Multi-purpose open-end monitoring procedures for multivariate observations based on the empirical distribution function

Mark Holmes[1], **Ivan Kojadinovic**[2] **and Alex Verhoijsen**[1,2]

[1]*School of Mathematics & Statistics*
The University of Melbourne
Parkville, VIC 3010, Australia

[2]*CNRS / Université de Pau et des Pays de l'Adour / E2S UPPA*
Laboratoire de mathématiques et applications IPRA, UMR 5142
B.P. 1155, 64013 Pau Cedex, France

Abstract: We propose nonparametric open-end sequential testing procedures that can detect all types of changes in the contemporary distribution function of multivariate observations. Their asymptotic properties are theoretically investigated under stationarity and under alternatives to stationarity. Monte Carlo experiments reveal their good finite-sample behavior in the case of continuous univariate observations. A short data example concludes the work.

MSC 2010 subject classifications: Primary 62L99, 62E20; secondary 62G10.
Keywords and phrases: asymptotic results, change-point detection, open-end monitoring, sequential testing, theoretical quantile estimation.

1. Introduction

From an historical perspective, monitoring is often associated with *control charts* also known as *Shewart charts*. Such graphical tools central to *statistical process control* (see, e.g., Lai, 2001; Montgomery, 2007, for an overview) are usually calibrated in terms of the so-called *average run length* (ARL) controlling how many monitoring steps are necessary on average before the data generating process is declared *out of control*. The fact that the conclusion that the probabilistic properties of the monitored observations have changed is reached with probability one could be regarded as the major drawback of this type of procedure. To remedy this situation, Chu, Stinchcombe and White (1996) have proposed to treat the issue of monitoring from the point of view of statistical testing. The main advantage is that, when observations arise from a stationary time series, monitoring procedures *à la* Chu, Stinchcombe and White (1996) will lead to the conclusion that a change has occurred in the data generating process only with a small probability α controlled by the user.

In addition to being of the latter type, the monitoring procedures investigated in this work are nonparametric and deal with d-dimensional observations, $d \geq 1$. In that respect, as we continue, we use the superscript $^{[\ell]}$ to denote the ℓth coordinate of a vector (for instance, $\boldsymbol{x} = (x^{[1]}, \dots, x^{[d]}) \in \mathbb{R}^d$). As is customary in the literature, we assume that we have at hand $m \geq 1$ observations from the initial data generating process. Monitoring starts immediately thereafter. To be more precise, we assume that we have at our disposal a stretch $\boldsymbol{X}_i = (X_i^{[1]}, \dots, X_i^{[d]})$, $i \in \{1, \dots, m\}$, from a d-dimensional stationary time series with unknown contemporary distribution function (d.f.) F given by $F(\boldsymbol{x}) = \mathbb{P}(X_1^{[1]} \le x^{[1]}, \dots, X_1^{[d]} \le x^{[d]}) = \mathbb{P}(\boldsymbol{X}_1 \le \boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^d$. These available observations will be referred to as the *learning sample* as we continue. Once the monitoring starts, new observations $\boldsymbol{X}_{m+1}, \boldsymbol{X}_{m+2}, \dots$ arrive sequentially and the aim is to issue an alarm as soon as possible if there is evidence that the contemporary distribution of the most recent observations is no longer equal to F.

Most approaches in the literature are of a *closed-end* nature: the monitoring eventually stops if stationarity is not rejected after the arrival of observation $\boldsymbol{X}_n$, $n > m$. Our focus in this work is on the more difficult scenario in which the monitoring can in principle continue indefinitely: this is called the *open-end* setting in this literature. From a practical perspective, it can be argued that the fact that the *monitoring horizon* n does not need to be specified is a great advantage of open-end procedures. As discussed for instance in Remark 2.2 of Gösmann, Kley and Dette (2021), the price to pay for open-endness is a significantly more complicated theoretical setting.

The null hypothesis of the sequential testing procedures studied in this work is thus

$$H_0 : \boldsymbol{X}_1, \dots, \boldsymbol{X}_m, \boldsymbol{X}_{m+1}, \boldsymbol{X}_{m+2}, \dots, \text{ is a stretch from a stationary time series.} \tag{1.1}$$

When $d = 1$, starting from the work of Gösmann, Kley and Dette (2021), Holmes and Kojadinovic (2021) have recently introduced a detection procedure that is particularly sensitive to changes in the mean. Because it uses the *retrospective cumulative sum (CUSUM) statistic* as *detector*, it turns out to be superior to existing procedures as long as changes do not occur at the very beginning of the monitoring. As noted in Section 6 of the latter reference, this approach can be adapted to obtain alternative procedures that are particularly sensitive for instance to changes in the variance or some other moments. The goal of this work is to generalize the method of Holmes and Kojadinovic (2021) in order to obtain open-end monitoring procedures that can be sensitive simultaneously to all types of changes in the d.f. F. Although such procedures already exist in a closed-end setting (see, e.g., Kojadinovic and Verdier, 2021), to the best of our knowledge, they are unavailable in the open-end setting.

This paper is organized as follows. In the second section, starting from the work of Holmes and Kojadinovic (2021), we propose a detector that can be sensitive to all types of changes in the contemporary d.f. of multivariate observations. We additionally introduce a suitable threshold function and study the asymptotics of the resulting monitoring procedure under H_0 in (1.1) and under sequences of alternatives to H_0. In the third section, we focus on the case of continuous observations and provide additional asymptotic results under the null. In the fourth section, using *asymptotic regression models*, we address the estimation of high quantiles of the distributions appearing in the asymptotic results under H_0 which are necessary in practice to carry out the sequential tests. The fifth section reports the results of numerous Monte Carlo experiments restricted to the case of continuous univariate observations whose aim is to study the behavior of the monitoring procedures under H_0 and under alternatives to H_0. A data example and concluding remarks are gathered in the last section.

Unless mentioned otherwise, all convergences are as $m \to \infty$. Also, as there are many discrete intervals appearing in this work, we will conveniently use the notation $[\![j, k]\!]$, $[\![j, k[\![$,

$\rrbracket j, k\rrbracket$, and $\rrbracket j, k\llbracket$ for the sets of integers $\{j, \dots, k\}$, $\{j, \dots, k-1\}$, $\{j-1, \dots, k\}$, and $\{j-1, \dots, k-1\}$, respectively.

2. A first detector, the threshold function and related asymptotics

2.1. A first detector and the threshold function

One of the two main ingredients of a monitoring procedure *à la* Chu, Stinchcombe and White (1996) is a statistic, also called a *detector*, that is potentially computed after the arrival of every new observation $\boldsymbol{X}_k$, $k > m$. This statistic is typically positive and quantifies some type of departure from stationary. Once computed, it is compared to a positive threshold, possibly also depending on k. If greater, evidence against H_0 in (1.1) is deemed significant and the monitoring stops. Otherwise, a new observation is collected.

For sensitivity to changes in the mean of univariate ($d = 1$) observations, Holmes and Kojadinovic (2021) used among others the detector

$$R_m(k) = \max_{j \in \llbracket m,k \llbracket} \frac{j(k-j)}{m^{\frac{3}{2}}} |\bar{X}^{[1]}_{1:j} - \bar{X}^{[1]}_{j+1:k}|, \qquad k \geq m+1, \tag{2.1}$$

where $\bar{X}^{[1]}_{j:k} = \frac{1}{k-j+1} \sum_{i=j}^{k} X^{[1]}_j$, $1 \leq j \leq k$. Some thought reveals that, for any fixed $k \geq m+1$, $R_m(k)$ is akin to the so-called *retrospective CUSUM statistic* frequently used in offline change-point detection tests (see, e.g., Csörgő and Horváth, 1997; Aue and Horváth, 2013).

When $d \geq 1$, to be sensitive to changes in the d.f. at a fixed point $\boldsymbol{x} = (x^{[1]}, \dots, x^{[d]}) \in \mathbb{R}^d$, a straightforward adaptation of the previous approach would be to compute (2.1) from the stretch of univariate observations $\mathbf{1}(\boldsymbol{X}_1 \leq \boldsymbol{x}), \dots, \mathbf{1}(\boldsymbol{X}_m \leq \boldsymbol{x}), \mathbf{1}(\boldsymbol{X}_{m+1} \leq \boldsymbol{x}), \dots, \mathbf{1}(\boldsymbol{X}_k \leq \boldsymbol{x}), \dots$, where inequalities between vectors are to be understood componentwise. The detector R_m can then be equivalently expressed as

$$E^{\boldsymbol{x}}_m(k) = \max_{j \in \llbracket m,k \llbracket} \frac{j(k-j)}{m^{\frac{3}{2}}} |F_{1:j}(\boldsymbol{x}) - F_{j+1:k}(\boldsymbol{x})|, \qquad k \geq m+1, \tag{2.2}$$

where, for any integers $j, k \geq 1$,

$$F_{j:k}(\boldsymbol{x}) = \begin{cases} \dfrac{1}{k-j+1} \displaystyle\sum_{i=j}^{k} \mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{x}), & \text{if } j \leq k, \\ 0, & \text{otherwise}, \end{cases} \tag{2.3}$$

is the empirical d.f. of $\boldsymbol{X}_j, \dots, \boldsymbol{X}_k$ evaluated at $\boldsymbol{x}$. Our aim is to extend the previous approach using $p \geq 1$ points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_p$ in $\mathbb{R}^d$, where the integer p and the points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_p$ are chosen by the user. Let $\mathcal{P} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_p)$ and, for any $i \in \mathbb{N}$, let $\boldsymbol{Y}^{\mathcal{P}}_i = \big(\mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{x}_1), \dots, \mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{x}_p)\big)$, which is a p-dimensional random vector. Combining the approach of Gösmann, Kley and Dette (2021) with the one of Holmes and Kojadinovic (2021), the first detector considered in this work is defined by

$$D^{\mathcal{P}}_m(k) = \max_{j \in \llbracket m,k \llbracket} \frac{j(k-j)}{m^{\frac{3}{2}}} \|\boldsymbol{F}^{\mathcal{P}}_{1:j} - \boldsymbol{F}^{\mathcal{P}}_{j+1:k}\|_{(\Sigma^{\mathcal{P}}_m)^{-1}}, \qquad k \geq m+1, \tag{2.4}$$

where, for any integers $j, k \geq 1$, $\boldsymbol{F}_{j:k}^{\mathcal{P}} = \big(F_{j:k}(\boldsymbol{x}_1), \dots, F_{j:k}(\boldsymbol{x}_p)\big) \in \mathbb{R}^p$, $\Sigma_m^{\mathcal{P}}$ is an estimator (based on $\boldsymbol{Y}_1^{\mathcal{P}}, \dots, \boldsymbol{Y}_m^{\mathcal{P}}$) of the long-run $p \times p$ covariance matrix

$$\Sigma^{\mathcal{P}} = \mathrm{Cov}(\boldsymbol{Y}_1^{\mathcal{P}}, \boldsymbol{Y}_1^{\mathcal{P}}) + 2 \sum_{i=2}^{\infty} \mathrm{Cov}(\boldsymbol{Y}_1^{\mathcal{P}}, \boldsymbol{Y}_i^{\mathcal{P}}) \tag{2.5}$$

of the p-dimensional time series $\big(\boldsymbol{Y}_i^{\mathcal{P}}\big)_{i \in \mathbb{N}}$ and, for any $\boldsymbol{y} \in \mathbb{R}^p$, $\|\boldsymbol{y}\|_M = \sqrt{(\boldsymbol{y}^\top M \boldsymbol{y})/p}$ denotes a weighted norm of $\boldsymbol{y}$ induced by a $p \times p$ positive-definite matrix M and the integer p. Note that (2.4) can be equivalently rewritten as

$$D_m^{\mathcal{P}}(k) = \max_{j \in [\![m,k [\![} \frac{j(k-j)}{m^{\frac{3}{2}}} \|\bar{\boldsymbol{Y}}_{1:j}^{\mathcal{P}} - \bar{\boldsymbol{Y}}_{j+1:k}^{\mathcal{P}}\|_{(\Sigma_m^{\mathcal{P}})^{-1}}, \qquad k \geq m+1, \tag{2.6}$$

where, for any integers $1 \leq j \leq k$, $\bar{\boldsymbol{Y}}_{j:k}^{\mathcal{P}} = \frac{1}{k-j+1} \sum_{i=j}^{k} \boldsymbol{Y}_i^{\mathcal{P}}$.

The second key ingredient of a monitoring procedure is a *threshold function*. In the considered open-end setting, given a significance level $\alpha \in (0, \frac{1}{2})$, the aim is to define a deterministic function $w : [1, \infty) \to (0, \infty)$ such that (ideally) under H_0 in (1.1),

$$\mathbb{P}\Big(D_m^{\mathcal{P}}(k) \leq w(k/m) \text{ for all } k > m\Big) = \mathbb{P}\left(\sup_{k>m} \frac{D_m^{\mathcal{P}}(k)}{w(k/m)} \leq 1\right) = 1 - \alpha, \tag{2.7}$$

where the supremum is over integers $k > m$. Because the detector $D_m^{\mathcal{P}}$ in (2.4) or in (2.6) can be regarded as a multivariate generalization of the detector R_m in (2.1), one can use the same reasoning as in Section 2 of Holmes and Kojadinovic (2021) to suggest that a meaningful threshold function in the considered case is

$$w(t) = q_{p,\eta}^{(1-\alpha)} t^{\frac{3}{2}+\eta}, \qquad t \in [1, \infty), \tag{2.8}$$

where η is a positive real parameter and $q_{p,\eta}^{(1-\alpha)} > 0$ is the $(1-\alpha)$-quantile of $\mathcal{L}_{p,\eta}$, the weak limit of $\sup_{k>m}(m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k)$ under H_0, assuming that $\mathcal{L}_{p,\eta}$ is continuous. In that case, under H_0, by the Portmanteau theorem,

$$\begin{aligned} \lim_{m\to\infty} \mathbb{P}\left(\sup_{k>m} \frac{D_m^{\mathcal{P}}(k)}{w(k/m)} \leq 1\right) &= \lim_{m\to\infty} \mathbb{P}\left(\sup_{k>m} (m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k) \leq q_{p,\eta}^{(1-\alpha)}\right) \\ &= \mathbb{P}(\mathcal{L}_{p,\eta} \leq q_{p,\eta}^{(1-\alpha)}) = 1 - \alpha, \end{aligned} \tag{2.9}$$

which can be regarded as an "asymptotic version" of (2.7). As far as η is concerned, as a consequence of Proposition 2.9 below, it needs to be chosen strictly positive so that $\mathcal{L}_{p,\eta}$ is almost surely finite. In practice, we follow the recommendation made in Holmes and Kojadinovic (2021) and set η to 0.001.

Remark 2.1. Proceeding for instance along the lines of Horváth et al. (2004), Fremdt (2015), Kirch and Weber (2018), Gösmann, Kley and Dette (2021) or Holmes and Kojadinovic (2021), instead of w in (2.8), one could alternatively consider as threshold function $\tilde{w}$ defined by $\tilde{w}(t) = \bar{w}_\gamma(t) w(t)$, $t \in [1, \infty)$, where

$$\bar{w}_\gamma(t) = \max\left\{\left(\frac{t-1}{t}\right)^{\gamma}, \epsilon\right\}, \qquad t \in [1, \infty),$$

with $\gamma \geq 0$ a real parameter and $\epsilon > 0$ a technical constant that can be taken very small in practice. The multiplication of a candidate threshold function by $\bar{w}_\gamma$ was initially considered in Horváth et al. (2004) and Aue and Horváth (2004) for the so-called *ordinary CUSUM* detector in order to study, under suitable alternatives, the limiting distribution of the detection delay (the time delay after which the detector exceeds the threshold function). From a practical perspective, as discussed in Holmes and Kojadinovic (2021), an appropriate choice of $\gamma \geq 0$ may improve the finite-sample performance of the sequential test at the beginning of the monitoring. The multiplication of a candidate threshold function by $\bar{w}_\gamma$ does not however affect the asymptotics of the underlying monitoring procedure.

2.2. Asymptotics under the null

One of the first assumptions required to be able to study the asymptotics (as $m \to \infty$, of the monitoring procedure based on $D_m^{\mathcal{P}}$ in (2.4) and w in (2.8)) concerns the long-run covariance matrix $\Sigma^{\mathcal{P}}$ in (2.5) of the p-dimensional time series $\left(\boldsymbol{Y}_i^{\mathcal{P}}\right)_{i\in\mathbb{N}}$, under the null. As we shall see later in this section, it will be necessary to consider both its inverse $(\Sigma^{\mathcal{P}})^{-1}$ and its square root $(\Sigma^{\mathcal{P}})^{\frac{1}{2}}$. The following assumption guarantees that these two matrices exist (and are unique).

Condition 2.2 (On the long-run covariance matrix $\Sigma^{\mathcal{P}}$)**.** *Under H_0 in (1.1), the long-run covariance matrix $\Sigma^{\mathcal{P}}$ in (2.5) of the p-dimensional (stationary) time series $\left(\boldsymbol{Y}_i^{\mathcal{P}}\right)_{i\in\mathbb{N}}$ exists and is positive-definite.*

Remark 2.3. Under H_0 and when the time series $(\boldsymbol{X}_i)_{i\in\mathbb{N}}$ consists of independent observations, the $p \times p$ elements of $\Sigma^{\mathcal{P}}$ in (2.5) are simply $\mathrm{Cov}\{\mathbf{1}(\boldsymbol{X}_1 \leq \boldsymbol{x}_i), \mathbf{1}(\boldsymbol{X}_1 \leq \boldsymbol{x}_j)\}$, $i, j \in [\![1, p]\!]$. Hence, in the case of serially independent observations, by definition of positive-definiteness, Condition 2.2 will hold if the points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_p$ appearing in $\mathcal{P}$ are chosen such that any linear combination of the $\mathbf{1}(\boldsymbol{X}_1 \leq \boldsymbol{x}_i)$, $i \in [\![1, p]\!]$, has a strictly positive variance. A necessary condition for this is that $\boldsymbol{x}_1, \dots, \boldsymbol{x}_p$ all belong to the support of $\boldsymbol{X}_1$ and are all distinct. Since the law of $\boldsymbol{X}_1$ is unknown, the user could in practice rely on the learning sample $\boldsymbol{X}_1, \dots, \boldsymbol{X}_m$ to choose $\boldsymbol{x}_1, \dots, \boldsymbol{x}_p$. If the learning sample seems to be a stretch from a discrete time series, a natural possibility consists of choosing $\boldsymbol{x}_1, \dots, \boldsymbol{x}_p$ from a subset of frequently occurring observations. The choice of $\mathcal{P}$ when the observations in the learning sample seem to arise from a continuous time series will be discussed in Section 3.

