth section to the three proposed classes of detectors based on differences of empirical d.f.s. To do so, the consistency of a dependent multiplier bootstrap for the detectors is proven under strong mixing. The fifth section presents a summary of large-scale Monte-Carlo experiments demonstrating the good finite-sample properties of the resulting sequential testing procedures. An application on real financial data concludes the article.

Auxiliary results and proofs are deferred to a sequence of appendices. 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 detectors and their asymptotics

After defining three classes of detectors based on empirical distribution functions and noticing that they are margin-free under H_0 in (1.1), we study their asymptotics under the null and sequences of alternatives related to H_1 in (1.3).

2.1. Detectors based on empirical distribution functions

Let $F_{j:k}$ be the empirical d.f. computed from the stretch $\boldsymbol{X}_j, \ldots, \boldsymbol{X}_k$ of available observations. More formally, for any integers $j, k \geq 1$ and $\boldsymbol{x} \in \mathbb{R}^d$, let

$$F_{j:k}(\boldsymbol{x}) = \begin{cases} \dfrac{1}{k-j+1} \displaystyle\sum_{i=j}^{k} \mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{x}), & \text{if } j \leq k, \\ 0, & \text{otherwise,} \end{cases} \tag{2.1}$$

where the inequalities $\boldsymbol{X}_i \leq \boldsymbol{x}$ are to be understood componentwise and $\mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{x})$ is equal to 1 (resp. 0) if all (resp. some) of the d underlying inequalities are true (resp. false).

After the kth observation has arrived, the available data take the form of the stretch $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k$. If we were in the context of a posteriori change-point detection, a prototypical test statistic would be the maximally selected Kolmogorov–Smirnov-type statistic

$$R_k = \max_{1 \leq j \leq k-1} \frac{j(k-j)}{k^{3/2}} \sup_{\boldsymbol{x} \in \mathbb{R}^d} |F_{1:j}(\boldsymbol{x}) - F_{j+1:k}(\boldsymbol{x})|, \tag{2.2}$$

practically considered for instance in Gombay and Horváth (1999) or Holmes, Kojadinovic and Quessy (2013). The intuition behind R_k is the following: every $j \in \{1, \ldots, k-1\}$ is treated as a potential break point in the sequence and the maximum

in (2.2) implies that R_k will be large as soon as the difference between $F_{1:j}$ and $F_{j+1:k}$ is large for some j. The weighting $j(k-j)/k^{3/2}$ ensures that R_k converges in distribution under stationarity as $k \to \infty$. As explained for instance in Csörgő and Szyszkowicz (1994a) in the case of independent observations, replacing for instance these weights simply by $\sqrt{k}$ would result in a statistic that diverges in probability to ∞ under stationarity. The part $j(k-j)$ in the weighting favors however the detection of potential break points in the middle of the sequence. Test statistics that are more sensitive to changes at the beginning or at the end of the sequence but still converge in distribution under stationarity can be obtained by considering weights of the form $j(k-j)/\{k^{3/2}q(j/k)\}$ for some suitable strictly positive function q on $(0,1)$; see, e.g., Csörgő and Szyszkowicz (1994b,a), Csörgő, Horváth and Szyszkowicz (1997) and Csörgő and Horváth (1997).

Going back to the setting of sequential change-point detection considered in this work, a first meaningful modification of (2.2) is to restrict the maximum over j to $j \in \{m, \ldots, k-1\}$ since a change cannot occur at the beginning of the sequence given that $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ is the learning sample known to be a stretch from a stationary sequence. Another modification is a rescaling consisting of replacing $k^{3/2}$ by $m^{3/2}$ in the weighting. The latter is made solely for asymptotic reasons as shall become clear in Sections 2.3 and 2.4. These modifications essentially lead to the first detectors considered in this work:

$$R_{m,q}(k) = \max_{m \le j \le k-1} \frac{j(k-j)}{m^{3/2}q(j/m, k/m)} \sup_{\boldsymbol{x} \in \mathbb{R}^d} |F_{1:j}(\boldsymbol{x}) - F_{j+1:k}(\boldsymbol{x})|, \qquad k \in \{m+1, \ldots, n\}. \tag{2.3}$$

In the previous display, q is a strictly positive function whose role is to potentially give more weight to recent observations. In the sequel, we consider the parametric form

$$q(s,t) = \max\{s^\gamma(t-s)^\gamma, \delta\}, \qquad 0 \le s \le t, \tag{2.4}$$

where $\delta \in (0,1)$ is a small constant and γ is a parameter in $[0, 1/2]$. If $\gamma = 0$, q is the constant function 1 and $R_{m,q}(k)$ is then a straightforward adaptation of R_k in (2.2) to sequential change-point detection. In that case, the general form of $R_{m,q}(k)$ can also be heuristically justified through a likelihood ratio approach; see Section 2 in Dette and Gösmann (2019). When $\gamma = 1/2$, as shall be discussed in Remark 2.6 using asymptotic arguments, $R_{m,q}(k)$ can be regarded, under H_0 in (1.1), as a maximum of random variables with, approximately, the same mean and variance. This heuristically implies that all the potential break points $j \in \{m, \ldots, k-1\}$ are given roughly the same weight in the computation of $R_{m,q}(k)$ unlike in the case $\gamma = 0$ in which potential break points closest to $\lfloor k/2 \rfloor$ are given more weight. Hence, for certain types of alternatives to H_0 in (1.1), choosing $\gamma \in (0, 1/2]$ might accelerate the detection of the corresponding change in the contemporary distribution of the underlying time series. Examples of such alternatives will be given in Section 5 in which the results of numerous Monte Carlo experiments are summarized.

Two similar Cramér–von Mises-like detectors are also considered in this work. They

are defined, for $k \in \{m+1, \ldots, n\}$, by

$$
\begin{aligned}
S_{m,q}(k) &= \max_{m \leq j \leq k-1} \int_{\mathbb{R}^d} \left[\frac{j(k-j)}{m^{3/2}q(j/m, k/m)} \{ F_{1:j}(\boldsymbol{x}) - F_{j+1:k}(\boldsymbol{x}) \} \right]^2 \, \mathrm{d}F_{1:k}(\boldsymbol{x}) \\
&= \max_{m \leq j \leq k-1} \frac{1}{k} \sum_{i=1}^{k} \left[\frac{j(k-j)}{m^{3/2}q(j/m, k/m)} \{ F_{1:j}(\boldsymbol{X}_i) - F_{j+1:k}(\boldsymbol{X}_i) \} \right]^2,
\end{aligned}
\tag{2.5}
$$

and

$$
T_{m,q}(k) = \frac{1}{m} \sum_{j=m}^{k-1} \frac{1}{k} \sum_{i=1}^{k} \left[\frac{j(k-j)}{m^{3/2}q(j/m, k/m)} \{ F_{1:j}(\boldsymbol{X}_i) - F_{j+1:k}(\boldsymbol{X}_i) \} \right]^2.
\tag{2.6}
$$

Remark 2.1. The detectors in (2.3) and (2.5) are related to those used in Ross and Adams (2012, pp 104–106). The latter are also based on differences of empirical d.f.s but deal only with independent univariate observations. The analogue of (2.3) is apparently defined as a maximum of the quantities $\sup_{\boldsymbol{x} \in \mathbb{R}} |F_{1:j}(\boldsymbol{x}) - F_{j+1:k}(\boldsymbol{x})|$, $j \in \{m, \ldots, k-1\}$, previously normalized using an empirical probability integral transformation, probably based on simulations. The analogue of (2.5) takes the form of a maximum of the quantities $\sum_{i=1}^{k} \{ F_{1:j}(\boldsymbol{X}_i) - F_{j+1:k}(\boldsymbol{X}_i) \}^2$, $j \in \{m, \ldots, k-1\}$, after centering and scaling. The asymptotics of these detectors were not studied. Given that the detectors of Ross and Adams (2012) are distribution-free (see also Section 2.2 hereafter), that their approach assumes serially independent observations and is based on simulations, the absence of asymptotic theory is not problematic.

Remark 2.2. The detectors in (2.3), (2.5) and (2.6) are almost of the Page-CUSUM type considered initially in Fremdt (2015); see also Kirch and Weber (2018). For instance, in the case of (2.3), the adaption of the latter construction to the present setting would have instead involved a maximum of the quantities $\sup_{\boldsymbol{x} \in \mathbb{R}^d} |F_{1:m}(\boldsymbol{x}) - F_{j+1:k}(\boldsymbol{x})|$, $j \in \{m, \ldots, k-1\}$. As explained in Dette and Gösmann (2019), the use of such detectors may result in a loss of power in the case of a small learning sample and a rather late change point. In the Monte Carlo experiments carried out in Dette and Gösmann (2019), Page-CUSUM detectors were always outperformed by their analogues of type (2.3), which is why we do not consider them in this work.

Remark 2.3. In addition to the detectors (2.3), (2.5) and (2.6), we also considered in our Monte Carlo experiments the following natural competitors which are straightforward adaptions of the so-called CUSUM construction considered for instance in Horváth et al. (2004) and Aue et al. (2012b). Their Kolmogorov–Smirnov versions and Cramér–von Mises versions are respectively given, for any $k \in \{m+1, \ldots, n\}$, by

$$
P_m(k) = \frac{m(k-m)}{m^{3/2}} \sup_{\boldsymbol{x} \in \mathbb{R}^d} |F_{1:m}(\boldsymbol{x}) - F_{m+1:k}(\boldsymbol{x})|,
\tag{2.7}
$$

$$
Q_m(k) = \frac{1}{k} \sum_{i=1}^{k} \left[\frac{m(k-m)}{m^{3/2}} \{ F_{1:m}(\boldsymbol{X}_i) - F_{m+1:k}(\boldsymbol{X}_i) \} \right]^2.
\tag{2.8}
$$

The asymptotic theory for these detectors being simpler than for the detectors (2.3), (2.5) and (2.6), it will not be stated in the forthcoming sections for the sake of readability.

2.2. The detectors are margin-free under the null

The detectors defined previously are actually margin-free under H_0 in (1.1), a property that shall be exploited in the forthcoming sections to carry out the corresponding sequential change-point detection procedures.

Recall that $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are assumed to be continuous random vectors. Saying that the detectors are margin-free under the null means that they do not depend on the d univariate margins $F_1, \ldots, F_d$ of F (the unknown d.f. of $\boldsymbol{X}_1$) or, equivalently, that they can alternatively be written in terms of the unobservable random vectors $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n$ defined from $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ through marginal probability integral transformations:

$$\boldsymbol{U}_i = (F_1(X_{i1}), \ldots, F_d(X_{id})). \tag{2.9}$$

Notice that we can recover the $\boldsymbol{X}_i$ from the $\boldsymbol{U}_i$ by marginal quantile transformations:

$$\boldsymbol{X}_i = (F_1^-(U_{i1}), \ldots, F_d^-(U_{id})), \tag{2.10}$$

where, for any univariate d.f. G, G^{-1} denotes its associated quantile function defined by

$$G^{-1}(y) = \inf\{x \in \mathbb{R} : G(x) \geq y\}, \qquad y \in [0, 1], \tag{2.11}$$

with the convention that the infimum of the empty set is ∞.

To verify that the detectors are margin-free under the null, for any integers $j, k \geq 1$ and $\boldsymbol{u} \in [0, 1]^d$, let

$$C_{j:k}(\boldsymbol{u}) = \begin{cases} \dfrac{1}{k - j + 1} \displaystyle\sum_{i=j}^{k} \mathbf{1}(\boldsymbol{U}_i \leq \boldsymbol{u}), & \text{if } j \leq k, \\ 0, & \text{otherwise,} \end{cases} \tag{2.12}$$

be the analogue of $F_{j:k}$ in (2.1) based on the $\boldsymbol{U}_i$ in (2.9). For any $j \in \{1, \ldots, d\}$, by (right) continuity of F_j, we have that $\mathbf{1}\{F_j^-(u) \leq x\} = \mathbf{1}\{u \leq F_j(x)\}$ for all $u \in [0, 1]$ and $x \in \mathbb{R}$; see, e.g., Proposition 1 (5) in Embrechts and Hofert (2013). The latter property combined with (2.10) implies that, under H_0 in (1.1), for any $i \in \{1, \ldots, n\}$,

$$\mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{x}) = \mathbf{1}\{\boldsymbol{U}_i \leq \boldsymbol{F}(\boldsymbol{x})\}, \qquad \boldsymbol{x} \in \mathbb{R}^d, \tag{2.13}$$

where $\boldsymbol{F}(\boldsymbol{x}) = (F_1(x_1), \ldots, F_d(x_d))$. Hence, for any $k \in \{m + 1, \ldots, n\}$,

$$\begin{aligned}
R_{m,q}(k) &= \max_{m \leq j \leq k-1} \frac{j(k-j)}{m^{3/2} q(j/m, k/m)} \sup_{\boldsymbol{x} \in \mathbb{R}^d} |C_{1:j}\{\boldsymbol{F}(\boldsymbol{x})\} - C_{j+1:k}\{\boldsymbol{F}(\boldsymbol{x})\}|, \\
&= \max_{m \leq j \leq k-1} \frac{j(k-j)}{m^{3/2} q(j/m, k/m)} \sup_{\boldsymbol{u} \in [0,1]^d} |C_{1:j}(\boldsymbol{u}) - C_{j+1:k}(\boldsymbol{u})|
\end{aligned}$$

using again the continuity of $F_1, \ldots, F_d$. Similarly,

$$S_{m,q}(k) = \max_{m \leq j \leq k-1} \frac{1}{k} \sum_{i=1}^{k} \left[\frac{j(k-j)}{m^{3/2} q(j/m, k/m)} \{C_{1:j}(\boldsymbol{U}_i) - C_{j+1:k}(\boldsymbol{U}_i)\} \right]^2,$$

$$T_{m,q}(k) = \frac{1}{m} \sum_{j=m}^{k-1} \frac{1}{k} \sum_{i=1}^{k} \left[\frac{j(k-j)}{m^{3/2} q(j/m, k/m)} \{C_{1:j}(\boldsymbol{U}_i) - C_{j+1:k}(\boldsymbol{U}_i)\} \right]^2.$$

In the case of univariate independent observations, the margin-free property under the null implies that the detectors are distribution-free under the null. When $d > 1$, this is not true anymore as the null distribution of the detectors depends on the *copula* C associated with F. The latter is merely the d.f. of the random vector $\boldsymbol{U}_1$ obtained through (2.9). Equivalently, C is a d-dimensional d.f. with standard uniform margins further uniquely defined (see Sklar, 1959) through the relationships

$$F(\boldsymbol{x}) = C\{F_1(x_1), \ldots, F_d(x_d)\}, \qquad \boldsymbol{x} \in \mathbb{R}^d,$$

and

$$C(\boldsymbol{u}) = F\{F_1^-(u_1), \ldots, F_d^-(u_d)\}, \qquad \boldsymbol{u} \in [0, 1]^d.$$

Remark 2.4. To be able to handle both the univariate and the multivariate situations, in the rest of the paper, we adopt the convention that C is the copula associated with F when $d > 1$ and merely the identity function when $d = 1$.

2.3. Asymptotics of the detectors under the null

As shall become clear in the forthcoming sections, the knowledge of the asymptotic behavior of the detectors under H_0 in (1.1) is instrumental in showing the asymptotic validity of the corresponding sequential change-point detection procedures. To study these asymptotics, we follow Dette and Gösmann (2019), among others, and set $n = \lfloor m(T + 1) \rfloor$ for some fixed real number $T > 0$. This will imply that, in the asymptotics, as the size m of the learning sample goes to infinity, the maximum number of new observations considered in the monitoring increases proportionally.

Let $\Delta = \{(s, t) \in [0, T + 1]^2 : s \leq t\}$ and let

$$\lambda_m(s, t) = (\lfloor mt \rfloor - \lfloor ms \rfloor)/m, \qquad (s, t) \in \Delta. \tag{2.14}$$

Then, for any $(s, t) \in \Delta$ and $\boldsymbol{u} \in [0, 1]^d$, let

$$\mathbb{G}_m(s, t, \boldsymbol{u}) = \sqrt{m}\lambda_m(0, s)\lambda_m(s, t)\{C_{1:\lfloor ms \rfloor}(\boldsymbol{u}) - C_{\lfloor ms \rfloor + 1:\lfloor mt \rfloor}(\boldsymbol{u})\} \tag{2.15}$$

where $C_{1:\lfloor ms \rfloor}$ and $C_{\lfloor ms \rfloor + 1:\lfloor mt \rfloor}$ are generically defined by (2.12), and let

$$\mathbb{G}_{m,q}(s, t, \boldsymbol{u}) = \frac{\mathbb{G}_m(s, t, \boldsymbol{u})}{q\{\lambda_m(0, s), \lambda_m(0, t)\}}, \tag{2.16}$$

where q is defined in (2.4). Notice that, with the definitions adopted thus far, $\mathbb{G}_m(s, s, \cdot) = \mathbb{G}_{m,q}(s, s, \cdot) = 0$ for all $s \in [0, T + 1]$.

For any $k \in \{m + 1, \ldots, n\}$ with $n = \lfloor m(T + 1) \rfloor$ and any $j \in \{m, \ldots, k - 1\}$, there exists $(s, t) \in \Delta \cap [1, T + 1]^2$ such that $k = \lfloor mt \rfloor$ and $j = \lfloor ms \rfloor$. We can thus write $R_{m,q}(k)$ as

$$
\begin{aligned}
R_{m,q}(\lfloor mt \rfloor) &= \max_{m \leq j \leq k-1} \frac{j(k - j)}{m^{3/2}q(j/m, k/m)} \sup_{\boldsymbol{u} \in [0,1]^d} |C_{1:j}(\boldsymbol{u}) - C_{j+1:k}(\boldsymbol{u})| \\
&= \sup_{s \in [1,t]} \sup_{\boldsymbol{u} \in [0,1]^d} \frac{\sqrt{m}\lambda_m(0, s)\lambda_m(s, t)|C_{1:\lfloor ms \rfloor}(\boldsymbol{u}) - C_{\lfloor ms \rfloor + 1:\lfloor mt \rfloor}(\boldsymbol{u})|}{q\{\lambda_m(0, s), \lambda_m(0, t)\}} \\
&= \sup_{s \in [1,t]} \sup_{\boldsymbol{u} \in [0,1]^d} |\mathbb{G}_{m,q}(s, t, \boldsymbol{u})|.
\end{aligned}
\tag{2.17}
$$

Similarly, it can be verified that

$$S_{m,q}(k) = S_{m,q}(\lfloor mt \rfloor) = \sup_{s \in [1,t]} \int_{[0,1]^d} \{\mathbb{G}_{m,q}(s,t,\boldsymbol{u})\}^2 \mathrm{d}C_{1:\lfloor mt \rfloor}(\boldsymbol{u}), \tag{2.18}$$

$$T_{m,q}(k) = T_{m,q}(\lfloor mt \rfloor) = \int_1^t \int_{[0,1]^d} \{\mathbb{G}_{m,q}(s,t,\boldsymbol{u})\}^2 \mathrm{d}C_{1:\lfloor mt \rfloor}(\boldsymbol{u}) \mathrm{d}s. \tag{2.19}$$

As we continue, we adopt the convention that $R_{m,q}(m) = S_{m,q}(m) = T_{m,q}(m) = 0$. Furthermore, given a set $\mathcal{S}$, the space of all bounded real-valued functions on $\mathcal{S}$ equipped with the uniform metric is denoted by $\ell^\infty(\mathcal{S})$. The main purpose of this section is to study the asymptotics under the null of the elements $\mathbb{R}_{m,q}, \mathbb{S}_{m,q}$ and $\mathbb{T}_{m,q}$ of $\ell^\infty([1, T+1])$ defined respectively, for any $t \in [1, T+1]$, by

$$\mathbb{R}_{m,q}(t) = R_{m,q}(\lfloor mt \rfloor), \qquad \mathbb{S}_{m,q}(t) = S_{m,q}(\lfloor mt \rfloor), \qquad \mathbb{T}_{m,q}(t) = T_{m,q}(\lfloor mt \rfloor). \tag{2.20}$$

Specifically, we will provide conditions under which they converge weakly in the sense of Definition 1.3.3 in van der Vaart and Wellner (2000) under H_0 in (1.1). Throughout the paper, this mode of convergence will be denoted by the arrow '$\rightsquigarrow$' and all convergences will be for $m \to \infty$ unless mentioned otherwise.

From the expressions given in (2.17), (2.18) and (2.19), we see that, under the null, the detectors studied in this work are functionals of $\mathbb{G}_{m,q}$ in (2.16), and thus of $\mathbb{G}_m$ in (2.15). Under stationarity, the latter is in turn a functional of the *sequential empirical process* defined, for any $s \in [0, T+1]$ and $\boldsymbol{u} \in [0,1]^d$, by

$$\mathbb{B}_m(s,\boldsymbol{u}) = \frac{1}{\sqrt{m}} \sum_{i=1}^{\lfloor ms \rfloor} \{\mathbf{1}(\boldsymbol{U}_i \le \boldsymbol{u}) - C(\boldsymbol{u})\} = \sqrt{m}\lambda_m(0,s)\{C_{1:\lfloor ms \rfloor}(\boldsymbol{u}) - C(\boldsymbol{u})\}. \tag{2.21}$$

Indeed, under H_0 in (1.1), for any $(s,t) \in \Delta$ and $\boldsymbol{u} \in [0,1]^d$,

$$\begin{aligned}
\mathbb{G}_m(s,t,\boldsymbol{u}) &= \sqrt{m}\lambda_m(0,s)\lambda_m(s,t)\{C_{1:\lfloor ms \rfloor}(\boldsymbol{u}) - C(\boldsymbol{u}) - C_{\lfloor ms \rfloor+1:\lfloor mt \rfloor}(\boldsymbol{u}) + C(\boldsymbol{u})\} \\
&= \lambda_m(s,t)\mathbb{B}_m(s,\boldsymbol{u}) - \lambda_m(0,s) \times \sqrt{m}\lambda_m(s,t)\{C_{\lfloor ms \rfloor+1:\lfloor mt \rfloor}(\boldsymbol{u}) - C(\boldsymbol{u})\} \\
&= \{\lambda_m(0,t) - \lambda_m(0,s)\}\mathbb{B}_m(s,\boldsymbol{u}) - \lambda_m(0,s)\{\mathbb{B}_m(t,\boldsymbol{u}) - \mathbb{B}_m(s,\boldsymbol{u})\} \\
&= \lambda_m(0,t)\mathbb{B}_m(s,\boldsymbol{u}) - \lambda_m(0,s)\mathbb{B}_m(t,\boldsymbol{u}).
\end{aligned} \tag{2.22}$$

In the forthcoming asymptotic results, we shall assume that the underlying stationary sequence $(\boldsymbol{X}_i)_{i \in \mathbb{Z}}$ (or, equivalently, the corresponding unobservable stationary sequence $(\boldsymbol{U}_i)_{i \in \mathbb{Z}}$ defined through (2.9)) is *strongly mixing*. Denote by $\mathcal{F}_j^k$ the σ-field generated by $(\boldsymbol{X}_i)_{j \le i \le k}$, $j, k \in \mathbb{Z} \cup \{-\infty, +\infty\}$, and recall that the strong mixing coefficients corresponding to the stationary sequence $(\boldsymbol{X}_i)_{i \in \mathbb{Z}}$ are then defined by $\alpha_0^{\boldsymbol{X}} = 1/2$,

$$\alpha_r^{\boldsymbol{X}} = \sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_r^{+\infty}} \left|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)\right|, \qquad r \in \mathbb{N}, \, r > 0,$$

and that the sequence $(\boldsymbol{X}_i)_{i \in \mathbb{Z}}$ is said to be *strongly mixing* if $\alpha_r^{\boldsymbol{X}} \to 0$ as $r \to \infty$. The following result is proven in Appendix A.

Proposition 2.5. *Assume that H_0 in* (1.1) *holds and that, additionally, $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ is a stretch from a stationary sequence $(\boldsymbol{X}_i)_{i \in \mathbb{Z}}$ of continuous d-dimensional random vectors whose strong mixing coefficients satisfy $\alpha_r^{\boldsymbol{X}} = O(r^{-a})$ for some $a > 1$ as $r \to \infty$. Then, $\mathbb{B}_m \rightsquigarrow \mathbb{B}_C$ in $\ell^\infty([0, T+1] \times [0,1]^d)$, where $\mathbb{B}_C$ is a tight centered Gaussian process with covariance function*

$$\text{Cov}\{\mathbb{B}_C(s, \boldsymbol{u}), \mathbb{B}_C(t, \boldsymbol{v})\} = (s \wedge t)\Gamma(\boldsymbol{u}, \boldsymbol{v}), \qquad s, t \in [0, T+1], \boldsymbol{u}, \boldsymbol{v} \in [0,1]^d,$$

with $\wedge$ the minimum operator and

$$\Gamma(\boldsymbol{u}, \boldsymbol{v}) = \sum_{i \in \mathbb{Z}} \text{Cov}\{\mathbf{1}(\boldsymbol{U}_0 \leq \boldsymbol{u}), \mathbf{1}(\boldsymbol{U}_i \leq \boldsymbol{v})\}. \tag{2.23}$$

As a consequence, $\mathbb{G}_m \rightsquigarrow \mathbb{G}_C$ and $\mathbb{G}_{m,q} \rightsquigarrow \mathbb{G}_{C,q}$ in $\ell^\infty(\Delta \times [0,1]^d)$, where $\mathbb{G}_m$ and $\mathbb{G}_{m,q}$ are defined in (2.15) *and* (2.16), *respectively, and, for any $(s,t) \in \Delta$ and $\boldsymbol{u} \in [0,1]^d$,*

$$\mathbb{G}_C(s, t, \boldsymbol{u}) = t\mathbb{B}_C(s, \boldsymbol{u}) - s\mathbb{B}_C(t, \boldsymbol{u}) \qquad and \qquad \mathbb{G}_{C,q}(s, t, \boldsymbol{u}) = \frac{\mathbb{G}_C(s, t, \boldsymbol{u})}{q(s, t)}. \tag{2.24}$$

It follows that $\mathbb{R}_{m,q} \rightsquigarrow \mathbb{R}_{C,q}$, $\mathbb{S}_{m,q} \rightsquigarrow \mathbb{S}_{C,q}$ and $\mathbb{T}_{m,q} \rightsquigarrow \mathbb{T}_{C,q}$ in $\ell^\infty([1, T+1])$, where

$$\mathbb{R}_{C,q}(t) = \sup_{s \in [1,t]} \sup_{\boldsymbol{u} \in [0,1]^d} |\mathbb{G}_{C,q}(s, t, \boldsymbol{u})|,$$

$$\mathbb{S}_{C,q}(t) = \sup_{s \in [1,t]} \int_{[0,1]^d} \{\mathbb{G}_{C,q}(s, t, \boldsymbol{u})\}^2 \mathrm{d}C(\boldsymbol{u}), \tag{2.25}$$

$$\mathbb{T}_{C,q}(t) = \int_1^t \int_{[0,1]^d} \{\mathbb{G}_{C,q}(s, t, \boldsymbol{u})\}^2 \mathrm{d}C(\boldsymbol{u})\mathrm{d}s, \qquad t \in [1, T+1].$$

Furthermore, for any interval $[t_1, t_2] \subset [1, T+1]$ such that $t_2 > 1$, the distributions of $\sup_{t \in [t_1, t_2]} \mathbb{R}_{C,q}(t)$, $\sup_{t \in [t_1, t_2]} \mathbb{S}_{C,q}(t)$ and $\sup_{t \in [t_1, t_2]} \mathbb{T}_{C,q}(t)$ are absolutely continuous with respect to the Lebesgue measure.

