Journal of Multivariate Analysis 111 (2012) 136–159

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Change-point analysis in increasing dimension

Moritz Jirak

Institute of Statistics, Graz University of Technology, Münzgrabenstraße 11, A-8010 Graz, Austria

ARTICLE INFO

Article history:
Received 21 June 2011
Available online 29 May 2012

AMS 2000 subject classifications:
60F05
60F15
62M10

Keywords:
Brownian bridge
Change in parameters
Weakly dependent processes
Strong invariance principle
Increasing dimension

ABSTRACT

Let $\{\mathbf{Y}_k,\ k \in \mathbb{Z}\}$ be a d-dimensional stationary process, and $\mathbf{g}_{(d)} = (g_1(\mathbf{Y}_1, \ldots \mathbf{Y}_n), \ldots, g_d(\mathbf{Y}_1, \ldots \mathbf{Y}_n))^t$ be a collection of estimators for some parameter $\boldsymbol{\Psi}_{(d)} \in \mathbb{R}^d$. Based on the weighted CUSUM process, we discuss several procedures to detect possible changes in $\boldsymbol{\Psi}_{(d)}$, where we explicitly allow $d = d_n$ to increase with the sample size n. It is demonstrated that an increase in d_n (as n increases) may both lead to a loss or gain in power for testing procedures.

1. Introduction

Let $Y_1, Y_2, \ldots, Y_n$ denote some collected observations. Structural stability is a very important topic in statistics and econometrics, excellent surveys can be found in Banerjee and Urga [5] and Perron [39], for deeper mathematical insights we refer to Csörgo and Horváth [14] and Csörgő and Horváth [16]. Many authors studied testing for the stability of the mean $\mu_i = \mathbb{E}(Y_i)$, $1 \le i \le n$ in case of independent and dependent observations, whereas others considered tests for a change in variance or some other parameters, see for instance [2,4,3,7,23,26,28] and the references therein. A very popular method to detect possible changes are so called CUSUM statistics, which are based on the CUSUM process defined by

$$S_n(t) \equiv \begin{cases} n^{-1/2} \sum_{i=1}^{[(n+1)t]} \left(Y_i - \overline{Y}_n\right), & \text{if } 0 \le t < 1, \\ 0, & \text{if } t = 1, \end{cases} \tag{1.1}$$

where $\overline{X}_n = n^{-1} \sum_{i=1}^{n} Y_i$. Usually, the point where the statistic reaches its maximum is considered as the change point, if the test statistic exceeds a certain critical value. Naturally, these quantiles arise from the asymptotic distribution of the CUSUM process $S_n(t)$. If a functional limit theorem holds for the process $M_n(t) = n^{-1/2} \sum_{i=1}^{[nt]} Y_i$, i.e.

$$M_n(t) \xrightarrow{\mathbb{D}[0,1]} \sigma W_t, \quad 0 \le t \le 1,$$

where W_t is a Brownian motion and $\mathbb{D}[0, 1]$ stands for the space of càdlàg functions on $[0, 1]$, then it follows for instance that

$$\sup_{0 \le t \le 1} \sigma^{-1} \left|S_n(t)\right| \xrightarrow{w} \sup_{0 \le t \le 1} \left|B_t\right|, \tag{1.2}$$

E-mail address: m0ritz@yahoo.com.

0047-259X/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2012.05.007

where B_t denotes a Brownian Bridge, and $\stackrel{w}{\to}$ stands for weak convergence. It is well established in the literature (cf. [16,15]) that a weight function $v(t)$ will increase the power of testing procedures against certain alternatives. In particular the weight function $w(t)^{1/2} = \sqrt{t(1-t)}$ has received considerable attention due to its natural scaling property of the Brownian Bridge, we refer to [16,14] and the references there for more details on the subject. A theoretical drawback of weight functions is that usually a functional limit theorem is no longer sufficient to determine the asymptotic distribution, more refined methods need to be used, such as strong or almost sure approximations, often also called strong invariance principles, for details see [9,35,36,42,50,57] and the references there. Based on these methods, under appropriate assumptions, one may obtain that

$$\sup_{0<t<1} \sigma^{-1} \frac{\left|M_n(t)\right|}{v(t)} \stackrel{w}{\to} \sup_{0<t<1} \frac{\left|B_t\right|}{v(t)}. \tag{1.3}$$

In higher dimension, testing for changes is a much less studied problem, in particular in nearly most cases the dimension d is fixed, see for instance [37,22,27,16,14] and the references there. Notable exceptions are [2,6], which are based on a principal component approach, and Aue et al. [4], who established a framework to test for changes in the covariance structure for d-dimensional processes. Among other things, they also studied the *sequential limit* of their test statistics $\Lambda_{n,d}$, $\Omega_{n,d}$, i.e; they showed that as the dimension d increases, it holds that

$$\Lambda_d^* \stackrel{w}{\to} \mathcal{N}(0,1), \qquad \Omega_d^* \stackrel{w}{\to} \mathcal{N}(0,1),$$

where $\mathcal{N}(0,1)$ denotes a standard normal random variable, and Λ_d^*, Ω_d^* correspond to the weak limits of $\lim_{n\to\infty} \Lambda_{n,d}$ resp. $\lim_{n\to\infty} \Omega_{n,d}$, and n denotes the sample size of the underlying process. The problem of handling a statistic that depends on more than one quantity that tends to infinity arises naturally whenever dealing with high dimensional data and parameters, and the concept of the *sequential limit* is one way to resolve it. However, its lack is obvious. The sequential limit only implies the existence of a sequence $d_n \to \infty$ such that

$$\Omega_{n,d_n} \stackrel{w}{\to} \mathcal{N}(0,1), \qquad \Omega_{n,d_n} \stackrel{w}{\to} \mathcal{N}(0,1) \quad \text{as } n \to \infty,$$