As shall become clearer in the forthcoming paragraphs, studying the asymptotics under the null (of the monitoring procedure as $m \to \infty$) actually amounts to establishing the weak limit of $\sup_{k>m}(m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k)$ under H_0 in (1.1). The following assumption on the time series $\left(\boldsymbol{Y}_i^{\mathcal{P}}\right)_{i\in\mathbb{N}}$ (a type of "strong approximation" condition) is typical of the kinds of assumptions in the sequential change-point literature; see, e.g., Assumption 2.3 in Gösmann, Kley and Dette (2021), Condition 3.1 in Holmes and Kojadinovic (2021) and the corresponding discussions in these references. Let $\|\cdot\|_2$ denote the Euclidean norm.

Condition 2.4 (Approximation)**.** *There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which:*

- $(\boldsymbol{Y}_i^{\mathcal{P}})_{i\in\mathbb{N}}$ *is a p-dimensional stationary time series satisfying Condition 2.2,*

- *for each $m \in \mathbb{N}$, $\boldsymbol{W}_{1,m}$ and $\boldsymbol{W}_{2,m}$ are two independent p-dimensional standard Brownian motions,*

such that, for some $0 < \xi < \frac{1}{2}$,

$$\sup_{k>m} \frac{1}{(k-m)^{\xi}} \left\| \sum_{i=m+1}^{k} \{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}} \boldsymbol{W}_{1,m}(k-m) \right\|_2 = O_{\mathbb{P}}(1) \tag{2.10}$$

and

$$\frac{1}{m^{\xi}} \left\| \sum_{i=1}^{m} \{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}} \boldsymbol{W}_{2,m}(m) \right\|_2 = O_{\mathbb{P}}(1). \tag{2.11}$$

We use the notation '$\rightsquigarrow$' to denote convergence in distribution (weak convergence) and I_p to denote the $p \times p$ identity matrix. The following result, proven in Appendix A, can be regarded as a multivariate extension of Theorem 3.3 of Holmes and Kojadinovic (2021).

Theorem 2.5. *Fix $\eta > 0$. Under Condition 2.4, if $\Sigma_m^{\mathcal{P}} \xrightarrow{\mathbb{P}} \Sigma^{\mathcal{P}}$ then*

$$\sup_{k>m} (m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k) \rightsquigarrow \mathcal{L}_{p,\eta} = \sup_{1 \le s \le t < \infty} t^{-\frac{3}{2}-\eta} \|t\boldsymbol{W}(s) - s\boldsymbol{W}(t)\|_{I_p},$$

where $D_m^{\mathcal{P}}$ is defined in (2.4) and $\boldsymbol{W}$ is a p-dimensional standard Brownian motion. In addition, the limiting random variable $\mathcal{L}_{p,\eta}$ is almost surely finite.

Note that in Theorem 2.5 the supremum on the left is over integers k while the supremum on the right is over real numbers s, t.

Remark 2.6. It is important to notice that the limiting random variable $\mathcal{L}_{p,\eta}$ depends neither on the dimension d of the observations $\boldsymbol{X}_1, \dots, \boldsymbol{X}_m, \boldsymbol{X}_{m+1}, \dots$, nor on the user-chosen points $\mathcal{P} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_p)$ involved in the definition of $D_m^{\mathcal{P}}$. It only depends on the integer p and on the real η. As shall become clearer below, the previous observation has an important practical consequence: the monitoring procedure could be used as soon as it is possible to compute or estimate quantiles of $\mathcal{L}_{p,\eta}$ for the chosen parameters p and η. This important aspect will be investigated in Section 4 in more detail.

For a given serial dependence scenario under H_0 in (1.1), it is hoped that Condition 2.4 will hold for many different vectors of points $\mathcal{P} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_p)$. The following proposition shows that this is for instance the case when the time series $(\boldsymbol{X}_i)_{i\in\mathbb{N}}$ is *strongly mixing* under H_0. Given a time series $(\boldsymbol{Z}_i)_{i\in\mathbb{N}}$ and for any $j, k \in \mathbb{N} \cup \{+\infty\}$, denote by $\mathcal{M}_j^k$ the σ-field generated by $(\boldsymbol{Z}_i)_{j\le i\le k}$ and recall that the strong mixing coefficients corresponding to $(\boldsymbol{Z}_i)_{i\in\mathbb{N}}$ are defined by

$$\alpha_r^{\boldsymbol{Z}} = \sup_{k\in\mathbb{N}} \sup_{A\in\mathcal{M}_1^k, B\in\mathcal{M}_{k+r}^{+\infty}} \left|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)\right|, \qquad r \in \mathbb{N}.$$

The sequence $(\boldsymbol{Z}_i)_{i\in\mathbb{N}}$ is then said to be *strongly mixing* if $\alpha_r^{\boldsymbol{Z}} \to 0$ as $r \to \infty$.

The following result is proven in Appendix B.

Proposition 2.7. *Assume that the time series $(\boldsymbol{X}_i)_{i\in\mathbb{N}}$ is stationary and strong mixing, and that its strong mixing coefficients satisfy $\alpha_r^{\boldsymbol{X}} = O(r^{-a})$ as $r \to \infty$ with $a > 3$. Then, Condition 2.4 holds for all vectors of points $\mathcal{P}$ such that Condition 2.2 holds.*

The previous proposition leads to the following immediate corollary of Theorem 2.5.

Corollary 2.8. *Assume that the time series $(\boldsymbol{X}_i)_{i\in\mathbb{N}}$ is stationary and strong mixing, and that its strong mixing coefficients satisfy $\alpha_r^{\boldsymbol{X}} = O(r^{-a})$ as $r \to \infty$ with $a > 3$. Then, for any fixed $\eta > 0$ and any vector of points $\mathcal{P}$ such that Condition 2.2 holds,*

$$\sup_{k>m}(m/k)^{\frac{3}{2}+\eta}D_m^{\mathcal{P}}(k) \rightsquigarrow \mathcal{L}_{p,\eta} = \sup_{1\le s\le t<\infty} t^{-\frac{3}{2}-\eta}\|t\boldsymbol{W}(s) - s\boldsymbol{W}(t)\|_{I_p}.$$

The strong mixing conditions in the previous corollary are for instance satisfied (with much to spare) when $(\boldsymbol{X}_i)_{i\in\mathbb{N}}$ is a stationary vector ARMA process with absolutely continuous innovations (see Mokkadem, 1988).

The following result, proven in Appendix B, can be regarded as a multivariate extension of Proposition 3.4 of Holmes and Kojadinovic (2021). It shows that imposing that η is strictly positive in Theorem 2.5 and Corollary 2.8 is necessary and sufficient for ensuring that the limiting random variable $\mathcal{L}_{p,\eta}$ is almost surely finite.

Proposition 2.9. *For any fixed $M > 0$,*

$$\mathbb{P}\Big(\sup_{1\le s\le t<\infty} t^{-\frac{3}{2}}\|t\boldsymbol{W}(s) - s\boldsymbol{W}(t)\|_{I_p} \ge M\Big) = 1.$$

The next proposition, also proven in Appendix B, shows that the weak limit appearing in Theorem 2.5 and Corollary 2.8 is absolutely continuous. The proof is an application of Theorem 7.1 of Davydov and Lifshits (1984) together with an argument allowing us to reduce the problem to compact sets.

Proposition 2.10. *For any $\eta > 0$ and $p \in \mathbb{N}$, $\mathcal{L}_{p,\eta}$ is an absolutely continuous random variable.*

Let us finally explain how Theorem 2.5 can be used to carry out the monitoring in practice for a chosen vector of points $\mathcal{P}$ for which Condition 2.2 is assumed to hold. Given a significance level $\alpha \in (0, \frac{1}{2})$, suppose that we are able to compute $q_{p,\eta}^{(1-\alpha)}$, the $(1-\alpha)$-quantile of $\mathcal{L}_{p,\eta}$. Then, under H_0 in (1.1) and Condition 2.4, from the Portmanteau theorem, (2.9) holds. Hence, for large m, we can expect that, under H_0 and Condition 2.4,

$$\mathbb{P}\Big(D_m^{\mathcal{P}}(k) > q_{p,\eta}^{(1-\alpha)}(k/m)^{\frac{3}{2}+\eta} \text{ for some } k \ge m+1\Big) \simeq \alpha.$$

In practice, after the arrival of observation $\boldsymbol{X}_k$, $k > m$, $D_m^{\mathcal{P}}(k)$ is computed from $\boldsymbol{X}_1, \dots, \boldsymbol{X}_k$ and compared to the threshold $q_{p,\eta}^{(1-\alpha)}(k/m)^{\frac{3}{2}+\eta}$ (or, equivalently, $(m/k)^{\frac{3}{2}+\eta}D_m^{\mathcal{P}}(k)$ is computed and compared to $q_{p,\eta}^{(1-\alpha)}$). If greater, the null hypothesis is rejected and the monitoring stops. Otherwise, $\boldsymbol{X}_{k+1}$ is collected and the previous iteration is repeated using the $k+1$ available observations.

2.3. Asymptotics under alternatives

To complement the previously stated asymptotic results, it is necessary to study the asymptotics of the monitoring procedure under sequences of alternatives to H_0 in (1.1). Because it is based on the detector $D_m^{\mathcal{P}}$ in (2.4), the studied monitoring procedure is expected to be particularly sensitive to alternative hypotheses of the form

$$H_1 : \exists\, k^\star \geq m \text{ and } \ell \in [\![1,p]\!] \text{ such that } \mathbb{P}(\boldsymbol{X}_1 \leq \boldsymbol{x}_\ell) = \cdots = \mathbb{P}(\boldsymbol{X}_{k^\star} \leq \boldsymbol{x}_\ell) \neq \mathbb{P}(\boldsymbol{X}_{k^\star+1} \leq \boldsymbol{x}_\ell) = \mathbb{P}(\boldsymbol{X}_{k^\star+2} \leq \boldsymbol{x}_\ell) = \dots$$

corresponding to a change in the d.f. at one or more of the chosen evaluation points. Note that this can be interpreted as a change in mean since by rewriting in terms of the univariate time series $\big(Y_i^{\mathcal{P},[\ell]}\big)_{i\in\mathbb{N}} = \big(\mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{x}_\ell)\big)_{i\in\mathbb{N}}$, we get the following equivalent statement

$$H_1 : \exists\, k^\star \geq m \text{ and } \ell \in [\![1,p]\!] \text{ such that } \mathbb{E}(Y_1^{\mathcal{P},[\ell]}) = \cdots = \mathbb{E}(Y_{k^\star}^{\mathcal{P},[\ell]}) \neq \mathbb{E}(Y_{k^\star+1}^{\mathcal{P},[\ell]}) = \mathbb{E}(Y_{k^\star+2}^{\mathcal{P},[\ell]}) = \dots. \tag{2.12}$$

As already mentioned at the beginning of Section 2, monitoring procedures designed to be particularly sensitive to changes in the mean were studied in Holmes and Kojadinovic (2021). Theorem 3.7 and Condition 3.7 in the latter reference specifically provide conditions under which, for a sequence of alternatives to H_0 related to H_1 in (2.12), $\sup_{k>m}(m/k)^{\frac{3}{2}+\eta}E_m^{\boldsymbol{x}_\ell}(k) \xrightarrow{\mathbb{P}} \infty$, where $E_m^{\boldsymbol{x}_\ell}$ is defined as in (2.2) with $\boldsymbol{x} = \boldsymbol{x}_\ell$. For the sake of brevity, we do not restate these conditions with the notation used in this work as they are lengthy to write. Very roughly speaking, they imply that for "early" or "late" changes in the d.f. of the observations at $\boldsymbol{x}_\ell$, the scaled detector $k \mapsto (m/k)^{\frac{3}{2}+\eta}E_m^{\boldsymbol{x}_\ell}(k)$ will end up exceeding any fixed threshold provided m is sufficiently large. The following result, proven in Appendix C, shows that, as $\boldsymbol{x}_\ell \in \mathcal{P}$, the same will hold for the scaled detector $k \mapsto (m/k)^{\frac{3}{2}+\eta}D_m^{\mathcal{P}}(k)$, where $D_m^{\mathcal{P}}$ is defined in (2.4).

Proposition 2.11. *Let $\eta > 0$ and assume that for some $\boldsymbol{x}_\ell \in \mathcal{P}$, $\sup_{k>m}(m/k)^{\frac{3}{2}+\eta}E_m^{\boldsymbol{x}_\ell}(k) \xrightarrow{\mathbb{P}} \infty$. Then, if $\Sigma_m^{\mathcal{P}} \xrightarrow{\mathbb{P}} \Sigma^{\mathcal{P}}$, where $\Sigma^{\mathcal{P}}$ is positive-definite and $\Sigma_m^{\mathcal{P}}$ is positive-definite almost surely for all $m \in \mathbb{N}$,*

$$\sup_{k>m}(m/k)^{\frac{3}{2}+\eta}D_m^{\mathcal{P}}(k) \xrightarrow{\mathbb{P}} \infty.$$

3. The case of continuous observations: practical implementation and additional asymptotic results under the null

Prior to using the monitoring procedure based on the detector $D_m^{\mathcal{P}}$ in (2.4), the user needs to choose the points $\mathcal{P} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_p)$. Following the discussion initiated in Remark 2.3, using the learning sample $\boldsymbol{X}_1, \dots, \boldsymbol{X}_m$ to do so seems meaningful. As mentioned in the latter remark, when the observations are discrete, a natural possibility consists of choosing $\boldsymbol{x}_1, \dots, \boldsymbol{x}_p$ from a subset of frequently occurring observations. We focus in this section on the more complicated situation when the learning sample seems to be a stretch from a continuous time series. After suggesting a general strategy for choosing $\mathcal{P}$ and instantiating it when $d = 1$, we study the asymptotics of the resulting monitoring procedure under the null.

3.1. On the choice of the points $\mathcal{P}$ in the case of continuous observations

Fix $\eta > 0$ and assume that m is large. Having Theorem 2.5 as well as Remark 2.6 and Corollary 2.8 in mind, one can hope that, under H_0 in (1.1), $\sup_{k>m}(m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k)$ has roughly the same distribution as the random variable $\mathcal{L}_{p,\eta}$ for all vectors of points $\mathcal{P}$ such that Condition 2.2 holds. One natural possibility is then to choose the vector of points $\mathcal{P}$ such that the coordinates of each of the p points are empirical quantiles computed from the coordinate samples of the learning sample $\boldsymbol{X}_1, \dots, \boldsymbol{X}_m$. As we continue, for any $1 \le j \le k$, let $F_{j:k}^{[1]}, \dots, F_{j:k}^{[d]}$ be the d univariate margins of $F_{j:k}$ defined in (2.3). Also, for any univariate d.f. H, let H^{-1} denote its associated quantile function (generalized inverse) defined by $H^{-1}(y) = \inf\{x \in \mathbb{R} : H(x) \ge y\}$, $y \in [0,1]$, with the convention that $\inf \emptyset = \infty$. Specifically, one natural possibility for the points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_p$ would be to assume that, for any $i \in [\![1,p]\!]$, $\boldsymbol{x}_i = \boldsymbol{\mathcal{X}}_{i,m}$, where

$$\boldsymbol{\mathcal{X}}_{i,m} = \big(F_{1:m}^{[1],-1}(\pi_i^{[1]}), \dots, F_{1:m}^{[d],-1}(\pi_i^{[d]})\big), \qquad i \in [\![1,p]\!], \tag{3.1}$$

for some $\pi_i^{[1]}, \dots, \pi_i^{[d]} \in (0,1)$. This reduces to a statement about order statistics of the d coordinate samples of $\boldsymbol{X}_1, \dots, \boldsymbol{X}_m$. For example, if for some $\ell \in [\![1,d]\!]$, $j \in [\![1,m]\!]$ is such that $(j-1)/m < \pi_i^{[\ell]} \le j/m$, then $F_{1:m}^{[\ell],-1}(\pi_i^{[\ell]}) = X_{(j)}^{[\ell]}$, which is the jth order statistic of $X_1^{[\ell]}, \dots, X_m^{[\ell]}$. Let $\mathcal{P}_m = (\boldsymbol{\mathcal{X}}_{1,m}, \dots, \boldsymbol{\mathcal{X}}_{p,m})$. According to the notation considered in (2.4), the resulting detector is then $D_m^{\mathcal{P}_m}$.

When $d = 1$, a natural instantiation of the previous strategy for choosing $\mathcal{P}$ consists of setting $\pi_i^{[1]} = i/(p+1)$, $i \in [\![1,p]\!]$, which implies that $\mathcal{X}_{i,m}^{[1]} = F_{1:m}^{[1],-1}\big(i/(p+1)\big)$, $i \in [\![1,p]\!]$, that is, the $\mathcal{X}_{i,m}^{[1]}$'s are merely taken as the $i/(p+1)$-empirical quantiles of the learning sample $X_1^{[1]}, \dots, X_m^{[1]}$. As we will see from the results of the univariate Monte Carlo experiments reported in Section 5, the previous choice seem to lead to powerful multi-purpose open-end monitoring procedures.

3.2. Additional asymptotic results under the null

The asymptotic results stated in Sections 2.2 and 2.3 concern the monitoring procedure based on the detector $D_m^{\mathcal{P}}$ in (2.4) where the chosen evaluation points $\mathcal{P} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_p)$ are fixed (they are not allowed to change in the asymptotics with m). To fully asymptotically justify the use of the detector $D_m^{\mathcal{P}_m}$ introduced in the case of continuous observations in Section 3.1 based on $\mathcal{P}_m = (\boldsymbol{\mathcal{X}}_{1,m}, \dots, \boldsymbol{\mathcal{X}}_{p,m})$ with $\boldsymbol{\mathcal{X}}_{i,m}$ given by (3.1), one needs an analogue of Theorem 2.5 in which the points at which the empirical d.f.s are evaluated are allowed to change with m.