Combined with a generic result on the threshold estimation procedure to be stated in Section 3 and additional bootstrap consistency results to be stated in Section 4, the last claims of Proposition 2.5 constitute a first step in proving that the derived change-point detection procedures hold their level asymptotically.

Remark 2.6. Proposition 2.5 can be used to heuristically justify the form of the weight function q in (2.4) appearing in the expression of the detectors (2.3), (2.5) and (2.6). From (2.24), for any $(s, t) \in \Delta$ and $\boldsymbol{u} \in [0,1]^d$, we obtain that

$$\text{Var}\{\mathbb{G}_C(s, t, \boldsymbol{u})\} = t^2 \text{Var}\{\mathbb{B}_C(s, \boldsymbol{u})\} + s^2 \text{Var}\{\mathbb{B}_C(t, \boldsymbol{u})\} - 2st\text{Cov}\{\mathbb{B}_C(s, \boldsymbol{u}), \mathbb{B}_C(t, \boldsymbol{u})\}$$
$$= (st^2 + s^2t - 2s^2t)\Gamma(\boldsymbol{u}, \boldsymbol{u}) = st(t - s)\Gamma(\boldsymbol{u}, \boldsymbol{u}).$$

As a consequence, for any $1 \leq s < t \leq T+1$ and $\boldsymbol{u} \in [0,1]^d$,

$$\text{Var}\{s^{-1/2}(t-s)^{-1/2}\mathbb{G}_C(s, t, \boldsymbol{u})\} = t\Gamma(\boldsymbol{u}, \boldsymbol{u}).$$

Under the conditions of the proposition, when $\gamma = 1/2$ and m is large, we could then expect that, very roughly, $\text{Var}\{\mathbb{G}_{m,q}(s, t, \boldsymbol{u})\} \approx \text{Var}\{s^{-1/2}(t - s)^{-1/2}\mathbb{G}_C(s, t, \boldsymbol{u})\}$ does not depend on s and thus regard the quantities $\sup_{\boldsymbol{u}\in[0,1]^d} \mathbb{G}_{m,q}(j/m, t, \boldsymbol{u})$, $j \in \{m, \ldots, \lfloor mt \rfloor - 1\}$, appearing in the expression of $R_{m,q}(\lfloor mt \rfloor)$ in (2.17) as random variables with, approximately, the same mean and variance. The latter conveys the intuition that, when $\gamma = 1/2$, all the potential break points $j \in \{m, \ldots, \lfloor mt \rfloor - 1\}$ are given roughly the same weight in the computation of $R_{m,q}(\lfloor mt \rfloor)$. This is, of course, only approximately true because of the presence of the constant δ in the expression of the weight function q in (2.4). In practice, the setting $\gamma = 1/2$ might accelerate the detection of certain types of changes.

2.4. Asymptotics of the detectors under alternatives related to H_1

Under alternatives related to H_1 in (1.3), the detectors are not margin-free anymore. As we shall see in the forthcoming proposition, their asymptotic behavior is then a consequence of that of the process

$$\mathbb{H}_{m,q}(s, t, \boldsymbol{x}) = \frac{\mathbb{H}_m(s, t, \boldsymbol{x})}{q\{\lambda_m(0, s), \lambda_m(0, t)\}}, \qquad (s, t) \in \Delta, \boldsymbol{x} \in \mathbb{R}^d,$$

where

$$\mathbb{H}_m(s, t, \boldsymbol{x}) = \sqrt{m}\lambda_m(0, s)\lambda_m(s, t)\{F_{1:\lfloor ms \rfloor}(\boldsymbol{x}) - F_{\lfloor ms \rfloor+1:\lfloor mt \rfloor}(\boldsymbol{x})\}, \qquad (2.26)$$

the empirical d.f.s $F_{1:\lfloor ms \rfloor}$ and $F_{\lfloor ms \rfloor+1:\lfloor mt \rfloor}$ are generically defined by (2.1), and q is defined in (2.4).

The following result is proven in Appendix A. The arrow '$\xrightarrow{\mathbb{P}}$' in its statement denotes convergence in probability.

Proposition 2.7. *Assume that, for some $c \in [1, T+1)$, $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{\lfloor mc \rfloor}$, denoted equivalently by $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{\lfloor mc \rfloor}$, is a stretch from a stationary sequence $(\boldsymbol{Y}_i)_{i\in\mathbb{Z}}$ of continuous d-dimensional random vectors with contemporary d.f. F whose strong mixing coefficients satisfy $\alpha_r^{\boldsymbol{Y}} = O(r^{-a})$ for some $a > 1$ as $r \to \infty$, and that $\boldsymbol{X}_{\lfloor mc \rfloor+1}, \ldots, \boldsymbol{X}_{\lfloor m(T+1) \rfloor}$, denoted equivalently by $\boldsymbol{Z}_{\lfloor mc \rfloor+1}, \ldots, \boldsymbol{Z}_{\lfloor m(T+1) \rfloor}$, is a stretch from a stationary sequence $(\boldsymbol{Z}_i)_{i\in\mathbb{Z}}$ of continuous d-dimensional random vectors with contemporary d.f. $G \neq F$ whose strong mixing coefficients satisfy $\alpha_r^{\boldsymbol{Z}} = O(r^{-b})$ for some $b > 1$ as $r \to \infty$. Then, $m^{-1/2}\mathbb{H}_m \xrightarrow{\mathbb{P}} K_c$ in $\ell^\infty(\Delta \times \mathbb{R}^d)$, where*

$$K_c(s, t, \boldsymbol{x}) = (s \wedge c)\{(t \vee c) - (s \vee c)\}\{F(\boldsymbol{x}) - G(\boldsymbol{x})\}, \qquad (s, t) \in \Delta, \boldsymbol{x} \in \mathbb{R}^d. \quad (2.27)$$

Consequently, $m^{-1/2}\mathbb{H}_{m,q} \xrightarrow{\mathbb{P}} K_{c,q}$ in $\ell^\infty(\Delta \times \mathbb{R}^d)$, where $K_{c,q}(s, t, \boldsymbol{x}) = K_c(s, t, \boldsymbol{x})/q(s, t)$, $(s, t) \in \Delta$, $\boldsymbol{x} \in \mathbb{R}^d$, and

$$\sup_{t\in[1,T+1]} \mathbb{R}_{m,q}(t) \xrightarrow{\mathbb{P}} \infty, \quad \sup_{t\in[1,T+1]} \mathbb{S}_{m,q}(t) \xrightarrow{\mathbb{P}} \infty \quad and \quad \sup_{t\in[1,T+1]} \mathbb{T}_{m,q}(t) \xrightarrow{\mathbb{P}} \infty.$$

Combined with a generic result on the threshold estimation procedure to be stated in the forthcoming section and additional results on the asymptotic validity of adequate resampling methods to be stated in Section 4, the three last claims of the previous result will be instrumental in showing that the derived change-point detection procedures have asymptotic power one under sequences of alternatives related to H_1.

3. A generic threshold estimation procedure

In the studied context, the second ingredient of a sequential change-point detection procedure is a set of strictly positive thresholds to which detectors will be compared. In this section, we consider a generic threshold estimation procedure that can be employed with any type of detector, and provide conditions under which it is asymptotically valid. The derived results will be applied in the next section to establish the asymptotic validity of sequential change-point detection procedures based on the detectors studied in Section 2.

3.1. A constant probability of false alarm at each step

Within the context of closed-end monitoring from time $m + 1$ to time n, let $D_m(k)$, $k \in \{m + 1, \ldots, n\}$, be arbitrary detectors. As discussed in the introduction, it seems natural to choose the corresponding thresholds $w_m(k)$, $k \in \{m+1, \ldots, n\}$, so that, under H_0 in (1.1), the probability of rejection of H_0 is the same at every step $k \in \{m+1, \ldots, n\}$ of the procedure. Adapting the reasoning of Anatolyev and Kosenok (2018) to our context, the latter requirement consists of choosing the $w_m(k)$, $k \in \{m+1, \ldots, n\}$, such that, under stationarity, for all $k \in \{m + 1, \ldots, n\}$,

$$\mathbb{P}\{D_m(m + 1) \leq w_m(m + 1), \ldots, D_m(k) \leq w_m(k)\} = 1 - \frac{k - m}{n - m}\alpha, \qquad (3.1)$$

where $\alpha \in (0, 1/2)$ is the desired significance level of the sequential testing procedure, or, equivalently, such that, under H_0, for all $k \in \{m + 1, \ldots, n\}$,

$$\mathbb{P}\{\exists\, i \in \{m + 1, \ldots, k\} \text{ s.t. } D_m(i) > w_m(i)\} = \frac{k - m}{n - m}\alpha.$$

Notice that the previous display translates mathematically our requirement that the probability of false alarm be proportional to the number of monitoring steps.

Interestingly enough, the previous way of choosing the thresholds $w_m(k)$, $k \in \{m + 1, \ldots, n\}$, is strongly related to an approach used in SPC and possibly first appearing in Margavio et al. (1995) (see also, e.g., Hawkins and Zamba, 2005; Ross, 2014). It consists of choosing the $w_m(k)$, $k \in \{m+1, \ldots, n\}$, such that, under stationarity, for some small

$\xi_m > 0$,

$$\begin{cases} \mathbb{P}\{D_m(m+1) > w_m(m+1)\} = \xi_m, \\[2mm] \text{and, for all } k \in \{m+2,\ldots,n\}, \\[2mm] \mathbb{P}\{D_m(k) > w_m(k) \mid D_m(m+1) \le w_m(m+1),\ldots,D_m(k-1) \le w_m(k-1)\} = \xi_m. \end{cases} \tag{3.2}$$

We then obtain that, under H_0, for all $k \in \{m+2,\ldots,n\}$,

$$\begin{aligned} \mathbb{P}\{&D_m(m+1) \le w_m(m+1),\ldots,D_m(k) \le w_m(k)\} \\ =&\mathbb{P}\{D_m(k) \le w_m(k) \mid D_m(m+1) \le w_m(m+1),\ldots,D_m(k-1) \le w_m(k-1)\} \\ &\times \mathbb{P}\{D_m(m+1) \le w_m(m+1),\ldots\ldots,D_m(k-1) \le w_m(k-1)\} \\ =&(1-\xi_m) \times \mathbb{P}\{D_m(k-1) \le w_m(k-1) \mid D_m(m+1) \le w_m(m+1),\ldots \\ &\ldots,D_m(k-2) \le w_m(k-2)\} \\ &\times \mathbb{P}\{D_m(m+1) \le w_m(m+1),\ldots,D_m(k-2) \le w_m(k-2)\} \\ =&\cdots = (1-\xi_m)^{k-m} \end{aligned} \tag{3.3}$$

with the convention that $D_m(m) = w_m(m) = 0$.

Given a desired significance level $\alpha \in (0,1/2)$, a simple way to ensure that (1.4) holds under H_0 is to choose ξ_m such that $1-\alpha = (1-\xi_m)^{n-m}$, that is, $\xi_m = 1-(1-\alpha)^{1/(n-m)}$. As one can see from (3.2), $w_m(m+1)$ is then a quantile of order $(1-\alpha)^{1/(n-m)}$ of $D_m(m+1)$ under stationarity and, for any $k \in \{m+2,\ldots,n\}$, $w_m(k)$ is a quantile of order $(1-\alpha)^{1/(n-m)}$ of $D_m(k)$ conditionally on $D_m(m+1) \le w_m(m+1),\ldots,D_m(k-1) \le w_m(k-1)$ under stationarity. When $\alpha \in \{0.01,0.05,0.1\}$ as is typically the case, the first-order approximation $\xi_m \simeq 1 - (1 - \frac{\alpha}{n-m}) = \frac{\alpha}{n-m}$ turns out to be precise up to at least two decimals. Similarly, for all $k \in \{m+2,\ldots,n\}$,

$$\begin{aligned} \mathbb{P}\{D_m(m+1) \le w_m(m+1),\ldots,D_m(k) \le w_m(k)\} &= (1-\xi_m)^{k-m} \\ &\simeq 1 - (k-m)\xi_m \simeq 1 - \frac{k-m}{n-m}\alpha, \end{aligned}$$

where it can be verified that the resulting approximation is precise up to at least two decimals. In other words, for typical values of α, choosing the thresholds $w_m(k)$, $k \in \{m+1,\ldots,n\}$ such that (3.2) holds with $\xi_m = 1 - (1-\alpha)^{1/(n-m)}$ is almost equivalent to choosing the thresholds such that (3.1) holds (some thought reveals that the latter equivalence could be made to hold exactly by allowing ξ_m in (3.2) to change with k).

Given the precision of the aforementioned first-order approximations for typical values of α, for the sake of a simplicity, we shall base our threshold estimation procedure on (3.2). Before we discuss the estimation of the thresholds and its validity, let us give an alternative view of (3.2). In Sections 2.3 and 2.4 in which n was taken equal to $\lfloor m(T+1)\rfloor$, we saw that the asymptotic results for the detectors are given in terms of elements of $\ell^\infty([1,T+1])$. With the convention that $D_m(m) = w_m(m) = 0$, another equivalent

way of looking at sequential change-point detection procedures of the considered type is then to consider that the piecewise constant *detector function* $\mathbb{D}_m$ defined by $\mathbb{D}_m(t) = D_m(\lfloor mt \rfloor)$, $t \in [1, T+1]$, is compared to the piecewise constant *threshold function* τ_m defined by $\tau_m(t) = w_m(\lfloor mt \rfloor)$, $t \in [1, T+1]$. Let $s_k = (m+k)/m$, $k \in \{0, \ldots, n-m\}$ and define the intervals $J_k = [s_k, s_{k+1})$, $k \in \{0, \ldots, n-m-1\}$, and $J_{n-m} = [s_{n-m}, T+1]$. Some thought reveals that (3.2) is then equivalent to choosing the threshold function τ_m such that, under H_0 in (1.1), for any $k \in \{1, \ldots, n-m\}$,

$$\mathbb{P}\{\exists\, t \in J_k \text{ s.t. } \mathbb{D}_m(t) > \tau_m(t) \mid \mathbb{D}_m(t) \le \tau_m(t), \forall\, t \in J_0 \cup \cdots \cup J_{k-1}\}$$
$$= 1 - (1-\alpha)^{1/(n-m)}. \quad (3.4)$$

3.2. A formulation compatible with asymptotic validity results

With $n = \lfloor m(T+1) \rfloor$, the threshold setting procedure as given in (3.2) or (3.4) makes no sense asymptotically since the number of (conditional) probabilities tends to infinity as $m \to \infty$. A natural solution consists of keeping the number of probabilities fixed, or, equivalently, of considering a time grid that does not depend on m. Let $p \ge 1$ and let $t_0 = 1 < t_1 < \cdots < t_p = T+1$ be a fixed uniformly spaced time grid such that $T/p > 1/m$ (a condition that will always be satisfied for m large enough). Let τ_m be a piecewise constant threshold function taking the value $g_{i,m}$ on the interval $I_i = [t_{i-1}, t_i)$, $i \in \{1, \ldots, p-1\}$, and $g_{p,m}$ on the interval $I_p = [t_{p-1}, t_p]$. Mimicking (3.4), the aim is then to choose τ_m such that, under H_0 in (1.1), for any $i \in \{1, \ldots, p\}$,

$$\mathbb{P}\{\exists\, t \in I_i \text{ s.t. } \mathbb{D}_m(t) > \tau_m(t) \mid \mathbb{D}_m(t) \le \tau_m(t), \forall\, t \in I_0 \cup \cdots \cup I_{i-1}\} = 1 - (1-\alpha)^{1/p}, \quad (3.5)$$

with the convention that $I_0 = \emptyset$. Some thought reveals that the formulation in (3.5) is equivalent to choosing τ_m such that, under H_0 in (1.1),

$$\begin{cases} \mathbb{P}\left\{ \sup_{t \in I_1} \mathbb{D}_m(t) > g_{1,m} \right\} = 1 - (1-\alpha)^{1/p}, \\[2mm] \text{and, for all } i \in \{2, \ldots, p\}, \\[2mm] \mathbb{P}\left\{ \sup_{t \in I_i} \mathbb{D}_m(t) > g_{i,m} \;\middle|\; \sup_{t \in I_1} \mathbb{D}_m(t) \le g_{1,m}, \ldots, \sup_{t \in I_{i-1}} \mathbb{D}_m(t) \le g_{i-1,m} \right\} = 1 - (1-\alpha)^{1/p}. \end{cases}$$
$$(3.6)$$

In other words, $g_{1,m}$ is a quantile of order $(1-\alpha)^{1/p}$ of $\sup_{t \in I_1} \mathbb{D}_m(t)$ under stationarity and $g_{i,m}$, $i \in \{2, \ldots, p\}$, is a quantile of order $(1-\alpha)^{1/p}$ of $\sup_{t \in I_i} \mathbb{D}_m(t)$ given that $\sup_{t \in I_1} \mathbb{D}_m(t) \le g_{1,m}, \ldots, \sup_{t \in I_{i-1}} \mathbb{D}_m(t) \le g_{i-1,m}$ under stationarity. Notice that the suprema in (3.6) are actually maxima since $\mathbb{D}_m$ is a piecewise constant function.

Remark 3.1. A further generalization of (3.5) or, equivalently (3.6), would be to consider that τ_m is not necessarily piecewise constant but only defined up to a multiplicative constant on each of the intervals I_i, $i \in \{1, \ldots, p\}$. For instance, it could have one of the parametric forms considered in Dette and Gösmann (2019, Section 5), among others. For the sake of simplicity, we shall not however consider such an extension in this work.

3.3. Estimation of the threshold function

As we continue, we shall focus on the threshold setting procedure as formulated in (3.5) or, equivalently, (3.6), mostly because its asymptotic validity can be studied. To estimate the threshold function τ_m in (3.5), or, equivalently, the $g_{i,m}$, $i \in \{1, \ldots, p\}$, in (3.6), it is thus necessary to be able to compute, at least approximately, the distribution of the p-dimensional random vector

$$\left(\sup_{t \in I_1} \mathbb{D}_m(t), \ldots, \sup_{t \in I_p} \mathbb{D}_m(t) \right). \tag{3.7}$$

3.3.1. Monte Carlo estimation and asymptotic validity

Assume that the observations to be monitored are univariate and independent, and that $\mathbb{D}_m$ is distribution-free under H_0^{ind} in (1.2). Notice that the latter implies that so is the random vector (3.7). To obtain a Monte Carlo estimate of the distribution of (3.7), it then suffices to consider a large integer M, generate M independent samples $U_1^{[s]}, \ldots, U_n^{[s]}$, $s \in \{1, \ldots, M\}$, of size n from the standard uniform distribution and compute the corresponding realizations $\mathbb{D}_m^{[s]}$, $s \in \{1, \ldots, M\}$, of $\mathbb{D}_m$. The latter can be used to obtain a Monte Carlo estimate τ_m^M of the threshold function τ_m. More formally, let

$$g_{i,m}^M = F_{\mathbb{D}_m,i}^{M,-1}\{(1-\alpha)^{1/p}\}, \qquad i \in \{1, \ldots, p\},$$

where $F_{\mathbb{D}_m,1}^M$ is the empirical d.f. of the sample $\sup_{t \in I_1} \mathbb{D}_m^{[1]}(t), \ldots, \sup_{t \in I_1} \mathbb{D}_m^{[M]}(t)$, for any $i \in \{2, \ldots, p\}$ and $x \in \mathbb{R}$,

$$F_{\mathbb{D}_m,i}^M(x) = \frac{\sum_{s=1}^M \mathbf{1}\left\{ \sup_{t \in I_i} \mathbb{D}_m^{[s]}(t) \leq x, \sup_{t \in I_1} \mathbb{D}_m^{[s]}(t) \leq g_{1,m}^M, \ldots, \sup_{t \in I_{i-1}} \mathbb{D}_m^{[s]}(t) \leq g_{i-1,m}^M \right\}}{\sum_{s=1}^M \mathbf{1}\left\{ \sup_{t \in I_1} \mathbb{D}_m^{[s]}(t) \leq g_{1,m}^M, \ldots, \sup_{t \in I_{i-1}} \mathbb{D}_m^{[s]}(t) \leq g_{i-1,m}^M \right\}},$$

and $F_{\mathbb{D}_m,i}^{M,-1}$, $i \in \{1, \ldots, p\}$, are the associated quantile functions generically defined by (2.11). Notice that, in this particular case, the resulting estimate τ_m^M of the threshold function τ_m does not at all depend on the learning sample.

By taking a sufficiently large M, the Monte Carlo estimates $g_{i,m}^M$, $i \in \{1, \ldots, p\}$, can be made arbitrarily close to the quantiles $g_{i,m} = F_{\mathbb{D}_m,i}^{-1}\{(1-\alpha)^{1/p}\}$, $i \in \{1, \ldots, p\}$, where $F_{\mathbb{D}_m,1}$ is the d.f. of $\sup_{t \in I_1} \mathbb{D}_m(t)$, and $F_{\mathbb{D}_m,i}$, $i \in \{2, \ldots, p\}$, is the d.f. of $\sup_{t \in I_i} \mathbb{D}_m(t)$ given that $\sup_{t \in I_j} \mathbb{D}_m(t) \leq g_{j,m}$ for all $j \in \{1, \ldots, i-1\}$. Interestingly enough more can be said as a consequence of the fact that Monte Carlo simulation can be regarded as a particular resampling scheme. As shall become clear in the next section, the general result stated in Theorem 3.3 hereafter can actually be used to show the asymptotic validity of the Monte Carlo based threshold estimation procedure when both m and M tend to infinity, under both H_0^{ind} in (1.2) and sequences of alternatives related to H_1 in (1.3). This is discussed in more detail in Remark 3.4 below.

3.3.2. Bootstrap-based estimation and asymptotic validity

In settings in which $\mathbb{D}_m$ is not distribution-free anymore, a natural alternative is to rely on a resampling scheme making use of the available learning sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ known to be under H_0 in (1.1). Specifically, let B be a large integer and suppose that we have available *bootstrap replicates* $\mathbb{D}_m^{[b]}$, $b \in \{1, \ldots, B\}$, of $\mathbb{D}_m$ computed from $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ and depending on additional sources of randomness involved in the resampling scheme. Mimicking the previous situation in which $\mathbb{D}_m$ was distribution-free, let

$$g_{j,m}^B = F_{\mathbb{D}_{m,j}}^{B,-1}\{(1-\alpha)^{1/p}\}, \qquad j \in \{1, \ldots, p\},$$

where $F_{\mathbb{D}_{m,1}}^B$ is the empirical d.f. of the sample $\sup_{t \in I_1} \mathbb{D}_m^{[1]}(t), \ldots, \sup_{t \in I_1} \mathbb{D}_m^{[B]}(t)$, and, for any $j \in \{2, \ldots, p\}$ and $x \in \mathbb{R}$,

$$F_{\mathbb{D}_{m,j}}^B(x) = \frac{\sum_{b=1}^B \mathbf{1}\left\{\sup_{t \in I_j} \mathbb{D}_m^{[b]}(t) \leq x, \sup_{t \in I_1} \mathbb{D}_m^{[b]}(t) \leq g_{1,m}^B, \ldots, \sup_{t \in I_{j-1}} \mathbb{D}_m^{[b]}(t) \leq g_{j-1,m}^B\right\}}{\sum_{b=1}^B \mathbf{1}\left\{\sup_{t \in I_1} \mathbb{D}_m^{[b]}(t) \leq g_{1,m}^B, \ldots, \sup_{t \in I_{j-1}} \mathbb{D}_m^{[b]}(t) \leq g_{j-1,m}^B\right\}}.$$

As we shall see below, the main result of this section is that, essentially, as soon as the underlying resampling scheme for $\mathbb{D}_m$ is consistent, the above bootstrap-based version of the threshold setting procedure (3.6) is asymptotically valid in the sense that, under H_0, $\mathbb{P}(\mathbb{D}_m \leq \tau_m^B) \to 1 - \alpha$ as $m, B \to \infty$, where τ_m^B is the estimated bootstrap-based piecewise constant threshold function taking the value $g_{i,m}^B$ on the interval $I_i = [t_{i-1}, t_i)$, $i \in \{1, \ldots, p-1\}$, and $g_{p,m}^B$ on the interval $I_p = [t_{p-1}, t_p]$.

Remark 3.2. Following for instance van der Vaart and Wellner (2000, Section 3.6) or Kosorok (2008, Section 2.2.3), a resampling scheme for $\mathbb{D}_m$ is typically considered consistent, if, informally, "$\mathbb{D}_m^{[1]}$ converges weakly to the weak limit of $\mathbb{D}_m$ in $\ell^\infty([0, T+1])$ conditionally on $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ in probability". A rigorous definition of the underlying mode of convergence is more subtle than that of weak convergence. From Lemma 3.1 of Bücher and Kojadinovic (2019), the aforementioned validity statement is actually equivalent to the joint unconditional weak convergence of $\mathbb{D}_m$ and two bootstrap replicates to independent copies of the same limit. Throughout the paper, all our bootstrap asymptotic validity results will take that form.

The following general result is proved in Appendix C.