(cf. [33]), which we refer to as the *joint limit*. Recently, a number of authors have studied the joint limit of various statistics (mainly involving the covariance structure), see for instance [31–33,51,55]. The aim of this paper is to contribute to this scenery by reconsidering the approach of Aue et al. [4], in particular, we wish to establish a joint limit result with an explicit relation between the sample size n and the dimension $d = d_n$ as an increasing function in n. Moreover, we will also consider possible weighted versions of the underlying statistics. A particularly interesting result is that the power of the tests may vary with the dimension d_n, we may both encounter a decrease or increase in power.

The setting is quite general and includes a large class of d-dimensional stationary processes $\left\{\mathbf{Y_k}\right\}_{k\in\mathbb{Z}} = \left(Y_{k,1},\ldots,Y_{k,d}\right)^t$, such as ARMA, ARCH and GARCH models. In this context, the dependence structure is very important. To motivate the general notation, consider the following example. Let $\left\{\boldsymbol{\epsilon}_k\right\}_{k\in\mathbb{Z}}$ be a d-dimensional IID sequence. A d-dimensional linear process $\left\{\mathbf{Y}_k\right\}_{k\in\mathbb{Z}}$ can then be constructed by letting

$$\mathbf{Y}_k = \sum_{i=0}^{\infty} \mathbf{A}_i \boldsymbol{\epsilon}_{k-i},$$

where $\mathbf{A}_i$ is a sequence of $d \times d$ dimensional matrices with absolutely summable components. We may thus write $Y_{k,h} = g_h\left(\boldsymbol{\epsilon}_k, \boldsymbol{\epsilon}_{k-1}, \ldots\right)$, for some function $g_h(\cdot)$. This notion allows for the very powerful concept of *m-dependent approximations*, which we will explain more fully in Section 2. If one is interested in studying changes in the underlying covariance structure, then one has to deal with processes of the type of $X_{k,i,j} = Y_{k,r}Y_{k+i,r+j} = g_{i,j}\left(\boldsymbol{\epsilon}_{k+i}, \boldsymbol{\epsilon}_{k+i-1}, \ldots\right)$, which increases the level of complexity. More generally, we may say that we have the collection of stationary processes $X_{k,h} = g_h\left(\boldsymbol{\epsilon}_{k+d}, \boldsymbol{\epsilon}_{k+d-1}, \ldots\right)$, $k \in \mathbb{Z}$ and $1 \le h \le d$, with

$$\mathbb{E}\left(X_{k,h}\right) = \mu_h, \qquad \boldsymbol{\mu}_d = \left(\mu_1, \ldots, \mu_d\right)^t \in \mathbb{R}^d. \tag{1.4}$$

We then wish to test for changes in the vector $\boldsymbol{\mu}_d$, which may reflect any kind of multivariate statistic that fits into this framework.

The paper is structured as follows. In Section 2 we present the setting and some notation, alongside some comments and remarks. The main results are given in Section 3, together with some examples. Numerical illustrations are provided in Section 4. Based on a general approximation result, the proofs are given in Section 5. In Section 6, the approximation results are presented, which may have interest in themselves.

2. Notation and preliminary remarks

Throughout this paper, we will use the following notation. For a random variable X, we denote with $\left\|X\right\|_p = \mathbb{E}\left(|X|^p\right)^{1/p}$ the L^p norm. For a vector $\mathbf{x} \in \mathbb{R}^d$, we put $\left|\mathbf{x}\right|_2$ for the usual euclidian norm, and $\max\left|\mathbf{x}\right|$ for the maximum norm. Let

$\mathbf{A} = (a_{i,j})_{\{1\le i\le r, 1\le j\le s\}}$, $r, s \in \mathbb{N}$ be an $r \times s$ matrix. Then we denote with

$$\max|\mathbf{A}| = \max\{\mathbf{A}\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^s, \max|\mathbf{x}| = 1\} = \max_{1\le i\le r} \sum_{j=1}^{s} |a_{i,j}| \tag{2.1}$$

the usual induced matrix norm. We will now discuss the dependence structure in more detail. To this end, let $\{X_{k,h}\}_{k,h\ge 1}$ be a collection of random variables such that for each h_0, $\{X_{k,h_0}\}_{k\ge 1}$ is a stationary sequence, and we put

$$\phi_{i,h} = \mathrm{Cov}(X_{k,h}, X_{k+i,h}), \quad k \in \mathbb{Z}, 1 \le h \le d.$$

Given an $\mathbb{R}^{\mathbb{Z}}$ valued sequence $\{\boldsymbol{\epsilon}_k\}_{k\in\mathbb{Z}}$ of independent and identically distributed random variables, we define the following two σ-algebras.