Let $F^{[1]}, \dots, F^{[d]}$ be the d univariate margins of F, where $F(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{X}_1 \le \boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^d$. Furthermore, for any $i \in [\![1,p]\!]$, recall that $\pi_i^{[1]}, \dots, \pi_i^{[d]} \in (0,1)$ are d fixed probabilities chosen by the user and set $\boldsymbol{x}_i = \big(F^{[1],-1}(\pi_i^{[1]}), \dots, F^{[d],-1}(\pi_i^{[d]})\big)$. In addition, write $\mathcal{P} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_p)$ and notice that the points in $\mathcal{P}$ are unobservable since $F^{[1]}, \dots, F^{[d]}$ are unknown. The p-dimensional random vectors $\boldsymbol{Y}_i^{\mathcal{P}} = \big(\mathbf{1}(\boldsymbol{X}_i \le \boldsymbol{x}_1), \dots, \mathbf{1}(\boldsymbol{X}_i \le \boldsymbol{x}_p)\big)$ are also unobservable. Since $\mathcal{P}_m = (\boldsymbol{\mathcal{X}}_{1,m}, \dots, \boldsymbol{\mathcal{X}}_{p,m})$ with $\boldsymbol{\mathcal{X}}_{i,m}$ given by (3.1) is an estimator of $\mathcal{P}$, the long-run covariance matrix $\Sigma^{\mathcal{P}}$ in (2.5) of the unobservable p-dimensional time series $\big(\boldsymbol{Y}_i^{\mathcal{P}}\big)_{i \in \mathbb{N}}$ may still be estimated from the sample $\boldsymbol{Y}_1^{\mathcal{P}_m}, \dots, \boldsymbol{Y}_m^{\mathcal{P}_m}$ which is a proxy for the sample $\boldsymbol{Y}_1^{\mathcal{P}}, \dots, \boldsymbol{Y}_m^{\mathcal{P}}$.

We first state a condition under which the monitoring procedure based on $D_m^{\mathcal{P}_m}$ and the unobservable monitoring procedure based on $D_m^{\mathcal{P}}$ in (2.4) are asymptotically equivalent.

Condition 3.1 (For the asymptotic equivalence of $D_m^{\mathcal{P}_m}$ and $D_m^{\mathcal{P}}$)**.** *For any $i \in [\![1, p]\!]$,*

$$\sup_{k>m} k^{-\frac{1}{2}} \max_{j\in[\![1,k]\!]} j|F_{1:j}(\boldsymbol{\mathcal{X}}_{i,m}) - F(\boldsymbol{\mathcal{X}}_{i,m}) - F_{1:j}(\boldsymbol{x}_i) + F(\boldsymbol{x}_i)| = o_{\mathbb{P}}(1). \tag{3.2}$$

The following result, proven in Appendix D, shows that the previous condition can be satisfied under the null and *absolute regularity*. Given a time series $(\boldsymbol{Z}_i)_{i\in\mathbb{N}}$, recall that, for $j, k \in \mathbb{N}\cup\{+\infty\}$, $\mathcal{M}_j^k$ denotes the σ-field generated by $(\boldsymbol{Z}_i)_{j\leq i\leq k}$, that the absolute regularity coefficients corresponding to $(\boldsymbol{Z}_i)_{i\in\mathbb{N}}$ are defined by

$$\beta_r^{\boldsymbol{Z}} = \mathbb{E}\left\{\sup_{k\in\mathbb{N}} \sup_{B\in\mathcal{M}_{k+r}^{+\infty}} \left|\mathbb{P}(B \mid \mathcal{M}_1^k) - \mathbb{P}(B)\right|\right\}, \qquad r \in \mathbb{N}, \tag{3.3}$$

and that the sequence $(\boldsymbol{Z}_i)_{i\in\mathbb{N}}$ is said to be *absolutely regular* if $\beta_r^{\boldsymbol{Z}} \to 0$ as $r \to \infty$. Also, note that absolute regularity is known to imply strong mixing (see, e.g., Dehling and Philipp, 2002) and that an independent sequence is clearly absolutely regular since in this case for every k and B as in (3.3), $\mathbb{P}(B|\mathcal{M}_1^k) = \mathbb{P}(B)$ almost surely.

Proposition 3.2 (Condition 3.1 can hold under the null)**.** *Assume that the underlying time series $(\boldsymbol{X}_i)_{i\in\mathbb{N}}$ is stationary and absolutely regular, and that its absolute regularity coefficients satisfy $\beta_r^{\boldsymbol{X}} = O(r^{-a})$ as $r \to \infty$ with $a > 1$. Then, if the d univariate margins $F^{[1]}, \dots, F^{[d]}$ of F are continuous and if, for each $i \in [\![1, p]\!]$, $\boldsymbol{\mathcal{X}}_{i,m} \xrightarrow{\mathbb{P}} \boldsymbol{x}_i$, Condition 3.1 holds.*

Remark 3.3. In the statement of Proposition 3.2, it is assumed that, for each $i \in [\![1, p]\!]$, $\boldsymbol{\mathcal{X}}_{i,m} \xrightarrow{\mathbb{P}} \boldsymbol{x}_i$. This condition can actually be dispensed with provided additional conditions of the true unobservable quantile functions $F^{[1],-1}, \dots, F^{[d],-1}$ are assumed instead. Indeed, from Rio (1998), we know that the condition on the absolute regularity coefficients in Proposition 3.2 implies that, for any $\ell \in [\![1, d]\!]$, $F_{1:m}^{[\ell]} \xrightarrow{\mathbb{P}} F^{[\ell]}$ in $\ell^\infty(\mathbb{R})$, where $\ell^\infty(\mathbb{R})$ denotes the space of bounded functions on $\mathbb{R}$ equipped with the uniform metric. From Lemma 21.2 in van der Vaart (1998), this is then equivalent to the fact that $F_{1:m}^{[\ell],-1}(\pi) \xrightarrow{\mathbb{P}} F^{[\ell],-1}(\pi)$ at every $\pi \in (0, 1)$ at which $F^{[\ell],-1}$ is continuous. Consequently, the condition that, for any $i \in [\![1, p]\!]$, $\boldsymbol{\mathcal{X}}_{i,m} \xrightarrow{\mathbb{P}} \boldsymbol{x}_i$ could be replaced by the condition that, for any $\ell \in [\![1, d]\!]$, $F^{[\ell],-1}$ is continuous at $\pi_i^{[\ell]}$, for all $i \in [\![1, p]\!]$.

The next proposition, also proven in Appendix D, states that, under Conditions 2.4 and 3.1, the monitoring procedures based on $D_m^{\mathcal{P}_m}$ and $D_m^{\mathcal{P}}$ are asymptotically equivalent.

Proposition 3.4. *Under Conditions 2.4 and 3.1, and if $\Sigma_m^{\mathcal{P}} \xrightarrow{\mathbb{P}} \Sigma^{\mathcal{P}}$ and $\Sigma_m^{\mathcal{P}_m} \xrightarrow{\mathbb{P}} \Sigma^{\mathcal{P}}$, for any $\eta > 0$,*

$$\sup_{k>m} (m/k)^{\frac{3}{2}+\eta} |D_m^{\mathcal{P}_m}(k) - D_m^{\mathcal{P}}(k)| = o_{\mathbb{P}}(1),$$

and, consequently,

$$\sup_{k>m} (m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}_m}(k) \rightsquigarrow \mathcal{L}_{p,\eta} = \sup_{1\leq s\leq t<\infty} t^{-\frac{3}{2}-\eta} \|t\boldsymbol{W}(s) - s\boldsymbol{W}(t)\|_{I_p}.$$

The last claim of the previous proposition suggests to carry out the monitoring procedure based on the detector $D_m^{\mathcal{P}_m}$ exactly as the procedure based on the detector $D_m^{\mathcal{P}}$ when the points $\mathcal{P}$ are hand-picked by the user (see the last paragraph of Section 2.2).

4. Estimation of high quantiles of the limiting distribution

From the two previous sections, we know that, to carry out the studied monitoring procedures, it is necessary to be able to accurately estimate high quantiles of the random variable $\mathcal{L}_{p,\eta}$, $p \geq 1$, $\eta > 0$, appearing first in the statement of Theorem 2.5. The underlying estimation problem was empirically solved in Holmes and Kojadinovic (2021) for $p = 1$ using *asymptotic regression modeling*. We chose to use the same approach when $p > 1$ and refer the reader to Section 4 of the aforementioned reference where the motivation for this way of proceeding is explained in detail.

Fix $\eta > 0$, $p \geq 1$, $\alpha \in (0, \frac{1}{2})$ and let $q_{p,\eta}^{(1-\alpha)}$ be the $(1-\alpha)$-quantile of $\mathcal{L}_{p,\eta}$. Furthermore, let $d = 1$, let $\big(X_i^{[1]}\big)_{i \in \mathbb{N}}$ be an infinite sequence of independent standard normals and let $\mathcal{P} = (x_1^{[1]}, \dots, x_p^{[1]})$ such that $x_i^{[1]} = \Phi^{-1}\big(i/(p+1)\big)$, $i \in [\![1, p]\!]$, where Φ is d.f. of the standard normal. Then, according to Theorem 2.5, for large m, the distribution of $\sup_{k>m} (m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k)$ should be close to that of $\mathcal{L}_{p,\eta}$. If, for a given realization of $\big(X_i^{[1]}\big)_{i \in \mathbb{N}}$, we could compute the corresponding realization of $\sup_{k>m} (m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k)$, $q_{p,\eta}^{(1-\alpha)}$ could be estimated by $\hat{q}_{p,\eta}^{(1-\alpha)}$, the $(1-\alpha)$-empirical quantile of a large sample of realizations of $\sup_{k>m} (m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k)$. As this is not possible because of the supremum over $k > m$, the idea taken from Holmes and Kojadinovic (2021) is to model the relationship between $r \in [\![9, 16]\!]$ and $\hat{q}_{p,\eta,r}^{(1-\alpha)}$, the $(1-\alpha)$-empirical quantile of $\sup_{k \in]\!]m, m+2^r]\!]} (m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k)$, using an asymptotic regression model. To begin with, using a computer grid, we computed the empirical quantiles $\hat{q}_{p,\eta,r}^{(1-\alpha)}$, $r \in [\![9, 16]\!]$, from 10,000 simulated trajectories of the scaled detector $k \mapsto (m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k)$ for $k \in]\!]m, m + 2^{16}]\!]$ and $m = 500$. In a next step, an asymptotic regression model was fitted to the points $(r, \hat{q}_{p,\eta,r}^{(1-\alpha)})$, $r \in [\![9, 16]\!]$. The considered model is a three-parameter model with mean function

$$f(x) = \beta_1 + (\beta_2 - \beta_1)\{1 - \exp(-x/\beta_3)\},$$

where $y = \beta_2$ is the equation of the upper horizontal asymptote of f. Its fitting was carried out using the R package `drc` (Ritz et al., 2015). A candidate estimate $\hat{q}_{p,\eta}^{(1-\alpha)}$ of $q_{p,\eta}^{(1-\alpha)}$, the $(1-\alpha)$-quantile of $\mathcal{L}_{p,\eta}$, is then the resulting estimate of the parameter β_2. The previous steps were carried out for $p \in \{2, 5, 10, 20\}$, $\alpha \in \{0.01, 0.05, 0.1\}$ and $\eta = 0.001$ (following the practical recommendation made in Holmes and Kojadinovic (2021)), and can be visualized in the first four panels of Figure 1 for $\alpha = 0.05$. The corresponding estimates of the quantiles $q_{p,\eta}^{(1-\alpha)}$ are given in columns two to five of Table 1.

In a last step, to be able to carry out the monitoring procedures for some values of p different than those in $\{2, 5, 10, 20\}$, we fitted asymptotic regression models to the points $(\log(p), 2 - \hat{q}_{p,\eta}^{(1-\alpha)})$, $p \in \{2, 5, 10, 20\}$, for $\alpha \in \{0.01, 0.05, 0.1\}$ and $\eta = 0.001$. The estimates of the parameters β_1, β_2 and β_3 are reported in the last three columns of Table 1. We make no claim regarding the theoretical adequacy of this type of model. The aim is only to be able

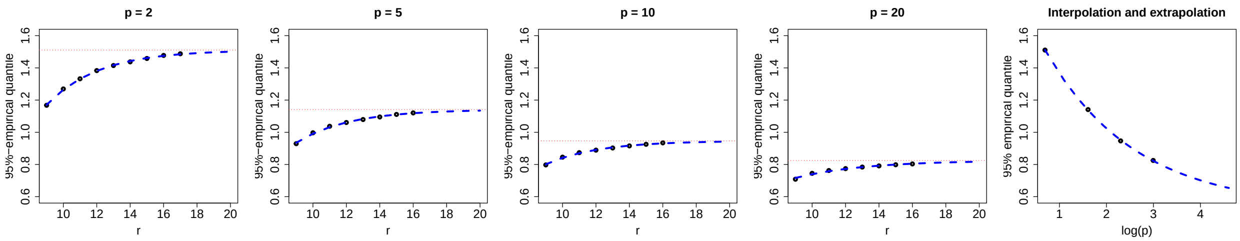

FIG 1. *First four panels: for $\alpha = 0.05$, $\eta = 0.001$ and $p \in \{2, 5, 10, 20\}$, scatter plots of $\{(r, \hat{q}^{(1-\alpha)}_{p,\eta,r})\}_{r \in [\![9,16]\!]}$, corresponding fitted asymptotic regression models (dashed blue curves) and estimates of the upper horizontal asymptotes (dotted red lines) which are candidate estimates of $q^{(1-\alpha)}_{p,\eta}$, the $(1-\alpha)$-quantile of $\mathcal{L}_{p,\eta}$. Fifth panel: scatter plot of $\{(\log(p), \hat{q}^{(1-\alpha)}_{p,\eta})\}_{p \in \{2,5,10,20\}}$ and corresponding transformed fitted asymptotic regression model (dashed blue curve) that could be used to interpolate (resp. extrapolate) the value of $\hat{q}^{(1-\alpha)}_{p,\eta}$ for $p \in [2, 20]$ (resp. for p slightly larger than 20).*

TABLE 1

Columns 2 to 5: for $\eta = 0.001$, $p \in \{2, 5, 10, 20\}$ and $\alpha \in \{0.01, 0.05, 0.1\}$, estimates $\hat{q}^{(1-\alpha)}_{p,\eta}$ of the $(1-\alpha)$-quantiles $q^{(1-\alpha)}_{p,\eta}$ of the distribution of $\mathcal{L}_{p,\eta}$. Columns 6 to 8: corresponding estimates of the parameters of the transformed asymptotic regression models that can be used to interpolate (resp. extrapolate) the value of $\hat{q}^{(1-\alpha)}_{p,\eta}$ for $p \in [2, 20]$ (resp. for p slightly larger than 20).

	p				$p \notin \{2, 5, 10, 20\}$		
$1-\alpha$	2	5	10	20	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
0.99	1.654	1.234	1.010	0.860	-0.126	1.535	2.080
0.95	1.511	1.141	0.946	0.825	0.060	1.475	1.921
0.90	1.450	1.099	0.921	0.806	0.140	1.462	1.870

to interpolate (resp. extrapolate) the value of $\hat{q}^{(1-\alpha)}_{p,\eta}$ for $p \in [2, 20]$ (resp. for p slightly larger than 20). Note that, since $\mathcal{L}_{p,\eta} \overset{as}{\to} 0$ as $p \to \infty$ (as a consequence of the definition of the norm $\|\cdot\|_{I_p}$), a fitted two-parameter submodel with β_2 fixed to 2 is expected to behave better for large p. We have nonetheless decided to keep the fitted three parameter model because its accuracy for values of p slightly larger than 20 was found to be better in our Monte Carlo experiments reported in the forthcoming section.

5. Monte Carlo experiments

We chose to restrict our numerical experiments to one-dimensional ($d = 1$) continuous observations and investigated in detail the behavior of the monitoring procedure described in Section 3.1 in the case when the p real points at which the (univariate) empirical d.f.s are evaluated are empirical quantiles of order $\pi_i^{[1]} = i/(p+1)$, $i \in [\![1, p]\!]$, computed from the learning sample $X_1^{[1]}, \dots, X_m^{[1]}$. Specifically, the underlying detector is $D_m^{\mathcal{P}_m}$ defined as in (2.4) with $\mathcal{P} = \mathcal{P}_m = (\mathcal{X}^{[1]}_{1,m}, \dots, \mathcal{X}^{[1]}_{p,m})$ and $\mathcal{X}^{[1]}_{i,m} = F^{[1],-1}_{1:m}\big(i/(p+1)\big)$, $i \in [\![1, p]\!]$.

From the definition of the detector $D_m^{\mathcal{P}_m}$, we see that the unknown long-run covariance matrix is estimated from the sample $\boldsymbol{Y}_1^{\mathcal{P}_m}, \dots, \boldsymbol{Y}_m^{\mathcal{P}_m}$. In practice, for $\Sigma_m^{\mathcal{P}_m}$, we used the estimator of Andrews (1991) based on the quadratic spectral kernel with automatic bandwidth selection as implemented in the function `lrvar()` of the R package `sandwich` (Zeileis, 2004).

The resulting estimates of the unknown long-run covariance matrix were always found to be positive-definite. In all our experiments, the sequential tests were carried out at the $\alpha = 5\%$ nominal level.

5.1. Experiments for the procedure based on $D_m^{\mathcal{P}_m}$ under the null

Following Holmes and Kojadinovic (2021), to investigate the empirical levels of the sequential test based on $D_m^{\mathcal{P}_m}$, we considered 9 data generating models, denoted M1, ..., M9. Models M1, ..., M5 are AR(1) models with normal innovations whose autoregressive parameter is equal to 0, 0.1, 0.3, 0.5 and 0.7, respectively. Model M6 generates independent observations from the Student t distribution with 5 degrees of freedom. Model M7 (resp. M8) is a GARCH(1,1) model with normal innovations and parameters $\omega = 0.012$, $\beta = 0.919$ and $\alpha = 0.072$ (resp. $\omega = 0.037$, $\beta = 0.868$ and $\alpha = 0.115$) to mimic SP500 (resp. DAX) daily log-returns following Jondeau, Poon and Rockinger (2007). Model M9 is the nonlinear autoregressive model used in Paparoditis and Politis (2001, Section 3.3) with underlying generating equation $X_i = 0.6 \sin(X_{i-1}) + \epsilon_i$, where the ϵ_i are independent standard normal innovations.