Theorem 3.3. *Assume that, under H_0 in* (1.1),

$$(\mathbb{D}_m, \mathbb{D}_m^{[1]}, \mathbb{D}_m^{[2]}) \rightsquigarrow (\mathbb{D}_F, \mathbb{D}_F^{[1]}, \mathbb{D}_F^{[2]}) \tag{3.8}$$

in $\{\ell^\infty([1, T+1])\}^3$, *where* $\mathbb{D}_F$ *is the weak limit of* $\mathbb{D}_m$ *and* $\mathbb{D}_F^{[1]}$ *and* $\mathbb{D}_F^{[2]}$ *are independent copies of* $\mathbb{D}_F$. *Assume furthermore that* $\left(\sup_{t \in I_1} \mathbb{D}_F(t), \ldots, \sup_{t \in I_p} \mathbb{D}_F(t)\right)$ *has a continuous d.f. Then, under H_0 in* (1.1), *as* $m, B \to \infty$,

$$\mathbb{P}\left\{\sup_{t \in I_1} \mathbb{D}_m(t) \leq g_{1,m}^B\right\} \to (1-\alpha)^{1/p}, \tag{3.9}$$

and, for any $i \in \{2, \ldots, p\}$,

$$\mathbb{P}\left\{\sup_{t \in I_i} \mathbb{D}_m(t) \leq g_{i,m}^B \;\Big|\; \sup_{t \in I_1} \mathbb{D}_m(t) \leq g_{1,m}^B, \ldots, \sup_{t \in I_{i-1}} \mathbb{D}_m(t) \leq g_{i-1,m}^B\right\} \to (1-\alpha)^{1/p}.$$

(3.10)

As a consequence, on one hand, under H_0, $\mathbb{P}(\mathbb{D}_m \leq \tau_m^B) \to 1 - \alpha$ as $m, B \to \infty$ and, on the other hand, when $\sup_{t \in [1,T+1]} \mathbb{D}_m(t) \xrightarrow{\mathbb{P}} \infty$,

$$\mathbb{P}\{\exists\, t \in [1, T+1] \text{ s.t. } \mathbb{D}_m(t) > \tau_m^B(t)\} \to 1 \qquad as\ m, B \to \infty.$$

Remark 3.4. Consider the Monte Carlo estimation setting of Section 3.3.1 in which the observations to be monitored are univariate independent and $\mathbb{D}_m$ is distribution-free. Then, the weak convergence $\mathbb{D}_m \rightsquigarrow \mathbb{D}_F$ in $\ell^\infty([1, T+1])$ under H_0 immediately implies (3.8), where $\mathbb{D}_m^{[1]}$ and $\mathbb{D}_m^{[2]}$ are (independent) Monte Carlo replicates of $\mathbb{D}_m$. Hence, as a consequence of Theorem 3.3, the asymptotic validity under H_0^{ind} in (1.2) of the sequential change-point detection procedure based on $\mathbb{D}_m$ and the Monte Carlo estimated threshold function τ_m^M defined in Section 3.3.1 is an immediate corollary of the weak convergence under the null of $\mathbb{D}_m$ if $\left(\sup_{t \in I_1} \mathbb{D}_F(t), \ldots, \sup_{t \in I_p} \mathbb{D}_F(t)\right)$ has a continuous d.f.

4. Threshold function estimation for the detectors based on empirical d.f.s

The aim of this section is to apply the generic results of the previous section to estimate the threshold functions for the empirical d.f.-based detector functions $\mathbb{R}_{m,q}$, $\mathbb{S}_{m,q}$ and $\mathbb{T}_{m,q}$ defined in (2.20). We distinguish two situations for the observations to be monitored: the independent univariate case and the possibly multivariate, time series case.

4.1. Monte Carlo estimation in the independent univariate case

As verified in Section 2.2, the detector functions $\mathbb{R}_{m,q}$, $\mathbb{S}_{m,q}$ and $\mathbb{T}_{m,q}$ defined in (2.20) are margin-free under H_0 in (1.1). In the univariate case, they are thus distribution-free. When dealing with independent univariate observations, one can therefore proceed exactly as explained in Section 3.3.1 to estimate the corresponding threshold functions. Furthermore, from Proposition 2.5, Remark 3.4 and Proposition 2.7, we know that the assumptions of Theorem 3.3 are satisfied. The latter then implies that the corresponding sequential change-point detection procedures are asymptotically valid both under H_0^{ind} in (1.2) and sequences of alternatives related to H_1 in (1.3).

4.2. A dependent multiplier bootstrap in the time series case

When the monitored observations are multivariate or exhibit serial dependence, the approach considered in Section 4.1 is not meaningful anymore. Having the asymptotic

results of Sections 2.3 and 3.3.2 in mind, our aim in the considered time series context is to define suitable bootstrap replicates of $\mathbb{B}_m$ in (2.21) such that, following Remark 3.2, $\mathbb{B}_m$ and two of its replicates jointly weakly converge to independent copies of the process $\mathbb{B}_C$ defined in Proposition 2.5. Subsequently defining corresponding bootstrap replicates of the detectors functions $\mathbb{R}_{m,q}$, $\mathbb{S}_{m,q}$ and $\mathbb{T}_{m,q}$ defined in (2.20) will lead to asymptotically valid corresponding sequential change-point detection procedures.

Following Bühlmann (1993, Section 3.3) and Bücher and Kojadinovic (2016), we opted for a *dependent multiplier bootstrap* in the considered time series context. In the rest of the paper, we say that a sequence of random variables $(\xi_{i,m})_{i\in\mathbb{Z}}$ is a *dependent multiplier sequence* if:

(M1) The sequence $(\xi_{i,m})_{i\in\mathbb{Z}}$ is stationary, independent of the available learning sample $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_m$ and satisfies $\mathbb{E}(\xi_{0,m}) = 0$, $\mathbb{E}(\xi_{0,m}^2) = 1$ and $\sup_{m\geq 1}\mathbb{E}(|\xi_{0,m}|^\nu) < \infty$ for all $\nu \geq 1$.

(M2) There exists a sequence $\ell_m \to \infty$ of strictly positive constants such that $\ell_m = o(m)$ and the sequence $(\xi_{i,m})_{i\in\mathbb{Z}}$ is ℓ_m-dependent, i.e., $\xi_{i,m}$ is independent of $\xi_{i+h,m}$ for all $h > \ell_m$ and $i \in \mathbb{N}$.

(M3) There exists a function $\varphi : \mathbb{R} \to [0,1]$, symmetric around 0, continuous at 0, satisfying $\varphi(0) = 1$ and $\varphi(x) = 0$ for all $|x| > 1$ such that $\mathbb{E}(\xi_{0,m}\xi_{h,m}) = \varphi(h/\ell_m)$ for all $h \in \mathbb{Z}$.

Let $(\xi_{i,m}^{[b]})_{i\in\mathbb{Z}}$, $b \in \mathbb{N}$, be independent copies of the same dependent multiplier sequence. If we had a learning sample of size $n = \lfloor m(T+1)\rfloor$, following Bücher and Kojadinovic (2016), a natural definition of a *dependent multiplier replicate* of $\mathbb{B}_m$ in (2.21) would be

$$\check{\mathbb{B}}_m^{[b]}(s,\boldsymbol{u}) = \frac{1}{\sqrt{m}}\sum_{i=1}^{\lfloor ms\rfloor}\xi_{i,m}^{[b]}\{\mathbf{1}(\boldsymbol{U}_i \leq \boldsymbol{u}) - C_{1:n}(\boldsymbol{u})\}, \qquad s \in [0,T+1], \boldsymbol{u} \in [0,1]^d, b \in \mathbb{N},$$
(4.1)

where $C_{1:n}$ is generically defined by (2.12). Since threshold functions need to be estimated prior to the beginning of the monitoring and the learning sample is only of size m, we consider a time-rescaled version of $\check{\mathbb{B}}_m^{[b]}$ in which, roughly, $m' = \lfloor(m/n)m\rfloor \simeq m/(T+1)$ and m play the role of m and n, respectively. Hence, in the considered context, our definition of a dependent multiplier replicate of $\mathbb{B}_m$ is

$$\hat{\mathbb{B}}_m^{[b]}(s,\boldsymbol{u}) = \frac{1}{\sqrt{m'}}\sum_{i=1}^{\lfloor m's\rfloor}\xi_{i,m}^{[b]}\{\mathbf{1}(\boldsymbol{U}_i \leq \boldsymbol{u}) - C_{1:m}(\boldsymbol{u})\}, \qquad s \in [0,T+1], \boldsymbol{u} \in [0,1]^d, b \in \mathbb{N},$$
(4.2)

thereby translating the fact that we can only rely on functionals computed from the learning sample to approximate the variability of the detector functions under the null.

From the two previous displays, we see that the multipliers act as random weights and that the bandwidth ℓ_m defined in Assumption (M2) plays a role somehow similar to that of the *block length* in the block bootstrap of Künsch (1989). Note that, in our Monte Carlo experiments to be presented in Section 5, ℓ_m was estimated from the learning sample $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_m$ as explained in detail in Section 5.1 of Bücher and Kojadinovic

[(2016)](#) while corresponding dependent multiplier sequences were generated using the so-called *moving average approach* based on an initial standard normal random sample and Parzen's kernel as precisely described in Section 5.2 of the same reference.

The latter construction based on a time-rescaling suggests to form a dependent multiplier replicate of $\mathbb{G}_m$ in [(2.22)](#) as

$$\hat{\mathbb{G}}_m^{[b]}(s,t,\boldsymbol{u}) = \lambda_{m'}(0,t)\hat{\mathbb{B}}_m^{[b]}(s,\boldsymbol{u}) - \lambda_{m'}(0,s)\hat{\mathbb{B}}_m^{[b]}(t,\boldsymbol{u}), \qquad (s,t) \in \Delta, \boldsymbol{u} \in [0,1]^d, b \in \mathbb{N},$$

with its weighted version being

$$\hat{\mathbb{G}}_{m,q}^{[b]}(s,t,\boldsymbol{u}) = \frac{\hat{\mathbb{G}}_m^{[b]}(s,t,\boldsymbol{u})}{q\{\lambda_{m'}(0,s),\lambda_{m'}(0,t)\}}, \qquad (s,t) \in \Delta, \boldsymbol{u} \in [0,1]^d, b \in \mathbb{N},$$

where $\lambda_{m'}$ is defined as in [(2.14)](#). Finally, for any $b \in \mathbb{N}$ and $t \in [1, T+1]$, let

$$\hat{\mathbb{R}}_{m,q}^{[b]}(t) = \sup_{s\in[1,t]} \sup_{\boldsymbol{u}\in[0,1]^d} |\hat{\mathbb{G}}_{m,q}^{[b]}(s,t,\boldsymbol{u})|,$$

$$\hat{\mathbb{S}}_{m,q}^{[b]}(t) = \sup_{s\in[1,t]} \int_{[0,1]^d} \{\hat{\mathbb{G}}_{m,q}^{[b]}(s,t,\boldsymbol{u})\}^2 \mathrm{d}C_{1:\lfloor m't\rfloor}(\boldsymbol{u}), \qquad (4.3)$$

$$\hat{\mathbb{T}}_{m,q}^{[b]}(t) = \int_1^t \int_{[0,1]^d} \{\hat{\mathbb{G}}_{m,q}^{[b]}(s,t,\boldsymbol{u})\}^2 \mathrm{d}C_{1:\lfloor m't\rfloor}(\boldsymbol{u})\mathrm{d}s$$

be dependent multiplier replicates of $\mathbb{R}_{m,q}$, $\mathbb{S}_{m,q}$ and $\mathbb{T}_{m,q}$, respectively, defined in [(2.20)](#), where $C_{1:\lfloor m't\rfloor}$ is defined generically by [(2.12)](#).

The definitions given in [(4.3)](#) hide the fact that the proposed dependent multipliers replicates of the detector functions $\mathbb{R}_{m,q}$, $\mathbb{S}_{m,q}$ and $\mathbb{T}_{m,q}$ actually depend on the learning sample $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_m$. To verify that this is the case, for any $b \in \mathbb{N}$, $(s,t) \in \Delta$ and $\boldsymbol{x} \in \mathbb{R}^d$, let $\hat{\mathbb{F}}_m^{[b]}(s,\boldsymbol{x}) = \hat{\mathbb{B}}_m^{[b]}\{s, \boldsymbol{F}(\boldsymbol{x})\}$, where $\boldsymbol{F}(\boldsymbol{x}) = (F_1(x_1),\ldots,F_d(x_d))$,

$$\hat{\mathbb{H}}_m^{[b]}(s,t,\boldsymbol{x}) = \hat{\mathbb{G}}_m^{[b]}\{s,t,\boldsymbol{F}(\boldsymbol{x})\} = \lambda_{m'}(0,t)\hat{\mathbb{F}}_m^{[b]}(s,\boldsymbol{x}) - \lambda_{m'}(0,s)\hat{\mathbb{F}}_m^{[b]}(t,\boldsymbol{x})$$

and

$$\hat{\mathbb{H}}_{m,q}^{[b]}(s,t,\boldsymbol{x}) = \hat{\mathbb{G}}_{m,q}^{[b]}\{s,t,\boldsymbol{F}(\boldsymbol{x})\} = \frac{\hat{\mathbb{H}}_m^{[b]}(s,t,\boldsymbol{x})}{q\{\lambda_{m'}(0,s),\lambda_{m'}(0,t)\}}.$$

Since [(2.13)](#) always holds for all $i \in \{1,\ldots,m\}$, we immediately obtain that, for any $b \in \mathbb{N}$,

$$\hat{\mathbb{F}}_m^{[b]}(s,\boldsymbol{x}) = \frac{1}{\sqrt{m'}} \sum_{i=1}^{\lfloor m's\rfloor} \xi_{i,m}^{[b]}\{\mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{x}) - F_{1:m}(\boldsymbol{x})\}, \qquad s \in [0, T+1], \boldsymbol{x} \in \mathbb{R}^d,$$

where $F_{1:m}$ is generically defined by [(2.1)](#), and furthermore that, for any $t \in [1, T+1]$,

$$\hat{\mathbb{R}}_{m,q}^{[b]}(t) = \sup_{s\in[1,t]} \sup_{\boldsymbol{x}\in\mathbb{R}^d} |\hat{\mathbb{H}}_{m,q}^{[b]}(s,t,\boldsymbol{x})|,$$

$$\hat{\mathbb{S}}_{m,q}^{[b]}(t) = \sup_{s\in[1,t]} \int_{\mathbb{R}^d} \{\hat{\mathbb{H}}_{m,q}^{[b]}(s,t,\boldsymbol{x})\}^2 \mathrm{d}F_{1:\lfloor m't\rfloor}(\boldsymbol{x}),$$

$$\hat{\mathbb{T}}_{m,q}^{[b]}(t) = \int_1^t \int_{\mathbb{R}^d} \{\hat{\mathbb{H}}_{m,q}^{[b]}(s,t,\boldsymbol{x})\}^2 \mathrm{d}F_{1:\lfloor m't\rfloor}(\boldsymbol{x})\mathrm{d}s.$$

The following result is proven in Appendix D.

Proposition 4.1. *Assume that H_0 in* (1.1) *holds and that, additionally, $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ is a stretch from a stationary sequence $(\boldsymbol{X}_i)_{i \in \mathbb{Z}}$ of continuous d-dimensional random vectors whose strong mixing coefficients satisfy $\alpha_r^{\boldsymbol{X}} = O(r^{-a})$ for some $a > 3 + 3d/2$ as $r \to \infty$. If $\ell_m = O(m^{1/2-\varepsilon})$ for some $0 < \varepsilon < 1/2$, then*

$$(\mathbb{B}_m, \hat{\mathbb{B}}_m^{[1]}, \hat{\mathbb{B}}_m^{[2]}) \rightsquigarrow (\mathbb{B}_C, \mathbb{B}_C^{[1]}, \mathbb{B}_C^{[2]})$$

in $\{\ell^\infty([0, T+1] \times [0,1]^d)\}^3$, where $\mathbb{B}_m$ is defined in (2.21), *$\hat{\mathbb{B}}_m^{[1]}$ and $\hat{\mathbb{B}}_m^{[2]}$ are defined in* (4.2), *$\mathbb{B}_C$ is the weak limit of $\mathbb{B}_m$ defined in Proposition* 2.5, *and $\mathbb{B}_C^{[1]}$ and $\mathbb{B}_C^{[2]}$ are independent copies of $\mathbb{B}_C$.*

As a consequence,

$$(\mathbb{R}_{m,q}, \hat{\mathbb{R}}_{m,q}^{[1]}, \hat{\mathbb{R}}_{m,q}^{[2]}) \rightsquigarrow (\mathbb{R}_{C,q}, \mathbb{R}_{C,q}^{[1]}, \mathbb{R}_{C,q}^{[2]}), \qquad (\mathbb{S}_{m,q}, \hat{\mathbb{S}}_{m,q}^{[1]}, \hat{\mathbb{S}}_{m,q}^{[2]}) \rightsquigarrow (\mathbb{S}_{C,q}, \mathbb{S}_{C,q}^{[1]}, \mathbb{S}_{C,q}^{[2]}),$$
$$(\mathbb{T}_{m,q}, \hat{\mathbb{T}}_{m,q}^{[1]}, \hat{\mathbb{T}}_{m,q}^{[2]}) \rightsquigarrow (\mathbb{T}_{C,q}, \mathbb{T}_{C,q}^{[1]}, \mathbb{T}_{C,q}^{[2]}), \qquad\qquad in \; \{\ell^\infty([1, T+1])\}^3,$$

where $\hat{\mathbb{R}}_{m,q}^{[b]}$, $\hat{\mathbb{S}}_{m,q}^{[b]}$ and $\hat{\mathbb{T}}_{m,q}^{[b]}$, $b \in \{1,2\}$, are defined in (4.3), *$\mathbb{R}_{C,q}$, $\mathbb{S}_{C,q}$ and $\mathbb{T}_{C,q}$ are defined in* (2.25), *and $\mathbb{R}_{C,q}^{[b]}$, $\mathbb{S}_{C,q}^{[b]}$ and $\mathbb{T}_{C,q}^{[b]}$, $b \in \{1,2\}$, are independent copies of $\mathbb{R}_{C,q}$, $\mathbb{S}_{C,q}$ and $\mathbb{T}_{C,q}$, respectively.*

The last claims of the previous proposition along with the last claim of Proposition 2.5 and Proposition 2.7 are the assumptions of Theorem 3.3 for $\mathbb{D}_m \in \{\mathbb{R}_{m,q}, \mathbb{S}_{m,q}, \mathbb{T}_{m,q}\}$. It follows that the sequential change-point detection procedures based on these detector functions carried out as explained in Section 3.3.2 using the above dependent multiplier replicates are asymptotically valid under H_0 in (1.1) and sequences of alternatives related to H_1 in (1.3). Note that, in practice, since in the considered approach $m' = \lfloor (m/n)m \rfloor$ and m play the role of m and n, respectively, the largest possible value for p, the number of steps of the estimated threshold function τ_m^B, is $m - m'$ and, in this case, each of the p estimated thresholds covers approximately $\lfloor n/m \rfloor$ time steps in the monitoring.

5. Monte Carlo experiments

Large-scale Monte Carlo experiments were carried out to investigate the finite-sample properties of the studied sequential change-point detection procedures. The aim was in particular to try to answer the following questions:

- How well do the procedures hold their level, in particular, when the threshold functions are estimated using the dependent multiplier bootstrap of Section 4.2?
- What is the influence of the number of steps p of the estimated threshold function (see Sections 3.2 and 3.3) on the distribution of the false alarms?
- What is the effect of p on the power and the *mean detection delay* (the latter is the expectation under H_1 of the difference between the time at which the change was detected and the time $k^\star$ at which the change really occurred)?

TABLE 5.1

Percentages of rejection of H_0 in (1.1) for the sequential change-point detection procedures when based on the Monte Carlo estimation method with $M = 10^5$ as described in Sections 3.3.1 and 4.1 for $p \in \{1, 2, 4, 10, 50\}$, $m \in \{50, 100\}$ and $n = 2m$. The rejection percentages are computed from 10^4 samples of size $n = 2m$ generated from the standard uniform distribution.

		$T_{m,q}$			$S_{m,q}$			$R_{m,q}$				
p	m	$\gamma=0$	$\gamma=0.25$	$\gamma=0.5$	$\gamma=0$	$\gamma=0.25$	$\gamma=0.5$	$\gamma=0$	$\gamma=0.25$	$\gamma=0.5$	Q_m	P_m
1	50	5.2	5.2	5.1	4.9	5.0	4.9	4.7	4.9	4.7	5.2	5.2
	100	4.9	4.9	4.6	4.8	4.8	5.0	4.9	4.8	4.7	5.1	5.0
2	50	4.9	5.1	5.0	4.8	5.2	5.1	4.9	4.9	4.9	5.1	4.4
	100	4.9	4.8	4.9	4.9	4.7	4.9	4.9	4.8	5.1	5.0	4.6
4	50	4.9	4.9	5.1	4.6	4.9	5.3	4.6	4.9	5.0	5.1	5.2
	100	5.0	5.0	5.0	4.8	4.9	5.3	4.9	4.9	4.9	5.0	4.9
10	50	5.2	5.1	5.0	4.9	4.9	5.2	4.6	4.7	5.0	5.0	4.8
	100	5.0	5.1	5.1	4.9	4.9	5.1	5.0	5.1	5.0	4.9	4.6
50	50	5.0	5.1	5.1	4.8	4.9	5.1	4.3	4.6	4.9	4.9	4.5
	100	5.0	4.9	5.1	4.8	4.9	5.0	5.0	4.9	4.8	4.9	4.7

- What is the effect of the parameter γ appearing in the expression of the weight function q defined in (2.4) on the power and mean detection delay?
- How do the detectors $R_{m,q}$, $S_{m,q}$, $T_{m,q}$, P_m and Q_m defined in (2.3), (2.5), (2.6), (2.7) and (2.8), respectively, compare in terms of power and mean detection delay?
- How do the derived procedures compare with similar, more specialized procedures in terms of power and mean detection delay?

We tried to answer these questions in detail in the univariate independent case when the estimation of the threshold functions of the sequential change-point detection procedures can be rightfully so based on the Monte Carlo approach described in Sections 3.3.1 and 4.1. When the observations to be monitored are not univariate or independent so that resampling as described in Section 4.2 is needed for the estimation of the threshold functions, we essentially investigated how well the procedures hold their level depending on the underlying data generating mechanism. Although many other questions could be formulated given the complexity of the problem, the following experiments should allow the reader to grasp the main finite-sample properties of the studied procedures.

5.1. Monte Carlo estimation in the independent univariate case

As already discussed in Section 3.3.1, the estimation of the threshold functions when monitoring independent univariate observations can be made arbitrarily precise by increasing the number M of Monte Carlo samples. We used the setting $M = 10^5$ in our experiments and estimated all the rejection percentages from 10^4 samples. The change-point detection procedures were always carried out at the $\alpha = 5\%$ nominal level.

5.1.1. Under the null

Unsurprisingly, all the studied tests were found to hold their level very well as can for instance be seen by inspecting the rejection percentages reported in Table 5.1. Further-
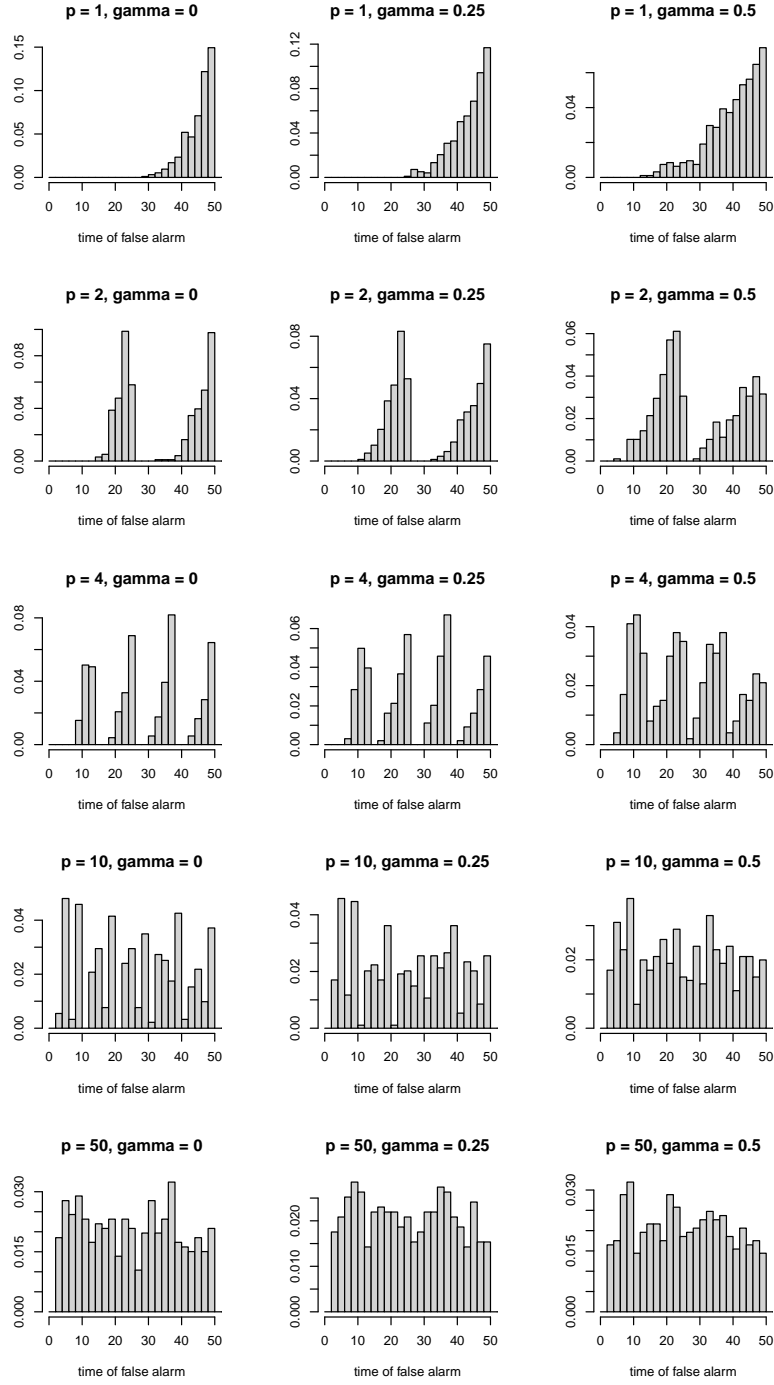
FIG 5.1. *For the procedure based on $R_{m,q}$ in (2.3) and Monte Carlo estimation with $M = 10^5$, distribution of the false alarms (that is, of the time of rejection of H_0 in (1.1)) for $m = 50$ and $n = 100$ obtained from 10^4 samples of size n from the standard uniform distribution. The histograms for $S_{m,q}$ in (2.5) and $T_{m,q}$ in (2.6) are very similar.*
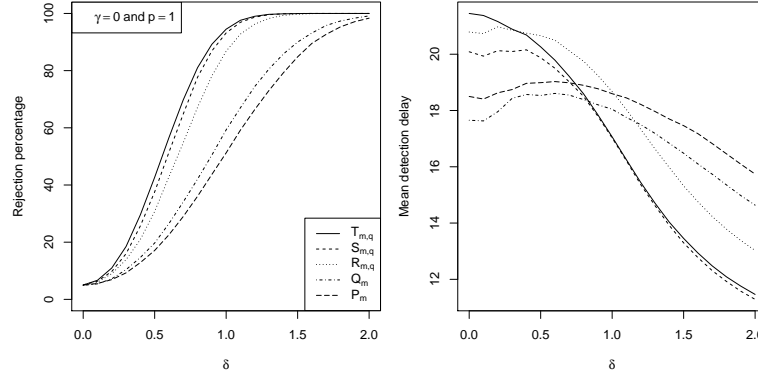
FIG 5.2. *Left: estimated rejection probabilities of H_0^{ind} in* (1.2) *under H_1 in* (1.3) *with $m = 50$, $k^\star = 75$, $n = 100$, F the d.f. of the standard normal and G the d.f. of the $N(\delta, 1)$. Right: corresponding mean detection delays. The value of γ in* (2.4) *is 0. The number of steps in the threshold functions is $p = 1$.*

more, as could have been expected given the fact that the studied detector functions have a tendency to be increasing on average, it was observed that the setting $p = 1$ resulted in a concentration of false alarms at the end of the monitoring period, while the larger p, the more uniform the distribution of the false alarms over the monitoring period. These unsurprising findings are for instance illustrated in Figure 5.1 for the procedure based on $R_{m,q}$ in (2.3).