$$\mathcal{F}_k = \sigma(\boldsymbol{\epsilon}_j, j \le k), \qquad \mathcal{F}_{k-m}^{k+m} = \sigma(\boldsymbol{\epsilon}_j, k-m \le j \le k+m). \tag{2.2}$$

We will always assume that $\{X_{k,h}\}_{k,h\ge 1}$ is adapted to $\mathcal{F}_{k+d}$, more specifically, we assume that $X_{k,h}$ is $\mathcal{F}_{k+d}$ measurable for each $k, h \ge 1$. Hence we implicitly assume that $X_{k,h}$ can be written as a function

$$X_{k,h} = g_h(\boldsymbol{\epsilon}_{k+d}, \boldsymbol{\epsilon}_{k+d-1}, \ldots).$$

For convenience, we will write $g_h(\xi_{k+d})$, with $\xi_k = (\boldsymbol{\epsilon}_k, \boldsymbol{\epsilon}_{k-1}, \ldots)$. The class of processes that fits into this framework is large, and contains a variety of linear and nonlinear processes including ARCH, GARCH and related processes, see for instance [21,40,46,47]. A very nice feature of the representation given above is that it allows to give simple, yet very efficient and general dependence conditions. Following Wu [52], let $\{\boldsymbol{\epsilon}_k'\}_{k\in\mathbb{Z}}$ be an independent copy of $\{\boldsymbol{\epsilon}_k\}_{k\in\mathbb{Z}}$ on the same probability space, and define the 'filters' $\xi_k^{(m,')}$, $\xi_k^{(m,*)}$ as

$$\xi_k^{(m,')} = (\boldsymbol{\epsilon}_k, \boldsymbol{\epsilon}_{k-1}, \ldots, \boldsymbol{\epsilon}_{k-m}', \boldsymbol{\epsilon}_{k-m-1}, \ldots)$$

and

$$\xi_k^{(m,*)} = (\boldsymbol{\epsilon}_k, \boldsymbol{\epsilon}_{k-1}, \ldots, \boldsymbol{\epsilon}_{k-m}', \boldsymbol{\epsilon}_{k-m-1}', \boldsymbol{\epsilon}_{k-m-2}', \ldots).$$

Note that in $\xi_k^{(m,')}$, only the element $\boldsymbol{\epsilon}_{k-m}$ is replaced, but in $\xi_k^{(m,*)}$ the whole past up to $k-m$ gets replaced. We put $\xi_k' = \xi_k^{(k,')} = (\boldsymbol{\epsilon}_k, \boldsymbol{\epsilon}_{k-1}, \ldots, \boldsymbol{\epsilon}_0', \boldsymbol{\epsilon}_{-1}, \ldots)$ and $\xi_k^* = \xi_k^{(k,*)} = (\boldsymbol{\epsilon}_k, \boldsymbol{\epsilon}_{k-1}, \ldots, \boldsymbol{\epsilon}_0, \boldsymbol{\epsilon}_{-1}', \boldsymbol{\epsilon}_{-2}' \cdots)$. In analogy, we put $X_{k,h}^{(m,')} = g_h(\xi_{k+d}^{(m,')})$ and $X_{k,h}^{(m,*)} = g_h(\xi_{k+d}^{(m,*)})$, in particular we have $X_{k,h}' = X_{k,h}^{(k+d,')}$ and $X_{k,h}^* = X_{k,h}^{(k+d,*)}$.

As a dependence measure, one may now consider the quantities $\|X_{k,h} - X_{k,h}'\|_p$ or $\|X_{k,h} - X_{k,h}^*\|_p$, $p \ge 1$. For example, if we define the linear processes $X_{k,h} = \sum_{i=0}^{\infty} \alpha_{i,h}\boldsymbol{\epsilon}_{k-i}$, the condition

$$\sum_{k=0}^{\infty} \|X_{k,h} - X_{k,h}'\|_2 < \infty \tag{2.3}$$

amounts to $\sum_{i=0}^{\infty} |\alpha_{i,h}| < \infty$ if $\mathbb{E}(\boldsymbol{\epsilon}_0^2) < \infty$. Dependence conditions of type (2.3) are often quite general and easy to verify in many cases, see for instance [8,17,19,50] and the references there.

A very important feature of the above representation is that it allows to quantify approximations with m-dependent variables. To this end, for $d \le m$ let

$$Y_{k,h}^{(\le m)} = \mathbb{E}(X_{k,h} \mid \mathcal{F}_{k-m}^{k+m}), \qquad Y_{k,h}^{(>m)} = X_{k,h} - Y_{k,h}^{(\le m)}. \tag{2.4}$$

Then one can show (cf. Proposition 5.3), that

$$\|Y_{k,h}^{(>m)}\|_p \le C \sum_{i=0}^{\infty} \|X_{m+i,h} - X_{m+i,h}'\|_p^2,$$

hence we can quantify the approximation error in terms of the dependence measures.