For each of the nine models, the probability of rejection of H_0 in (1.1) was estimated from 1000 samples of size $n = m + 5000$ with $m \in \{200, 400, 800, 1600\}$ and for $p \in \{2, 5, 10, 20\}$. The empirical levels are reported in Table 2. As one can see, for any fixed p, reassuringly, the empirical levels decrease as m increases. For $m \geq 800$, they are all below the 5% nominal level except for Models M5, M7 and M8. The latter is not so surprising for Model M5 (the AR(1) with autoregressive parameter 0.7) and highlights the difficulty of the estimation of the long-run covariance matrix $\Sigma^{\mathcal{P}}$ using the estimator $\Sigma_m^{\mathcal{P}_m}$ in the case of strong serial dependence. It may be slightly more surprising for the GARCH(1,1) models M7 and M8 for which a very large learning sample seems necessary to obtain a reasonably good estimate of $\Sigma^{\mathcal{P}}$. The fact that for the other models, the empirical levels are all below the 5% nominal level when $m \geq 800$ is a consequence of the fact that they are underestimated in all settings. Indeed, the monitoring was stopped after 5000 steps whereas it would theoretically be necessary to monitor "indefinitely" to estimate empirical levels accurately. For any fixed m, we see that increasing p tends in general to increase the empirical level. This is again a consequence of the difficulty of the estimation of the long-run covariance matrix $\Sigma^{\mathcal{P}}$ which is a $p \times p$ matrix.

In a second experiment, we restricted our attention to independent observations and investigated the influence of the contemporary distribution on the empirical levels. Table 3 reports rejection percentages of H_0 computed from 1000 random samples of size $n = m + 5000$ from either the standard normal distribution, the exponential distribution, the centered and scaled exponential distribution (so that it has expectation zero and variance one) or the scaled Student t distribution with 3 degrees of freedom (so that it has variance one). As one can see, the interplay between m and p is the same as in the previous experiment. Furthermore, the empirical levels seem hardly unaffected by the choice of the contemporary distribution.

In a third experiment, we briefly investigated the effect of the randomness of the points $\mathcal{X}_{1,m}^{[1]}, \dots, \mathcal{X}_{p,m}^{[1]}$ on the empirical levels. Instead of using the definition $\mathcal{X}_{i,m}^{[1]} = F_{1:m}^{[1],-1}\big(i/(p+1)\big)$, $i \in [\![1, p]\!]$, we considered the definition $\mathcal{X}_{i,m}^{[1]} = \Phi^{-1}\big(i/(p+1)\big)$, $i \in [\![1, p]\!]$, where Φ is the d.f. of the standard normal distribution, and estimated the rejection percentages from 1000

Table 2
Percentages of rejection of H_0 in (1.1) for the procedure based on $D_m^{\mathcal{P}_m}$ considered in Section 3.1. The rejection percentages are computed from 1000 samples of size $n = m + 5000$ generated from the time series models M1 – M9.

Model	m	$p = 2$	$p = 5$	$p = 10$	$p = 20$
M1	200	4.1	6.2	10.4	29.7
	400	2.5	3.1	6	9
	800	1.7	1.6	2.7	2.1
	1600	1	1.2	1	1.6
M2	200	4.2	6.3	12.7	31.5
	400	3.3	3	4.7	9.1
	800	1.8	2.9	2.2	3.8
	1600	1.3	1.1	1.4	1
M3	200	4.1	8	14.2	34.2
	400	2.6	3.6	4.1	8.1
	800	2.3	2.3	3	2.9
	1600	1.4	1.1	1.3	1.1
M4	200	6.2	7.9	16.4	37.5
	400	4	3.8	5.3	10.7
	800	2.8	3.1	2.7	2.4
	1600	1.9	1.2	1.4	0.8
M5	200	15.1	15.8	26.4	53.5
	400	8.5	8.8	12.6	18.9
	800	6.1	4.1	5.6	5.3
	1600	3.7	1.9	2.4	1.3
M6	200	5	5.4	11.1	29.8
	400	3.4	2.8	4.7	8.3
	800	1.8	2.2	2.4	2.3
	1600	1.5	1.1	1	1.2
M7	200	13.6	25	33.6	51.6
	400	10.3	17	20.2	23.1
	800	8.3	11.8	13.6	10.5
	1600	5.6	7.6	8.1	4.5
M8	200	10.8	18.1	27.7	47.2
	400	8.2	11.6	15.1	18.2
	800	7.4	8.5	9.1	6.5
	1600	4.4	5.3	4.6	3.5
M9	200	4.6	7	15	35.1
	400	3.1	2.5	5.9	9.4
	800	2.4	1.8	2.8	2.5
	1600	1.4	1.8	1.5	0.9

random samples of size $n = m + 5000$ from the standard normal distribution. The results are reported in Table 4. By comparing the results with the first horizontal block of Table 3, we see that, somehow surprisingly, for the same value of m, the use of empirical quantiles for $\mathcal{X}_{1,m}^{[1]}, \dots, \mathcal{X}_{p,m}^{[1]}$ leads to better empirical levels than when the true quantiles are used.

TABLE 3

Percentages of rejection of H_0 in (1.1) for the procedure based on $D_m^{\mathcal{P}_m}$ considered in Section 3.1. The rejection percentages are computed from 1000 random samples of size $n = m + 5000$ from either the standard normal distribution, the exponential distribution, the centered and scaled exponential distribution or the scaled Student t distribution with 3 degrees of freedom.

Distribution	m	$p=2$	$p=5$	$p=10$	$p=20$
standard normal	200	4.1	6.2	10.4	29.7
	400	2.5	3.1	6	9
	800	1.7	1.6	2.7	2.1
exponential	200	4.9	5.5	12	29
	400	2.8	3.2	5.2	7.9
	800	2.1	1.6	2.2	1.8
centered and scaled exponential	200	4.9	5.5	12	29
	400	2.8	3.2	5.2	8.1
	800	2.1	1.6	2.2	1.8
scaled Student t with 3 degrees of freedom	200	4.5	5.1	11.5	29.2
	400	1.7	2.4	3.8	8.8
	800	1.7	1.9	2.3	1.8

TABLE 4

Percentages of rejection of H_0 in (1.1) for the procedure based on $D_m^{\mathcal{P}}$ in (2.4) with $\mathcal{P} = \big(\Phi^{-1}\big(1/(p+1)\big), \dots, \Phi^{-1}\big(p/(p+1)\big)\big)$. The rejection percentages are computed from 1000 random samples of size $n = m + 5000$ from the standard normal distribution.

m	$p=2$	$p=5$	$p=10$	$p=20$
200	4.5	9.7	22.6	61.3
400	2.8	4.0	9.2	18.0
800	2.0	2.4	3.0	4.9

As a last experiment under the null, we investigated the quality of the model fitted at the end of Section 4 to extrapolate the values of the quantiles of the distribution of $\mathcal{L}_{p,\eta}$ for $p > 20$ and $\eta = 0.001$. Using $m = 1000$ and 1000 random samples of size $n = m + 5000$ from the standard normal distribution, the rejection percentages for $p = 30$ (resp. 40, 50) were found to be 1.7% (resp. 1.8%, 3.9%), thereby suggesting that the quality of the model for extrapolating the values of the quantiles may be deteriorating when $p > 40$.

5.2. Experiments for the procedure based on $D_m^{\mathcal{P}_m}$ under alternatives

In order to understand the behavior of the monitoring procedure based on the detector $D_m^{\mathcal{P}_m}$ considered in Section 3.1 under alternatives to H_0 in (1.1), we considered successively changes in the expectation of the $X_i^{[1]}$'s, in their variance and in their d.f. (while keeping their expectation and variance constant). As a first experiment, we studied the finite-sample behavior of the sequential test under a change in the expectation of an AR(1) model with autoregressive parameter equal to 0.3 (Model M3). Specifically, to estimate rejection percentages, we generated 1000 samples of size $n = m + 5000$ from Model M3 with $m = 800$ and, for each sample, added a positive offset of δ to all observations after position $m + k$ with $k \in \{0, 500, 1000, 2000\}$. The results are reported in Figure 2. Notice that only the

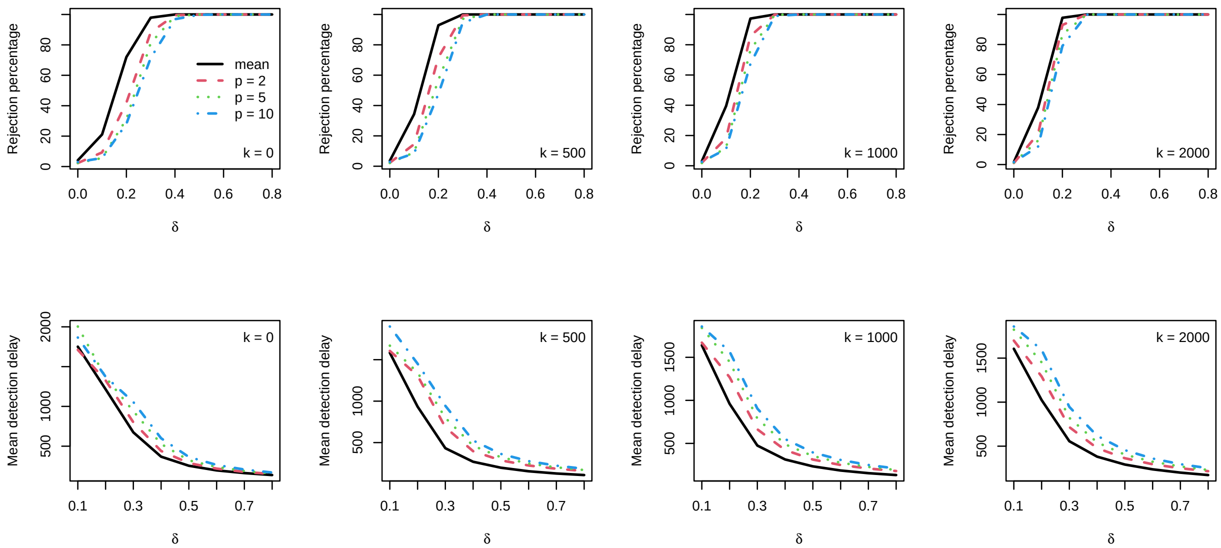

FIG 2. *Rejection percentages of H_0 in (1.1) (first row) and corresponding mean detection delays (second row) for the procedure based R_m in (2.1) (solid line) and for the procedure $D_m^{\mathcal{P}_m}$ considered in Section 3.1 with $p \in \{2, 5, 10\}$ (dash, dotted and dash-dotted lines) estimated from 1000 samples of size $n = m + 5000$ from Model M3 with $m = 800$ such that, for each sample, a positive offset of δ was added to all observations after position k.*

exceedences (of the detectors with respect to their thresholds) after position $m+k$ are taken into account when calculating the rejection percentages.

As one can see from the top row of graphs in Figure 2, as expected, the power of all procedures increases as δ increases. Furthermore, for the procedure based on $D_m^{\mathcal{P}_m}$ and a fixed value of the offset δ, increasing the number p of evaluation points slightly lowers the power of the test. We also see that the procedure based on R_m in (2.1) is always the most powerful. This was to be expected as the latter was specifically designed to be sensitive to changes in the mean. Similarly, from the second row of plots in Figure 2, we see that mean detection delays are smallest for the procedure based on R_m and increase for the procedure based on $D_m^{\mathcal{P}_m}$ as p increases. Note finally that the power of every procedure becomes larger as the time k at which the offset δ is added becomes closer to half of $n = m + 5000$. This is a consequence of using of CUSUM statistics to define the detectors.

As a second experiment, we considered a change in the variance of independent centered observations. To estimate the power of the sequential test, we generated 1000 samples of size $n = m + 5000$ with $m = 800$ such that, observations up to position $m + k$ with $k \in \{0, 500, 1000, 2000\}$ are from a standard normal distribution while observations after position $m + k$ are for the $N(0, \sigma^2)$ distribution. The results are reported in Figure 3. As expected, the power of the procedure based on $D_m^{\mathcal{P}_m}$ increases as σ deviates further away from one. In contrast to the first experiment however, the rejection percentages (resp. mean detection delays) increase (resp. decrease) as the number of evaluation points p increases. Notice that the improvement as p increases from 5 to 10 appears to be rather small.

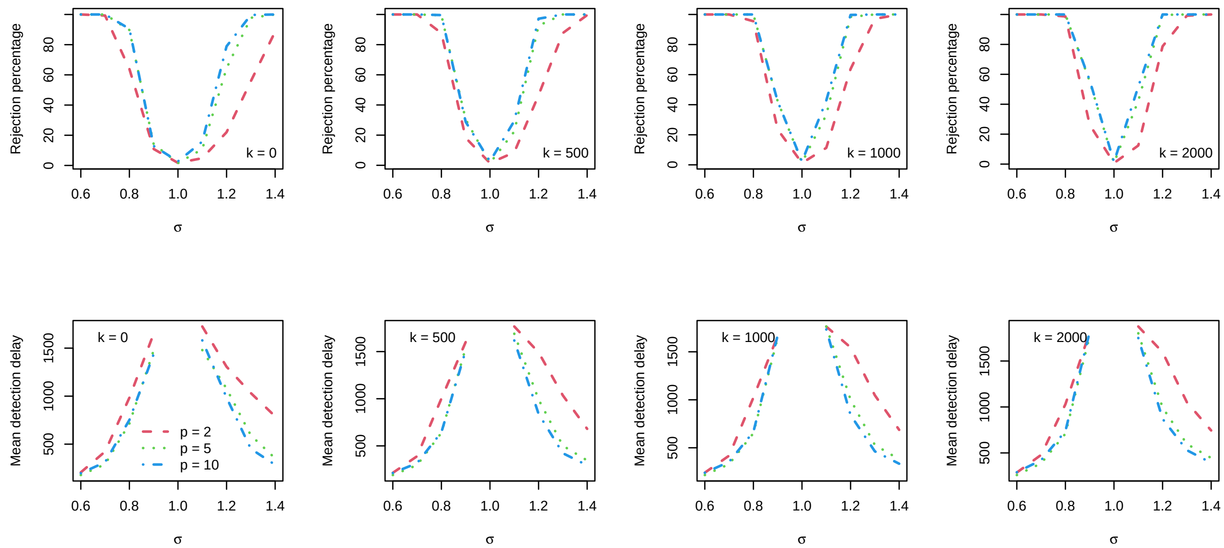

FIG 3. *Rejection percentages of* H_0 *in* (1.1) *and corresponding mean detection delays for the procedure based on* $D_m^{\mathcal{P}_m}$ *with* $p \in \{2, 5, 10\}$ *estimated from 1000 random samples of size* $n = m + 5000$ *with* $m = 800$ *such that observations up to position* $m + k$ *with* $k \in \{0, 500, 1000, 2000\}$ *are from a standard normal distribution while observations after position* $m + k$ *are for the* $N(0, \sigma^2)$ *distribution.*

As a final experiment, we considered a change in the contemporary distribution of independent observations that keeps the expectation and the variance constant. To estimate the rejection percentages, we generated 1000 samples of size $n = m + 5000$ with $m = 800$ such that observations up to position $m + k$ with $k \in \{0, 500, 1000, 2000\}$ are from a scaled Student t distribution with 3 degrees of freedom (where the scaling is performed so that the variance is equal to one) while observations after position $m + k$ are from the scaled Student t distribution with $\nu \in [\![3, 10]\!]$ degrees of freedom. The results are reported in Figure 4. As expected, the power of the procedure based on $D_m^{\mathcal{P}_m}$ increases as ν increases. Furthermore, as in the previous experiment, the rejection percentages (resp. mean detection delays) are larger (resp. smaller) when $p \in \{5, 10\}$. Somehow surprisingly however, the results seem slightly better when $p = 5$.

6. Data example and concluding remarks

Let us briefly illustrate how the procedure based on the detector $D_m^{\mathcal{P}_m}$ considered in Section 3.1 could be used for monitoring changes in the contemporary distribution of daily log-returns of a financial index. In our fictitious example, the learning sample (whose stationarity would need to be tested) corresponds to 1257 trading days of the NASDAQ for the period January 3rd 2012 – December 30th 2016. Monitoring starts on the first trading day of 2017. The corresponding daily log-returns are represented in Figure 5, where the dashed vertical line represents the start of the monitoring.

The sample paths of $k \mapsto (m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}_m}(k)$ are represented in Figure 6 for $p \in \{2, 5, 10, 20\}$.