5.1.2. Change in mean

To answer the aforementioned questions related to the behavior of the procedures under H_1 in (1.3), we first considered a simple experiment consisting of a change in the expectation of a normal distribution. Specifically, m, $k^\star$, n, F and G in H_1 were taken equal to 50, 75, 100, the d.f. of the standard normal and the d.f. of the $N(\delta, 1)$, respectively. The left graph in Figure 5.2 displays the estimated rejection percentages for the five detectors $R_{m,q}$, $S_{m,q}$, $T_{m,q}$, P_m and Q_m with γ in (2.4) set to zero and $p = 1$. The right graph represents the corresponding mean detection delays which were estimated only from the samples for which neither a false alarm was obtained (which occurs when the detector function becomes larger than the threshold function before the time of change $k^\star = 75$) nor the change was undetected (which occurs when the detector function remains below the threshold function during the entire monitoring period). Because the number of steps in the threshold function was set to $p = 1$, the left graph of Figure 5.2 is directly comparable with the top left graph given in Figure 1 of Dette and Gösmann (2019). An inspection of the latter seems to indicate that the powers of the procedures based on $R_{m,q}$, $S_{m,q}$ and $T_{m,q}$ are not substantially different from those of the mean-specialized procedures considered in Section 5.1 of Dette and Gösmann (2019), even though the detectors $R_{m,q}$, $S_{m,q}$ and $T_{m,q}$ are not specifically designed to be sensitive to changes in the expectation. The graphs are not substantially different for other values of γ and p. Overall, the procedures based on $R_{m,q}$, $S_{m,q}$ and $T_{m,q}$ were always observed
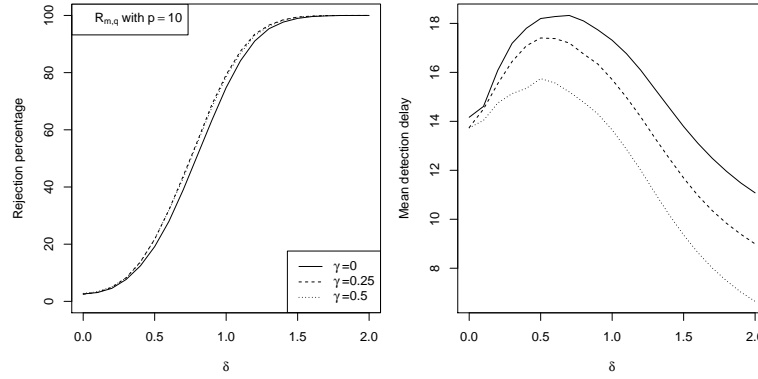
FIG 5.3. *Left: estimated rejection probabilities of H_0^{ind} in (1.2) under H_1 in (1.3) with $m = 50$, $k^\star = 75$, $n = 100$, F the d.f. of the standard normal and G the d.f. of the $N(\delta, 1)$ for the procedure based on $R_{m,q}$ with $p = 10$. Right: corresponding mean detection delays.*
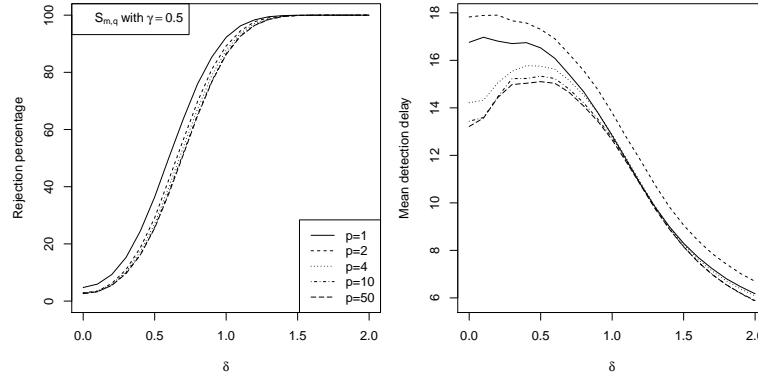

FIG 5.4. *Left: estimated rejection probabilities of H_0^{ind} in (1.2) under H_1 in (1.3) with $m = 50$, $k^\star = 75$, $n = 100$, F the d.f. of the standard normal and G the d.f. of the $N(\delta, 1)$ for the procedure based on $S_{m,q}$ with $\gamma = 0.5$. Right: corresponding mean detection delays.*

to be more powerful and superior in terms of mean detection delay than those based on P_m and Q_m. The latter is in full accordance with the empirical observations of Dette and Gösmann (2019) for more specialized procedures. Note that the procedure based on $T_{m,q}$ seems the most powerful for the alternative under consideration.

Figure 5.3 highlights the influence of the parameter γ in (2.4) on the rejection percentages and the mean detection delays of the procedure based on $R_{m,q}$ with $p = 10$. The graphs are very similar for other values of p or for the procedures based on $S_{m,q}$ and $T_{m,q}$. The conclusion is the same for all three procedures. For a change in expectation, while the parameter γ does not seem to affect the power of the procedures much, it has a clear influence on the mean detection delay: the greater γ, the shorter the mean detection delay. As we shall see, this behavior is not true for all types of alternatives.

Figure 5.4 displays the influence of the number of steps p of the threshold function
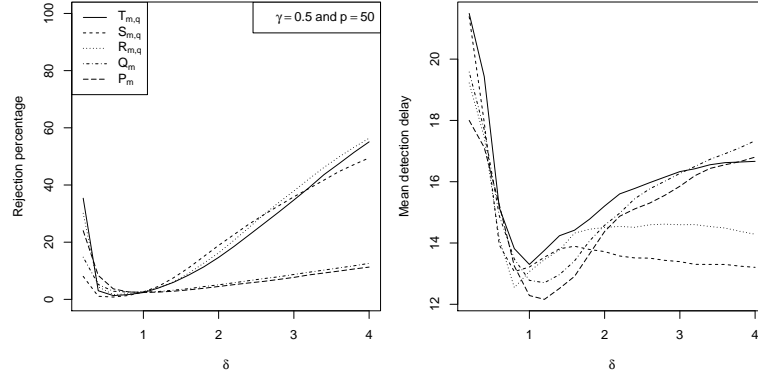
Fig 5.5. *Left: estimated rejection probabilities of H_0^{ind} in (1.2) under H_1 in (1.3) with $m = 50$, $k^\star = 75$, $n = 100$, F the d.f. of the standard normal and G the d.f. of the $N(0, \delta^2)$. Right: corresponding mean detection delays. The value of γ in (2.4) is 0.5. The number of steps in the threshold functions is $p = 50$.*

on the rejection percentages and the mean detection delays of the procedure based on $S_{m,q}$ with $\gamma = 0.5$. The graphs are not qualitatively different for other values of γ or for the procedures based on $R_{m,q}$ and $T_{m,q}$. Overall, the procedures with $p = 1$ have the highest rejection percentages. The latter is due to the fact that, because $k^\star = 75$, detections occur mostly at the end of the monitoring period, and, at the end of the monitoring interval, by construction, the threshold functions for $p = 1$ are below the corresponding threshold functions obtained for larger values of p.

5.1.3. Change in variance

The setting is similar to that of the previous experiment except that, this time, it is the variance of the normal distribution that changes from 1 to δ^2. The left graph in Figure 5.5 displays the estimated rejection percentages for the procedures based on $R_{m,q}$, $S_{m,q}$, $T_{m,q}$, P_m and Q_m with $\gamma = 0.5$ and $p = 50$. The graph on the right represents the corresponding mean detection delays. Again, the procedures based on $R_{m,q}$, $S_{m,q}$ and $T_{m,q}$ appear to be substantially more powerful and superior in terms of mean detection delay than those based on P_m and Q_m. The conclusion remains true for all values of γ and p. The influence of γ and p is of the same nature as in the case of a change in mean: the greater γ, the shorter the mean detection delay and the lower p, the higher the power.

5.1.4. Change in distribution

As a final experiment for independent univariate observations, we considered a change in distribution that keeps the expectation and the variance constant. Specifically, F and G in H_1 were taken equal to the d.f. of the $N(1, 2)$ and the d.f. of the Gamma distribution whose shape and rate parameters are both equal to $1/2$, respectively. The
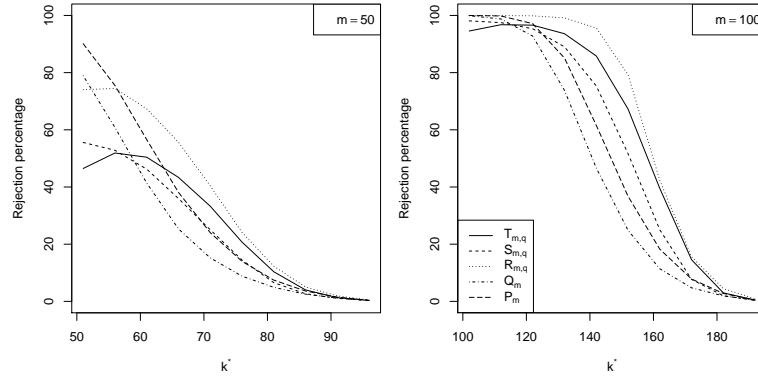
FIG 5.6. *Estimated rejection probabilities of H_0^{ind} in* (1.2) *under H_1 in* (1.3) *with F the d.f. of the $N(1,2)$ and G the d.f. of the Gamma distribution whose shape and rate parameters are both equal to 1/2. Left: $m = 50$ and $n = 100$. Right: $m = 100$ and $n = 200$. The value of γ in* (2.4) *is 0 and the number of steps in the threshold functions is $p = 50$.*

FIG 5.7. *Left: estimated rejection probabilities of H_0^{ind} in* (1.2) *under H_1 in* (1.3) *with $m = 50$, $n = 100$, F the d.f. of the $N(1,2)$ and G the d.f. of the Gamma distribution whose shape and rate parameters are both equal to 1/2 for the procedure based on $T_{m,q}$ with $p = 10$. Right: corresponding mean detection delays.*

parameter m was taken to be in $\{50, 100\}$, n was set to $2m$ and the parameter $k^\star$ to $\lfloor nc \rfloor$ with $c \in \{0.51, 0.56, \ldots, 0.96\}$.

Figure 5.6 shows the rejection percentages of H_0^{ind} in (1.2) against $k^\star$ for the procedures based on $R_{m,q}$, $S_{m,q}$, $T_{m,q}$, P_m and Q_m with $\gamma = 0$ and $p = 50$. The graphs for other values of γ and p are very similar. As one can see, the procedures based on P_m and Q_m appear to be the most powerful when the change occurs at the beginning of the monitoring period. The latter could have been expected from the definition of the underlying detectors and the simulation results of Dette and Gösmann (2019). As $k^\star$ increases, the procedures based on $R_{m,q}$ and $T_{m,q}$ become more powerful.

Figure 5.7 displays the influence of the parameter γ on the rejection percentages of the procedure based on $T_{m,q}$ with $p = 10$. The graphs are not qualitatively different
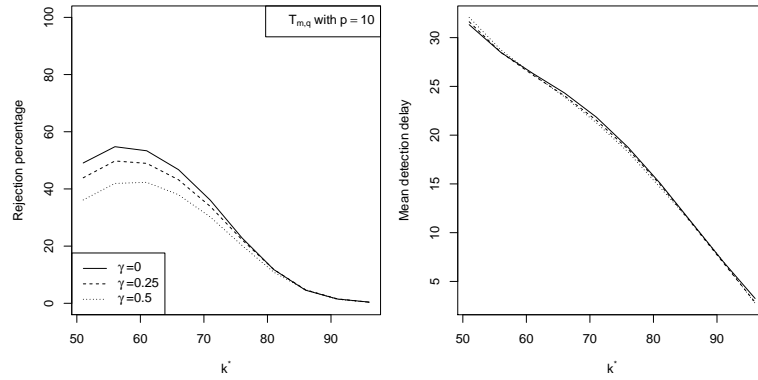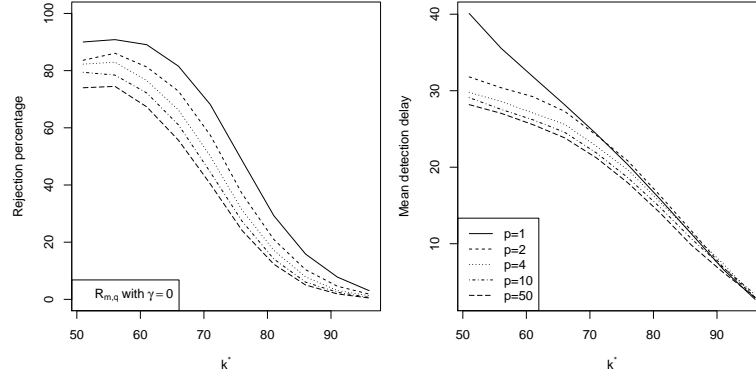
FIG 5.8. *Left: estimated rejection probabilities of H_0^{ind} in (1.2) under H_1 in (1.3) with $m = 50$, $n = 100$, F the d.f. of the $N(1,2)$ and G the d.f. of the Gamma distribution whose shape and rate parameters are both equal to 1/2 for the procedure based on $R_{m,q}$ with $\gamma = 0$. Right: corresponding mean detection delays.*

for other values of p and m or for the procedures based on $R_{m,q}$ and $S_{m,q}$. As one can see, unlike for a change in expectation, γ has hardly no influence on the mean detection delay and it is the setting $\gamma = 0$ that leads to the highest rejection percentages.

Finally, Figure 5.8 shows the influence of p for the procedure based on $R_{m,q}$ with $\gamma = 0$. The graphs are not qualitatively different for other values of γ or for the procedures based on $S_{m,q}$ and $T_{m,q}$. As in previous experiments, the setting $p = 1$ leads in the highest rejection percentages. However, when the change occurs in the first third of the monitoring period, the mean detection delay for $p = 1$ is clearly substantially larger than for $p > 1$. Additional simulations show that the larger the monitoring period, the more pronounced this phenomenon.

5.2. Dependent multiplier bootstrap-based estimation in the time series case

The threshold function estimation approach based on the dependent multiplier bootstrap described in Section 4.2 can in principle be used as soon as the observations to be monitored are either multivariate or serially dependent. We used the setting $B = 2000$ in our experiments and estimated all the rejection percentages from 1000 samples at the $\alpha = 5\%$ nominal level.

5.2.1. Under the null

One of the most important practical aspects is to assess how well the procedures hold their level when based on the dependent multiplier bootstrap. To attempt to answer this question, we conducted extensive simulations in the univariate case. For $m \in \{50, 100, 200, 400\}$ and $T \in \{0.5, 1, 2, 3\}$, we generated samples of size $n = \lfloor m(T + 1) \rfloor$ from an AR(1) model with normal innovations and autoregressive parameter $\beta \in$

$\{0, 0.1, 0.3, 0.5\}$, and estimated the levels of the procedures based on $R_{m,q}$, $S_{m,q}$ and $T_{m,q}$ with $\gamma \in \{0, 0.25, 0.5\}$ and $p \in \{1, 2, 4\}$. The rejection percentages of H_0 in (1.1) for $T_{m,q}$ are given in Table 5.2 (the missing entries in the table correspond to parameter settings for which computations took too long given our computer cluster resources). As one can see, unsurprisingly, the larger β, the more liberal the procedure based on $T_{m,q}$ tends to be. This phenomenon is particularly visible for $\beta \in \{0.3, 0.5\}$. Reassuringly however, for a given β, T, p and γ, the estimated levels seem to get closer to the 5% nominal level as m increases. Hence, unsurprisingly, the stronger the serial dependence, the larger m needs to be so that the procedure can be expected to hold its nominal level. For $\beta \in \{0.3, 0.5\}$ in particular and keeping T, γ and m fixed, we also see that the larger p, the more liberal the procedure based on $T_{m,q}$ tends to be. The latter could be explained by the fact that, as p increases, more thresholds need to be estimated, and that, except for the first, all the estimated thresholds are conditional empirical quantiles: the precision of the estimation of a threshold thus critically depends on the precision of the previously estimated thresholds with respect to which conditioning is performed. In other words, the fact that the empirical levels tend to become higher when p increases could be explained by an error propagation effect. Finally, for β, T, p and m fixed, we also see that the larger γ, the lower the rejection percentages tend to be. Overall, for $\beta \geq 0.1$, the procedure based on $T_{m,q}$ holds its level best for $\gamma = 0.5$.

The conclusions for the procedures based on $R_{m,q}$ and $S_{m,q}$ are very similar, with the exception that the effect of γ seems "stronger": while the empirical levels are better for $\gamma = 0.25$ than for $\gamma = 0$, the procedures become way too conservative for $\gamma = 0.5$. The latter effect might be due to the fact the detectors $R_{m,q}$ and $S_{m,q}$ involve maxima (unlike $T_{m,q}$ which involves means) and to our too low setting of the constant δ in the definition of the weight function (2.4). The latter was arbitrarily set to 10^{-4} (and had *de facto* no effect in our Monte Carlo experiments given the values of m that we considered).

A similar experiment was conducted by generating samples from a GARCH(1,1) model with parameters $\omega = 0.012$, $\beta = 0.919$ and $\alpha = 0.072$ to mimic SP500 daily log-returns following Jondeau, Poon and Rockinger (2007). The empirical levels for the procedure based on $T_{m,q}$ are reported in Table 5.3 and appear to be closest to the 5% nominal overall when $\gamma = 0.5$.

Finally, a bivariate experiment with independent observations consisting of generating samples of size $n = 2m$ for $m \in \{50, 100, 200\}$ from a normal copula with a Kendall's tau of $\tau \in \{-0.6, -0.3, 0, 0.3, 0.6\}$ was carried out. The empirical levels for the procedure based on $T_{m,q}$ are reported in Table 5.4. The effect of γ appears as in the previous experiments. For fixed γ and p, it can also be observed that the procedure has a tendency of being too conservative in the case of strong negative dependence but, reassuringly, the agreement with the 5% nominal level seems to improve as m increases.

5.2.2. Change in the copula parameter

To grasp further the finite-sample behavior of the procedures in the case of bivariate independent observations, we simulated a change in the parameter of a normal copula. The left (resp. right) graph in Figure 5.9 displays the estimated rejection probabilities
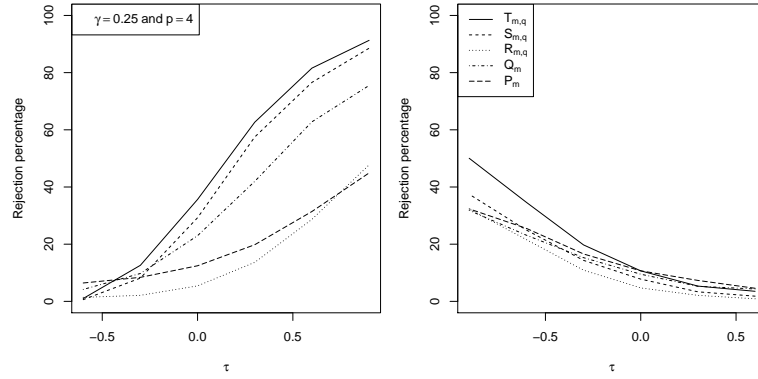
FIG 5.9. *Left: estimated rejection probabilities of H_0^{ind} in (1.2) under H_1 in (1.3) with $m = 50$, $k^\star = 75$, $n = 100$, F the bivariate normal copula with a Kendall's tau of -0.6 and G the bivariate normal copula with a Kendall's tau of τ. Right: estimated rejection probabilities of H_0^{ind} in (1.2) under H_1 in (1.3) with $m = 50$, $k^\star = 75$, $n = 100$, F the bivariate normal copula with a Kendall's tau of 0.6 and G the bivariate normal copula with a Kendall's tau of τ. The value of γ in (2.4) is 0.25. The number of steps in the threshold functions is $p = 4$.*

of H_0^{ind} in (1.2) for the procedures based on $R_{m,q}$, $S_{m,q}$, $T_{m,q}$, P_m and Q_m with $\gamma = 0.25$ and $p = 4$ under H_1 in (1.3) with $m = 50$, $k^\star = 75$, $n = 100$, F the bivariate normal copula with a Kendall's tau of -0.6 (resp. 0.6) and G the bivariate normal copula with a Kendall's tau of $\tau \in \{-0.6, 0.3, 0, 0.3, 0.6, 0.9\}$ (resp. $\tau \in \{-0.9, -0.6, -0.3, 0, 0.3, 0.6\}$). As one can notice, the procedure based on $T_{m,q}$ (resp. P_m) is always among the most (resp. least) powerful ones. Graphs for other values of γ and p are not qualitatively different. As for all previous experiments, we observed that the smaller p, the more powerful the procedures. For this experiment, the parameter γ appeared to have a rather small impact on the rejection percentages of $T_{m,q}$.

6. Data examples

To illustrate the use of the proposed sequential change-point detection tests, we considered two fictitious scenarios, the first (resp. second) of a univariate (resp. bivariate) nature based on closing quotes of the NASDAQ composite index (resp. Microsoft and Intel stocks) for the period 2019-01-02 – 2020-04-11. The latter were obtained using the `get.hist.quote()` function of the `tseries` R package (Trapletti and Hornik, 2019). In both scenarios, it was assumed that, on the last day of 2019, one wished to monitor the (univariate or bivariate) daily log-returns for a change in contemporary distribution possibly using the stretch of $m = 251$ (univariate of bivariate) log-returns of 2019 as learning sample. The latter decision was confirmed after the tests of Bücher, Fermanian and Kojadinovic (2019) implemented in the functions `stDistAutocop()` and `cpDist()` of the `npcp` R package (Kojadinovic, 2020) provided no evidence against stationarity for the two candidate learning samples. Notice that, unsurprisingly, the rank-based test of serial independence of Genest and Rémillard (2004) implemented in the function
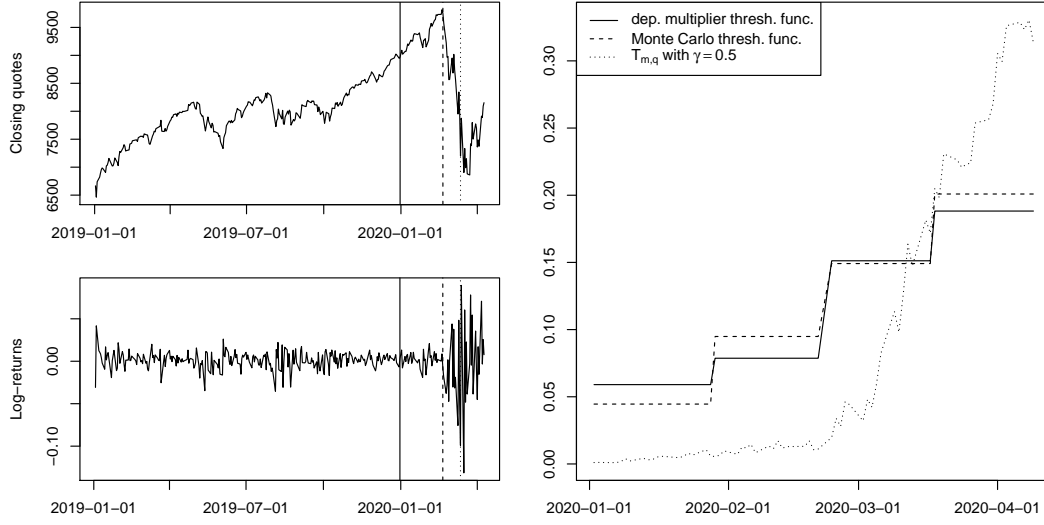
FIG 6.1. *Left: closing quotes and corresponding daily log-returns of the NASDAQ composite index for the period 2019-01-02 – 2020-04-11. The solid vertical line represents the beginning of the monitoring. The dotted vertical line represent the date (2020-03-12) at which the detector function based on $T_{m,q}$ with $\gamma = 0.5$ first exceeded the two threshold functions on the right. The dashed vertical line corresponds to the estimated date of change (2020-02-21). Right: the dotted line represents the detector function based on $T_{m,q}$ with $\gamma = 0.5$. The solid (resp. dashed) line represents the threshold function with $p = 4$ steps obtained using the dependent multiplier bootstrap (resp. Monte Carlo estimation).*

`serialIndepTest()` of the `copula` R package ([Hofert et al., 2018](#)) provided weak evidence against the serial independence of the squared component time series.

The closing quotes and corresponding daily log-returns of the NASDAQ composite index are represented in the left panel of Figure 6.1. The solid vertical lines mark the beginning of the monitoring. The dotted line in the right panel represents the detector function based on $T_{m,q}$ with $\gamma = 0.5$. The latter was chosen given its overall good performance in our Monte Carlo experiments, both in terms of empirical level, power and mean detection delay. In the right panel, the solid line represents the threshold function with $p = 4$ steps estimated using the dependent multiplier bootstrap with $B = 10^5$, while the dashed line represents the threshold function with $p = 4$ steps estimated using Monte Carlo with $M = 10^5$. Note that the latter did not at all use the learning sample as it is computed under the assumption that the observations are serially independent. The relative proximity of the two threshold functions could be explained by the fact that, although present, serial dependence in the learning sample is probably very weak. The detector function exceeded the two threshold functions at the same date (2020-03-12), which is marked by the dotted vertical line in the left panel of Figure 6.1 and corresponds to the 49th daily log-return of 2020. Given the definition of $T_{m,q}$ in (2.6) and having that of $S_{m,q}$ in (2.5) in mind, a possible estimate of a point
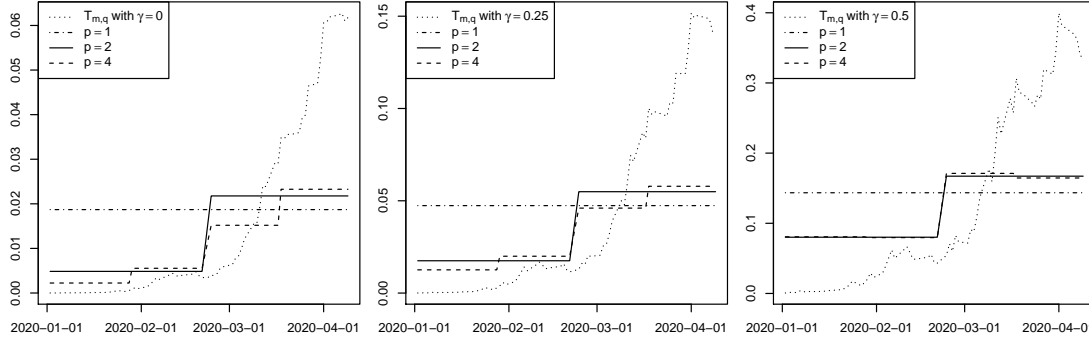
FIG 6.2. *Monitoring of the bivariate daily log-returns of the Microsoft and Intel stocks for the period 2020-01-01 – 2020-04-11 using the 2019 bivariate log-returns as learning sample, and the procedure based on $T_{m,q}$ with $\gamma \in \{0, 0.25, 0.5\}$ and a threshold function with $p \in \{1, 2, 4\}$ steps estimated using the dependent multiplier bootstrap. The estimated dates of change do not depend on p and are the 2020-01-25 for $\gamma = 0$, the 2020-02-15 for $\gamma = 0.25$ and the 2020-02-20 for $\gamma = 0.5$.*

of change for an exceedance at position $k = m + 49 = 300$ is given by

$$\operatorname{argmax}_{m \leq j \leq k-1} \frac{1}{k} \sum_{i=1}^{k} \left[\frac{j(k-j)}{m^{3/2}q(j/m, k/m)} \{F_{1:j}(\boldsymbol{X}_i) - F_{j+1:k}(\boldsymbol{X}_i)\} \right]^2 + 1, \qquad (6.1)$$

which returned 286 and corresponds to the date 2020-02-21. The latter is marked by a dashed vertical line in the left panel of Figure 6.1 and corresponds to the beginning of the sharp decrease of the NASDAQ composite index as a consequence of the Covid-19 pandemic.