In order to formulate the relevant statistics for the change point analysis, introduce the following notation. Put $w(t) = t(1-t)$, $0 \le t \le 1$, for the weight function and

$$S_h^{(n,l)} = \sum_{k=1}^{l} X_{k,h}, \qquad M_{t,h}^{(n)} = n^{-1/2}\left(\sum_{k=1}^{[nt]} X_{k,h} - t\sum_{k=1}^{n} X_{k,h}\right), \tag{2.5}$$

for the partial sum and the corresponding CUSUM statistic. For $0 < l < 1/2$, we denote the weighted version of $M_{t,h}^{(n)}$ with

$$Z_h^{(n,l)} = \sup_{l\le t\le 1-l}\left(\frac{|M_{t,h}^{(n)}|}{\sqrt{t(1-t)}}\right). \tag{2.6}$$

We use the abbreviation $S_h^{(n)} = S_h^{(n,n)}$, and we denote the corresponding random vectors with $\mathbf{S}^{(n)} = \big(S_1^{(n)}, S_2^{(n)}, \ldots, S_{d_n}^{(n)}\big)^t$, $\mathbf{M}_t^{(n)} = \big(M_{t,1}^{(n)}, M_{t,2}^{(n)}, \ldots, M_{t,d_n}^{(n)}\big)^t$, and $\mathbf{Z}_l^{(n)} = \big(Z_1^{(n,l)}, Z_2^{(n,l)}, \ldots, Z_{d_n}^{(n,l)}\big)^t$. We formally define the variance as

$$\psi_h = \lim_n n^{-1}\mathbb{E}\big(S_h^{(n)} S_h^{(n)}\big), \tag{2.7}$$

and for $1 \le i \le j \le d_n$ the sample correlation

$$\rho_{i,j}^{(n)} = \mathrm{Cov}\big(S_i^{(n)} S_j^{(n)}\big)\left(\mathrm{Var}\big(S_i^{(n)}\big)\mathrm{Var}\big(S_j^{(n)}\big)\right)^{-1/2}. \tag{2.8}$$

In addition, let $\mathbf{B}_t^{(*)} = \big(B_{t,1}^{(*)}, B_{t,2}^{(*)}, \ldots, B_{t,d_n}^{(*)}\big)^t$, where $\big\{B_{t,h}^{(*)}\big\}_{0\le t\le 1, 1\le h\le d_n}$ are independent Brownian Bridges. Similarly, put $\mathbf{B}_t = \big(B_{t,1}, B_{t,2}, \ldots, B_{t,d_n}\big)^t$, where $B_{t,h} = W_{t,h} - tW_{1,h}$, $1 \le h \le d_n$ is a sequence of Brownian Bridges, where the d_n-dimensional Brownian motion $\mathbf{W}_t^{(d_n)} = \big(W_{t,1}, W_{t,2}, \ldots, W_{t,d_n}\big)^t$ has the covariance matrix $\boldsymbol{\Gamma}_{\mathbf{W}^{(d_n)}}$.

3. Main results

3.1. Approximations and limit theorems

In this section, we present approximation results that can be used to detect changes in time series. To this end, denote with $\boldsymbol{\Gamma}_{\mathbf{S}^{(n)}}$ the covariance matrix of the vector $n^{-1}\mathbf{S}^{(n)}$, defined in (2.5), and let $\widehat{\boldsymbol{\Gamma}}_{\mathbf{S}^{(n)}}$ be a consistent estimator. Let $d_n \ge 1$ be a positive, monotone increasing sequence. We will work under the following assumption.

Assumption 3.1. Assume that for $p > 8$, $d_n = \mathcal{O}\big((\log n)^{\delta}\big)$, $\delta > 0$

(i) $\sup_h \|X_{1,h}\|_p < \infty$, $\mathbb{E}\big(X_{k,h}\big) = \mu_h$,
(ii) $\max_{1\le h\le d_n}\big\|X_{k,h} - X'_{k,h}\big\|_p = \mathcal{O}\big(k^{-\beta}\big)$, where $\beta > (3+\sqrt{3})(2\sqrt{3}-2)^{-1} \approx 3.232\cdots$,
(iii) $P\big(\max\big|\widehat{\boldsymbol{\Gamma}}_{\mathbf{S}^{(n)}} - \boldsymbol{\Gamma}_{\mathbf{S}^{(n)}}\big| \ge n^{-\chi}\big) = \mathcal{O}\big(1\big)$, $\chi > 0$,
(iv) The smallest eigenvalue $\sigma_{\min}\big(\boldsymbol{\Gamma}_{\mathbf{S}^{(n)}}\big)$ of the matrix $\boldsymbol{\Gamma}_{\mathbf{S}^{(n)}}$ satisfies $\sigma_{\min}\big(\boldsymbol{\Gamma}_{\mathbf{S}^{(n)}}\big)^{-1} = \mathcal{O}\big(d_n^{\kappa}\big)$, $\kappa > 0$.