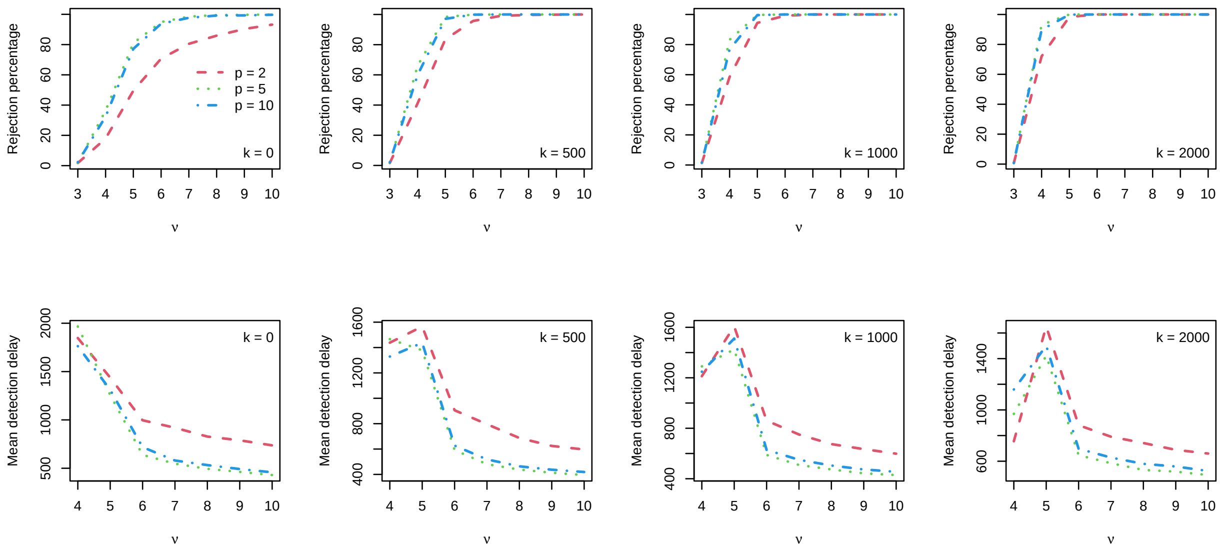

FIG 4. *Rejection percentages of H_0 in (1.1) and corresponding mean detection delays for the procedure based on $D_m^{\mathcal{P}_m}$ with $p \in \{2, 5, 10\}$ estimated from 1000 random samples of size $n = m+5000$ with $m = 800$ such that observations up to position $m+k$ with $k \in \{0, 500, 1000, 2000\}$ are from the scaled Student t distribution with $\nu = 3$ degrees of freedom while observations after position $m+k$ are from the scaled Student t distribution with $\nu \in [\![3, 10]\!]$ degrees of freedom.*

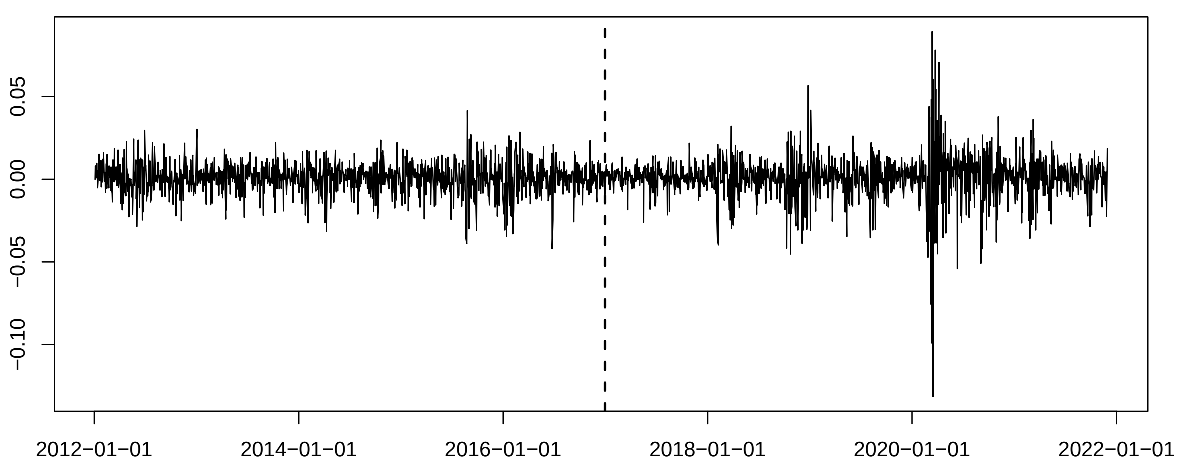

FIG 5. *Daily log-returns computed from closing quotes of the NASDAQ composite index from 2012 to 2021. The learning sample in our fictitious data example corresponds to the period 2012 – 2016. Monitoring starts on the first trading day of 2017, which is represented by a dashed vertical line.*

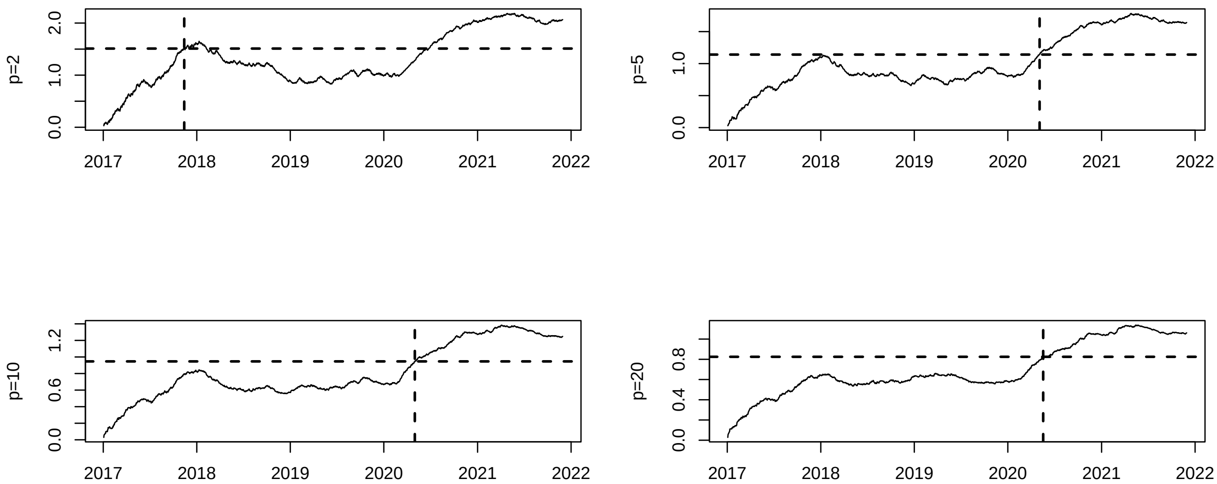

FIG 6. *Sample paths of* $k \mapsto (m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}_m}(k)$ *for* $p = 2, 5, 10$ *and* 20. *The horizontal dashed lines represent the 95%-quantiles given in the second row of Table 1. The vertical dashed lines mark the first time the scaled versions of the detectors exceed the corresponding quantiles.*

The horizontal dashed line in each panel represent one of the 95%-quantiles given in the second row of Table 1. The dashed vertical lines mark first time threshold exceedences. As one can see, times of exceedences correspond either to November 2017 (only for $p = 2$) or May 2020. The changes in distribution corresponding to the latter date seem more significant given that all the scaled detectors remain above the corresponding 95%-quantiles after that date. This is confirmed by an inspection of Figure 5 which reveals indeed a period of very high volatility during the first months of 2020.

We end this section by stating a few remarks:

- An implementation of the studied multi-purpose open-end monitoring procedure will be made available in the very near future in the R package `npcp` (Kojadinovic and Verhoijsen, 2021).
- For the monitoring of continuous multivariate observations, the strategy for choosing the evaluation points proposed in Section 3.1 was not completely defined and requires further research. One possibility would be to choose the points according to the *depth* of the observations (a concept initially proposed by Tukey, 1975, in the bivariate case with numerous extensions since then) in the learning sample.
- The type of monitoring procedure used in this work could also be used to detect changes in the serial dependence. For instance, to detect such changes "at lag 1" from an initial sequence of univariate observations $Z_1, \dots, Z_m, \dots$, the $\boldsymbol{X}_i$ could be formed as $\boldsymbol{X}_i = (Z_i, Z_{i+1})$.
- Additional numerical experiments would be required to fully understand the behavior of the procedure based on $D_m^{\mathcal{P}_m}$ in the case of continuous multivariate observations,

whether the observations are truly multivariate or obtained as mentioned in the previous point.

Appendix A: Proof of Theorem 2.5

The proof of Theorem 2.5 is based on four lemmas which we prove first. Throughout the remainder of the proof, we let

$$\tilde{D}_m^{\mathcal{P}}(k) = \max_{j \in [\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} \|\bar{\boldsymbol{Y}}_{1:j}^{\mathcal{P}} - \bar{\boldsymbol{Y}}_{j+1:k}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}}, \qquad k \geq m+1, \tag{A.1}$$

which is the version of the detector in (2.6), in which the estimated long-run variance $\Sigma_m^{\mathcal{P}}$ is replaced by the true long-run variance $\Sigma^{\mathcal{P}}$.

On the probability space of Condition 2.4 (assuming that this condition holds), we may define

$$\bar{D}_m^{\mathcal{P}}(k) = \frac{1}{\sqrt{m}} \max_{j \in [\![m,k[\![} \Bigg\| \frac{k}{m} (\Sigma^{\mathcal{P}})^{\frac{1}{2}} \{\boldsymbol{W}_{2,m}(m) + \boldsymbol{W}_{1,m}(j-m)\} - \frac{j}{m} (\Sigma^{\mathcal{P}})^{\frac{1}{2}} \{\boldsymbol{W}_{2,m}(m) + \boldsymbol{W}_{1,m}(k-m)\} \Bigg\|_{(\Sigma^{\mathcal{P}})^{-1}}, \tag{A.2}$$

where we recall that for each $m \in \mathbb{N}$, $\boldsymbol{W}_{1,m}$ and $\boldsymbol{W}_{2,m}$ are independent p-dimensional standard Brownian motions.

Lemma A.1. *Assume that Condition 2.4 holds. Then, for any $\eta > 0$,*

$$\sup_{k>m} \left(\frac{m}{k}\right)^{\frac{3}{2}+\eta} \left|\tilde{D}_m^{\mathcal{P}}(k) - \bar{D}_m^{\mathcal{P}}(k)\right| = o_{\mathbb{P}}(1).$$

Proof. Fix $\eta > 0$. Applying the reverse triangle inequality for suprema

$$|\sup_{x \in A} f(x) - \sup_{x \in A} g(x)| \leq \sup_{x \in A} |f(x) - g(x)| \tag{A.3}$$

to the maximum over $j \in [\![m,k[\![$ below gives

$$\begin{aligned}
&\left|\tilde{D}_m^{\mathcal{P}}(k) - \bar{D}_m^{\mathcal{P}}(k)\right| \\
&\leq \frac{1}{\sqrt{m}} \max_{j \in [\![m,k[\![} \Bigg| \frac{j(k-j)}{m} \|\bar{\boldsymbol{Y}}_{1:j}^{\mathcal{P}} - \bar{\boldsymbol{Y}}_{j+1:k}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}} - \Bigg\| \frac{k}{m} (\Sigma^{\mathcal{P}})^{\frac{1}{2}} \{\boldsymbol{W}_{2,m}(m) + \boldsymbol{W}_{1,m}(j-m)\} \\
&\qquad - \frac{j}{m} (\Sigma^{\mathcal{P}})^{\frac{1}{2}} \{\boldsymbol{W}_{2,m}(m) + \boldsymbol{W}_{1,m}(k-m)\} \Bigg\|_{(\Sigma^{\mathcal{P}})^{-1}} \Bigg| \\
&\leq \frac{1}{\sqrt{m}} \max_{j \in [\![m,k[\![} \Bigg\| \frac{j(k-j)}{m} \{\bar{\boldsymbol{Y}}_{1:j}^{\mathcal{P}} - \bar{\boldsymbol{Y}}_{j+1:k}^{\mathcal{P}}\} - \frac{k}{m} (\Sigma^{\mathcal{P}})^{\frac{1}{2}} \{\boldsymbol{W}_{2,m}(m) + \boldsymbol{W}_{1,m}(j-m)\} \\
&\qquad + \frac{j}{m} (\Sigma^{\mathcal{P}})^{\frac{1}{2}} \{\boldsymbol{W}_{2,m}(m) + \boldsymbol{W}_{1,m}(k-m)\} \Bigg\|_{(\Sigma^{\mathcal{P}})^{-1}} = U_m^{\mathcal{P}}(k).
\end{aligned}$$

Next, we rewrite $j(k-j)\{\bar{\boldsymbol{Y}}^{\mathcal{P}}_{1:j} - \bar{\boldsymbol{Y}}^{\mathcal{P}}_{j+1:k}\}$ as

$$\begin{aligned}
j(k-j)\left\{\frac{1}{j}\sum_{i=1}^{j}\boldsymbol{Y}_i^{\mathcal{P}} - \frac{1}{k-j}\sum_{i=j+1}^{k}\boldsymbol{Y}_i^{\mathcal{P}}\right\} &= (k-j)\sum_{i=1}^{j}\{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - j\sum_{i=j+1}^{k}\{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} \\
&= k\sum_{i=1}^{j}\{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - j\sum_{i=1}^{k}\{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} \\
&= k\sum_{i=1}^{m}\{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} + k\sum_{i=m+1}^{j}\{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} \\
&\quad - j\sum_{i=1}^{m}\{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - j\sum_{i=m+1}^{k}\{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\}.
\end{aligned}$$

Hence, using the triangle inequality, we have that

$$\begin{aligned}
\sup_{k>m}\left(\frac{m}{k}\right)^{\frac{3}{2}+\eta}\left|\tilde{D}_m^{\mathcal{P}}(k) - \bar{D}_m^{\mathcal{P}}(k)\right| &\le \sup_{k>m}\left(\frac{m}{k}\right)^{\frac{3}{2}+\eta} U_m^{\mathcal{P}}(k) \\
&\le I_m^{\mathcal{P}} + (I_m^{\mathcal{P}})' + J_m^{\mathcal{P}} + (J_m^{\mathcal{P}})',
\end{aligned}$$

where (using the convention that empty sums are equal to 0)

$$\begin{aligned}
I_m^{\mathcal{P}} &= \frac{1}{\sqrt{m}}\sup_{k>m}\left(\frac{m}{k}\right)^{\frac{1}{2}+\eta}\max_{j\in[\![m,k[\![}\left\|\sum_{i=1}^{m}\{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}}\boldsymbol{W}_{2,m}(m)\right\|_{(\Sigma^{\mathcal{P}})^{-1}}, \\
(I_m^{\mathcal{P}})' &= \frac{1}{\sqrt{m}}\sup_{k>m}\left(\frac{m}{k}\right)^{\frac{1}{2}+\eta}\max_{j\in[\![m,k[\![}\left\|\sum_{i=m+1}^{j}\{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}}\boldsymbol{W}_{1,m}(j-m)\right\|_{(\Sigma^{\mathcal{P}})^{-1}}, \\
J_m^{\mathcal{P}} &= \frac{1}{\sqrt{m}}\sup_{k>m}\left(\frac{m}{k}\right)^{\frac{3}{2}+\eta}\max_{j\in[\![m,k[\![}\frac{j}{m}\left\|\sum_{i=1}^{m}\{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}}\boldsymbol{W}_{2,m}(m)\right\|_{(\Sigma^{\mathcal{P}})^{-1}}, \\
(J_m^{\mathcal{P}})' &= \frac{1}{\sqrt{m}}\sup_{k>m}\left(\frac{m}{k}\right)^{\frac{3}{2}+\eta}\max_{j\in[\![m,k[\![}\frac{j}{m}\left\|\sum_{i=m+1}^{k}\{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}}\boldsymbol{W}_{1,m}(k-m)\right\|_{(\Sigma^{\mathcal{P}})^{-1}}.
\end{aligned}$$

It suffices to show that $I_m^{\mathcal{P}}$, $(I_m^{\mathcal{P}})'$, $J_m^{\mathcal{P}}$, and $(J_m^{\mathcal{P}})'$ are $o_{\mathbb{P}}(1)$. Since $(m/k)\times(j/m) = j/k \le 1$ when $m \le j < k$, we have that $J_m^{\mathcal{P}} \le I_m^{\mathcal{P}}$ and $(J_m^{\mathcal{P}})' \le (I_m^{\mathcal{P}})'$. So it suffices to consider $I_m^{\mathcal{P}}$ and $(I_m^{\mathcal{P}})'$. Let $0 < \xi < \frac{1}{2}$ be as in Condition 2.4. Then,

$$\begin{aligned}
I_m^{\mathcal{P}} &= \frac{1}{m^{\xi}}\left\|\sum_{i=1}^{m}\{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}}\boldsymbol{W}_{2,m}(m)\right\|_{(\Sigma^{\mathcal{P}})^{-1}} m^{\xi-\frac{1}{2}}\sup_{k>m}\left(\frac{m}{k}\right)^{\frac{1}{2}+\eta} \\
&\le \frac{1}{m^{\xi}}\left\|\sum_{i=1}^{m}\left\{\boldsymbol{Y}_{i,m}^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\right\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}}\boldsymbol{W}_{2,m}(m)\right\|_{(\Sigma^{\mathcal{P}})^{-1}} m^{\xi-\frac{1}{2}}.
\end{aligned}$$

Hence, by equivalence of norms on $\mathbb{R}^p$, (2.11) and the fact that $\xi < \frac{1}{2}$, $I_m^{\mathcal{P}}$ converges in probability to zero as $m \to \infty$. Next, since the norm in $(I_m^{\mathcal{P}})'$ is zero when $j = m$ and using the fact that $j - m < k - m < k$, we see that $(I_m^{\mathcal{P}})'$ is equal to

$$\frac{1}{\sqrt{m}} \sup_{k>m} \left(\frac{m}{k}\right)^{\frac{1}{2}+\eta} \max_{j \in]\!]m,k[\![} \left\| \sum_{i=m+1}^{j} \{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}} \boldsymbol{W}_{1,m}(j-m) \right\|_{(\Sigma^{\mathcal{P}})^{-1}}$$

$$\leq \frac{1}{m^{\frac{1}{2}-\xi}} \sup_{k>m} \left(\frac{m}{k}\right)^{\frac{1}{2}-\xi+\eta} \max_{j \in]\!]m,k[\![} \frac{1}{(j-m)^{\xi}} \left\| \sum_{i=m+1}^{j} \{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}} \boldsymbol{W}_{1,m}(j-m) \right\|_{(\Sigma^{\mathcal{P}})^{-1}}$$

$$\leq \frac{1}{m^{\frac{1}{2}-\xi}} \sup_{j>m} \frac{1}{(j-m)^{\xi}} \left\| \sum_{i=m+1}^{j} \{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}} \boldsymbol{W}_{1,m}(j-m) \right\|_{(\Sigma^{\mathcal{P}})^{-1}} \sup_{k>m} \left(\frac{m}{k}\right)^{\frac{1}{2}-\xi+\eta}.$$

The supremum over k above is less than 1 since $\xi < \frac{1}{2}$. The supremum over j is bounded in probability by (2.10) and equivalence of norms. Since $m^{-\frac{1}{2}+\xi} \to 0$, this proves that $(I_m^{\mathcal{P}})'$ converges to 0 in probability. □

Given p-dimensional independent standard Brownian motions $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ and $\eta > 0$, define the random function D_η by

$$D_\eta(s,t) = t^{-\frac{3}{2}-\eta} \left\| (t-s)\boldsymbol{W}_2(1) + t\boldsymbol{W}_1(s-1) - s\boldsymbol{W}_1(t-1) \right\|_{I_p}, \quad 1 \leq s \leq t < \infty. \tag{A.4}$$

Lemma A.2. *For any $\eta > 0$, the function D_η is almost surely bounded and uniformly continuous.*

Proof. Fix $\eta > 0$. For $i \in [\![1,p]\!]$ and $1 \leq s \leq t < \infty$, define $A^{[i]}(s,t)$ to be the i-th coordinate of

$$\boldsymbol{A}(s,t) = t^{-\frac{3}{2}-\eta} \{(t-s)\boldsymbol{W}_2(1) + t\boldsymbol{W}_1(s-1) - s\boldsymbol{W}_1(t-1)\}, \quad 1 \leq s \leq t < \infty,$$

and note that $D_\eta(s,t) = \|\boldsymbol{A}(s,t)\|_{I_p}$. From Lemma A.2 in Holmes and Kojadinovic (2021), we know that $|A^{[i]}|$ is almost surely bounded and uniformly continuous for each i. It follows that each $A^{[i]}$ is also bounded and uniformly continuous, almost surely. Now note that

$$|D_\eta(s,t) - D_\eta(s',t')| = \Big| \|\boldsymbol{A}(s,t)\|_{I_p} - \|\boldsymbol{A}(s',t')\|_{I_p} \Big| \leq \|\boldsymbol{A}(s,t) - \boldsymbol{A}(s',t')\|_{I_p}$$
$$= \sqrt{\frac{1}{p} \sum_{i=1}^{p} \{A^{[i]}(s,t) - A^{[i]}(s',t')\}^2}.$$

So D_η is almost surely bounded and uniformly continuous since each $A^{[i]}$ is. □

Lemma A.3. *For any $\eta > 0$,*

$$\sup_{k>m} \left(\frac{m}{k}\right)^{\frac{3}{2}+\eta} \bar{D}_m^{\mathcal{P}}(k) \rightsquigarrow \sup_{1 \leq s \leq t < \infty} D_\eta(s,t),$$

where $\bar{D}_m^{\mathcal{P}}$ is defined in (A.2) and D_η is defined in (A.4).