Figure 6.2 describes the monitoring of the bivariate daily log-returns of the Microsoft and Intel stocks using the procedure based on $T_{m,q}$ with $\gamma \in \{0, 0.25, 0.5\}$ and a threshold function with $p \in \{1, 2, 4\}$ steps estimated using the dependent multiplier bootstrap. All the dates of exceedance are between the 2020-03-12 ($\gamma = 0$ and $p = 2$) and the 2020-03-09 ($\gamma = 0.5$ and $p = 1$). The estimated dates of change turn out not to depend on p and are the 2020-01-25 for $\gamma = 0$, the 2020-02-15 for $\gamma = 0.25$ and the 2020-02-20 for $\gamma = 0.5$. This effect of γ on (6.1) was to be expected: larger values of γ give more weight to potential break points close to k.

7. Concluding remarks

In the context of closed-end sequential change-point detection, it can be argued (see Anatolyev and Kosenok, 2018) that it is desirable that the underlying threshold function is such that the probability of false alarm remains approximately constant over the monitoring period. In this work, the asymptotic validity of the bootstrap-based estimation of such a threshold function was established for generic detectors. The latter was applied to sequential change-point tests involving detectors based on differences of empirical d.f.s that can be either simulated or resampled using a dependent multiplier

bootstrap depending on whether univariate independent or multivariate serially dependent observations are monitored. The proposed detectors are adaptations of statistics used in a posteriori change-point testing and include a weight function in the spirit of Csörgő and Szyszkowicz (1994a) that can be used to give more importance to recent observations.

Extensive Monte Carlo experiments were used to investigate the finite-sample properties of the resulting sequential change-point tests. Among the proposed detectors, none led to a uniformly better testing procedure. When based on the dependent multiplier bootstrap, the procedure based on $T_{m,q}$ in (2.6) was observed to have the best behavior, overall, in terms of empirical level, power and mean detection delay. In the case of univariate independent observations, when the threshold function can be estimated using Monte Carlo simulation, the number of step p of the threshold function can be chosen as large as the number of monitoring steps. However, in the time series case, when the estimation of the threshold function is based on the dependent multiplier bootstrap, p should not be taken too large because of an error propagation effect.

As already hinted at in Section 3, a straightforward extension of the generic results on the estimation of the threshold function consists of allowing the conditional probability in (3.5) to change with the underlying monitoring sub-interval (or, equivalently, to have monitoring sub-intervals of different lengths). The choice of the p conditional probabilities (or, equivalently, of the p monitoring sub-intervals) could for instance be carried out according to the user's prior knowledge. In future work, we also plan to investigate the validity of additional bootstraps for monitoring multivariate time series. The current and future theoretical results would further need to be complemented by additional Monte Carlo simulations, involving in particular multivariate experiments. Such finite-sample investigations are however a real computational challenge given the complexity and cost of execution of such change-point detection procedures.

Appendix A: Proofs of Propositions 2.5 and 2.7

Proof of Proposition 2.5. From Theorem 1 of Bücher (2015), we have that $\mathbb{B}_n \rightsquigarrow \mathbb{B}_C$ in $\ell^\infty([0,1]^{d+1})$, where $n = \lfloor m(T+1) \rfloor$. Let ψ be the map from $\ell^\infty([0,1]^{d+1})$ to $\ell^\infty([0,T+1] \times [0,1]^d)$ defined, for any $f \in \ell^\infty([0,1]^{d+1})$, by

$$\psi(f)(s, \boldsymbol{u}) = \sqrt{T+1} f\{s/(T+1), \boldsymbol{u}\}, \qquad s \in [0, T+1], \boldsymbol{u} \in [0,1]^d. \tag{A.1}$$

It is straightforward to verify that ψ is a continuous map which immediately implies by the continuous mapping theorem that $\psi(\mathbb{B}_n) \rightsquigarrow \psi(\mathbb{B}_C)$ in $\ell^\infty([0,T+1] \times [0,1]^d)$. Furthermore, it is easy to check that $\psi(\mathbb{B}_n) - \mathbb{B}_m = o_\mathbb{P}(1)$ and that $\psi(\mathbb{B}_C)$ is a tight centered Gaussian process such that, for any $s, t \in [0, T+1]$ and $\boldsymbol{u}, \boldsymbol{v} \in [0,1]^d$,

$$\begin{aligned} \mathrm{Cov}\{\psi(\mathbb{B}_C)(s, \boldsymbol{u}), \psi(\mathbb{B}_C)(t, \boldsymbol{v})\} &= (T+1)\mathrm{Cov}[\mathbb{B}_C\{s/(T+1), \boldsymbol{u}\}, \mathbb{B}_C\{t/(T+1), \boldsymbol{v}\}] \\ &= (s \wedge t)\Gamma(\boldsymbol{u}, \boldsymbol{v}) = \mathrm{Cov}\{\mathbb{B}_C(s, \boldsymbol{u}), \mathbb{B}_C(t, \boldsymbol{v})\}, \end{aligned}$$

where Γ is defined in (2.23). It follows that $\mathbb{B}_m \rightsquigarrow \mathbb{B}_C$ in $\ell^\infty([0, T+1] \times [0,1]^d)$.

It remains to show the subsequent claims. Under H_0, (2.22) holds and the continuous mapping theorem immediately implies that $\mathbb{G}_m \rightsquigarrow \mathbb{G}_C$ in $\ell^\infty(\Delta \times [0,1]^d)$ and, then, that $\mathbb{G}_{m,q} \rightsquigarrow \mathbb{G}_{C,q}$ in $\ell^\infty(\Delta \times [0,1]^d)$ since $\sup_{s \in [0,T+1]} |\lambda_m(0,s) - s| \le 1/m$, the function q in (2.4) is continuous on Δ and $\sup_{(s,t) \in \Delta}\{q(s,t)\}^{-1} < \infty$. The continuous mapping theorem further straightforwardly implies that $\mathbb{R}_{m,q} \rightsquigarrow \mathbb{R}_{C,q}$ in $\ell^\infty([1,T+1])$.

Let $\mathbb{J}_{m,q}(s,t) = \int_{[0,1]^d}\{\mathbb{G}_{m,q}(s,t,\boldsymbol{u})\}^2 dC_{1:\lfloor mt \rfloor}(\boldsymbol{u})$ for any $1 \le s \le t \le T+1$. In order to prove that $\mathbb{S}_{m,q} \rightsquigarrow \mathbb{S}_{C,q}$ and $\mathbb{T}_{m,q} \rightsquigarrow \mathbb{T}_{C,q}$ in $\ell^\infty([1,T+1])$, we shall first prove that $\mathbb{J}_{m,q} \rightsquigarrow \mathbb{J}_{C,q}$ in $\ell^\infty(\Delta \cap [1,T+1]^2)$, where $\mathbb{J}_{C,q}(s,t) = \int_{[0,1]^d}\{\mathbb{G}_{C,q}(s,t,\boldsymbol{u})\}^2 dC(\boldsymbol{u})$ for any $1 \le s \le t \le T+1$.

We start by showing that the finite-dimensional distributions of $\mathbb{J}_{m,q}$ converge weakly to those of $\mathbb{J}_{C,q}$. Let $p \in \mathbb{N}$, $p > 1$, and $(s_1,t_1),\ldots,(s_p,t_p) \in \Delta \cap [1,T+1]^2$ be arbitrary. The result is proven if we show that

$$\big(\mathbb{J}_{m,q}(s_1,t_1),\ldots,\mathbb{J}_{m,q}(s_p,t_p)\big) \rightsquigarrow \big(\mathbb{J}_{C,q}(s_1,t_1),\ldots,\mathbb{J}_{C,q}(s_p,t_p)\big). \tag{A.2}$$

From the already proven weak convergence of $\mathbb{B}_m$ to $\mathbb{B}_C$ in $\ell^\infty([1,T+1] \times [0,1]^d)$, we obtain that

$$\sup_{\substack{s \in [1,T+1] \\ \boldsymbol{u} \in [0,1]^d}} |C_{1:\lfloor ms \rfloor}(\boldsymbol{u}) - C(\boldsymbol{u})| \le \frac{1}{\sqrt{m}} \times \sup_{s \in [1,T+1]} \frac{1}{\lambda_m(0,s)} \times \sup_{\substack{s \in [1,T+1] \\ \boldsymbol{u} \in [0,1]^d}} |\mathbb{B}_m(s,\boldsymbol{u})| \overset{\mathbb{P}}{\to} 0,$$

since $\sup_{s \in [1,T+1]}\{\lambda_m(0,s)\}^{-1} \le 1$ and $\sup_{s \in [1,T+1], \boldsymbol{u} \in [0,1]^d} |\mathbb{B}_m(s,\boldsymbol{u})| = O_\mathbb{P}(1)$. From the fact that $\mathbb{G}_{m,q} \rightsquigarrow \mathbb{G}_{C,q}$ in $\ell^\infty\{(\Delta \cap [1,T+1]^2) \times [0,1]^d\}$, we then have that

$$\big(\mathbb{G}_{m,q}(s_1,t_1,\cdot),\ldots,\mathbb{G}_{m,q}(s_p,t_p,\cdot),C_{1:\lfloor mt_1 \rfloor},\ldots,C_{1:\lfloor mt_p \rfloor}\big)$$
$$\rightsquigarrow \big(\mathbb{G}_{C,q}(s_1,t_1,\cdot),\ldots,\mathbb{G}_{C,q}(s_p,t_p,\cdot),C,\ldots,C\big)$$

in $\{\ell^\infty([0,1]^d)\}^{2p}$. The latter can be combined with the fact that C is continuous, $\mathbb{G}_{C,q}$ has continuous sample paths with probability one, Lemma 1 in Kojadinovic, Segers and Yan (2011) and the continuous mapping theorem to obtain (A.2).

It remains to show that the process $\mathbb{J}_{m,q}$ is asymptotically tight (see, e.g., van der Vaart and Wellner, 2000, Section 1.5). From Section 2.1.2 and Problem 2.1.5 in the same reference, the latter is shown if, for every sequence $\delta_m \downarrow 0$,

$$\sup_{\substack{(s,t),(s',t') \in \Delta \cap [1,T+1]^2 \\ |s-s'|<\delta_m, |t-t'|<\delta_m}} |\mathbb{J}_{m,q}(s,t) - \mathbb{J}_{m,q}(s',t')| \overset{\mathbb{P}}{\to} 0. \tag{A.3}$$

The supremum on the left-hand side of the previous display is smaller than $I_m + J_m$,

where

$$I_m = \sup_{\substack{(s,t),(s',t')\in\Delta\cap[1,T+1]^2 \\ |s-s'|<\delta_m,|t-t'|<\delta_m}} \left| \int_{[0,1]^d} \{\mathbb{G}_{m,q}(s,t,\boldsymbol{u})\}^2 \mathrm{d}C_{1:\lfloor mt\rfloor}(\boldsymbol{u}) \right.$$

$$\left. - \int_{[0,1]^d} \{\mathbb{G}_{m,q}(s',t',\boldsymbol{u})\}^2 \mathrm{d}C_{1:\lfloor mt\rfloor}(\boldsymbol{u}) \right|,$$

$$J_m = \sup_{\substack{(s,t),(s',t')\in\Delta\cap[1,T+1]^2 \\ |s-s'|<\delta_m,|t-t'|<\delta_m}} \left| \int_{[0,1]^d} \{\mathbb{G}_{m,q}(s',t',\boldsymbol{u})\}^2 \mathrm{d}C_{1:\lfloor mt\rfloor}(\boldsymbol{u}) \right.$$

$$\left. - \int_{[0,1]^d} \{\mathbb{G}_{m,q}(s',t',\boldsymbol{u})\}^2 \mathrm{d}C_{1:\lfloor mt'\rfloor}(\boldsymbol{u}) \right|.$$

Now,

$$I_m \leq \sup_{\substack{(s,t),(s',t')\in\Delta\cap[1,T+1]^2 \\ |s-s'|<\delta_m,|t-t'|<\delta_m}} \int_{[0,1]^d} \left| \{\mathbb{G}_{m,q}(s,t,\boldsymbol{u})\}^2 - \{\mathbb{G}_{m,q}(s',t',\boldsymbol{u})\}^2 \right| \mathrm{d}C_{1:\lfloor mt\rfloor}(\boldsymbol{u})$$

$$\leq \sup_{\substack{(s,t),(s',t')\in\Delta\cap[1,T+1]^2 \\ |s-s'|<\delta_m,|t-t'|<\delta_m}} \sup_{\boldsymbol{u}\in[0,1]^d} \left| \{\mathbb{G}_{m,q}(s,t,\boldsymbol{u})\}^2 - \{\mathbb{G}_{m,q}(s',t',\boldsymbol{u})\}^2 \right| \int_{[0,1]^d} \mathrm{d}C_{1:\lfloor mt\rfloor}(\boldsymbol{u})$$

$$\leq \sup_{\substack{(s,t),(s',t')\in\Delta\cap[1,T+1]^2 \\ |s-s'|<\delta_m,|t-t'|<\delta_m}} \sup_{\substack{\boldsymbol{u},\boldsymbol{u}'\in[0,1]^d \\ \|\boldsymbol{u}-\boldsymbol{u}'\|\leq\delta_m}} \left| \{\mathbb{G}_{m,q}(s,t,\boldsymbol{u})\}^2 - \{\mathbb{G}_{m,q}(s',t',\boldsymbol{u}')\}^2 \right| \xrightarrow{\mathbb{P}} 0$$

by the asymptotic uniform equicontinuity in probability of $\mathbb{G}_{m,q}^2$. Concerning J_m, we have

$$J_m \leq 2 \sup_{\substack{(s,t)\in\Delta\cap[1,T+1]^2 \\ t'\in[1,T+1],0\leq t'-t<\delta_m}} \left| \frac{1}{\lfloor mt\rfloor} \sum_{i=1}^{\lfloor mt\rfloor} \{\mathbb{G}_{m,q}(s,t,\boldsymbol{U}_i)\}^2 - \frac{1}{\lfloor mt'\rfloor} \sum_{i=1}^{\lfloor mt'\rfloor} \{\mathbb{G}_{m,q}(s,t,\boldsymbol{U}_i)\}^2 \right|$$

$$\leq 2 \sup_{\substack{(s,t)\in\Delta\cap[1,T+1]^2 \\ t'\in[1,T+1],0\leq t'-t<\delta_m}} \sum_{i=1}^{\lfloor mt\rfloor} \{\mathbb{G}_{m,q}(s,t,\boldsymbol{U}_i)\}^2 \left(\frac{1}{\lfloor mt\rfloor} - \frac{1}{\lfloor mt'\rfloor} \right)$$

$$+ 2 \sup_{\substack{(s,t)\in\Delta\cap[1,T+1]^2 \\ t'\in[1,T+1],0\leq t'-t<\delta_m}} \frac{1}{\lfloor mt'\rfloor} \sum_{i=\lfloor mt\rfloor+1}^{\lfloor mt'\rfloor} \{\mathbb{G}_{m,q}(s,t,\boldsymbol{U}_i)\}^2$$

$$\leq 4 \sup_{\substack{t,t'\in[1,T+1] \\ 0\leq t'-t<\delta_m}} \frac{\lfloor mt'\rfloor - \lfloor mt\rfloor}{\lfloor mt'\rfloor} \times \sup_{\substack{(s,t)\in\Delta\cap[1,T+1]^2 \\ \boldsymbol{u}\in[0,1]^d}} \{\mathbb{G}_{m,q}(s,t,\boldsymbol{u})\}^2 \leq \delta_m \times O_{\mathbb{P}}(1) + o_{\mathbb{P}}(1),$$

since $\lfloor mt'\rfloor \geq m$ for all $t' \in [1, T+1]$,

$$\sup_{t,t'\in[0,T+1]} \left| \frac{\lfloor mt'\rfloor - \lfloor mt\rfloor}{m} - (t'-t) \right| \leq 2 \sup_{t\in[0,T+1]} \left| \frac{\lfloor mt\rfloor}{m} - t \right| \leq \frac{2}{m} \to 0,$$

and the fact that $\sup_{(s,t)\in\Delta\cap[1,T+1]^2, \boldsymbol{u}\in[0,1]^d}\{\mathbb{G}_{m,q}(s,t,\boldsymbol{u})\}^2$ converges weakly by the continuous mapping theorem. Hence, (A.3) holds and therefore $\mathbb{J}_{m,q} \rightsquigarrow \mathbb{J}_{C,q}$ in $\ell^\infty(\Delta \cap [1,T+1]^2)$. The fact that $\mathbb{S}_{m,q} \rightsquigarrow \mathbb{S}_{C,q}$ and $\mathbb{T}_{m,q} \rightsquigarrow \mathbb{T}_{C,q}$ in $\ell^\infty([1,T+1])$ is finally and again an immediate consequence of the continuous mapping theorem.

Finally, we have to show that, for any $[t_1,t_2] \subset [1,T+1]$ such that $t_2 > 1$, the distributions of $\sup_{t\in[t_1,t_2]}\mathbb{R}_{C,q}(t)$, $\sup_{t\in[t_1,t_2]}\mathbb{S}_{C,q}(t)$ and $\sup_{t\in[t_1,t_2]}\mathbb{T}_{C,q}(t)$ are absolutely continuous with respect to the Lebesgue measure. To show the latter, we adapt the proof of Proposition 3.3 of Bücher, Fermanian and Kojadinovic (2019) to the current setting. Let $\mathcal{C}(\mathcal{S})$ denote the space of all continuous real-valued functions on $\mathcal{S}$ equipped with the uniform metric. Since the sample paths of $\mathbb{B}_C$ are elements of $\mathcal{C}([0,T+1]\times[0,1]^d)$ with probability one and q in (2.4) is continuous, the sample paths of $\mathbb{G}_{C,q}$ are elements of $\mathcal{C}(\Delta \times [0,1]^d)$ with probability one. Fix $[t_1,t_2] \subset [1,T+1]$ such that $t_2 > 1$ and let ϑ_1, ϑ_2 and ϑ_3 be the maps from $\mathcal{C}(\Delta \times [0,1]^d)$ to $\mathbb{R}$ defined, for any $f \in \mathcal{C}(\Delta \times [0,1]^d)$, by

$$\vartheta_1(f) = \sup_{t\in[t_1,t_2]} \sup_{s\in[1,t]} \sup_{\boldsymbol{u}\in[0,1]^d} |f(s,t,\boldsymbol{u})|,$$

$$\vartheta_2(f) = \sup_{t\in[t_1,t_2]} \sup_{s\in[1,t]} \left\{ \int_{[0,1]^d} \{f(s,t,\boldsymbol{u})\}^2 \mathrm{d}C(\boldsymbol{u}) \right\}^{1/2},$$

$$\vartheta_3(f) = \sup_{t\in[t_1,t_2]} \left\{ \int_1^t \int_{[0,1]^d} \{f(s,t,\boldsymbol{u})\}^2 \mathrm{d}C(\boldsymbol{u})\mathrm{d}s \right\}^{1/2}.$$

Then, we have that $\sup_{t\in[t_1,t_2]}\mathbb{R}_{C,q}(t) = \vartheta_1(\mathbb{G}_{C,q})$, $\sup_{t\in[t_1,t_2]}\mathbb{S}_{C,q}(t) = \{\vartheta_2(\mathbb{G}_{C,q})\}^2$ and $\sup_{t\in[t_1,t_2]}\mathbb{T}_{C,q}(t) = \{\vartheta_3(\mathbb{G}_{C,q})\}^2$ and, to show the desired result, it suffices to prove that the distributions of $\vartheta_1(\mathbb{G}_{C,q})$, $\vartheta_2(\mathbb{G}_{C,q})$ and $\vartheta_2(\mathbb{G}_{C,q})$, denoted respectively by $\mathcal{L}\{\vartheta_1(\mathbb{G}_{C,q})\}$, $\mathcal{L}\{\vartheta_2(\mathbb{G}_{C,q})\}$ and $\mathcal{L}\{\vartheta_2(\mathbb{G}_{C,q})\}$, are absolutely continuous with respect to the Lebesgue measure. Since the maps ϑ_1, ϑ_2 and ϑ_3 are continuous and convex, from Theorem 7.1 in Davydov and Lifshits (1984), we obtain that, for any $i \in \{1,2,3\}$, $\mathcal{L}\{\vartheta_i(\mathbb{G}_{C,q})\}$ is concentrated on $[a_i,\infty)$ and absolutely continuous on (a_i,∞), where

$$a_i = \inf\{\vartheta_i(f) : f \text{ belongs to the support of } \mathcal{L}(\mathbb{G}_{C,q})\}.$$

By Lemma 1.2 (e) in Dereich et al. (2003), we have that, for any $\varepsilon > 0$, $\mathbb{P}\{\vartheta_1(\mathbb{G}_{C,q}) \leq \varepsilon\} > 0$. Hence, for any $\varepsilon > 0$,

$$\mathbb{P}\{\vartheta_3(\mathbb{G}_{C,q}) \leq \varepsilon\} \geq \mathbb{P}\{\vartheta_2(\mathbb{G}_{C,q}) \leq \varepsilon\} \geq \mathbb{P}\{\vartheta_1(\mathbb{G}_{C,q}) \leq \varepsilon\} > 0.$$

It follows that, for any $i \in \{1,2,3\}$ and any $\varepsilon > 0$, there exists functions f in the support of $\mathcal{L}(\mathbb{G}_{C,q})$ such that $\vartheta_i(f) \leq \varepsilon$, which implies that $a_1 = a_2 = a_3 = 0$.

To conclude, it remains to show that $\mathcal{L}\{\vartheta_1(\mathbb{G}_{C,q})\}$, $\mathcal{L}\{\vartheta_2(\mathbb{G}_{C,q})\}$ and $\mathcal{L}\{\vartheta_3(\mathbb{G}_{C,q})\}$ have no atom at 0. For any $f \in \mathcal{C}(\Delta \times [0,1]^d)$, we have that $\vartheta_2(f) = \vartheta_3(f) = 0$ if and only if $f(s,t,\boldsymbol{u}) = 0$ for all $(s,t) \in \Delta \cap ([1,t_2] \times [t_1,t_2])$ and all $\boldsymbol{u}$ in the support of the distribution induced by C. Let $\boldsymbol{u}^* \in (0,1)^d$ be an arbitrary point in the latter support such that $\mathrm{Var}\{\mathbb{B}_C(1,\boldsymbol{u}^*)\} = \Gamma(\boldsymbol{u}^*,\boldsymbol{u}^*) > 0$, where Γ is defined in (2.23), and let $(s^*,t^*) \in$

$[1, t_2] \times [t_1, t_2]$ such that $s^* < t^*$. Then, $\mathrm{Var}\{\mathbb{G}_C(s^*, t^*, \boldsymbol{u}^*)\} = s^* t^* (t^* - s^*) \Gamma(\boldsymbol{u}^*, \boldsymbol{u}^*) > 0$, which implies that $\mathrm{Var}\{\mathbb{G}_{C,q}(s^*, t^*, \boldsymbol{u}^*)\} = s^* t^* (t^* - s^*) \Gamma(\boldsymbol{u}^*, \boldsymbol{u}^*) \{q(s^*, t^*)\}^{-2} > 0$. The proof is complete since

$$\mathbb{P}\{\vartheta_1(\mathbb{G}_{C,q}) = 0\} = \mathbb{P}\{\vartheta_2(\mathbb{G}_{C,q}) = 0\} = \mathbb{P}\{\vartheta_3(\mathbb{G}_{C,q}) = 0\} \leq \mathbb{P}\{\mathbb{G}_{C,q}(s^*, t^*, \boldsymbol{u}^*) = 0\} = 0.$$

$\square$

Proof of Proposition 2.7. Let $\mathbb{K}_m(s, t, \boldsymbol{x}) = m^{-1/2} \mathbb{H}_m(s, t, \boldsymbol{x}) - K_c(s, t, \boldsymbol{x})$, $(s, t) \in \Delta$, $\boldsymbol{x} \in \mathbb{R}^d$. The first claim is proven if

$$\sup_{\substack{(s,t) \in \Delta \\ \boldsymbol{x} \in \mathbb{R}^d}} |\mathbb{K}_m(s, t, \boldsymbol{x})| \xrightarrow{\mathbb{P}} 0. \tag{A.4}$$

The supremum on the left-hand side of the previous display is equal to

$$\max \left\{ \sup_{\substack{0 \leq s \leq t \leq c \\ \boldsymbol{x} \in \mathbb{R}^d}} |\mathbb{K}_m(s, t, \boldsymbol{x})|, \sup_{\substack{0 \leq s \leq c \leq t \leq T+1 \\ \boldsymbol{x} \in \mathbb{R}^d}} |\mathbb{K}_m(s, t, \boldsymbol{x})|, \sup_{\substack{c \leq s \leq t \leq T+1 \\ \boldsymbol{x} \in \mathbb{R}^d}} |\mathbb{K}_m(s, t, \boldsymbol{x})| \right\}. \tag{A.5}$$

To prove (A.4), we shall show that each of the three suprema in the previous display converge in probability to zero. Notice first that

$$K_c(s, t, \boldsymbol{x}) = \begin{cases} 0, & \text{if } 0 \leq s \leq t \leq c, \\ s(t - c)\{F(\boldsymbol{x}) - G(\boldsymbol{x})\}, & \text{if } 0 \leq s \leq c \leq t \leq T+1, \\ c(t - s)\{F(\boldsymbol{x}) - G(\boldsymbol{x})\}, & \text{if } c \leq s \leq t \leq T+1. \end{cases} \tag{A.6}$$

Furthermore, for any $(s, t) \in \Delta$, $\boldsymbol{x} \in \mathbb{R}^d$ and $H \in \{F, G\}$, let

$$\mathbb{F}_{m,H}(s, t, \boldsymbol{x}) = \sqrt{m} \lambda_m(s, t)\{F_{\lfloor ms \rfloor + 1 : \lfloor mt \rfloor}(\boldsymbol{x}) - H(\boldsymbol{x})\} = \mathbb{F}_{m,H}^\circ(t, \boldsymbol{x}) - \mathbb{F}_{m,H}^\circ(s, \boldsymbol{x}),$$

where $\mathbb{F}_{m,H}^\circ(s, \boldsymbol{x}) = \sqrt{m} \lambda_m(0, s)\{F_{1:\lfloor ms \rfloor}(\boldsymbol{x}) - H(\boldsymbol{x})\}$, λ_m is defined in (2.14) and $F_{1:\lfloor ms \rfloor}$ is generically defined by (2.1). By proceeding as in Section 2.2, it can be verified that, under H_0 in (1.1), $\mathbb{F}_{m,F}^\circ(s, \boldsymbol{x}) = \mathbb{B}_m\{s, \boldsymbol{F}(\boldsymbol{x})\}$ for all $s \in [0, T+1]$ and $\boldsymbol{x} \in \mathbb{R}^d$, where $\mathbb{B}_m$ is defined in (2.21) and $\boldsymbol{F}(\boldsymbol{x}) = (F_1(x_1), \ldots, F_d(x_d))$. By the continuous mapping theorem, it thus immediately follows that, under H_0, $\mathbb{F}_{m,F}$ converges weakly in $\ell^\infty(\Delta \times \mathbb{R}^d)$ to a tight limit. Some thought then reveals that, under the conditions of the proposition, $\mathbb{F}_{m,F}$ (resp. $\mathbb{F}_{m,G}$) converges weakly in $\ell^\infty\{(\Delta \cap [0, c]^2) \times \mathbb{R}^d\}$ (resp. in $\ell^\infty\{(\Delta \cap [c, T+1]^2) \times \mathbb{R}^d\}$) to a tight limit.