Let us briefly discuss the assumptions. As already mentioned, conditions (i), (ii) are very general dependence conditions, which allow for a large class of processes. Examples are provided in Section 3.2, which consist of linear and nonlinear time series. Also note that one may weaken the moment assumption by strengthening the dependence condition (cf. Aue et al. [4]). This is accomplished by considering the transformation $U_{k,h} = |X_{k,h}|^{\rho}$, and then using $U_{k,h}$ instead. Indeed one obtains that for $\rho \in (0, 1]$ it holds that

$$\big\|U_{k,h} - U'_{k,h}\big\|_p \le \big\|X_{k,h} - X'_{k,h}\big\|_{p\,\rho}^{\rho}, \tag{3.1}$$

which is a stronger condition for the dependence assumption. For example, let $X_k = \sum_{i=0}^{\infty}\alpha_i\epsilon_{k-i}$ with $\sum_{i=0}^{\infty}|\alpha_i| < \infty$, where ϵ_k is an IID sequence with $\mathbb{E}\big(|\epsilon_k|^p\big) < \infty$, but $\mathbb{E}\big(|\epsilon_k|^{p+\delta}\big) = \infty$ for $\delta > 0$, $p \ge 1$. Clearly, we have $\|X_k - X'_k\|_p = \mathcal{O}\big(|\alpha_k|\big)$. Consider now $U_k = \sqrt{|X_k|}$. Then we obtain from (3.1) that

$$\big\|U_k - U'_k\big\|_{2p} \le \big\|X_k - X'_k\big\|_p^{1/2} = \mathcal{O}\big(\sqrt{|\alpha_k|}\big),$$

which is a slower decay rate than $|\alpha_k|$, since $|\alpha_k| \to 0$ as $k \to \infty$.

Condition (iii) is a mild convergence assumption for potential covariance estimators. Put $\mathbf{X}_k^{(d_n)} = \big(X_{k,1}, \ldots, X_{k,d_n}\big)^t$, and note that if $d_n = d$ remains fixed, then by Aue et al. [4, Theorem A.2], it holds that

$$n^{-1/2}\mathbf{S}^{(n)} \xrightarrow{w} \mathbf{W}_t^{(d)},$$

where the covariance matrix $\boldsymbol{\Gamma}_{\mathbf{W}^{(d)}} = \big\{\gamma_{|i-j|}\big\}_{1\le, i,j\le d}$ satisfies

$$\boldsymbol{\Gamma}_{\mathbf{W}^{(d)}} = \sum_{k\in\mathbb{Z}}\mathrm{Cov}\big(\mathbf{X}_k^{(n)}, \mathbf{X}_0^{(n)}\big).$$

Under very general conditions (cf. [4,13]), it holds that Bartlett-type estimators satisfy the relation $\big|\widehat{\boldsymbol{\Gamma}}_{\mathbf{S}^{(n)}} - \boldsymbol{\Gamma}_{\mathbf{W}^{(d)}}\big| = \mathcal{O}_p(1)$ for fixed d. Using similar arguments and conditions (i), (ii), one may show that (iii) is indeed valid for some $\chi > 0$. This issue will be more thoroughly discussed in Section 3.2.

Condition (iv) is needed to control the approximation error for the inverse of the covariance matrix $\boldsymbol{\Gamma}_{\mathbf{S}^{(n)}}$. Note that for fixed d, condition (iv) is redundant since every regular positive definite matrix $\boldsymbol{\Gamma}_{\mathbf{W}^{(d)}}$ has strictly positive eigenvalues. Unfortunately, this is no longer sufficient in case of a strictly increasing sequence d_n, and in general, no simple lower bounds are available for $\sigma_{\min}\big(\boldsymbol{\Gamma}_{\mathbf{S}^{(n)}}\big)$. Using the well-known Gershgorin Theorem (cf. [49]), it follows that $\sigma_{\min}\big(\boldsymbol{\Gamma}_{\mathbf{S}^{(n)}}\big) > 0$, uniformly in n, provided that $\psi_h = C$ for all $1 \leq h \leq d_n$ (one may weaken this condition) and

$$\limsup_{n\to\infty} \max_{1\leq h\leq d_n} \sum_{\substack{i=1,\\ i\neq h}}^{d_n} \big|\rho_{h,i}^{(n)}\big| < 1. \tag{3.2}$$

This, however, is a rather restrictive condition. Another well known approach in this context is based on the following Lemma, which is a special case of Section 5.2 in Grenander and Szego [24] (see also [56]).

Lemma 3.2. *Let f be a continuous symmetric function on $[-\pi, \pi]$ with m and M being its minimum and maximum. Define $\gamma_h = \int_{-\pi}^{\pi} f(x)e^{-\sqrt{-1}hx}dx$, and the $d \times d$ matrix $\boldsymbol{\Gamma}_{\mathbf{W}^{(d)}} = \big\{\gamma_{|i-j|}\big\}_{1\leq, i,j\leq d}$. Then*

$$2\pi m \leq \sigma_{\min}\big(\boldsymbol{\Gamma}_{\mathbf{W}^{(d)}}\big) \leq \sigma_{\max}\big(\boldsymbol{\Gamma}_{\mathbf{W}^{(d)}}\big) \leq 2\pi M,$$

where $\sigma_{\min}\big(\boldsymbol{\Gamma}_{\mathbf{S}^{(n)}}\big)$ and $\sigma_{\min}\big(\boldsymbol{\Gamma}_{\mathbf{S}^{(n)}}\big)$ denote the minimum and maximum eigenvalues.

In many cases it turns out that the function $f(x)$ is the spectral density function of some underlying process. Unfortunately, proving that $m > 0$ in this case is very difficult in general. In special cases (cf. Example 3.15) this can be accomplished though.

Under Assumption 3.1, the following approximation results are valid.