Proof. Fix $\eta > 0$. The random variable $\sup_{k>m}(m/k)^{\frac{3}{2}+\eta}\bar{D}_m^{\mathcal{P}}(k)$ is equal in distribution to

$$\frac{1}{\sqrt{m}}\sup_{k\geq m}\left(\frac{m}{k}\right)^{\frac{3}{2}+\eta}\max_{j\in[\![m,k]\!]}\left\|\frac{k}{m}\{\boldsymbol{W}_2(m)+\boldsymbol{W}_1(j-m)\}-\frac{j}{m}\{\boldsymbol{W}_2(m)+\boldsymbol{W}_1(k-m)\}\right\|_{I_p},$$

where we note that the norm is equal to zero when $j = k$, which has allowed us to include the cases $j = k$ and $k = m$ in the maximum and supremum, respectively. Using Brownian scaling, this is equal in distribution to

$$\sup_{k\geq m}\left(\frac{m}{k}\right)^{\frac{3}{2}+\eta}\max_{j\in[\![m,k]\!]}\left\|\frac{k}{m}\Big\{\boldsymbol{W}_2(1)+\boldsymbol{W}_1\Big(\frac{j}{m}-1\Big)\Big\}-\frac{j}{m}\Big\{\boldsymbol{W}_2(1)+\boldsymbol{W}_1\Big(\frac{k}{m}-1\Big)\Big\}\right\|_{I_p}.$$

In the above, j and k are integers. Letting $k = \lfloor mt \rfloor$ and $j = \lfloor ms \rfloor$ (where $s, t \in \mathbb{R}$ and $1 \leq s \leq t < \infty$), the above display becomes

$$\begin{aligned}&\sup_{t\geq 1}\left(\frac{m}{\lfloor mt\rfloor}\right)^{\frac{3}{2}+\eta}\\&\times\sup_{s\in[1,t]}\left\|\frac{\lfloor mt\rfloor}{m}\Big\{\boldsymbol{W}_2(1)+\boldsymbol{W}_1\Big(\frac{\lfloor ms\rfloor}{m}-1\Big)\Big\}-\frac{\lfloor ms\rfloor}{m}\Big\{\boldsymbol{W}_2(1)+\boldsymbol{W}_1\Big(\frac{\lfloor mt\rfloor}{m}-1\Big)\Big\}\right\|_{I_p}\\&=\sup_{1\leq s\leq t<\infty}D_\eta\left(\frac{\lfloor ms\rfloor}{m},\frac{\lfloor mt\rfloor}{m}\right).\end{aligned}$$

Using (A.3) with the supremum over (s,t), we have

$$\left|\sup_{1\leq s\leq t<\infty}D_\eta\left(\frac{\lfloor ms\rfloor}{m},\frac{\lfloor mt\rfloor}{m}\right)-\sup_{1\leq s\leq t<\infty}D_\eta(s,t)\right|\leq\sup_{1\leq s\leq t<\infty}\left|D_\eta\left(\frac{\lfloor ms\rfloor}{m},\frac{\lfloor mt\rfloor}{m}\right)-D_\eta(s,t)\right|,$$

which converges to zero almost surely as $m \to \infty$ since D_η is almost surely uniformly continuous and $t - 1/m < \frac{\lfloor mt \rfloor}{m} \leq t$. $\square$

For a $p \times p$ matrix A, denote the operator norm of A by

$$\|A\|_{\text{op}} = \inf\{c \geq 0 : \|A\boldsymbol{v}\|_2 \leq c\|\boldsymbol{v}\|_2,\ \text{for all } \boldsymbol{v} \in \mathbb{R}^p\}. \tag{A.5}$$

Lemma A.4. *Assume that Condition 2.4 holds and that $\Sigma_m^{\mathcal{P}} \xrightarrow{\mathbb{P}} \Sigma^{\mathcal{P}}$. Then, for any $\eta > 0$,*

$$\sup_{k>m}\left(\frac{m}{k}\right)^{\frac{3}{2}+\eta}|\tilde{D}_m^{\mathcal{P}}(k)-D_m^{\mathcal{P}}(k)|=o_{\mathbb{P}}(1),$$

where $\tilde{D}_m^{\mathcal{P}}$ is defined in (A.1) and $D_m^{\mathcal{P}}$ is defined in (2.4).

Proof. Fix $\eta > 0$. Applying (A.3) to the maximum over $j \in [\![m, k[\![$ and using the inequality $|a^{\frac{1}{2}} - b^{\frac{1}{2}}| \leq |a - b|^{\frac{1}{2}}$ for $a, b \geq 0$, we obtain that

$$\begin{aligned}&\sup_{k>m}\left(\frac{m}{k}\right)^{\frac{3}{2}+\eta}|\tilde{D}_m^{\mathcal{P}}(k)-D_m^{\mathcal{P}}(k)|\\&\quad\leq\sup_{k>m}\left(\frac{m}{k}\right)^{\frac{3}{2}+\eta}\max_{j\in[\![m,k[\![}\frac{j(k-j)}{m^{\frac{3}{2}}}\left|\|\bar{\boldsymbol{Y}}_{1:j}^{\mathcal{P}}-\bar{\boldsymbol{Y}}_{j+1:k}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}}-\|\bar{\boldsymbol{Y}}_{1:j}^{\mathcal{P}}-\bar{\boldsymbol{Y}}_{j+1:k}^{\mathcal{P}}\|_{(\Sigma_m^{\mathcal{P}})^{-1}}\right|\\&\quad\leq\sup_{k>m}\left(\frac{m}{k}\right)^{\frac{3}{2}+\eta}\max_{j\in[\![m,k[\![}\frac{j(k-j)}{p^{\frac{1}{2}}m^{\frac{3}{2}}}\left|\left(\bar{\boldsymbol{Y}}_{1:j}^{\mathcal{P}}-\bar{\boldsymbol{Y}}_{j+1:k}^{\mathcal{P}}\right)^\top\left((\Sigma^{\mathcal{P}})^{-1}-(\Sigma_m^{\mathcal{P}})^{-1}\right)\left(\bar{\boldsymbol{Y}}_{1:j}^{\mathcal{P}}-\bar{\boldsymbol{Y}}_{j+1:k}^{\mathcal{P}}\right)\right|^{\frac{1}{2}}.\end{aligned} \tag{A.6}$$

For $A \in \mathbb{R}^{p\times p}$ and $\boldsymbol{v} \in \mathbb{R}^p$, we have that $|\boldsymbol{v}^\top A\boldsymbol{v}| \le \|A\boldsymbol{v}\|_2\|\boldsymbol{v}\|_2$ by the Cauchy-Schwarz inequality. By (A.5), $\|A\boldsymbol{v}\|_2 \le \|A\|_{\mathrm{op}}\|\boldsymbol{v}\|_2$, so that $|\boldsymbol{v}^\top A\boldsymbol{v}|^{\frac{1}{2}} \le \|A\|_{\mathrm{op}}^{\frac{1}{2}}\|\boldsymbol{v}\|_2$. From (A.6), we therefore obtain

$$\sup_{k>m}\left(\frac{m}{k}\right)^{\frac{3}{2}+\eta}|\tilde{D}_m^{\mathcal{P}}(k) - D_m^{\mathcal{P}}(k)| \\ \le \frac{1}{p^{\frac{1}{2}}}\left\|(\Sigma^{\mathcal{P}})^{-1} - (\Sigma_m^{\mathcal{P}})^{-1}\right\|_{\mathrm{op}}^{\frac{1}{2}}\sup_{k>m}\left(\frac{m}{k}\right)^{\frac{3}{2}+\eta}\max_{j\in[\![m,k[\![}\frac{j(k-j)}{m^{\frac{3}{2}}}\left\|\bar{\boldsymbol{Y}}_{1:j}^{\mathcal{P}} - \bar{\boldsymbol{Y}}_{j+1:k}^{\mathcal{P}}\right\|_2.$$

It is well-known that the mapping that maps an invertible square matrix to its inverse is continuous. Since $\Sigma_m^{\mathcal{P}} \xrightarrow{\mathbb{P}} \Sigma^{\mathcal{P}}$ and Condition 2.2 holds, the continuous mapping theorem immediately implies that $(\Sigma_m^{\mathcal{P}})^{-1} \xrightarrow{\mathbb{P}} (\Sigma^{\mathcal{P}})^{-1}$, which in turn implies that $\|(\Sigma^{\mathcal{P}})^{-1} - (\Sigma_m^{\mathcal{P}})^{-1}\|_{\mathrm{op}} = o_{\mathbb{P}}(1)$ by equivalence of norms. Furthermore, from Lemmas A.1 and A.3, we have that $\sup_{k>m}(m/k)^{\frac{3}{2}+\eta}\tilde{D}_m^{\mathcal{P}}(k)$ converges weakly, which implies that it is bounded in probability. By equivalence of norms on $\mathbb{R}^p$, we immediately obtain that

$$\sup_{k>m}\left(\frac{m}{k}\right)^{\frac{3}{2}+\eta}\max_{j\in[\![m,k[\![}\frac{j(k-j)}{m^{\frac{3}{2}}}\left\|\bar{\boldsymbol{Y}}_{1:j}^{\mathcal{P}} - \bar{\boldsymbol{Y}}_{j+1:k}^{\mathcal{P}}\right\|_2 = O_{\mathbb{P}}(1)$$

and therefore the desired result. □

Proof of Theorem 2.5. From Lemmas A.1–A.4, we have that

$$\sup_{k>m}\left(\frac{m}{k}\right)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k) \rightsquigarrow \sup_{1\le s\le t<\infty} D_\eta(s,t),$$

where D_η is defined in (A.4). It remains to be shown that

$$\sup_{1\le s\le t<\infty} D_\eta(s,t) \stackrel{d}{=} \sup_{1\le s\le t<\infty} t^{-\frac{3}{2}-\eta}\|t\boldsymbol{W}(s) - s\boldsymbol{W}(t)\|_{I_p}. \tag{A.7}$$

Let $\boldsymbol{U}_p$ and $\boldsymbol{V}_p$ be two p-dimensional Gaussian processes defined, for any $1 \le s \le t$, by

$$\boldsymbol{U}_p(s,t) = (t-s)\boldsymbol{W}_2(1) + t\boldsymbol{W}_1(s-1) - s\boldsymbol{W}_1(t-1), \\ \boldsymbol{V}_p(s,t) = t\boldsymbol{W}(s) - s\boldsymbol{W}(t).$$

Since $\boldsymbol{W}_1$, $\boldsymbol{W}_2$, and $\boldsymbol{W}$ are p-dimensional standard Brownian motions, it follows that the coordinates of the Gaussian processes $\boldsymbol{U}_p$ and $\boldsymbol{V}_p$ are centered. Thus, to show that the random functions $\boldsymbol{U}_p$ and $\boldsymbol{V}_p$ are equal in distribution (which will immediately imply (A.7)), it suffices to establish equality of their covariance functions at any (s,t,s',t') with $1 \le s \le t$ and $1 \le s' \le t'$. On one hand, the covariance function of the Gaussian process $\boldsymbol{U}_p$ at (s,t,s',t') is

$$\begin{aligned}\mathbb{E}\{\boldsymbol{U}_p(s,t)\boldsymbol{U}_p(s',t')^\top\} &= \mathbb{E}\Big[\big((t-s)\boldsymbol{W}_2(1) + t\boldsymbol{W}_1(s-1) - s\boldsymbol{W}_1(t-1)\big) \\ &\quad\times\big((t'-s')\boldsymbol{W}_2(1) + t'\boldsymbol{W}_1(s'-1) - s'\boldsymbol{W}_1(t'-1)\big)^\top\Big] \\ &= \big[(t-s)(t'-s') + t't\{\min(s,s')-1\} - s't\{\min(s,t')-1\} \\ &\quad - t's\{\min(t,s')-1\} + s's\{\min(t,t')-1\}\big]I_p \\ &= \{t't\min(s,s') - s't\min(s,t') - t's\min(t,s') + s's\min(t,t')\}I_p,\end{aligned}$$

while, on the other hand, the covariance function of the Gaussian process $\boldsymbol{V}_p$ at (s,t,s',t') is

$$\begin{aligned}\mathbb{E}\{\boldsymbol{V}(s,t)\boldsymbol{V}(s',t')^\top\} &= \mathbb{E}\big[\big(t\boldsymbol{W}(s) - s\boldsymbol{W}(t)\big)\big(t'\boldsymbol{W}(s') - s'\boldsymbol{W}(t')\big)^\top\big] \\ &= \{tt'\min(s,s') - s't\min(s,t') - st'\min(t,s') + ss'\min(t,t')\}I_p,\end{aligned}$$

which concludes the proof. □

Appendix B: Proofs of Propositions 2.7, 2.9 and 2.10

Proof of Proposition 2.7. Let $\mathcal{P} \in (\mathbb{R}^d)^p$ be such that Condition 2.2 holds. Since $\boldsymbol{Y}_i^{\mathcal{P}} = \big(\mathbf{1}(\boldsymbol{X}_i \le \boldsymbol{x}_1), \dots, \mathbf{1}(\boldsymbol{X}_i \le \boldsymbol{x}_p)\big)$, it is immediate that $\alpha_r^{\boldsymbol{Y}} \le \alpha_r^{\boldsymbol{X}} = O(r^{-a})$ as $r \to \infty$. Furthermore, as all the components of the $\boldsymbol{Y}_i^{\mathcal{P}}$ are bounded in absolute value by 1, from Theorem 4 of Kuelbs and Philipp (1980), we can redefine the sequence $(\boldsymbol{Y}_i^{\mathcal{P}})_{i\in\mathbb{N}}$ on a new probability space together with a p-dimensional standard Brownian motion $\boldsymbol{W}$ such that, almost surely,

$$\left\| \sum_{i=1}^m \{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}}\boldsymbol{W}(m) \right\|_2 = O(m^{\frac{1}{2}-\lambda}),$$

for some $\lambda \in (0, \frac{1}{2})$ that depends on a, p and $\mathcal{P}$. Let $\xi \in (\frac{1}{2} - \lambda, \frac{1}{2})$. Then, almost surely,

$$\lim_{m\to\infty} \frac{1}{m^\xi} \left\| \sum_{i=1}^m \{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}}\boldsymbol{W}(m) \right\|_2 = 0. \tag{B.1}$$

Let $\boldsymbol{W}'$ be another p-dimensional standard Brownian motion independent of $\boldsymbol{W}$ and define $\boldsymbol{W}_2^{(m)}$ as

$$\boldsymbol{W}_2^{(m)}(s) = \begin{cases} \boldsymbol{W}(s) \text{ if } s \in [0,m], \\ \boldsymbol{W}'(s-m) + \boldsymbol{W}(m) \text{ otherwise,} \end{cases}$$

and $\boldsymbol{W}_1^{(m)}$ as $\boldsymbol{W}_1^{(m)}(s) = \boldsymbol{W}(m+s) - \boldsymbol{W}(m)$, $s \ge 0$. Then, for each $m \in \mathbb{N}$, $\boldsymbol{W}_1^{(m)}$ and $\boldsymbol{W}_2^{(m)}$ are independent p-dimensional standard Brownian motions.