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

$$\sup_{\substack{0 \leq s \leq t \leq c \\ \boldsymbol{x} \in \mathbb{R}^d}} |\mathbb{K}_m(s, t, \boldsymbol{x})| = m^{-1/2} \sup_{\substack{0 \leq s \leq t \leq c \\ \boldsymbol{x} \in \mathbb{R}^d}} |\mathbb{H}_m(s, t, \boldsymbol{x})| = o(1) \times O_\mathbb{P}(1) \xrightarrow{\mathbb{P}} 0,$$

since $\mathbb{H}_m$ converges weakly in $\ell^\infty\{(\Delta \cap [0, c]^2) \times \mathbb{R}^d\}$ to a tight limit as a consequence of the fact that, for any $0 \leq s \leq t \leq c$ and $\boldsymbol{x} \in \mathbb{R}^d$, $\mathbb{H}_m(s, t, \boldsymbol{x}) = \lambda_m(s, t)\mathbb{F}_{m,F}(0, s) - \lambda_m(0, s)\mathbb{F}_{m,F}(s, t)$ and the continuous mapping theorem.

Regarding the second supremum, for any $0 \leq s \leq c \leq t \leq T+1$ and $\boldsymbol{x} \in \mathbb{R}^d$, we have that

$$\lambda_m(s,t)F_{\lfloor ms \rfloor+1:\lfloor mt \rfloor}(\boldsymbol{x}) = \lambda_m(s,c)F_{\lfloor ms \rfloor+1:\lfloor mc \rfloor}(\boldsymbol{x}) + \lambda_m(c,t)F_{\lfloor mc \rfloor+1:\lfloor mt \rfloor}(\boldsymbol{x}).$$

Thus, on one hand,

$$m^{-1/2}\mathbb{H}_m(s,t,\boldsymbol{x})$$
$$= \lambda_m(0,s)\{\lambda_m(s,t)F_{1:\lfloor ms \rfloor}(\boldsymbol{x}) - \lambda_m(s,c)F_{\lfloor ms \rfloor+1:\lfloor mc \rfloor}(\boldsymbol{x}) - \lambda_m(c,t)F_{\lfloor mc \rfloor+1:\lfloor mt \rfloor}(\boldsymbol{x})\}.$$

On the other hand, from (A.6) and using the fact that $\sup_{(s,t) \in \Delta}|\lambda_m(s,t) - (t-s)| \leq 2/m$,

$$K_c(s,t,\boldsymbol{x}) = \lambda_m(0,s)\{\lambda_m(s,t)F(\boldsymbol{x}) - \lambda_m(s,c)F(\boldsymbol{x}) - \lambda_m(c,t)G(\boldsymbol{x})\} + O(1/m).$$

By the triangle inequality and using the fact that $\sup_{0 \leq s \leq t \leq T+1}|\lambda_m(s,t)| \leq T+1$, it then follows that

$$\sup_{\substack{0 \leq s \leq c \leq t \leq T+1 \\ \boldsymbol{x} \in \mathbb{R}^d}}|\mathbb{K}_m(s,t,\boldsymbol{x})| \leq m^{-1/2}(T+1)\left[\sup_{\substack{0 \leq s \leq c \\ \boldsymbol{x} \in \mathbb{R}^d}}|\mathbb{F}_{m,F}(0,s,\boldsymbol{x})|\right.$$
$$\left. + \sup_{\substack{0 \leq s \leq c \\ \boldsymbol{x} \in \mathbb{R}^d}}|\mathbb{F}_{m,F}(s,c,\boldsymbol{x})| + \sup_{\substack{c \leq t \leq T+1 \\ \boldsymbol{x} \in \mathbb{R}^d}}|\mathbb{F}_{m,G}(c,t,\boldsymbol{x})|\right] = o(1) \times O_{\mathbb{P}}(1).$$

Similarly, for the third supremum, for any $c \leq s \leq T+1$ and $\boldsymbol{x} \in \mathbb{R}^d$,

$$\lambda_m(0,s)F_{1:\lfloor ms \rfloor}(\boldsymbol{x}) = \lambda_m(0,c)F_{1:\lfloor mc \rfloor}(\boldsymbol{x}) + \lambda_m(c,s)F_{\lfloor mc \rfloor+1:\lfloor ms \rfloor}(\boldsymbol{x}),$$

and, hence, on one hand, for any $c \leq s \leq t \leq T+1$ and $\boldsymbol{x} \in \mathbb{R}^d$,

$$m^{-1/2}\mathbb{H}_m(s,t,\boldsymbol{x})$$
$$= \lambda_m(s,t)\{\lambda_m(0,c)F_{1:\lfloor mc \rfloor}(\boldsymbol{x}) + \lambda_m(c,s)F_{\lfloor mc \rfloor+1:\lfloor ms \rfloor}(\boldsymbol{x}) - \lambda_m(0,s)F_{\lfloor ms \rfloor+1:\lfloor mt \rfloor}(\boldsymbol{x})\},$$

while, on the other hand,

$$K_c(s,t,\boldsymbol{x}) = \lambda_m(s,t)\{\lambda_m(0,c)F(\boldsymbol{x}) + \lambda_m(c,s)G(\boldsymbol{x}) - \lambda_m(0,s)G(\boldsymbol{x})\} + O(1/m).$$

Finally, by the triangle inequality,

$$\sup_{\substack{c \leq s \leq t \leq T+1 \\ \boldsymbol{x} \in \mathbb{R}^d}}|\mathbb{K}_m(s,t,\boldsymbol{x})| \leq m^{-1/2}(T+1)\left[\sup_{\boldsymbol{x} \in \mathbb{R}^d}|\mathbb{F}_{m,F}(0,c,\boldsymbol{x})|\right.$$
$$\left. + \sup_{\substack{c \leq s \leq T+1 \\ \boldsymbol{x} \in \mathbb{R}^d}}|\mathbb{F}_{m,G}(c,s,\boldsymbol{x})| + \sup_{\substack{c \leq s \leq t \leq T+1 \\ \boldsymbol{x} \in \mathbb{R}^d}}|\mathbb{F}_{m,G}(s,t,\boldsymbol{x})|\right] = o(1) \times O_{\mathbb{P}}(1),$$

which completes the proof of (A.4).

Using the fact that $\sup_{s\in[0,T+1]}|\lambda_m(0,s)-s|\le 1/m$, the function q in (2.4) is continuous on Δ and $\sup_{(s,t)\in\Delta}\{q(s,t)\}^{-1}<\infty$, we obtain, from the continuous mapping theorem, that $\mathbb{H}_{m,q}\overset{\mathbb{P}}{\to}K_{c,q}$ in $\ell^\infty(\Delta\times\mathbb{R}^d)$. From (2.3) and (2.26), and proceeding as in (2.17), it is easy to verify that $\mathbb{R}_{m,q}(t)=\sup_{s\in[1,t]}\sup_{\boldsymbol{x}\in\mathbb{R}^d}|\mathbb{H}_m(s,t,\boldsymbol{x})|$, $t\in[1,T+1]$. Hence, again by the continuous mapping theorem, $m^{-1/2}\mathbb{R}_{m,q}\overset{\mathbb{P}}{\to}L_{c,q}$ in $\ell^\infty([1,T+1])$, where $L_{c,q}(t)=\sup_{s\in[1,t]}\sup_{\boldsymbol{x}\in\mathbb{R}^d}|K_{c,q}(s,t,\boldsymbol{x})|$, and, thus, $m^{-1/2}\sup_{t\in[1,T+1]}\mathbb{R}_{m,q}(t)\overset{\mathbb{P}}{\to}\sup_{t\in[1,T+1]}L_{c,q}(t)$. Since

$$\sup_{t\in[1,T+1]}L_{c,q}(t)=\sup_{1\le s\le t\le T+1}\frac{(s\wedge c)\{(t\vee c)-(s\vee c)\}}{q(s,t)}\sup_{\boldsymbol{x}\in\mathbb{R}^d}|F(\boldsymbol{x})-G(\boldsymbol{x})|>0,$$

we immediately obtain that $\sup_{t\in[1,T+1]}\mathbb{R}_{m,q}(t)\overset{\mathbb{P}}{\to}\infty$.

To show the two last remaining claims, we shall first prove that

$$\sup_{1\le s\le t\le T+1}\left|m^{-1}\int_{\mathbb{R}^d}\{\mathbb{H}_{m,q}(s,t,\boldsymbol{x})\}^2\mathrm{d}F_{1:\lfloor mt\rfloor}(\boldsymbol{x})-\int_{\mathbb{R}^d}\{K_{c,q}(s,t,\boldsymbol{x})\}^2\mathrm{d}H_t(\boldsymbol{x})\right|\overset{\mathbb{P}}{\to}0,$$
(A.7)

where, for any $t\in[1,T+1]$ and $\boldsymbol{x}\in\mathbb{R}^d$,

$$H_t(\boldsymbol{x})=\frac{c\wedge t}{t}F(\boldsymbol{x})+\frac{t-c\wedge t}{t}G(\boldsymbol{x}).\tag{A.8}$$

By the triangle inequality, the supremum on the left hand-side of (A.7) is smaller than I_m+J_m, where

$$I_m=\sup_{1\le s\le t\le T+1}\left|m^{-1}\int_{\mathbb{R}^d}\{\mathbb{H}_{m,q}(s,t,\boldsymbol{x})\}^2\mathrm{d}F_{1:\lfloor mt\rfloor}(\boldsymbol{x})-\int_{\mathbb{R}^d}\{K_{c,q}(s,t,\boldsymbol{x})\}^2\mathrm{d}F_{1:\lfloor mt\rfloor}(\boldsymbol{x})\right|$$

and

$$J_m=\sup_{1\le s\le t\le T+1}\left|\int_{\mathbb{R}^d}\{K_{c,q}(s,t,\boldsymbol{x})\}^2\mathrm{d}F_{1:\lfloor mt\rfloor}(\boldsymbol{x})-\int_{\mathbb{R}^d}\{K_{c,q}(s,t,\boldsymbol{x})\}^2\mathrm{d}H_t(\boldsymbol{x})\right|.$$

On one hand, some thought reveals that

$$I_m\le\sup_{\substack{1\le s\le t\le T+1\\ \boldsymbol{x}\in\mathbb{R}^d}}\left|\{m^{-1/2}\mathbb{H}_{m,q}(s,t,\boldsymbol{x})\}^2-\{K_{c,q}(s,t,\boldsymbol{x})\}^2\right|\overset{\mathbb{P}}{\to}0$$

as a consequence of the continuous mapping theorem. On the other hand, from (2.27), we have that

$$J_m\le\sup_{1\le s\le t\le T+1}\frac{[(s\wedge c)\{(t\vee c)-(s\vee c)\}]^2}{\{q(s,t)\}^2}$$
$$\times\sup_{1\le t\le T+1}\left|\int_{\mathbb{R}^d}\{F(\boldsymbol{x})-G(\boldsymbol{x})\}^2\mathrm{d}F_{1:\lfloor mt\rfloor}(\boldsymbol{x})-\int_{\mathbb{R}^d}\{F(\boldsymbol{x})-G(\boldsymbol{x})\}^2\mathrm{d}H_t(\boldsymbol{x})\right|.$$

Using (A.8), the second supremum on the right-hand side of the previous display is smaller than $\max\{K_m, L_m\}$, where

$$K_m = \sup_{1 \le t \le c} \left| \int_{\mathbb{R}^d} \{F(\boldsymbol{x}) - G(\boldsymbol{x})\}^2 \mathrm{d}F_{1:\lfloor mt \rfloor}(\boldsymbol{x}) - \int_{\mathbb{R}^d} \{F(\boldsymbol{x}) - G(\boldsymbol{x})\}^2 \mathrm{d}F(\boldsymbol{x}) \right|,$$

$$L_m = \sup_{c \le t \le T+1} \left| \int_{\mathbb{R}^d} \{F(\boldsymbol{x}) - G(\boldsymbol{x})\}^2 \mathrm{d}F_{1:\lfloor mt \rfloor}(\boldsymbol{x}) - \int_{\mathbb{R}^d} \{F(\boldsymbol{x}) - G(\boldsymbol{x})\}^2 \mathrm{d}H_t(\boldsymbol{x}) \right|.$$

From the assumptions on the strong mixing coefficients and Theorem 1.2 in Berbee (1987) (see also Rio, 2017, Chapter 3), the strong law of large numbers implies that, as $k \to \infty$,

$$M_k = \left| \frac{1}{k} \sum_{i=1}^{k} \{F(\boldsymbol{Y}_i) - G(\boldsymbol{Y}_i)\}^2 - \int_{\mathbb{R}^d} \{F(\boldsymbol{x}) - G(\boldsymbol{x})\}^2 \mathrm{d}F(\boldsymbol{x}) \right| \overset{as}{\to} 0,$$

where the arrow '$\overset{as}{\to}$' denotes almost sure convergence. Since the previous convergence is equivalent to the fact that $\limsup_{m \to \infty} M_m = 0$ almost surely, we obtain that $K_m = \sup_{m \le k \le \lfloor mc \rfloor} M_k \le \sup_{m \le k} M_k \overset{as}{\to} 0$. Using the fact that, for any $c \le t \le T+1$ and $\boldsymbol{x} \in \mathbb{R}^d$,

$$F_{1:\lfloor mt \rfloor}(\boldsymbol{x}) = \frac{\lfloor mc \rfloor}{\lfloor mt \rfloor} F_{1:\lfloor mc \rfloor}(\boldsymbol{x}) + \frac{\lfloor mt \rfloor - \lfloor mc \rfloor}{\lfloor mt \rfloor} F_{\lfloor mc \rfloor+1:\lfloor mt \rfloor}(\boldsymbol{x})$$

and that $\sup_{c \le t \le T+1} |\lfloor mc \rfloor / \lfloor mt \rfloor - c/t| = O(1/m)$, we obtain that

$$
\begin{aligned}
L_m \le {} & \left| \int_{\mathbb{R}^d} \{F(\boldsymbol{x}) - G(\boldsymbol{x})\}^2 \mathrm{d}F_{1:\lfloor mc \rfloor}(\boldsymbol{x}) - \int_{\mathbb{R}^d} \{F(\boldsymbol{x}) - G(\boldsymbol{x})\}^2 \mathrm{d}F(\boldsymbol{x}) \right| \\
& + \sup_{c \le t \le T+1} \frac{\lfloor mt \rfloor - \lfloor mc \rfloor}{\lfloor mt \rfloor} \left| \int_{\mathbb{R}^d} \{F(\boldsymbol{x}) - G(\boldsymbol{x})\}^2 \mathrm{d}F_{\lfloor mc \rfloor+1:\lfloor mt \rfloor}(\boldsymbol{x}) \right. \\
& \left. - \int_{\mathbb{R}^d} \{F(\boldsymbol{x}) - G(\boldsymbol{x})\}^2 \mathrm{d}G(\boldsymbol{x}) \right| + O(1/m).
\end{aligned}
$$

The first term on the right-hand side of the previous display is equal to $M_{\lfloor mc \rfloor}$ and thus converges to zero almost surely. The second term can be written as

$$\sup_{\lfloor mc \rfloor+1 \le k \le \lfloor m(T+1) \rfloor} \left| \frac{1}{k} \sum_{i=\lfloor mc \rfloor+1}^{k} \{F(\boldsymbol{Z}_i) - G(\boldsymbol{Z}_i)\}^2 - \frac{k - \lfloor mc \rfloor}{k} \int_{\mathbb{R}^d} \{F(\boldsymbol{x}) - G(\boldsymbol{x})\}^2 \mathrm{d}G(\boldsymbol{x}) \right|. \tag{A.9}$$

Letting

$$N_k = \left| \frac{1}{k} \sum_{i=1}^{k} \{F(\boldsymbol{Z}_i) - G(\boldsymbol{Z}_i)\}^2 - \int_{\mathbb{R}^d} \{F(\boldsymbol{x}) - G(\boldsymbol{x})\}^2 \mathrm{d}G(\boldsymbol{x}) \right|$$

and decomposing the sum in (A.9), by the triangle inequality, (A.9) is smaller than

$$\sup_{\lfloor mc \rfloor+1 \le k \le \lfloor m(T+1) \rfloor} N_k + N_{\lfloor mc \rfloor} \le \sup_{\lfloor mc \rfloor+1 \le k} N_k + N_{\lfloor mc \rfloor} \overset{as}{\to} 0,$$

since $N_k \overset{as}{\to} 0$ as $k \to \infty$.

Hence, $J_m \overset{as}{\to} 0$. It follows that (A.7) is proven and, from the continuous mapping theorem, we immediately obtain that $m^{-1}\mathbb{S}_{m,q} \overset{\mathbb{P}}{\to} M_{c,q}$ and $m^{-1}\mathbb{T}_{m,q} \overset{\mathbb{P}}{\to} N_{c,q}$ in $\ell^\infty([1, T+1])$, where

$$M_{c,q}(t) = \sup_{s\in[1,t]} \int_{\mathbb{R}^d} \{K_{c,q}(s,t,\boldsymbol{x})\}^2 \mathrm{d}H_t(\boldsymbol{x}) \text{ and } N_{c,q}(t) = \int_1^t \int_{\mathbb{R}^d} \{K_{c,q}(s,t,\boldsymbol{x})\}^2 \mathrm{d}H_t(\boldsymbol{x})\mathrm{d}s,$$

and then that

$$m^{-1} \sup_{t\in[1,T+1]} \mathbb{S}_{m,q}(t) \overset{\mathbb{P}}{\to} \sup_{t\in[1,T+1]} M_{c,q}(t),$$

$$m^{-1} \sup_{t\in[1,T+1]} \mathbb{T}_{m,q}(t) \overset{\mathbb{P}}{\to} \sup_{t\in[1,T+1]} N_{c,q}(t).$$

Let $\Lambda_F = \int_{\mathbb{R}^d} \{F(\boldsymbol{x}) - G(\boldsymbol{x})\}^2 \mathrm{d}F(\boldsymbol{x})$ and $\Lambda_G = \int_{\mathbb{R}^d} \{F(\boldsymbol{x}) - G(\boldsymbol{x})\}^2 \mathrm{d}G(\boldsymbol{x})$. Since F and G are continuous and $F \neq G$, $\Lambda_F > 0$ and $\Lambda_G > 0$. As a consequence, for all $t \in [1, T+1]$,

$$\Lambda_t = \frac{c \wedge t}{t}\Lambda_F + \frac{t - c \wedge t}{t}\Lambda_G > 0.$$

Furthermore, since $q(s,t) > 0$ for all $(s,t) \in \Delta$ and since, for all $c < t \leq T+1$,

$$\sup_{1\leq s\leq t} (s \wedge c)\{(t \vee c) - (s \vee c)\} = c(t-c) > 0,$$

we have, from the continuity of q, $t \mapsto \Lambda_t$ and $(s,t) \mapsto (s \wedge c)\{(t \vee c) - (s \vee c)\}$, that

$$\sup_{t\in[1,T+1]} M_{c,q}(t) \geq \sup_{t\in[c,T+1]} \Lambda_t \sup_{s\in[1,t]} \left[\frac{(s \wedge c)\{t - (s \vee c)\}}{q(s,t)}\right]^2 > 0,$$

and

$$\sup_{t\in[1,T+1]} N_{c,q}(t) \geq \sup_{t\in[c,T+1]} \Lambda_t \int_1^t \left[\frac{(s \wedge c)\{t - (s \vee c)\}}{q(s,t)}\right]^2 \mathrm{d}s > 0.$$

$\square$

Appendix B: Auxiliary lemmas for the proof of Theorem 3.3

This section, which is, to a large extent, notationally independent of the rest of the paper, provides the proofs of two lemmas, possibly of independent interest, necessary for showing Theorem 3.3.

Let $\boldsymbol{\mathcal{X}}_n$ denote available data. No assumptions are made on $\boldsymbol{\mathcal{X}}_n$ apart from measurability. To fix ideas, one can think of $\boldsymbol{\mathcal{X}}_n$ as a sequence of n multivariate serially dependent random vectors. Let $\boldsymbol{S}_n = \boldsymbol{S}_n(\boldsymbol{\mathcal{X}}_n)$ be a $\mathbb{R}^p$-valued statistic such that $\boldsymbol{S}_n = (S_{1,n}, \ldots, S_{p,n}) \rightsquigarrow \boldsymbol{S} = (S_1, \ldots, S_p)$ as $n \to \infty$, where the random vector $\boldsymbol{S}$ is assumed to have a continuous d.f. We additionally suppose that we have available *bootstrap*

replicates $\boldsymbol{S}_n^{[i]} = \boldsymbol{S}_n^{[i]}(\boldsymbol{\mathcal{X}}_n, \boldsymbol{\mathcal{W}}_n^{[i]})$ of $\boldsymbol{S}_n$, where the $\boldsymbol{\mathcal{W}}_n^{[i]}$, $i \in \mathbb{N}$, are n-dimensional independent and identically distributed random vectors representing the additional sources of randomness involved in the underlying bootstrap mechanism. We shall further assume that, as $n \to \infty$,

$$\left(\boldsymbol{S}_n, \boldsymbol{S}_n^{[1]}, \boldsymbol{S}_n^{[2]}\right) \rightsquigarrow \left(\boldsymbol{S}, \boldsymbol{S}^{[1]}, \boldsymbol{S}^{[2]}\right), \tag{B.1}$$

in $(\mathbb{R}^p)^3$, where $\boldsymbol{S}^{[1]}$ and $\boldsymbol{S}^{[2]}$ are independent copies of $\boldsymbol{S}$. Note that, from Lemma 2.2 of Bücher and Kojadinovic (2019), (B.1) is equivalent to the usual conditional bootstrap consistency statement, that is,

$$\sup_{\boldsymbol{x} \in \mathbb{R}^p} |\mathbb{P}(\boldsymbol{S}_n^{[1]} \leq \boldsymbol{x} \mid \boldsymbol{\mathcal{X}}_n) - \mathbb{P}(\boldsymbol{S}_n \leq \boldsymbol{x})| \overset{\mathbb{P}}{\to} 0, \qquad \text{as } n \to \infty.$$

Before stating and proving the two lemmas, we introduce some additional notation and list useful results. For any $q \in \{1, \ldots, p\}$, $\{j_1, \ldots, j_q\} \subset \{1, \ldots, p\}$, $x_{j_1}, \ldots, x_{j_q} \in \mathbb{R}$ and $B \in \mathbb{N}$, let

$$F_{\boldsymbol{S}, \{j_1, \ldots, j_q\}}(x_{j_1}, \ldots, x_{j_q}) = \mathbb{P}(S_{j_1} \leq x_{j_1}, \ldots, S_{j_q} \leq x_{j_q}),$$
$$F_{\boldsymbol{S}_n, \{j_1, \ldots, j_q\}}(x_{j_1}, \ldots, x_{j_q}) = \mathbb{P}(S_{j_1, n} \leq x_{j_1}, \ldots, S_{j_q, n} \leq x_{j_q}),$$
$$F_{\boldsymbol{S}_n, \{j_1, \ldots, j_q\}}^B(x_{j_1}, \ldots, x_{j_q}) = \frac{1}{B} \sum_{i=1}^{B} \mathbf{1}(S_{j_1, n}^{[i]} \leq x_{j_1}, \ldots, S_{j_q, n}^{[i]} \leq x_{j_q}).$$

Since $\boldsymbol{S}_n \rightsquigarrow \boldsymbol{S}$ in $\mathbb{R}^p$ as $n \to \infty$ and $\boldsymbol{S}$ has a continuous d.f., we have from Lemma 2.11 of van der Vaart (1998) that, for any $q \in \{1, \ldots, p\}$ and $\{j_1, \ldots, j_q\} \subset \{1, \ldots, p\}$, as $n \to \infty$,

$$\sup_{(x_{j_1}, \ldots, x_{j_q}) \in \mathbb{R}^q} |F_{\boldsymbol{S}_n, \{j_1, \ldots, j_q\}}(x_{j_1}, \ldots, x_{j_q}) - F_{\boldsymbol{S}, \{j_1, \ldots, j_q\}}(x_{j_1}, \ldots, x_{j_q})| \to 0. \tag{B.2}$$

Proceeding as in the proof of the aforementioned lemma, it can actually also be shown that, for any $q \in \{1, \ldots, p\}$ and $\{j_1, \ldots, j_q\} \subset \{1, \ldots, p\}$, as $n \to \infty$,

$$\sup_{(x_{j_1}, \ldots, x_{j_q}) \in \mathbb{R}^q} |\mathbb{P}(S_{j_1, n} < x_{j_1}, \ldots, S_{j_q, n} < x_{j_q}) - \mathbb{P}(S_{j_1} < x_{j_1}, \ldots, S_{j_q} < x_{j_q})| \to 0. \tag{B.3}$$

Combining (B.1) with Assertion (f) of Lemma 2.2 in Bücher and Kojadinovic (2019) and (B.2), we further obtain that, for any $q \in \{1, \ldots, p\}$ and $\{j_1, \ldots, j_q\} \subset \{1, \ldots, p\}$, as $n, B \to \infty$,

$$\sup_{(x_{j_1}, \ldots, x_{j_q}) \in \mathbb{R}^q} |F_{\boldsymbol{S}_n, \{j_1, \ldots, j_q\}}^B(x_{j_1}, \ldots, x_{j_q}) - F_{\boldsymbol{S}, \{j_1, \ldots, j_q\}}(x_{j_1}, \ldots, x_{j_q})| \overset{\mathbb{P}}{\to} 0. \tag{B.4}$$

Let $\xi \in (0, 1)$ be arbitrary. The following notation will also be used in the lemmas. Let

$$F_{\boldsymbol{S}, 1}(x) = F_{\boldsymbol{S}, \{1\}}(x) = \mathbb{P}(S_1 \leq x), \; x \in \mathbb{R}, \qquad g_1 = F_{\boldsymbol{S}, 1}^{-1}(1 - \xi),$$

and, recursively, for j successively equal to $2, \ldots, p$,

$$F_{\boldsymbol{S},j}(x) = \frac{F_{\boldsymbol{S},\{1,\ldots,j\}}(g_1, \ldots, g_{j-1}, x)}{F_{\boldsymbol{S},\{1,\ldots,j-1\}}(g_1, \ldots, g_{j-1})}, \; x \in \mathbb{R}, \qquad g_j = F_{\boldsymbol{S},j}^{-1}(1 - \xi). \tag{B.5}$$

Similarly, let

$$F_{\boldsymbol{S}_n,1}^B(x) = F_{\boldsymbol{S}_n,\{1\}}^B(x), \; x \in \mathbb{R}, \qquad g_{1,n}^B = F_{\boldsymbol{S}_n,1}^{B,-1}(1 - \xi),$$

and, recursively, for j successively equal to $2, \ldots, p$,

$$F_{\boldsymbol{S}_n,j}^B(x) = \frac{F_{\boldsymbol{S}_n,\{1,\ldots,j\}}^B(g_{1,n}^B, \ldots, g_{j-1,n}^B, x)}{F_{\boldsymbol{S}_n,\{1,\ldots,j-1\}}^B(g_{1,n}^B, \ldots, g_{j-1,n}^B)}, \; x \in \mathbb{R}, \qquad g_{j,n}^B = F_{\boldsymbol{S}_n,j}^{B,-1}(1 - \xi). \tag{B.6}$$

The following two lemmas are instrumental for proving Theorem 3.3.