Theorem 3.3. *Assume that Assumption 3.1 holds. Then on a possible larger probability space, we have that*

$$\left| \sup_{\lambda/n\leq t\leq 1-\lambda/n} \left| w(t)^{-1}\big(\mathbf{M}_t^{(n)}\big)^t \widehat{\boldsymbol{\Gamma}}_{\mathbf{S}^{(n)}}^{-1}\mathbf{M}_t^{(n)}\right| - \sup_{\lambda/n\leq t\leq 1-\lambda/n} w(t)^{-1}\sum_{h=1}^{d_n}\big(B_{t,h}^{(*)}\big)^2 \right| = \mathcal{O}_p\big(\sqrt{d_n}\big),$$

where $\lambda > 0$ and the dimension satisfies $d_n = \mathcal{O}\big((\log n)^{\delta}\big)$ for arbitrary $\delta > 0$.

Theorem 3.4. *Under the same conditions as in Theorem 3.3, it holds that*

$$\left| \frac{1}{n}\sum_{k=\lceil\lambda\rceil+1}^{\lceil n-\lambda\rceil} w(k/n)^{-1}\big(\mathbf{M}_{k/n}^{(n)}\big)^t \widehat{\boldsymbol{\Gamma}}_{\mathbf{S}^{(n)}}^{-1}\mathbf{M}_{k/n}^{(n)} - \sum_{h=1}^{d_n}\int_{\lambda/n}^{1-\lambda/n} w(t)^{-1}\big(B_{t,h}^{(*)}\big)^2 dt \right| = \mathcal{O}_p\big(\sqrt{d_n}\big).$$

Remark 3.5. Note that condition (iv) does not impose any conditions on δ, as long as we have $d_n = \mathcal{O}\big((\log n)^{\delta}\big)$. However, a closer look at the proof of Theorems 3.3 and 3.4 reveals that analogue versions are valid for $d_n = \mathcal{O}\big(n^{\delta}\big)$, for some $\delta > 0$, in which case condition (iv) does impose restrictions on the choice of δ. The bound for δ is however far from tractable, see in particular Theorem 6.4. Also, it is clear from the proofs that setting $w(t) \equiv 1$ will allow for a larger growth rate for d_n in this case.

Let us briefly elaborate on the weight function $w(t)$. In order to increase the power of test-statistics at edge points, weight functions have proven to be a very efficient method (cf. [7,16,14] and the references there). The specific choice $w(t)^{1/2} = \sqrt{t(1-t)}$ is particularly interesting, since it standardizes the Brownian Bridge $B_t = W_t - tW_1$, for details on this subject, we refer to [16,14]. In this context, the choice of λ in Theorem 5.4 is important. It is reported in [14] that choosing a sequence $\lambda = \lambda_n = (\log n)^{3/2}$ yields good results in practice. One may also work with weight functions $v(t)$ satisfying the following conditions.

- $v(t)$ is a function on (0, 1) increasing in a neighborhood of 0, and decreasing in a neighborhood of 1,
- $\inf_{c\leq t\leq 1-c} v(t) > 0$ for all $0 < c < 1/2$,
- the function $I(v, c) = \int_0^1 \frac{1}{t(t-1)} \exp(-\frac{cv^2(t)}{t(t-1)})dt$ is finite for some $0 < c < 1/2$.

It is then possible (cf. [7,15]) to establish analogue versions of Theorems 3.3 and 3.4, where $w(t)$ is replaced with $v(t)$, satisfying the conditions above. In particular, one may also choose $v(t) = 1$, i.e. no weight function at all.

Since d_n may become large as n increases, it is interesting to derive asymptotic expressions for the approximating quantities presented in Theorems 3.3 and 3.4. To this end, denote with

$$\Omega_n = \sup_{\lambda\leq t\leq 1-\lambda} w(t)^{-1}\big(\mathbf{M}_t^{(n)}\big)^t \widehat{\boldsymbol{\Gamma}}_{\mathbf{S}^{(n)}}^{-1}\mathbf{M}_t^{(n)}, \tag{3.3}$$

and

$$\Lambda_n = \frac{1}{n}\sum_{k=\lceil\lambda\rceil}^{\lceil n-\lambda\rceil} w(k/n)^{-1}\big(\mathbf{M}_{k/n}^{(n)}\big)^t \widehat{\boldsymbol{\Gamma}}_{\mathbf{S}^{(n)}}^{-1}\mathbf{M}_{k/n}^{(n)}. \tag{3.4}$$

If we have indeed $w(t) = v(t) = 1$, then Theorems 3.3 and 3.4 together with Remark 2.1 in [4] imply that both Ω_n and Λ_n, appropriately normalized, weakly converge to a standard Gaussian random variable. If we consider the weight function $w(t) = t(1-t)$, the following asymptotic expansion is valid.

Corollary 3.6. *Assume that Assumption 3.1 holds, and let* $0 < \lambda < 1$. *Then*

$$(2d_n)^{-1/2} \sup_{\lambda \le t \le 1-\lambda} \left| w(t)^{-1} \big(\mathbf{M}_t^{(n)}\big)^t \widehat{\boldsymbol{\Gamma}}_{\mathbf{S}^{(n)}}^{-1} \mathbf{M}_t^{(n)} - d_n \right| \xrightarrow{w} \sup_{0 \le s \le \log u(\lambda)} \big|V_s\big|, \qquad u^*(\lambda) = \frac{1-\lambda^2}{\lambda(1-\lambda)},$$

where V_s is a zero mean Ornstein–Uhlenbeck process with $\operatorname{Cov}\big(V(t), V(s)\big) = \exp(-|t-s|)$, *and we may choose* $d_n = \mathcal{O}\big((\log n)^{\delta}\big)$, *for arbitrary* $\delta > 0$.