For each $m \ge 0$, let

$$V_m = \sup_{k>m} \frac{1}{(k-m)^\xi} \left\| \sum_{i=m+1}^k \{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}}\boldsymbol{W}_1^{(m)}(k-m) \right\|_2$$

and note that the sequence $(V_m)_{m\ge0}$ consists of identically distributed random variables. Therefore, to show that (2.10) holds, it is sufficient to show that $V_0 = O_{\mathbb{P}}(1)$. From the previous definition, V_0 can be rewritten as

$$V_0 = \sup_{k>0} \frac{1}{k^\xi} \left\| \sum_{i=1}^k \{\boldsymbol{Y}_i^{\mathcal{P}} - \mathbb{E}(\boldsymbol{Y}_1^{\mathcal{P}})\} - (\Sigma^{\mathcal{P}})^{\frac{1}{2}}\boldsymbol{W}(k) \right\|_2,$$

and, from (B.1), it is the supremum of an almost surely convergent sequence. It follows that V_0 is an almost surely finite random variable, so $V_0 = O_{\mathbb{P}}(1)$ and therefore (2.10) holds. Finally, (B.1) and the definition of $\boldsymbol{W}_2^{(m)}$ immediately imply (2.11). □

Proof of Proposition 2.9. It is immediate that

$$\sqrt{p} \sup_{1\le s\le t<\infty} t^{-\frac{3}{2}} \|t\boldsymbol{W}(s) - s\boldsymbol{W}(t)\|_{I_p} \ge \sup_{1\le s\le t<\infty} t^{-\frac{3}{2}} |tW^{[1]}(s) - sW^{[1]}(t)|,$$

where $W^{[1]}$ is the first component of the p-dimensional Brownian motion $\boldsymbol{W}$. Hence, for any fixed $M > 0$,

$$\mathbb{P}\Big(\sup_{1\le s\le t<\infty} t^{-\frac{3}{2}} \|t\boldsymbol{W}(s) - s\boldsymbol{W}(t)\|_{I_p} \ge M\Big) \\ \ge \mathbb{P}\Big(p^{-1/2} \sup_{1\le s\le t<\infty} t^{-\frac{3}{2}} |tW^{[1]}(s) - sW^{[1]}(t)| \ge M\Big) = 1,$$

where the last equality is a consequence of Proposition 3.4 of Holmes and Kojadinovic (2021). □

Proof of Proposition 2.10. Fix $\eta > 0$ and $p \in \mathbb{N}$. Note first that $\boldsymbol{W}$ is a p-variate continuous Gaussian process, which implies that $\boldsymbol{W} \in C([0,\infty), \mathbb{R}^p)$ almost surely. For $v \ge 2$, let $f_v : C([0,v], \mathbb{R}^p) \to [0,\infty)$ be defined by

$$f_v(\boldsymbol{w}) = \sup_{1\le s\le t\le v} t^{-\frac{3}{2}-\eta} \|t\boldsymbol{w}(s) - s\boldsymbol{w}(t)\|_{I_p}, \qquad \boldsymbol{w} \in C([0,v], \mathbb{R}^p),$$

and, similarly, let $f : C([0,\infty), \mathbb{R}^p) \to [0,\infty)$ be defined by

$$f(\boldsymbol{w}) = \sup_{1\le s\le t<\infty} t^{-\frac{3}{2}-\eta} \|t\boldsymbol{w}(s) - s\boldsymbol{w}(t)\|_{I_p}, \qquad \boldsymbol{w} \in C([0,\infty), \mathbb{R}^p).$$

Notice that $\mathcal{L}_{p,\eta} = f(\boldsymbol{W})$ and that $f_v(\boldsymbol{w}) \le f(\boldsymbol{w})$ for every $v \ge 2$.

Next, for any $v \ge 2$, we equip $C([0,v], \mathbb{R}^p)$ with the uniform distance, which is then a separable (and locally convex) metric space. Furthermore, as we shall verify below, f_v is continuous and convex for any $v \ge 2$. We can then apply Theorem 7.1 of Davydov and Lifshits (1984) to obtain that, for any $v \ge 2$, the distribution of $f_v(\boldsymbol{W})$ is concentrated on $[0,\infty)$ and absolutely continuous on $(0,\infty)$. In addition, some thought reveals that, for any $v \ge 2$,

$$\mathbb{P}(f_v(\boldsymbol{W}) = 0) \le \mathbb{P}\big(f_2(\boldsymbol{W}) = 0\big) \le \mathbb{P}\big(\|2\boldsymbol{W}(1) - 1\boldsymbol{W}(2)\|_{I_p} = 0\big) = 0$$

since $2\boldsymbol{W}(1) - 1\boldsymbol{W}(2)$ is a centered multivariate normal random vector with covariance matrix $2I_p$. Hence, for any $v \ge 2$, the distribution of $f_v(\boldsymbol{W})$ has no atom at 0 and is therefore absolutely continuous.

Proof of the continuity of f_v: To show (uniform) continuity on $C([0,v], \mathbb{R}^p)$ (equipped with the uniform distance), let $\varepsilon > 0$ be given and let $\delta = \varepsilon/3$. Let $\boldsymbol{w}, \boldsymbol{w}' \in C([0,v], \mathbb{R}^p)$ be such that $\sup_{0\le t\le v} \|\boldsymbol{w}(t) - \boldsymbol{w}'(t)\|_{I_p} < \delta$. Then,

$$\begin{aligned} f_v(\boldsymbol{w}') &= \sup_{1\le s\le t\le v} t^{-\frac{3}{2}-\eta} \|t\boldsymbol{w}'(s) - s\boldsymbol{w}'(t)\|_{I_p} \\ &= \sup_{1\le s\le t\le v} t^{-\frac{3}{2}-\eta} \|\{t\boldsymbol{w}(s) - s\boldsymbol{w}(t)\} + t\{\boldsymbol{w}'(s) - \boldsymbol{w}(s)\} - s\{\boldsymbol{w}'(t) - \boldsymbol{w}(t)\}\|_{I_p} \\ &\le \sup_{1\le s\le t\le v} t^{-\frac{3}{2}-\eta} \Big\{ \|t\boldsymbol{w}(s) - s\boldsymbol{w}(t)\|_{I_p} + t\|\boldsymbol{w}'(s) - \boldsymbol{w}(s)\|_{I_p} + s\|\boldsymbol{w}'(t) - \boldsymbol{w}(t)\|_{I_p} \Big\} \\ &\le \sup_{1\le s\le t\le v} t^{-\frac{3}{2}-\eta} \|t\boldsymbol{w}(s) - s\boldsymbol{w}(t)\|_{I_p} + 2\delta. \end{aligned}$$

This shows that $f_v(\boldsymbol{w}') - f_v(\boldsymbol{w}) \le 2\delta < \varepsilon$. Similarly (or just by symmetry of the above argument), $f_v(\boldsymbol{w}) - f_v(\boldsymbol{w}') < \varepsilon$. Thus, $|f_v(\boldsymbol{w}) - f_v(\boldsymbol{w}')| < \varepsilon$ as required.

Proof of the convexity of f_v: To show convexity, let $\lambda \in (0,1)$ and $\boldsymbol{w}, \boldsymbol{w}' \in C([0,v], \mathbb{R}^p)$. Then $\lambda \boldsymbol{w} + (1-\lambda)\boldsymbol{w}' \in C([0,v], \mathbb{R}^p)$ and

$$\begin{aligned} f_v(\lambda \boldsymbol{w} + (1-\lambda)\boldsymbol{w}') &= \sup_{1\le s\le t\le v} t^{-\frac{3}{2}-\eta} \left\| t\{\lambda \boldsymbol{w}(s) + (1-\lambda)\boldsymbol{w}'(s)\} - s\{\lambda \boldsymbol{w}(t) + (1-\lambda)\boldsymbol{w}'(t)\} \right\|_{I_p} \\ &= \sup_{1\le s\le t\le v} t^{-\frac{3}{2}-\eta} \left\| \lambda\{t\boldsymbol{w}(s) - s\boldsymbol{w}(t)\} + (1-\lambda)\{t\boldsymbol{w}'(s) - s\boldsymbol{w}'(t)\} \right\|_{I_p} \\ &\le \lambda f_v(\boldsymbol{w}) + (1-\lambda) f_v(\boldsymbol{w}'), \end{aligned}$$

as required.

To complete the proof of the proposition, it suffices to fix $r, \varepsilon > 0$ and show that $\mathbb{P}(f(\boldsymbol{W}) = r) < \varepsilon$. Notice first that, for any $u \ge 1$, almost surely,

$$\sup_{(s,t):1\le s\le t, u\le t<\infty} t^{-\frac{3}{2}-\eta} \|t\boldsymbol{W}(s) - s\boldsymbol{W}(t)\|_{I_p} \le u^{-\frac{\eta}{2}} \sup_{1\le s\le t<\infty} t^{-\frac{3}{2}-\frac{\eta}{2}} \|t\boldsymbol{W}(s) - s\boldsymbol{W}(t)\|_{I_p}.$$

Since according to Theorem 2.5, $\mathcal{L}_{p,\frac{\eta}{2}}$ is almost surely finite, we see that

$$\sup_{(s,t):1\le s\le t, u\le t<\infty} t^{-\frac{3}{2}-\eta} \|t\boldsymbol{W}(s) - s\boldsymbol{W}(t)\|_{I_p} \xrightarrow{as} 0, \qquad \text{as } u \to \infty.$$

Hence, there exists $u_0 < \infty$ (depending on r, ε) such that

$$\mathbb{P}\left(\sup_{(s,t):1\le s\le t, u_0\le t<\infty} t^{-\frac{3}{2}-\eta} \|t\boldsymbol{W}(s) - s\boldsymbol{W}(t)\|_{I_p} \ge r/2 \right) < \varepsilon. \tag{B.2}$$

Now, if $f(\boldsymbol{W}) = r$, then either $f_{u_0}(\boldsymbol{W}) = r$ or the supremum (r) in the definition of f is attained for some s, t with $t > u_0$. Absolute continuity of $f_{u_0}(\boldsymbol{W})$ shows that the former has probability 0, while (B.2) shows that the latter has probability at most ε. □

Appendix C: Proof of Proposition 2.11

Proof of Proposition 2.11. Let A be a $p \times p$ symmetric positive-definite matrix. Then, there exits a diagonal matrix Δ with $\lambda_1 > 0, \dots, \lambda_p > 0$ on its diagonal and a $p \times p$ orthogonal matrix P such that the columns of P are the eigenvectors $\boldsymbol{e}_1, \dots, \boldsymbol{e}_p \in \mathbb{R}^p$ of A with corresponding eigenvalues $\lambda_1, \dots, \lambda_p$ and $A = P\Delta P^\top$. Starting from the definition of $\|\cdot\|_{A^{-1}}$ given below (2.5), for any $\boldsymbol{v} \in \mathbb{R}^p$,

$$\begin{aligned} \sqrt{p}\|\boldsymbol{v}\|_{A^{-1}} &= |\boldsymbol{v}^\top A^{-1} \boldsymbol{v}|^{\frac{1}{2}} = |\boldsymbol{v}^\top (P\Delta P^\top)^{-1} \boldsymbol{v}|^{\frac{1}{2}} = |\boldsymbol{v}^\top P \Delta^{-1} P^\top \boldsymbol{v}|^{\frac{1}{2}} \\ &\ge \min_{i\in[\![1,p]\!]} \lambda_i^{-\frac{1}{2}} \times \|P^\top \boldsymbol{v}\|_2 = \min_{i\in[\![1,p]\!]} \lambda_i^{-\frac{1}{2}} \times \sqrt{(P^\top \boldsymbol{v}) \cdot (P^\top \boldsymbol{v})} \\ &= \min_{i\in[\![1,p]\!]} \lambda_i^{-\frac{1}{2}} \times \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}} = \min_{i\in[\![1,p]\!]} \lambda_i^{-\frac{1}{2}} \times \|\boldsymbol{v}\|_2 \\ &\ge \min_{i\in[\![1,p]\!]} \lambda_i^{-\frac{1}{2}} \times \|\boldsymbol{v}\|_\infty, \end{aligned}$$

where we have used the fact that orthogonal matrices preserve the dot product.

Let ψ be the map from the set $\mathcal{S}_p$ of $p \times p$ symmetric positive-definite matrices to $(0, \infty)$ such that, for any $A \in \mathcal{S}_p$, $\psi(A) = \min_{i \in [\![1,p]\!]} \lambda_i^{-\frac{1}{2}}$. Since eigenvalue decomposition is a continuous operation and since the minimum is a continuous function, the map ψ is continuous. Hence, $\Sigma_m^{\mathcal{P}} \xrightarrow{\mathbb{P}} \Sigma^{\mathcal{P}}$ and the continuous mapping theorem imply that $\psi(\Sigma_m^{\mathcal{P}}) \xrightarrow{\mathbb{P}} \psi(\Sigma^{\mathcal{P}}) > 0$.

From the previous derivations and the assumptions of the proposition, we thus have that, for all $m \in \mathbb{N}$, $\|\cdot\|_{(\Sigma_m^{\mathcal{P}})^{-1}} \geq p^{-\frac{1}{2}} \psi(\Sigma_m^{\mathcal{P}}) \|\cdot\|_\infty$ almost surely. Therefore, for any $m \in \mathbb{N}$ and $k \geq m+1$, almost surely,

$$
\begin{aligned}
D_m^{\mathcal{P}}(k) &= \max_{j \in [\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} \|\bar{\boldsymbol{Y}}_{1:j}^{\mathcal{P}} - \bar{\boldsymbol{Y}}_{j+1:k}^{\mathcal{P}}\|_{(\Sigma_m^{\mathcal{P}})^{-1}} \\
&\geq p^{-\frac{1}{2}} \psi(\Sigma_m^{\mathcal{P}}) \max_{j \in [\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} \|\bar{\boldsymbol{Y}}_{1:j}^{\mathcal{P}} - \bar{\boldsymbol{Y}}_{j+1:k}^{\mathcal{P}}\|_\infty \geq p^{-\frac{1}{2}} \psi(\Sigma_m^{\mathcal{P}}) E_m^{\boldsymbol{x}_\ell}(k),
\end{aligned}
$$

which immediately implies that $\sup_{k>m} (m/k)^{\frac{3}{2}+\eta} D_m^{\mathcal{P}}(k) \xrightarrow{\mathbb{P}} \infty$ since $\psi(\Sigma_m^{\mathcal{P}}) \xrightarrow{\mathbb{P}} \psi(\Sigma^{\mathcal{P}}) > 0$ and $\sup_{k>m} (m/k)^{\frac{3}{2}+\eta} E_m^{\boldsymbol{x}_\ell}(k) \xrightarrow{\mathbb{P}} \infty$. □

Appendix D: Proofs of Propositions 3.2 and 3.4

Proof of Proposition 3.2. Let

$$R(n, \boldsymbol{x}) = \sum_{i=1}^n \{\mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{x}) - F(\boldsymbol{x})\}, \qquad n \in \mathbb{N}, \boldsymbol{x} \in \mathbb{R}^d.$$

From Theorem 3.1 part 2(b) in Dedecker, Merlevède and Rio (2014), without changing its distribution, the empirical process R can be redefined on a richer probability space on which there exists a *Kiefer process*, that is, a two-parameter centered continuous Gaussian process K with covariance function

$$
\begin{aligned}
\Gamma(s,t,\boldsymbol{x},\boldsymbol{y}) = \min(s,t) \Big[&\mathrm{Cov}\{\mathbf{1}(\boldsymbol{X}_1 \leq \boldsymbol{x}), \mathbf{1}(\boldsymbol{X}_1 \leq \boldsymbol{y})\} + \sum_{i=2}^\infty \mathrm{Cov}\{\mathbf{1}(\boldsymbol{X}_1 \leq \boldsymbol{x}), \mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{y})\} \\
&+ \sum_{i=2}^\infty \mathrm{Cov}\{\mathbf{1}(\boldsymbol{X}_i \leq \boldsymbol{x}), \mathbf{1}(\boldsymbol{X}_1 \leq \boldsymbol{y})\} \Big], \qquad s,t \in [0,\infty), \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d,
\end{aligned}
$$

and a random variable $C > 0$ such that, almost surely (a.s.),

$$\sup_{t \in [0,1]} \sup_{\boldsymbol{x} \in \mathbb{R}^d} |R(\lfloor nt \rfloor, \boldsymbol{x}) - K(\lfloor nt \rfloor, \boldsymbol{x})| \leq C n^{\frac{1}{2}-\lambda}, \qquad \text{for all } n \in \mathbb{N} \tag{D.1}$$

for some $\lambda \in (0, 1/2)$ only depending on d and a. Let $i \in [\![1, p]\!]$. Then,

$$\begin{aligned}
&\sup_{k>m} k^{-\frac{1}{2}} \max_{j\in[\![1,k]\!]} j|F_{1:j}(\boldsymbol{\mathcal{X}}_{i,m}) - F(\boldsymbol{\mathcal{X}}_{i,m}) - F_{1:j}(\boldsymbol{x}_i) + F(\boldsymbol{x}_i)| \\
&\quad = \sup_{k>m} k^{-\frac{1}{2}} \max_{j\in[\![1,k]\!]} |R(j, \boldsymbol{\mathcal{X}}_{i,m}) - R(j, \boldsymbol{x}_i)| \\
&\quad \le \sup_{k>m} k^{-\frac{1}{2}} \max_{j\in[\![1,k]\!]} |R(j, \boldsymbol{\mathcal{X}}_{i,m}) - K(j, \boldsymbol{\mathcal{X}}_{i,m})| + \sup_{k>m} k^{-\frac{1}{2}} \max_{j\in[\![1,k]\!]} |R(j, \boldsymbol{x}_i) - K(j, \boldsymbol{x}_i)| \\
&\qquad + \sup_{k>m} k^{-\frac{1}{2}} \max_{j\in[\![1,k]\!]} |K(j, \boldsymbol{\mathcal{X}}_{i,m}) - K(j, \boldsymbol{x}_i)|.
\end{aligned}$$

From (D.1), with probability one, the first two terms on the right-hand side of the last display are smaller than

$$2 \sup_{k>m} k^{-\frac{1}{2}} \sup_{t\in[0,1]} \sup_{\boldsymbol{x}\in\mathbb{R}^d} |R(\lfloor kt \rfloor, \boldsymbol{x}) - K(\lfloor kt \rfloor, \boldsymbol{x})| \le 2 \sup_{k>m} k^{-\frac{1}{2}} C k^{\frac{1}{2}-\lambda} \le 2Cm^{-\lambda} \overset{as}{\to} 0.$$

The third term is smaller than

$$I_m = \sup_{k>m} k^{-\frac{1}{2}} \max_{j\in[\![1,k]\!]} \sup_{\substack{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^d \\ \|\boldsymbol{x}-\boldsymbol{y}\|_\infty \le \Delta_m}} |K(j, \boldsymbol{x}) - K(j, \boldsymbol{y})|, \tag{D.2}$$

where $\Delta_m = \|\boldsymbol{\mathcal{X}}_{i,m} - \boldsymbol{x}_i\|_\infty$. To complete the proof, it thus suffices to show that $I_m \overset{\mathbb{P}}{\to} 0$.