Lemma B.1. *As $n, B \to \infty$,*

$$\sup_{x \in \mathbb{R}} |F_{\boldsymbol{S}_n,1}^B(x) - F_{\boldsymbol{S},1}(x)| \xrightarrow{\mathbb{P}} 0, \tag{B.7}$$

$$\mathbb{P}(S_{1,n} \le g_{1,n}^B) \to 1 - \xi, \tag{B.8}$$

$$F_{\boldsymbol{S},1}(g_{1,n}^B) \to 1 - \xi, \tag{B.9}$$

and, for any $j \in \{2, \ldots, p\}$,

$$\sup_{x \in \mathbb{R}} |F_{\boldsymbol{S}_n,j}^B(x) - F_{\boldsymbol{S},j}(x)| \xrightarrow{\mathbb{P}} 0, \tag{B.10}$$

$$\mathbb{P}(S_{j,n} \le g_{j,n}^B \mid S_{1,n} \le g_{1,n}^B, \ldots, S_{j-1,n} \le g_{j-1,n}^B) \to 1 - \xi, \tag{B.11}$$

$$F_{\boldsymbol{S},j}(g_{j,n}^B) \to 1 - \xi. \tag{B.12}$$

The statements (B.8) *and* (B.11) *with '$\le$' replaced by '$<$' hold as well.*

Lemma B.2. *Let $\boldsymbol{T}_n = \boldsymbol{T}_n(\boldsymbol{\mathcal{X}}_n) = (T_{1,n}, \ldots, T_{p,n})$ be a $\mathbb{R}^p$-valued statistic such that $\max_{1 \le j \le p} T_{j,n} \xrightarrow{\mathbb{P}} \infty$ as $n \to \infty$. Then*

$$\mathbb{P}(T_{1,n} \le g_{1,n}^B, \ldots, T_{p,n} \le g_{p,n}^B) \to 0 \qquad \text{as } n, B \to \infty.$$

Proof of Lemma B.1. The claims in (B.7) and (B.8) are immediate consequences of (B.4) and Lemma 4.2 in Bücher and Kojadinovic (2019), respectively. The claim (B.9) follows from the fact that $F_{\boldsymbol{S},1}(g_1) = 1 - \xi$, the triangle inequality, (B.2) and (B.8).

To prove (B.10), (B.11) and (B.12), we proceed by induction on j.

Proof of (B.10), (B.11) *and* (B.12) *for $j=2$:* By the triangle inequality, $|F_{\boldsymbol{S}_n,1}^B(g_{1,n}^B) - F_{\boldsymbol{S},1}(g_1)|$ is smaller than

$$|F_{\boldsymbol{S}_n,1}^B(g_{1,n}^B) - F_{\boldsymbol{S},1}(g_{1,n}^B)| + |F_{\boldsymbol{S},1}(g_{1,n}^B) - F_{\boldsymbol{S}_n,1}(g_{1,n}^B)| + |F_{\boldsymbol{S}_n,1}(g_{1,n}^B) - F_{\boldsymbol{S},1}(g_1)|.$$

The first term converges in probability to zero by (B.4) as $n, B \to \infty$. The second term converges to zero by (B.2) as $n, B \to \infty$. The third term converges to zero as

a consequence of (B.8) since $F_{\boldsymbol{S},1}(g_1) = 1 - \xi$. Hence, $|F_{\boldsymbol{S}_n,1}^B(g_{1,n}^B) - F_{\boldsymbol{S},1}(g_1)| \xrightarrow{\mathbb{P}} 0$ as $n, B \to \infty$. From (B.5) and (B.6), the latter implies that to prove that $\sup_{x \in \mathbb{R}} |F_{\boldsymbol{S}_n,2}^B(x) - F_{\boldsymbol{S},2}(x)| \xrightarrow{\mathbb{P}} 0$ as $n, B \to \infty$, it suffices to prove that

$$\sup_{x \in \mathbb{R}} |F_{\boldsymbol{S}_n,\{1,2\}}^B(g_{1,n}^B, x) - F_{\boldsymbol{S},\{1,2\}}(g_1, x)| \xrightarrow{\mathbb{P}} 0$$

as $n, B \to \infty$. From the triangle inequality, (B.2) and (B.4), the latter will hold if

$$\sup_{x \in \mathbb{R}} |F_{\boldsymbol{S}_n,\{1,2\}}(g_{1,n}^B, x) - F_{\boldsymbol{S}_n,\{1,2\}}(g_1, x)| \to 0 \tag{B.13}$$

as $n, B \to \infty$. For any $x \in \mathbb{R}$, we have

$$F_{\boldsymbol{S}_n,\{1,2\}}(g_{1,n}^B, x) - F_{\boldsymbol{S}_n,\{1,2\}}(g_1, x) = \mathbb{P}(S_{1,n} \geq g_1, S_{2,n} \leq x) - \mathbb{P}(S_{1,n} = g_1, S_{2,n} \leq x)$$
$$- \mathbb{P}(S_{1,n} \geq g_{1,n}^B, S_{2,n} \leq x) + \mathbb{P}(S_{1,n} = g_{1,n}^B, S_{2,n} \leq x). \tag{B.14}$$

Hence, by the triangle inequality,

$$\sup_{x \in \mathbb{R}} |F_{\boldsymbol{S}_n,\{1,2\}}(g_{1,n}^B, x) - F_{\boldsymbol{S}_n,\{1,2\}}(g_1, x)|$$
$$\leq \sup_{x \in \mathbb{R}} |\mathbb{P}(S_{1,n} \geq g_{1,n}^B, S_{2,n} \leq x) - \mathbb{P}(S_{1,n} \geq g_1, S_{2,n} \leq x)| + 2 \sup_{x \in \mathbb{R}} \mathbb{P}(S_{1,n} = x). \tag{B.15}$$

From the triangle inequality, (B.2) and (B.3), we obtain that the last supremum on the right-hand side of the previous display converges to zero a $n \to \infty$ since $\boldsymbol{S}$ has a continuous d.f. Hence, to show (B.13), it remains to show that the first supremum on the right converges to zero in probability as $n, B \to \infty$. From the right-continuity of $F_{\boldsymbol{S},1}$ and $F_{\boldsymbol{S}_n,1}^B$, we obtain that, for any $x \in \mathbb{R}$,

$$|\mathbb{P}(S_{1,n} \geq g_{1,n}^B, S_{2,n} \leq x) - \mathbb{P}(S_{1,n} \geq g_1, S_{2,n} \leq x)|$$
$$= |\mathbb{P}\{S_{1,n} \geq F_{\boldsymbol{S}_n,1}^{B,-1}(1 - \xi), S_{2,n} \leq x\} - \mathbb{P}\{S_{1,n} \geq F_{\boldsymbol{S},1}^{-1}(1 - \xi), S_{2,n} \leq x\}|$$
$$= |\mathbb{P}\{F_{\boldsymbol{S}_n,1}^B(S_{1,n}) \geq 1 - \xi, S_{2,n} \leq x\} - \mathbb{P}\{F_{\boldsymbol{S},1}(S_{1,n}) \geq 1 - \xi, S_{2,n} \leq x\}|.$$

Furthermore, using the fact that, for any $a, b, y \in \mathbb{R}$ and $\varepsilon > 0$,

$$|\mathbf{1}(y \leq a) - \mathbf{1}(y \leq b)| \leq \mathbf{1}(|y - a| \leq \varepsilon) + \mathbf{1}(|a - b| > \varepsilon), \tag{B.16}$$

we obtain that, for any $\varepsilon > 0$,

$$|\mathbb{P}\{F_{\boldsymbol{S}_n,1}^B(S_{1,n}) \geq 1 - \xi, S_{2,n} \leq x\} - \mathbb{P}\{F_{\boldsymbol{S},1}(S_{1,n}) \geq 1 - \xi, S_{2,n} \leq x\}|$$
$$= |\mathbb{E}[\mathbf{1}\{F_{\boldsymbol{S}_n,1}^B(S_{1,n}) \geq 1 - \xi, S_{2,n} \leq x\} - \mathbf{1}\{F_{\boldsymbol{S},1}(S_{1,n}) \geq 1 - \xi, S_{2,n} \leq x\}]|$$
$$\leq \mathbb{E}|\mathbf{1}\{F_{\boldsymbol{S}_n,1}^B(S_{1,n}) \geq 1 - \xi, S_{2,n} \leq x\} - \mathbf{1}\{F_{\boldsymbol{S},1}(S_{1,n}) \geq 1 - \xi, S_{2,n} \leq x\}| \tag{B.17}$$
$$\leq \mathbb{E}|\mathbf{1}\{F_{\boldsymbol{S}_n,1}^B(S_{1,n}) \geq 1 - \xi\} - \mathbf{1}\{F_{\boldsymbol{S},1}(S_{1,n}) \geq 1 - \xi\}|$$
$$\leq \mathbb{P}\{|F_{\boldsymbol{S},1}(S_{1,n}) - 1 + \xi| \leq \varepsilon\} + \mathbb{P}\{|F_{\boldsymbol{S}_n,1}^B(S_{1,n}) - F_{\boldsymbol{S},1}(S_{1,n})| > \varepsilon\}.$$

By the continuous mapping theorem and the Portmanteau theorem, the first probability on the right converges to $\mathbb{P}\{|F_{S_1}(S_1) - 1 + \xi| \leq \varepsilon\}$ as $n \to \infty$ and can be made arbitrarily small by decreasing ε. The second probability converges to zero as $n, B \to \infty$ by (B.7) for any $\varepsilon > 0$. Hence, (B.13) holds and so does (B.10) for $j = 2$.

Let us now prove (B.11) for $j = 2$. Since, from (B.8), $\mathbb{P}(S_{1,n} \leq g_{1,n}^B) \to P(S_1 \leq g_1) = 1 - \xi$, as $n, B \to \infty$, it remains to show that

$$\mathbb{P}(S_{1,n} \leq g_{1,n}^B, S_{2,n} \leq g_{2,n}^B) \to \mathbb{P}(S_1 \leq g_1, S_2 \leq g_2)$$

as $n, B \to \infty$. From the triangle inequality and (B.2), if suffices to prove that, as $n, B \to \infty$,

$$|F_{\boldsymbol{S}_n,\{1,2\}}(g_{1,n}^B, g_{2,n}^B) - F_{\boldsymbol{S}_n,\{1,2\}}(g_1, g_2)| \to 0.$$

The term on the left-hand side of the previous display is smaller than

$$|F_{\boldsymbol{S}_n,\{1,2\}}(g_{1,n}^B, g_{2,n}^B) - F_{\boldsymbol{S}_n,\{1,2\}}(g_1, g_{2,n}^B)| + |F_{\boldsymbol{S}_n,\{1,2\}}(g_1, g_{2,n}^B) - F_{\boldsymbol{S}_n,\{1,2\}}(g_1, g_2)|.$$

The first difference between absolute values is smaller than (B.15), which was already shown to converge to zero as $n, B \to \infty$. Proceeding as in (B.14), to show that the second difference between absolute values converges to zero as $n, B \to \infty$, it suffices to prove that

$$\begin{aligned}
|\mathbb{P}(S_{1,n} \leq g_1, S_{2,n} &\geq g_{2,n}^B) - \mathbb{P}(S_{1,n} \leq g_1, S_{2,n} \geq g_2)| \\
&= |\mathbb{P}\{S_{1,n} \leq g_1, S_{2,n} \geq F_{\boldsymbol{S}_n,2}^{B,-1}(1-\xi)\} - \mathbb{P}\{S_{1,n} \leq g_1, S_{2,n} \geq F_{\boldsymbol{S},2}^{-1}(1-\xi)\}| \to 0
\end{aligned}$$

as $n, B \to \infty$. Using again (B.16) and proceeding as in (B.17), the probability on the left can be shown to be smaller than

$$\mathbb{P}\{|F_{\boldsymbol{S},2}(S_{2,n}) - 1 + \xi| \leq \varepsilon\} + \mathbb{P}\{|F_{\boldsymbol{S}_n,2}^B(S_{2,n}) - F_{\boldsymbol{S},2}(S_{2,n})| > \varepsilon\}$$

for any $\varepsilon > 0$. Using the continuity of $F_{\boldsymbol{S},2}$ and the Portmanteau theorem, the first probability converges as $n \to \infty$ to a probability that can be made arbitrarily small by decreasing ε. The second probability converges to zero as $n, B \to \infty$ for any $\varepsilon > 0$ by (B.10) for $j = 2$, which was proven previously.

It remains to prove (B.12) for $j = 2$. From the triangle inequality, (B.2) and (B.13), we have that, as $n, B \to \infty$,

$$\sup_{x \in \mathbb{R}} |F_{\boldsymbol{S}_n,\{1,2\}}(g_{1,n}^B, x) - F_{\boldsymbol{S},\{1,2\}}(g_1, x)| \to 0.$$

Starting from (B.11) for $j = 2$, the desired result follows from the convergence in the previous display and (B.8).

Proof of (B.10), (B.11) *and* (B.12) *for all* $j \in \{3, \ldots, p\}$: As mentioned previously, we proceed by induction. Let $k \in \{3, \ldots, p\}$, assume that (B.10), (B.11) and (B.12) hold for all $j \in \{2, \ldots, k-1\}$ and let us show that they also hold for k.

For any $x \in \mathbb{R}$, let $F_{\boldsymbol{S}_n, 1}(x) = F_{\boldsymbol{S}_n, \{1\}}(x)$, and, for j successively equal to $2, \dots, p$, let

$$F_{\boldsymbol{S}_n, j}(x) = \frac{F_{\boldsymbol{S}_n, \{1, \dots, j\}}(g_{1,n}^B, \dots, g_{j-1,n}^B, x)}{F_{\boldsymbol{S}_n, \{1, \dots, j-1\}}(g_{1,n}^B, \dots, g_{j-1,n}^B)}.$$

Then, from the induction hypothesis for (B.11), as $n, B \to \infty$,

$$
\begin{aligned}
F_{\boldsymbol{S}_n, \{1, \dots, k-1\}}(g_{1,n}^B, \dots, g_{k-1}^B) &= F_{\boldsymbol{S}_n, k-1}(g_{k-1,n}^B) F_{\boldsymbol{S}_n, k-2}(g_{k-2,n}^B) \cdots F_{\boldsymbol{S}_n, 1}(g_{1,n}^B) \\
&\to F_{\boldsymbol{S}, k-1}(g_{k-1}) F_{\boldsymbol{S}, k-2}(g_{k-2}) \cdots F_{\boldsymbol{S}, 1}(g_1) = (1 - \xi)^{k-1} = F_{\boldsymbol{S}, \{1, \dots, k-1\}}(g_1, \dots, g_{k-1}).
\end{aligned}
\tag{B.18}
$$

Thus, from (B.4), as $n, B \to \infty$, $F_{\boldsymbol{S}_n, \{1, \dots, k-1\}}^B(g_{1,n}^B, \dots, g_{k-1}^B) \xrightarrow{\mathbb{P}} F_{\boldsymbol{S}, \{1, \dots, k-1\}}(g_1, \dots, g_{k-1})$. Hence, to show (B.10) for $j = k$, it suffices to prove that, as $n, B \to \infty$,

$$\sup_{x \in \mathbb{R}} |F_{\boldsymbol{S}_n, \{1, \dots, k\}}^B(g_{1,n}^B, \dots, g_{k-1,n}^B, x) - F_{\boldsymbol{S}, \{1, \dots, k\}}(g_1, \dots, g_{k-1}, x)| \xrightarrow{\mathbb{P}} 0.$$

From the triangle inequality, (B.2) and (B.4), the latter will hold if, as $n, B \to \infty$,

$$\sup_{x \in \mathbb{R}} |F_{\boldsymbol{S}_n, \{1, \dots, k\}}(g_{1,n}^B, \dots, g_{k-1,n}^B, x) - F_{\boldsymbol{S}_n, \{1, \dots, k\}}(g_1, \dots, g_{k-1}, x)| \to 0. \tag{B.19}$$

For any $x \in \mathbb{R}$, $|F_{\boldsymbol{S}_n, \{1, \dots, k\}}(g_{1,n}^B, \dots, g_{k-1,n}^B, x) - F_{\boldsymbol{S}_n, \{1, \dots, k\}}(g_1, \dots, g_{k-1}, x)|$ is smaller than the following sum of $k - 1$ terms:

$$
\begin{aligned}
&|F_{\boldsymbol{S}_n, \{1, \dots, k\}}(g_{1,n}^B, g_{2,n}^B, \dots, g_{k-1,n}^B, x) - F_{\boldsymbol{S}_n, \{1, \dots, k\}}(g_1, g_{2,n}^B, \dots, g_{k-1,n}^B, x)| \\
+&|F_{\boldsymbol{S}_n, \{1, \dots, k\}}(g_1, g_{2,n}^B, \dots, g_{k-1,n}^B, x) - F_{\boldsymbol{S}_n, \{1, \dots, k\}}(g_1, g_2, g_{3,n}^B, \dots, g_{k-1,n}^B, x)| \\
&\vdots \\
+&|F_{\boldsymbol{S}_n, \{1, \dots, k\}}(g_1, \dots, g_{k-2}, g_{k-1,n}^B, x) - F_{\boldsymbol{S}_n, \{1, \dots, k\}}(g_1, \dots, g_{k-1}, x)|.
\end{aligned}
$$

Proceeding as in (B.14), (B.15) and (B.17), the jth term, $j \in \{1, \dots, k-1\}$, can be shown to converge to zero as $n, B \to \infty$ as a consequence of the fact that

$$\mathbb{P}\{|F_{\boldsymbol{S}, j}(S_{j,n}) - 1 + \xi| \le \varepsilon\} + \mathbb{P}\{|F_{\boldsymbol{S}_n, j}^B(S_{j,n}) - F_{\boldsymbol{S}, j}(S_{j,n})| > \varepsilon\}$$

converges to zero as $n, B \to \infty$ followed by $\varepsilon \downarrow 0$ using the continuity of $F_{\boldsymbol{S}, j}$ and the induction hypothesis. Hence, (B.19) holds.

Let us now show (B.11) for $j = k$. From (B.18), it suffices to prove that, as $n, B \to \infty$,

$$|F_{\boldsymbol{S}_n, \{1, \dots, k\}}(g_{1,n}^B, \dots, g_{k,n}^B) - F_{\boldsymbol{S}, \{1, \dots, k\}}(g_1, \dots, g_k)| \to 0.$$

From the triangle inequality and (B.2), the latter will hold if, as $n, B \to \infty$,

$$|F_{\boldsymbol{S}_n, \{1, \dots, k\}}(g_{1,n}^B, \dots, g_{k,n}^B) - F_{\boldsymbol{S}_n, \{1, \dots, k\}}(g_1, \dots, g_k)| \to 0.$$

The difference between absolute values on the left of the previous display is smaller than

$$|F_{\boldsymbol{S}_n,\{1,\ldots,k\}}(g_{1,n}^B,\ldots,g_{k-1,n}^B,g_{k,n}^B) - F_{\boldsymbol{S}_n,\{1,\ldots,k\}}(g_1,\ldots,g_{k-1},g_{k,n}^B)|$$
$$+ |F_{\boldsymbol{S}_n,\{1,\ldots,k\}}(g_1,\ldots,g_{k-1},g_{k,n}^B) - F_{\boldsymbol{S}_n,\{1,\ldots,k\}}(g_1,\ldots,g_{k-1},g_k)|.$$

The first term converges to zero in probability as $n, B \to \infty$ as a consequence of (B.19). Proceeding as in (B.14), (B.15) and (B.17), the second term can be shown to converge to zero as $n, B \to \infty$ as a consequence of the fact that

$$\mathbb{P}\{|F_{\boldsymbol{S},k}(S_{k,n}) - 1 + \xi| \le \varepsilon\} + \mathbb{P}\{|F_{\boldsymbol{S}_n,k}^B(S_{k,n}) - F_{\boldsymbol{S},k}(S_{k,n})| > \varepsilon\}$$

converges to zero as $n, B \to \infty$ followed by $\varepsilon \downarrow 0$ using the continuity of $F_{\boldsymbol{S},k}$ and the already proven claim (B.10) for $j = k$.

It finally remains to show (B.12) for $j = k$. From the triangle inequality, (B.2) and (B.19), we have that, as $n, B \to \infty$,

$$\sup_{x \in \mathbb{R}} |F_{\boldsymbol{S}_n,\{1,\ldots,k\}}(g_{1,n}^B,\ldots,g_{k-1,n}^B,x) - F_{\boldsymbol{S},\{1,\ldots,k\}}(g_1,\ldots,g_{k-1},x)| \to 0.$$

Starting from (B.11) for $j = k$, the desired result follows by combining the latter convergence with (B.18).

The induction is thus complete. The fact that the statements (B.7) and (B.10) with '$\le$' replaced by '$<$' hold as well is a consequence of (B.2) and (B.3). $\qquad\square$

Proof of Lemma B.2. Let $M_n = \max_{1 \le j \le p} T_{j,n}$. Then,

$$\mathbb{P}(T_{1,n} \le g_{1,n}^B, \ldots, T_{p,n} \le g_{p,n}^B) \le \mathbb{P}\left(M_n \le \max_{1 \le j \le p} g_{j,n}^B\right)$$
$$\le \mathbb{P}(M_n \le g_{1,n}^B) + \cdots + \mathbb{P}(M_n \le g_{p,n}^B).$$

The proof is complete if we show that, for any $j \in \{1, \ldots, p\}$, $\mathbb{P}(M_n \le g_{j,n}^B) \to 0$ as $n, B \to \infty$. Let $j \in \{1, \ldots, p\}$ and recall the definition of the continuous d.f. $F_{\boldsymbol{S},j}$ defined in (B.5). Then,

$$\mathbb{P}(M_n \le g_{j,n}^B) \le \mathbb{P}\{F_{\boldsymbol{S},j}(M_n) \le F_{\boldsymbol{S},j}(g_{j,n}^B)\} = \mathbb{P}\{F_{\boldsymbol{S},j}(M_n) - F_{\boldsymbol{S},j}(g_{j,n}^B) \le 0\}.$$

Since $M_n \xrightarrow{\mathbb{P}} \infty$ as $n \to \infty$, we have that $F_{\boldsymbol{S},j}(M_n) \xrightarrow{\mathbb{P}} 1$ as $n \to \infty$. Furthermore, from Lemma B.1, $F_{\boldsymbol{S},j}(g_{j,n}^B) \to 1 - \xi$ as $n, B \to \infty$. Hence, $F_{\boldsymbol{S},j}(M_n) - F_{\boldsymbol{S},j}(g_{j,n}^B) \xrightarrow{\mathbb{P}} \xi > 0$ as $n, B \to \infty$, or, equivalently, for any $\varepsilon > 0$,

$$\mathbb{P}\{|F_{\boldsymbol{S},j}(M_n) - F_{\boldsymbol{S},j}(g_{j,n}^B) - \xi| > \varepsilon) \to 0$$

as $n, B \to \infty$. Hence, for any $0 < \varepsilon < \xi$,

$$\mathbb{P}(M_n \le g_{j,n}^B) \le \mathbb{P}\{F_{\boldsymbol{S},j}(M_n) - F_{\boldsymbol{S},j}(g_{j,n}^B) \le 0\} \le \mathbb{P}\{|F_{\boldsymbol{S},j}(M_n) - F_{\boldsymbol{S},j}(g_{j,n}^B) - \xi| > \varepsilon) \to 0$$

as $n, B \to \infty$, which completes the proof. $\qquad\square$

Appendix C: Proof of Theorem 3.3

Proof. Let $\boldsymbol{S}_m = \big(\sup_{t \in I_1} \mathbb{D}_m(t), \ldots, \sup_{t \in I_p} \mathbb{D}_m(t)\big)$ and

$$\boldsymbol{S}_m^{[b]} = \left(\sup_{t \in I_1} \mathbb{D}_m^{[b]}(t), \ldots, \sup_{t \in I_p} \mathbb{D}_m^{[b]}(t)\right), \qquad b \in \mathbb{N},$$

be the corresponding bootstrap replicates of $\boldsymbol{S}_m$. From (3.8) and the continuous mapping theorem, we immediately obtain that, under H_0, $(\boldsymbol{S}_m, \boldsymbol{S}_m^{[1]}, \boldsymbol{S}_m^{[2]}) \rightsquigarrow (\boldsymbol{S}, \boldsymbol{S}^{[1]}, \boldsymbol{S}^{[2]})$ in $(\mathbb{R}^p)^3$, where

$$\boldsymbol{S} = \left(\sup_{t \in I_1} \mathbb{D}_F(t), \ldots, \sup_{t \in I_p} \mathbb{D}_F(t)\right)$$

and $\boldsymbol{S}^{[1]}$ and $\boldsymbol{S}^{[2]}$ are independent copies of $\boldsymbol{S}$. Since $\boldsymbol{S}$ is assumed to have a continuous d.f., the assumptions of Appendix B are satisfied, and, therefore, Lemma B.1 with $n = m$ and $\xi = 1 - (1 - \alpha)^{1/p}$ implies that (3.9) and (3.10) hold. The fact that, under H_0, $\mathbb{P}(\mathbb{D}_m \leq \tau_m^B) \to 1 - \alpha$ as $m, B \to \infty$ follows from a decomposition similar to the one used in (3.3).