Unfortunately, there is no known simple formula for the distribution function of $\sup_{0 \le s \le \log u(\lambda)} \big|V_s\big|$. On the other hand, one may obtain the Laplace transform, and a numerical inversion gives rise to selected values of the distribution function of $\sup_{0 \le t \le \log u(\lambda)} \big|V_t\big|$ (cf. [18,34]). It turns out however (cf. [15]), that the following tail approximation of Vostrikova [48] works rather well even for moderate values of x.

Lemma 3.7. *For all fixed* $T > 0$ *we have*

$$P\left(\sup_{0 \le t \le T} \big|V(t)\big| \ge x\right) = \frac{x \exp(-x^2/2)}{\sqrt{2\pi}} \left(T - \frac{T}{x^2} + \frac{4}{x^4} + \mathcal{O}\big(x^{-4}\big)\right),$$

as $x \to \infty$.

In view of this result, it would be interesting to determine the limit distribution of Ω_n. Theorem A.3.2 in [15] suggests that

$$a_n \Omega_n - b_n \xrightarrow{w} \mathcal{G}, \quad \text{for appropriate sequences } a_n, b_n,$$

where $\mathcal{G}$ is an extreme value distribution. However, establishing this result is beyond the scope of the present paper and will be dealt with elsewhere. Fortunately, the situation is much simpler in case of Λ_n. We introduce the following quantity. Let

$$\sigma^2 = \int_{\mathbb{R}^3} x^2 y^2 \left(\frac{1}{2\pi(1-e^{-|u|})} \exp\left(-\frac{x^2+y^2-2e^{-|u|/2}xy}{2\pi(1-e^{-|u|})}\right) - \varphi(x)\varphi(y)\right) dxdydu,$$

where $\varphi(x)$ denotes the density of a standard Gaussian distribution function.

Corollary 3.8. *Assume that Assumption 3.1 holds, and put* $u_n = n^2/\lambda^2 - n/\lambda + 1$ *with* $\lambda > 0$. *Then*

$$(d_n \sigma^2 \log u_n)^{-1/2} \left(\frac{1}{n} \sum_{k=\lceil \lambda \rceil}^{\lceil n-\lambda \rceil} w(t)^{-1} \big(\mathbf{M}_{k/n}^{(n)}\big)^t \widehat{\boldsymbol{\Gamma}}_{\mathbf{S}^{(n)}}^{-1} \mathbf{M}_{k/n}^{(n)} - \sqrt{\pi} d_n \right) \xrightarrow{w} \mathcal{N}(0, 1),$$

and the dimension satisfies $d_n = \mathcal{O}\big((\log n)^{\delta}\big)$ *for arbitrary* $\delta > 0$.

3.2. Testing for change-points

In this section, we will discuss how to apply the results of the previous section for change-point detection. To this end, denote with $\boldsymbol{\mu}_k^{(n)} = \big(\mu_{k,1}, \dots \mu_{k,d_n}\big)^t = \big(\mathbb{E}(X_{k,1}), \dots \mathbb{E}(X_{k,d_n})\big)^t$ the vector of the means. To test for changes in $\boldsymbol{\mu}_k^{(n)}$, we formulate a null hypothesis as

$$\mathcal{H}_0 : \boldsymbol{\mu}_1^{(n)} = \boldsymbol{\mu}_2^{(n)} = \dots = \boldsymbol{\mu}_n^{(n)},$$

and the alternative

$$\mathcal{H}_A : \boldsymbol{\mu}_1^{(n)} = \boldsymbol{\mu}_2^{(n)} = \dots = \boldsymbol{\mu}_{k^*}^{(n)} \neq \boldsymbol{\mu}_{k^*+1}^{(n)} = \dots = \boldsymbol{\mu}_n^{(n)}$$

where $1 \le k^* \le n$ denotes the change-point. Theorem 3.3 and Corollaries 3.6 and 3.8 provide us with parameter free asymptotic expressions for Ω_n and Λ_n if $\mathcal{H}_0$ holds. A natural question is what one may expect of the power of the involved statistics. Note that both Ω_n and Λ_n essentially 'sum up' all *local discrepancies*, and can therefore be viewed as a *global quality measure*. This means that if a change affects all or most of the involved processes $M_{t,h}^{(n)}$, $1 \le h \le d_n$ then the power of tests based on Ω_n and Λ_n can be expected to be high. Contrarily, if only a few of the $M_{t,h}^{(n)}$ are affected, the power can be expected to be rather low. In order to be more precise, suppose that the alternative $\mathcal{H}_A$ is valid. As is common practice in the literature, we assume that $k^* = \lceil \tau n \rceil$, $\tau \in (0, 1)$ depends on n. Let

$$\mathbf{S}^{(n)} = \mathbf{S}^{(\le \lceil \tau n \rceil)} + \mathbf{S}^{(> \lceil \tau n \rceil)}, \tag{3.5}$$

where $\mathbf{S}^{(\le \lceil \tau n \rceil)}$ denotes the pre-change vector, and $\mathbf{S}^{(> \lceil \tau n \rceil)}$ the post-change vector, and define $\mathbf{M}_t^{(\le \lceil \tau n \rceil)}$, $\mathbf{M}_t^{(> \lceil \tau n \rceil)}$ in an analogue manner.