From Theorem 3.1 part 2(a) in Dedecker, Merlevède and Rio (2014), we know that the sample paths of the Gaussian process K are almost surely uniformly continuous with respect to the pseudo-metric ρ on $[0, \infty) \times \mathbb{R}^d$ defined by

$$\rho\big((s, \boldsymbol{x}), (t, \boldsymbol{y})\big) = |s - t| + \sum_{\ell=1}^{d} |F^{[\ell]}(x^{[\ell]}) - F^{[\ell]}(y^{[\ell]})|, \qquad s, t \in [0, \infty), \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d.$$

We will use this fact to show that $I_m \overset{\mathbb{P}}{\to} 0$. To this end, define $\Delta_m^* = \psi(\Delta_m)$, where

$$\psi(\delta) = \sup_{\substack{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^d \\ \|\boldsymbol{x}-\boldsymbol{y}\|_\infty \le \delta}} \sum_{\ell=1}^{d} |F^{[\ell]}(x^{[\ell]}) - F^{[\ell]}(y^{[\ell]})|, \qquad \delta \ge 0.$$

Since $\|\boldsymbol{x} - \boldsymbol{y}\|_\infty \le \Delta_m$ implies that $\rho\big((t, \boldsymbol{x}), (t, \boldsymbol{y})\big) \le \Delta_m^*$, we have that

$$I_m \le \sup_{k>m} k^{-\frac{1}{2}} \sup_{\substack{s,t\in[0,k] \\ s=t}} \sup_{\substack{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^d \\ \|\boldsymbol{x}-\boldsymbol{y}\|_\infty \le \Delta_m}} |K(s, \boldsymbol{x}) - K(t, \boldsymbol{y})| \le \sup_{k>m} k^{-\frac{1}{2}} J_m \le m^{-\frac{1}{2}} J_m,$$

where $J_m = \phi(\Delta_m^*)$ with

$$\phi(\delta) = \sup_{\substack{s,t\in[0,\infty),\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^d \\ \rho((s,\boldsymbol{x}),(t,\boldsymbol{y}))\le\delta}} |K(s, \boldsymbol{x}) - K(t, \boldsymbol{y})|, \qquad \delta \ge 0.$$

Let $\varepsilon > 0$. By almost sure uniform continuity of the sample paths of the process K, there exists $\delta_1 = \delta_1(\varepsilon) > 0$ such that, for all $0 \leq \delta \leq \delta_1$, $\phi(\delta) < \varepsilon$ almost surely. Since the d univariate margins $F^{[1]}, \dots, F^{[d]}$ of F are (uniformly) continuous, there exists $\delta_0 = \delta_0(\delta_1) > 0$ such that, for all $0 \leq \delta \leq \delta_0$, $\psi(\delta) < \delta_1$. Therefore

$$\begin{aligned}\mathbb{P}(J_m > \varepsilon) = \mathbb{P}\big(\phi(\Delta_m^*) > \varepsilon\big) &\leq \mathbb{P}\big(\phi(\Delta_m^*) > \varepsilon, \Delta_m^* \leq \delta_1\big) + \mathbb{P}(\Delta_m^* > \delta_1)\\ &= \mathbb{P}(\Delta_m^* > \delta_1)\\ &= \mathbb{P}\big(\psi(\Delta_m) > \delta_1\big)\\ &\leq \mathbb{P}\big(\psi(\Delta_m) > \delta_1, \Delta_m \leq \delta_0\big) + \mathbb{P}(\Delta_m > \delta_0)\\ &= \mathbb{P}(\Delta_m > \delta_0).\end{aligned}$$

Since $\Delta_m = \|\boldsymbol{\mathcal{X}}_{i,m} - \boldsymbol{x}_i\|_\infty \xrightarrow{\mathbb{P}} 0$ by assumption, this shows that $J_m \xrightarrow{\mathbb{P}} 0$ which completes the proof since $I_m \leq m^{-\frac{1}{2}} J_m$. □

Proof of Proposition 3.4. The second claim is immediate from the first claim and Theorem 2.5, so we need only prove the first claim.

First, we assert that, for any $\ell \in [\![1, p]\!]$,

$$\sup_{k>m} k^{-\frac{1}{2}} \max_{1\leq i<j\leq k} (j-i+1)|F_{i:j}(\boldsymbol{\mathcal{X}}_{\ell,m}) - F(\boldsymbol{\mathcal{X}}_{\ell,m}) - F_{i:j}(\boldsymbol{x}_\ell) + F(\boldsymbol{x}_\ell)| = o_{\mathbb{P}}(1). \tag{D.3}$$

Indeed, the left-hand side of (D.3) is equal to

$$\begin{aligned}&\sup_{k>m} k^{-\frac{1}{2}} \max_{1\leq i<j\leq k} \Big| \sum_{r=i}^{j} \{1(\boldsymbol{X}_r \leq \boldsymbol{\mathcal{X}}_{\ell,m}) - F(\boldsymbol{\mathcal{X}}_{\ell,m}) - 1(\boldsymbol{X}_r \leq \boldsymbol{x}_\ell) + F(\boldsymbol{x}_\ell)\} \Big|\\ &= \sup_{k>m} k^{-\frac{1}{2}} \max_{1\leq i<j\leq k} \Big| \sum_{r=1}^{j} \{1(\boldsymbol{X}_r \leq \boldsymbol{\mathcal{X}}_{\ell,m}) - F(\boldsymbol{\mathcal{X}}_{\ell,m}) - 1(\boldsymbol{X}_r \leq \boldsymbol{x}_\ell) + F(\boldsymbol{x}_\ell)\}\\ &\quad - \sum_{r=1}^{i-1} \{1(\boldsymbol{X}_r \leq \boldsymbol{\mathcal{X}}_{\ell,m}) - F(\boldsymbol{\mathcal{X}}_{\ell,m}) - 1(\boldsymbol{X}_r \leq \boldsymbol{x}_\ell) + F(\boldsymbol{x}_\ell)\} \Big|\\ &\leq 2 \sup_{k>m} k^{-\frac{1}{2}} \max_{j\in[\![1,k]\!]} j|F_{1:j}(\boldsymbol{\mathcal{X}}_{\ell,m}) - F(\boldsymbol{\mathcal{X}}_{\ell,m}) - F_{1:j}(\boldsymbol{x}_\ell) + F(\boldsymbol{x}_\ell)|,\end{aligned}$$

so the assertion holds by (3.2). Using the fact that all norms on $\mathbb{R}^p$ are equivalent, (D.3) implies that

$$\sup_{k>m} k^{-\frac{1}{2}} \max_{1\leq i<j\leq k} (j-i+1)\|\boldsymbol{F}_{i:j}^{\mathcal{P}_m} - \boldsymbol{F}^{\mathcal{P}_m} - \boldsymbol{F}_{i:j}^{\mathcal{P}} - \boldsymbol{F}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}} = o_{\mathbb{P}}(1), \tag{D.4}$$

where $\boldsymbol{F}^{\mathcal{P}_m} = \big(F(\boldsymbol{\mathcal{X}}_{1,m}), \dots, F(\boldsymbol{\mathcal{X}}_{p,m})\big)$ and $\boldsymbol{F}^{\mathcal{P}} = \big(F(\boldsymbol{x}_1), \dots, F(\boldsymbol{x}_p)\big)$ (similarly for $\boldsymbol{F}_{i:j}^{\cdot}$).

Recall the definition of $\tilde{D}_m^{\mathcal{P}}$ in (A.1) which can be rewritten as

$$\tilde{D}_m^{\mathcal{P}}(k) = \max_{j\in[\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} \|\boldsymbol{F}_{1:j}^{\mathcal{P}} - \boldsymbol{F}_{j+1:k}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}}, \qquad k \geq m+1,$$

and define the unobservable detector $\tilde{D}_m^{\mathcal{P}_m}$ by changing the norm in the definition of $D_m^{\mathcal{P}_m}$ as

$$\tilde{D}_m^{\mathcal{P}_m}(k) = \max_{j\in[\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} \|\boldsymbol{F}_{1:j}^{\mathcal{P}_m} - \boldsymbol{F}_{j+1:k}^{\mathcal{P}_m}\|_{(\Sigma^{\mathcal{P}})^{-1}}, \qquad k \geq m+1.$$

Using the reverse triangle inequality for the maximum norm and the norm $\|\cdot\|_{(\Sigma^{\mathcal{P}})^{-1}}$, we then obtain that, for any $k \geq m+1$,

$$\begin{aligned}
&|\tilde{D}_m^{\mathcal{P}_m}(k) - \tilde{D}_m^{\mathcal{P}}(k)| \\
&\leq \max_{j\in[\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} \big| \|\boldsymbol{F}_{1:j}^{\mathcal{P}_m} - \boldsymbol{F}^{\mathcal{P}_m} - \boldsymbol{F}_{j+1:k}^{\mathcal{P}_m} + \boldsymbol{F}^{\mathcal{P}_m}\|_{(\Sigma^{\mathcal{P}})^{-1}} - \|\boldsymbol{F}_{1:j}^{\mathcal{P}} - \boldsymbol{F}^{\mathcal{P}} - \boldsymbol{F}_{j+1:k}^{\mathcal{P}} + \boldsymbol{F}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}} \big| \\
&\leq \max_{j\in[\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} \|\boldsymbol{F}_{1:j}^{\mathcal{P}_m} - \boldsymbol{F}^{\mathcal{P}_m} - \boldsymbol{F}_{1:j}^{\mathcal{P}} + \boldsymbol{F}^{\mathcal{P}} - \boldsymbol{F}_{j+1:k}^{\mathcal{P}_m} + \boldsymbol{F}^{\mathcal{P}_m} + \boldsymbol{F}_{j+1:k}^{\mathcal{P}} - \boldsymbol{F}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}} \\
&\leq \max_{j\in[\![m,k[\![} \frac{j(k-j)}{m^{\frac{3}{2}}} \Big\{ \|\boldsymbol{F}_{1:j}^{\mathcal{P}_m} - \boldsymbol{F}^{\mathcal{P}_m} - \boldsymbol{F}_{1:j}^{\mathcal{P}} + \boldsymbol{F}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}} + \|\boldsymbol{F}_{j+1:k}^{\mathcal{P}_m} - \boldsymbol{F}^{\mathcal{P}_m} - \boldsymbol{F}_{j+1:k}^{\mathcal{P}} + \boldsymbol{F}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}} \Big\}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sup_{k>m} \left(\frac{m}{k}\right)^{\frac{3}{2}+\eta} |\tilde{D}_m^{\mathcal{P}_m}(k) - \tilde{D}_m^{\mathcal{P}}(k)| \\
&\leq \sup_{k>m} \left(\frac{m}{k}\right)^{\eta} k^{-\frac{3}{2}} \max_{j\in[\![m,k[\![} j(k-j) \|\boldsymbol{F}_{1:j}^{\mathcal{P}_m} - \boldsymbol{F}^{\mathcal{P}_m} - \boldsymbol{F}_{1:j}^{\mathcal{P}} + \boldsymbol{F}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}} \\
&\quad + \sup_{k>m} \left(\frac{m}{k}\right)^{\eta} k^{-\frac{3}{2}} \max_{j\in[\![m,k[\![} j(k-j) \|\boldsymbol{F}_{j+1:k}^{\mathcal{P}_m} - \boldsymbol{F}^{\mathcal{P}_m} - \boldsymbol{F}_{j+1:k}^{\mathcal{P}} + \boldsymbol{F}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}}.
\end{aligned}$$

We claim that both terms on the right hand side converge to 0 in probability. For example, since $j \leq k$ and $k > m$, the second term is at most

$$\begin{aligned}
&\sup_{k>m} k^{-\frac{1}{2}} \max_{j\in[\![m,k[\![} (k-j) \|\boldsymbol{F}_{j+1:k}^{\mathcal{P}_m} - \boldsymbol{F}^{\mathcal{P}_m} - \boldsymbol{F}_{j+1:k}^{\mathcal{P}} + \boldsymbol{F}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}} \\
&\leq \sup_{k>m} k^{-\frac{1}{2}} \max_{1\leq i<j\leq k} (j-i+1) \|\boldsymbol{F}_{i:j}^{\mathcal{P}_m} - \boldsymbol{F}^{\mathcal{P}_m} - \boldsymbol{F}_{i:j}^{\mathcal{P}} + \boldsymbol{F}^{\mathcal{P}}\|_{(\Sigma^{\mathcal{P}})^{-1}},
\end{aligned}$$

which converges to 0 in probability by (D.4). The first term is similar. Hence,

$$\sup_{k>m} \left(\frac{m}{k}\right)^{\frac{3}{2}+\eta} |\tilde{D}_m^{\mathcal{P}_m}(k) - \tilde{D}_m^{\mathcal{P}}(k)| = o_{\mathbb{P}}(1). \tag{D.5}$$

From Lemma A.4, we have that $\sup_{k>m}(m/k)^{\frac{3}{2}+\eta}|\tilde{D}_m^{\mathcal{P}}(k) - D_m^{\mathcal{P}}(k)| = o_{\mathbb{P}}(1)$. Therefore, it remains to prove that

$$\sup_{k>m} \left(\frac{m}{k}\right)^{\frac{3}{2}+\eta} |D_m^{\mathcal{P}_m}(k) - \tilde{D}_m^{\mathcal{P}_m}(k)| = o_{\mathbb{P}}(1).$$

Proceeding as in the proof of Lemma A.4, we have that

$$\begin{aligned}
&\sup_{k>m}\left(\frac{m}{k}\right)^{\frac{3}{2}+\eta}|\tilde{D}_m^{\mathcal{P}_m}(k)-D_m^{\mathcal{P}_m}(k)| \\
&\quad\le \sup_{k>m}\left(\frac{m}{k}\right)^{\frac{3}{2}+\eta}\max_{j\in[\![m,k[\![}\frac{j(k-j)}{m^{\frac{3}{2}}}\left|\|\boldsymbol{F}_{1:j}^{\mathcal{P}_m}-\boldsymbol{F}_{j+1:k}^{\mathcal{P}_m}\|_{(\Sigma^{\mathcal{P}})^{-1}}-\|\boldsymbol{F}_{1:j}^{\mathcal{P}_m}-\boldsymbol{F}_{j+1:k}^{\mathcal{P}_m}\|_{(\Sigma_m^{\mathcal{P}_m})^{-1}}\right| \\
&\quad\le \sup_{k>m}\left(\frac{m}{k}\right)^{\frac{3}{2}+\eta}\max_{j\in[\![m,k[\![}\frac{j(k-j)}{p^{\frac{1}{2}}m^{\frac{3}{2}}}\left|\left(\boldsymbol{F}_{1:j}^{\mathcal{P}_m}-\boldsymbol{F}_{j+1:k}^{\mathcal{P}_m}\right)^{\top}\left((\Sigma^{\mathcal{P}})^{-1}-(\Sigma_m^{\mathcal{P}_m})^{-1}\right)\left(\boldsymbol{F}_{1:j}^{\mathcal{P}_m}-\boldsymbol{F}_{j+1:k}^{\mathcal{P}_m}\right)\right|^{\frac{1}{2}} \\
&\quad\le \frac{1}{p^{\frac{1}{2}}}\left\|(\Sigma^{\mathcal{P}})^{-1}-(\Sigma_m^{\mathcal{P}_m})^{-1}\right\|_{\mathrm{op}}^{\frac{1}{2}}\sup_{k>m}\left(\frac{m}{k}\right)^{\frac{3}{2}+\eta}\max_{j\in[\![m,k[\![}\frac{j(k-j)}{m^{\frac{3}{2}}}\left\|\boldsymbol{F}_{1:j}^{\mathcal{P}_m}-\boldsymbol{F}_{j+1:k}^{\mathcal{P}_m}\right\|_2.
\end{aligned}$$

We claim that this converges to zero in probability (which completes the proof). By Condition 2.2 and the fact that $\Sigma_m^{\mathcal{P}_m}\xrightarrow{\mathbb{P}}\Sigma^{\mathcal{P}}$, we have that $\|(\Sigma^{\mathcal{P}})^{-1}-(\Sigma_m^{\mathcal{P}})^{-1}\|_{\mathrm{op}}=o_{\mathbb{P}}(1)$. To prove this final claim, it therefore suffices to show that

$$\sup_{k>m}\left(\frac{m}{k}\right)^{\frac{3}{2}+\eta}\max_{j\in[\![m,k[\![}\frac{j(k-j)}{m^{\frac{3}{2}}}\left\|\boldsymbol{F}_{1:j}^{\mathcal{P}_m}-\boldsymbol{F}_{j+1:k}^{\mathcal{P}_m}\right\|_2=O_{\mathbb{P}}(1).$$

Indeed, by equivalence of norms on $\mathbb{R}^p$, this follows from the fact that $\sup_{k>m}(m/k)^{\frac{3}{2}+\eta}\tilde{D}_m^{\mathcal{P}_m}(k)=O_{\mathbb{P}}(1)$, itself a consequence of (D.5), Lemma A.4, and Theorem 2.5. □

References

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

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

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

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.

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

Dedecker, J., Merlevède, F. and Rio, E. (2014). Strong approximation of the empirical distribution function for absolutely regular sequences in $\mathbb{R}^d$. *Electronic Journal of Probability* **19** 1 – 56.

Dehling, H. and Philipp, W. (2002). Empirical process techniques for dependent data. In *Empirical process techniques for dependent data* (H. Dehling, T. Mikosch and M. Sorensen, eds.) 1–113. Birkhäuser, Boston.

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

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

HOLMES, M. and KOJADINOVIC, I. (2021). Open-end nonparametric sequential change-point detection based on the retrospective CUSUM statistic. *Electron. J. Statist.*

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

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

KOJADINOVIC, I. and VERHOIJSEN, A. (2021). npcp: Some Nonparametric Tests for Change-Point Detection in Possibly Multivariate Observations R package version 0.2-3.

KUELBS, J. and PHILIPP, W. (1980). Almost Sure Invariance Principles for Partial Sums of Mixing B-Valued Random Variables. *The Annals of Probability* **8** 1003 – 1036.

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

MOKKADEM, A. (1988). Mixing properties of ARMA processes. *Stochastic Processes and Applications* **29** 309–315.

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

PAPARODITIS, E. and POLITIS, D. N. (2001). Tapered block bootstrap. *Biometrika* **88** 1105–1119.

RIO, E. (1998). Processus empiriques absolument réguliers et entropie universelle. *Probability Theory and Related Fields* **111** 585–608.

RITZ, C., BATY, F., STREIBIG, J. C. and GERHARD, D. (2015). Dose-Response Analysis Using R. *PLOS ONE* **10**.

TUKEY, J. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians* **2** 523–531.

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

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