To prove the last claim, it suffices to show that, when $\sup_{t \in [1, T+1]} \mathbb{D}_m(t) \xrightarrow{\mathbb{P}} \infty$, $\mathbb{P}(\mathbb{D}_m \leq \tau_m^B) \to 0$ as $m, B \to \infty$. Let $T_{i,m} = \sup_{t \in I_i} \mathbb{D}_m(t)$, $i \in \{1, \ldots, p\}$, which implies that $\sup_{t \in [1, T+1]} \mathbb{D}_m(t) = \max_{1 \leq i \leq p} T_{i,m} \xrightarrow{\mathbb{P}} \infty$. Lemma B.2 with $n = m$ then implies that, as $m, B \to \infty$,

$$\mathbb{P}(T_{1,m} \leq g_{1,m}^B, \ldots, T_{p,m} \leq g_{p,m}^B) = \mathbb{P}(\mathbb{D}_m \leq \tau_m^B) \to 0,$$

which completes the proof. $\qquad \square$

Appendix D: Proof of Proposition 4.1

Proof. Recall that $n = \lfloor m(T+1) \rfloor$ and, for any $b \in \mathbb{N}, s \in [0, 1]$ and $\boldsymbol{u} \in [0, 1]^d$, let

$$\tilde{\mathbb{B}}_n^{[b]}(s, \boldsymbol{u}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \xi_{i,m}^{[b]} \{\mathbf{1}(\boldsymbol{U}_i \leq \boldsymbol{u}) - C_{1:n}(\boldsymbol{u})\}.$$

From Corollary 2.2 of Bücher and Kojadinovic (2016), we have that

$$(\mathbb{B}_n, \tilde{\mathbb{B}}_n^{[1]}, \tilde{\mathbb{B}}_n^{[2]}) \rightsquigarrow (\mathbb{B}_C, \mathbb{B}_C^{[1]}, \mathbb{B}_C^{[2]})$$

in $\{\ell^\infty([0, 1]^{d+1})\}^3$. Proceeding as at the beginning of the proof of Proposition 2.5, from the continuous mapping theorem and with the map ψ defined in (A.1), we obtain that

$$\big(\psi(\mathbb{B}_n), \psi(\tilde{\mathbb{B}}_n^{[1]}), \psi(\tilde{\mathbb{B}}_n^{[2]})\big) \rightsquigarrow (\mathbb{B}_C, \mathbb{B}_C^{[1]}, \mathbb{B}_C^{[2]})$$

in $\{\ell^\infty([0, T+1] \times [0, 1]^d)\}^3$. Recall the definition of $\check{\mathbb{B}}_m^{[b]}$ in (4.1). Since $\psi(\mathbb{B}_n) - \mathbb{B}_m = o_\mathbb{P}(1)$ and $\psi(\tilde{\mathbb{B}}_n^{[b]}) - \check{\mathbb{B}}_m^{[b]} = o_\mathbb{P}(1)$, $b \in \{1, 2\}$, it follows that

$$(\mathbb{B}_m, \check{\mathbb{B}}_m^{[1]}, \check{\mathbb{B}}_m^{[2]}) \rightsquigarrow (\mathbb{B}_C, \mathbb{B}_C^{[1]}, \mathbb{B}_C^{[2]}) \tag{D.1}$$

in $\{\ell^\infty([0, T+1] \times [0, 1]^d)\}^3$. Let ϕ be the continuous map from $\ell^\infty([0, T+1] \times [0, 1]^d)$ to $\ell^\infty([0, T+1] \times [0, 1]^d)$ defined, for any $f \in \ell^\infty([0, T+1] \times [0, 1]^d)$, by

$$\phi(f)(s, \boldsymbol{u}) = \sqrt{T+1} f\{s/(T+1), \boldsymbol{u}\}, \qquad s \in [0, T+1], \boldsymbol{u} \in [0, 1]^d.$$

Furthermore, for any $b \in \mathbb{N}$, $s \in [0, T+1]$ and $\boldsymbol{u} \in [0, 1]^d$, let

$$\bar{\mathbb{B}}_m^{[b]}(s, \boldsymbol{u}) = \frac{1}{\sqrt{m'}} \sum_{i=1}^{\lfloor m's \rfloor} \xi_{i,m}^{[b]} \{\mathbf{1}(\boldsymbol{U}_i \leq \boldsymbol{u}) - C_{1:n}(\boldsymbol{u})\}.$$

Then, starting from (D.1) and using the continuous mapping theorem with the map ϕ as well as the fact that $\phi(\check{\mathbb{B}}_m^{[b]}) - \bar{\mathbb{B}}_m^{[b]} = o_\mathbb{P}(1)$ and $\phi(\mathbb{B}_C^{[b]})$ and $\mathbb{B}_C^{[b]}$ have the same distribution, $b \in \{1, 2\}$, we obtain that

$$(\mathbb{B}_m, \bar{\mathbb{B}}_m^{[1]}, \bar{\mathbb{B}}_m^{[2]}) \rightsquigarrow (\mathbb{B}_C, \mathbb{B}_C^{[1]}, \mathbb{B}_C^{[2]})$$

in $\{\ell^\infty([0, T+1] \times [0, 1]^d)\}^3$.

Fix $b \in \{1, 2\}$. To prove the first claim, it remains to show that

$$I_m = \sup_{s \in [0, T+1]} \sup_{\boldsymbol{u} \in [0, 1]^d} |\hat{\mathbb{B}}_m^{[b]}(s, \boldsymbol{u}) - \bar{\mathbb{B}}_m^{[b]}(s, \boldsymbol{u})| \xrightarrow{\mathbb{P}} 0, \tag{D.2}$$

where $\hat{\mathbb{B}}_m^{[b]}$ is defined in (4.2). Let $J_m = \sup_{\boldsymbol{u} \in [0,1]^d} \sqrt{m'} |C_{1:m}(\boldsymbol{u}) - C_{1:n}(\boldsymbol{u})|$ and

$$K_m = \sup_{s \in [0, T+1]} \frac{1}{m'} \Big| \sum_{i=1}^{\lfloor m's \rfloor} \xi_{i,m}^{[b]} \Big|,$$

and notice that $I_m \leq J_m \times K_m$. On one hand,

$$J_m \leq \sup_{\boldsymbol{u} \in [0,1]^d} \sqrt{m'} |C_{1:m}(\boldsymbol{u}) - C(\boldsymbol{u})| + \sup_{\boldsymbol{u} \in [0,1]^d} \sqrt{m'} |C_{1:n}(\boldsymbol{u}) - C(\boldsymbol{u})|$$

$$= O(1) \times \sup_{s \in [0, T+1]} \sup_{\boldsymbol{u} \in [0,1]^d} |\mathbb{B}_m(s, \boldsymbol{u})| = O_\mathbb{P}(1).$$

On the other hand, we have

$$K_m = O(1) \times m^{-1} \max_{1 \leq k \leq m} \Big| \sum_{i=1}^{k} \xi_{i,m}^{[b]} \Big|.$$

Let $Z_{k,m} = \sum_{i=1}^{k} \xi_{i,m}^{[b]}$, $k \in \{1, \ldots, m\}$, and let $\nu > 1$. Then,

$$\mathbb{E}\left\{ \left(m^{-1} \max_{1 \leq k \leq m} |Z_{k,m}| \right)^{\nu} \right\} = m^{-\nu} \mathbb{E}\left(\max_{1 \leq k \leq m} |Z_{k,m}|^{\nu} \right)$$
$$\leq m^{-\nu} \mathbb{E}\left(\sum_{k=1}^{m} |Z_{k,m}|^{\nu} \right) \leq m^{1-\nu} \max_{1 \leq k \leq m} \mathbb{E}\left(|Z_{k,m}|^{\nu} \right).$$

Proceeding as in the proof of Lemma D.2 in the supplementary material of Bücher and Kojadinovic (2016), for $\nu = 4$, using the fact that the sequence $(\xi_{i,m}^{[b]})_{i \in \mathbb{Z}}$ is stationary and ℓ_m-dependent,

$$\mathbb{E}\left(|Z_{k,m}|^4 \right) = \mathbb{E}\left(Z_{k,m}^4 \right) = \sum_{i_1,i_2,i_3,i_4=1}^{k} \mathbb{E}\left(\xi_{i_1,m}^{[b]} \xi_{i_2,m}^{[b]} \xi_{i_3,m}^{[b]} \xi_{i_4,m}^{[b]} \right)$$
$$\leq 4! \sum_{1 \leq i_1 \leq i_2 \leq i_3 \leq i_4 \leq m} \left| \mathbb{E}\left(\xi_{i_1,m}^{[b]} \xi_{i_2,m}^{[b]} \xi_{i_3,m}^{[b]} \xi_{i_4,m}^{[b]} \right) \right|$$
$$\leq 4! \, m \sum_{1 \leq i_2 \leq i_3 \leq i_4 \leq m} \left| \mathbb{E}\left(\xi_{1,m}^{[b]} \xi_{i_2,m}^{[b]} \xi_{i_3,m}^{[b]} \xi_{i4,m}^{[b]} \right) \right|$$
$$\leq 4! \, m \sum_{i_2=1}^{\ell_m} \sum_{i_3=i_2}^{m} \sum_{i_4=i_3}^{\ell_m} \left| \mathbb{E}\left(\xi_{1,m}^{[b]} \xi_{i_2,m}^{[b]} \xi_{i_3,m}^{[b]} \xi_{i_4,m}^{[b]} \right) \right|.$$

From Cauchy-Swcharz's inequality, there holds $|\mathbb{E}(\xi_{1,m}^{[b]} \xi_{i_2,m}^{[b]} \xi_{i_3,m}^{[b]} \xi_{i_4,m}^{[b]})| \leq \mathbb{E}\{(\xi_{1,m}^{[b]})^4\}$, and therefore $\max_{1 \leq k \leq m} \mathbb{E}\left(|Z_{k,m}|^4 \right) = O(m^2 \ell_m^2)$ since $\sup_{m \geq 1} \mathbb{E}\{(\xi_{1,m}^{[b]})^4\} < \infty$. It follows that

$$\mathbb{E}\left\{ \left(m^{-1} \max_{1 \leq k \leq m} |Z_{k,m}| \right)^4 \right\} = O(m^{-2\varepsilon}) \to 0,$$

which implies that $K_m \xrightarrow{\mathbb{P}} 0$ and therefore that (D.2) holds. The proof of the first claim is thus complete. The remaining claims are essentially consequences of the continuous mapping theorem and can be shown by proceeding as in the proof of Proposition 2.5. $\square$

Acknowledgments

References

ANATOLYEV, S. and KOSENOK, G. (2018). Sequential Testing with Uniformly Distributed Size. *Journal of Time Series Econometrics* **10** 1-35.

AUE, A., HORVÁTH, L., KÜHN, M. and STEINEBACH, J. (2012a). On the reaction time of moving sum detectors. *Journal of Statistical Planning and Inference* **142** 2271 - 2288.

AUE, A., HÖRMANN, S., HORVÁTH, L., HUŠKOVÁ, M. and STEINEBACH, J. G. (2012b). Sequential testing for the stability of high-frequency portfolio betas. *Econometric Theory* **28** 804–837.

BERBEE, H. (1987). Convergence rates in the strong law for bounded mixing sequences. *Probability Theory and Related Fields* **74** 255–270.

BÜCHER, A. (2015). A note on weak convergence of the sequential multivariate empirical process under strong mixing. *Journal of Theoretical Probability* **28** 1028–1037.

BÜCHER, A., FERMANIAN, J. D. and KOJADINOVIC, I. (2019). Combining cumulative sum change-point detection tests for assessing the stationarity of continuous univariate time series. *Journal of Time Series Analysis* **40** 124–150.

BÜCHER, A. and KOJADINOVIC, I. (2016). A dependent multiplier bootstrap for the sequential empirical copula process under strong mixing. *Bernoulli* **22** 927–968.

BÜCHER, A. and KOJADINOVIC, I. (2019). A note on conditional versus joint unconditional weak convergence in bootstrap consistency results. *Journal of Theoretical Probability* **32** 1145–1165.

BÜHLMANN, P. (1993). The blockwise bootstrap in time series and empirical processes, PhD thesis, ETH Zürich Diss. ETH No. 10354.

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.

CSÖRGŐ, M., HORVÁTH, L. and SZYSZKOWICZ, B. (1997). Integral tests for suprema of Kiefer processes with application. *Statistics and Risk Modeling* **15** 365–378.

CSÖRGŐ, M. and SZYSZKOWICZ, B. (1994a). *Applications of multi-time parameter processes to change-point analysis* In *Probability Theory and Mathematical Statistics: Proceedings of the Sixth Vilnius Conference* 159-222. VSP/TEV.

CSÖRGŐ, M. and SZYSZKOWICZ, B. (1994b). *Weighted multivariate empirical processes and contiguous change-point analysis.* In *Change-point problems. Lecture Notes–Monograph Series* **23** 93–98. Institute of Mathematical Statistics.

DAVYDOV, Y. A. and LIFSHITS, M. A. (1984). The fibering method in some probability problems. In *Probability theory. Mathematical statistics. Theoretical cybernetics, Vol. 22. Itogi Nauki i Tekhniki* 61–157, 204. Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Inform., Moscow. MR778385

DEREICH, S., FEHRINGER, F., MATOUSSI, A. and SCHEUTZOW, M. (2003). On the link between small ball probabilities and the quantization problem for Gaussian measures on Banach spaces. *J. Theoret. Probab.* **16** 249–265. MR1956830

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.

EMBRECHTS, P. and HOFERT, M. (2013). A note on generalized inverses. *Mathematical Methods of Operations Research* **77** 423–432.

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

GENEST, C. and RÉMILLARD, B. (2004). Tests of independence and randomness based

on the empirical copula process. *Test* **13** 335–369.

GOMBAY, E. and HORVÁTH, L. (1999). Change-points and bootstrap. *Environmetrics* **10** 725–736.

HAWKINS, D. M., QIU, P. and KANG, C. W. (2003). The Changepoint Model for Statistical Process Control. *Journal of Quality Technology* **35** 355-366.

HAWKINS, D. M. and ZAMBA, K. D. (2005). Statistical Process Control for Shifts in Mean or Variance using a Change Point Formulation. *Technometrics* **47** 164–173.

HOFERT, M., KOJADINOVIC, I., MÄCHLER, M. and YAN, J. (2018). copula: Multivariate dependence with copulas R package version 0.999-19.1.

HOLMES, M., KOJADINOVIC, I. and QUESSY, J.-F. (2013). Nonparametric tests for change-point detection à la Gombay and Horváth. *Journal of Multivariate Analysis* **115** 16–32.

HORVÁTH, L., HUŠKOVÁ, M., KOKOSZKA, P. and STEINEBACH, J. (2004). Monitoring changes in linear models. *Journal of Statistical Planning and Inference* **126** 225 - 251.

JONDEAU, E., POON, S. H. and ROCKINGER, M. (2007). *Financial modeling under non-Gaussian distributions.* Springer, London.

KIRCH, C. and STOEHR, C. (2019). Sequential change point tests based on U-statistics.

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., SEGERS, J. and YAN, J. (2011). Large-sample tests of extreme-value dependence for multivariate copulas. *The Canadian Journal of Statistics* **39** 703–720.

KOSOROK, M. R. (2008). *Introduction to empirical processes and semiparametric inference.* Springer, New York.

KÜNSCH, H. R. (1989). The jacknife and the bootstrap for general stationary observations. *The Annals of Statistics* **17** 1217–1241.

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

MARGAVIO, T. M., CONERLY, M. D., WOODALL, W. H. and DRAKE, L. G. (1995). Alarm rates for quality control charts. *Statistics & Probability Letters* **24** 219–224.

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

RIO, E. (2017). *Asymptotic theory of weakly dependent random processes.* Springer-Verlag Berlin Heidelberg.

ROSS, G. J. (2014). Sequential change detection in the presence of unknown parameters. *Statistics and Computing* **24** 1017–1030.

ROSS, G. J. (2015). Parametric and Nonparametric Sequential Change Detection in R: The cpm Package. *Journal of Statistical Software* **66** 1–20.

ROSS, G. J. and ADAMS, N. M. (2012). Two Nonparametric Control Charts for Detecting Arbitrary Distribution Changes. *Journal of Quality Technology* **44** 102-116.

ROSS, G. J., TASOULIS, D. K. and ADAMS, N. M. (2011). Nonparametric Monitoring of Data Streams for Changes in Location and Scale. *Technometrics* **53** 379-389.

SKLAR, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications*

de l'Institut de Statistique de l'Université de Paris **8** 229–231.

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

TRAPLETTI, A. and HORNIK, K. (2019). tseries: Time series analysis and computational finance R package version 0.10-47.

VAN DER VAART, A. W. (1998). *Asymptotic statistics*. Cambridge University Press.

VAN DER VAART, A. W. and WELLNER, J. A. (2000). *Weak convergence and empirical processes*. Springer, New York. Second edition.

TABLE 5.2

Percentages of rejection of H_0 in (1.1) for the procedure based on $T_{m,q}$ with $\gamma \in \{0, 0.25, 0.5\}$ and $p \in \{1, 2, 4\}$ when based on the dependent multtliplier bootstrap with $B = 2000$. The rejection percentages are computed from 1000 samples of size $\lfloor m(T+1) \rfloor$ generated from AR(1) models with autoregressive parameter $\beta \in \{0, 0.1, 0.3, 0.5\}$.

β	γ	m	$T=0.5$ $(n=1.5m)$			$T=1$ $(n=2m)$			$T=2$ $(n=3m)$			$T=3$ $(n=4m)$		
			$p{=}1$	$p{=}2$	$p{=}4$	$p{=}1$	$p{=}2$	$p{=}4$	$p{=}1$	$p{=}2$	$p{=}4$	$p{=}1$	$p{=}2$	$p{=}4$
0.0	0.00	50	4.4	6.0	5.4	6.6	5.9	5.8	3.6	3.7	2.9	5.2	4.9	4.4
		100	5.6	5.2	4.9	7.3	7.6	6.4	6.2	6.1	6.1	6.2	5.8	7.2
		200	6.1	5.2	5.5	5.1	5.1	6.1	5.1	5.3	6.2	6.9	6.2	5.2
		400	5.5	5.9	5.3	7.5	6.9	7.1						
	0.25	50	4.1	5.0	4.6	5.5	4.9	4.9	3.4	4.1	3.0	6.1	5.1	4.2
		100	5.0	4.6	4.3	6.5	6.2	6.1	6.2	5.6	5.8	6.2	5.8	6.7
		200	5.5	5.5	5.0	5.0	4.3	5.4	5.0	5.3	5.9	6.7	5.9	5.3
		400	4.8	5.9	4.9	7.3	6.7	6.2						
	0.50	50	2.8	3.1	2.7	3.6	2.5	2.9	2.9	2.5	3.0	5.1	3.2	3.3
		100	3.7	2.7	3.2	4.6	4.3	4.6	5.3	4.2	4.3	5.1	4.9	4.9
		200	3.9	3.0	3.1	3.6	2.6	3.4	4.2	4.7	4.6	5.9	4.8	3.6
		400	3.7	3.6	3.2	6.2	5.7	4.6						
0.1	0.00	50	4.3	4.8	6.7	7.3	8.1	9.2	4.3	4.9	3.8	4.6	5.6	4.3
		100	5.4	6.2	7.4	6.5	7.8	7.6	6.2	5.8	6.1	7.9	6.1	7.5
		200	6.5	7.0	7.5	6.2	6.9	7.9	6.3	6.7	6.4	7.4	8.4	8.6
		400	6.1	5.7	7.1	5.3	5.7	6.4						
	0.25	50	3.6	4.0	4.7	7.2	7.1	7.2	4.4	5.4	4.3	5.0	5.9	4.4
		100	5.4	5.6	6.1	5.5	6.9	6.5	6.0	5.2	4.9	7.7	5.8	7.0
		200	5.9	6.0	6.5	5.7	7.2	7.2	6.4	6.9	6.6	7.5	8.2	8.6
		400	5.8	6.0	6.6	5.6	5.0	6.3						
	0.50	50	2.7	2.5	2.9	5.0	5.0	4.9	4.0	4.4	4.5	4.1	4.6	3.5
		100	3.6	3.4	3.8	4.9	4.9	4.6	4.9	3.9	3.5	6.6	4.7	5.5
		200	3.7	3.7	3.9	5.3	5.2	5.1	5.8	5.8	5.8	7.0	6.9	6.5
		400	4.8	4.7	4.2	4.5	4.0	4.7						
0.3	0.00	50	9.3	9.3	10.0	10.8	11.7	12.9	6.4	6.5	6.5	5.9	6.9	6.2
		100	8.4	10.1	10.2	8.3	9.5	11.1	7.4	7.6	8.3	7.7	8.8	10.2
		200	7.4	7.7	8.5	7.0	8.0	10.1	7.2	6.8	8.9	7.1	8.6	8.9
		400	5.7	7.7	8.4	7.0	8.1	8.5						
	0.25	50	8.9	8.4	9.0	10.2	11.0	12.0	7.1	7.2	7.1	6.3	7.1	7.8
		100	8.3	9.3	9.8	8.3	9.5	10.3	7.5	7.7	8.1	7.8	9.2	10.3
		200	7.1	6.7	7.6	7.3	8.0	10.2	7.3	6.6	9.0	7.2	8.8	9.6
		400	5.9	6.8	7.6	7.1	7.9	7.8						
	0.50	50	5.7	6.0	6.5	9.3	9.1	10.2	7.4	6.5	6.6	6.5	7.6	8.0
		100	6.1	6.4	8.4	7.7	6.2	7.7	6.7	7.5	7.2	6.6	7.8	8.1
		200	5.0	4.9	5.1	6.3	6.4	8.1	6.4	5.3	6.9	6.0	6.2	7.9
		400	4.2	5.3	4.6	5.9	5.9	5.5						
0.5	0.00	50	9.6	11.7	13.5	12.8	13.2	17.3	9.8	9.8	9.1	8.5	11.7	11.7
		100	9.6	12.2	13.4	11.5	14.0	14.7	9.7	9.8	13.4	11.4	12.3	14.0
		200	8.6	10.0	12.4	7.5	8.7	10.9	8.2	10.5	11.1	8.4	9.1	10.5
		400	7.5	8.9	11.3	5.7	7.5	9.4						
	0.25	50	9.6	11.5	12.4	12.5	13.3	16.6	10.3	10.8	10.7	9.6	12.7	13.1
		100	9.2	11.0	12.2	11.3	13.4	14.0	10.2	10.2	13.4	11.5	12.3	14.3
		200	7.9	9.7	11.1	6.8	7.8	10.1	8.7	10.4	11.5	8.4	9.4	10.8
		400	6.7	8.0	10.6	5.6	7.7	8.8						
	0.50	50	7.7	9.6	10.2	10.9	10.9	15.1	9.7	10.3	10.5	9.7	12.3	13.5
		100	6.8	7.6	9.8	9.5	9.8	11.4	8.2	9.7	12.4	10.2	11.1	12.0
		200	5.8	6.8	8.7	6.1	5.8	7.1	7.5	8.5	9.7	7.6	8.6	9.7
		400	5.2	6.2	7.4	4.8	5.3	6.5						

TABLE 5.3

Percentages of rejection of H_0 in (1.1) for the procedure based on $T_{m,q}$ with $\gamma \in \{0, 0.25, 0.5\}$ and $p \in \{1, 2, 4\}$ when based on the dependent multtliplier bootstrap with $B = 2000$. The rejection percentages are computed from 1000 samples of size $\lfloor m(T + 1) \rfloor$ generated from a GARCH(1, 1) model with parameters $\omega = 0.012$, $\beta = 0.919$ and $\alpha = 0.072$ to mimick SP500 daily logreturns following Jondeau, Poon and Rockinger (2007).

γ	m	$T=0.5$ $(n=1.5m)$			$T=1$ $(n=2m)$			$T=2$ $(n=3m)$			$T=3$ $(n=4m)$		
		$p{=}1$	$p{=}2$	$p{=}4$	$p{=}1$	$p{=}2$	$p{=}4$	$p{=}1$	$p{=}2$	$p{=}4$	$p{=}1$	$p{=}2$	$p{=}4$
0.00	50	5.4	5.4	4.9	9.1	7.6	7.6	4.8	4.7	4.0	5.4	6.3	5.3
	100	6.6	7.5	7.6	7.9	8.6	7.2	9.6	8.4	9.2	10.5	10.3	10.0
	200	7.4	7.2	8.6	8.3	8.4	8.5	6.6	7.7	6.1	8.7	8.0	8.2
	400	6.7	7.5	8.4	7.0	7.0	8.4						
0.25	50	4.7	4.2	4.3	7.5	5.9	5.6	4.8	4.5	4.2	5.8	6.4	5.9
	100	5.9	6.3	6.3	7.2	6.9	6.7	9.6	8.8	8.4	10.2	9.8	9.2
	200	7.0	6.9	7.8	7.2	7.3	8.2	6.4	6.9	5.9	8.3	7.6	8.4
	400	6.2	7.9	7.9	6.9	6.7	8.4						
0.50	50	2.4	2.5	2.6	4.4	3.9	4.2	4.0	3.9	3.5	5.0	4.7	4.6
	100	4.1	4.2	4.5	5.0	4.1	3.9	6.6	5.9	6.0	7.9	7.5	7.0
	200	4.9	5.0	5.2	5.7	4.8	5.7	5.8	5.1	4.9	6.8	6.6	7.1
	400	5.1	5.4	5.1	6.3	5.9	5.9						

TABLE 5.4

Percentages of rejection of H_0 in (1.1) for the procedure based on $T_{m,q}$ with $\gamma \in \{0, 0.25, 0.5\}$ and $p \in \{1, 2, 4\}$ when based on the dependent multtliplier bootstrap with $B = 2000$. The rejection percentages are computed from 1000 samples of size $2m$ generated from a normal copula with a Kendall's tau of $\tau \in \{-0.6, -0.3, 0, 0.3, 0.6\}$.

τ	m	$\gamma = 0$			$\gamma = 0.25$			$\gamma = 0.5$		
		$p{=}1$	$p{=}2$	$p{=}4$	$p{=}1$	$p{=}2$	$p{=}4$	$p{=}1$	$p{=}2$	$p{=}4$
-0.6	50	3.1	1.9	1.5	2.4	1.1	1.1	0.7	0.3	0.8
	100	3.5	2.9	2.2	2.5	2.1	1.7	1.2	1.0	0.9
	200	4.0	3.2	3.1	3.0	2.4	2.1	0.9	1.0	1.0
-0.3	50	4.5	2.9	2.7	3.4	2.5	2.5	1.8	1.5	1.2
	100	4.0	4.0	3.8	3.1	3.0	2.9	1.5	0.9	1.4
	200	4.8	4.1	4.0	4.1	3.3	2.9	1.6	1.6	1.6
0.0	50	5.1	4.3	4.7	4.2	3.9	3.6	2.7	2.0	2.3
	100	7.0	5.8	4.7	5.7	4.7	3.7	2.6	2.2	2.0
	200	6.0	6.2	5.8	5.3	5.0	5.2	3.3	2.4	2.3
0.3	50	6.1	4.5	4.5	4.7	3.8	3.3	2.5	2.4	2.6
	100	5.0	5.5	4.9	4.4	4.8	3.9	2.3	2.5	2.5
	200	6.2	6.3	7.8	6.0	4.8	6.3	3.4	2.6	3.8
0.6	50	6.7	5.5	5.8	5.8	3.9	4.2	3.2	2.6	2.5
	100	7.7	7.6	7.8	7.0	6.6	7.0	4.3	4.0	4.8
	200	5.4	6.2	7.2	5.1	5.6	6.1	3.4	3.5	3.2