Theorem 3.9. *Assume that* $\limsup_{n\to\infty}\max_{1\le i\le d_n}\sum_{j=1}^{d_n}\left|\rho_{i,j}^{(n)}\right| < \infty$ *and* Assumption 3.1 *are both valid for* $\{X_{k,h}\}_{1\le h, 1\le k\le k^*}$ *and* $\{X_{k,h}\}_{1\le h, k^*<k}$. *Then*

$$\liminf_{n\to\infty}\chi_n^{-1}\sup_{\lambda\le t\le 1-\lambda}\left|w(t)^{-1}\big(\mathbf{M}_t^{(n)}\big)^t\widehat{\boldsymbol{\Gamma}}_{\mathbf{S}^{(n)}}^{-1}\mathbf{M}_t^{(n)}\right| > 0,$$

in probability, where $\chi_n = \mathcal{O}\big(n\big(\boldsymbol{\mu}_{k^*+1}^{(n)}\big)^t\boldsymbol{\mu}_{k^*+1}^{(n)}\big)$.

Remark 3.10. An analogue version is valid for Λ_n.

Note that $\big(\boldsymbol{\mu}_{k^*+1}^{(n)}\big)^t\boldsymbol{\mu}_{k^*+1}^{(n)}$ essentially gives the number of changes that have occurred. In particular, if $\big(\boldsymbol{\mu}_{k^*+1}^{(n)}\big)^t\boldsymbol{\mu}_{k^*+1}^{(n)} = \mathcal{O}\big(d_n^{1/2+\eta}\big)$, $\eta > 0$, then Theorem 3.9 together with Corollaries 3.6 and 3.8 shows that we have an *increase in power* as d_n increases. On the other hand, if $\big(\boldsymbol{\mu}_{k^*+1}^{(n)}\big)^t\boldsymbol{\mu}_{k^*+1}^{(n)} = \mathcal{O}\big(d_n^{1/2}\big)$, then we have a *decrease in power* as d_n increases. Simulation results given in Section 4 highlight this effect.

We now give examples on how to apply the results of the previous sections to test for changes in the underlying sequence. The focus lies on testing for changes in the mean and covariance structure. Given a d-dimensional time series $\{\mathbf{Y}_k\}_{k\in\mathbb{Z}} = \big(Y_{k,1},\ldots,X_{k,d}\big)^t$, we thus want to test for

(i) changes in the mean $\boldsymbol{\mu}_d$, where $\boldsymbol{\mu}_d = \big(\mu_1,\ldots,\mu_d\big)^t\in\mathbb{R}^d$, $\mathbb{E}\big(X_{k,h}\big) = \mu_h$,
(ii) changes in the covariance matrix $\sum_d = \big\{\mathrm{Cov}\big(X_{k,i},X_{k,j}\big)\big\}_{1\le i,j,\le d}$.

To this end, let $\mathfrak{d} = d(d+1)/2$, $\mathcal{I}_{\mathfrak{d}} = \big\{(i,j)\mid 1\le i,j\le d\big\}$. Put

- $X_{k,h}^{(1)} = Y_{k,i}-\mu_i$, $1\le i\le d$,
- $X_{k,h}^{(2)} = Y_{k,i}Y_{k,j}-\mathbb{E}\big(Y_{k,i}Y_{k,j}\big)$ for all $(i,j)\in\mathcal{I}_d$, $1\le h\ \le\mathfrak{d}$.

For the usage of Corollaries 3.6 and 3.8, we need to reconsider Assumption 3.1, in particular, we require an estimator $\widehat{\boldsymbol{\Gamma}}_{\mathbf{W}^{(d)}}$ for the covariance matrix $\boldsymbol{\Gamma}_{\mathbf{W}^{(d)}}$. The literature (cf. [1,13,25]) provides many potential candidates to estimate the long run covariance matrix $\boldsymbol{\Gamma}_{\mathbf{W}^{(d)}} = \big\{\gamma_{|i-j|}\big\}_{1\le i,j\le d}$. A popular estimator for γ_h is Bartlett's estimator, or more general, estimators of the form

$$\widehat{\gamma}_h^2 = \sum_{|j|\le l_n}\omega(k/r)\widehat{\gamma}_{j,h} \tag{3.6}$$

with weight function $\omega(x)$, where $\gamma_{j,h} = \mathbb{E}\big(X_{0,h}X_{j,h}\big)$ and $\widehat{\gamma}_{j,h} = n^{-1}\sum_{k=1}^{n-j}X_{k,h}X_{k+j,h}$. Considering the triangular weight function $\omega(x) = 1-|x|$ for $|x|\le 1$ and $\omega(x) = 0$ for $|x|>1$, one recovers the Bartlett estimator in (3.6). One may also use the plain plug in estimate

$$\widehat{\gamma}_h^2 = \widehat{\gamma}_{0,h} + 2\sum_{i=1}^{l_n}\widehat{\gamma}_{i,h}, \tag{3.7}$$

see for instance [43,44]. In particular, based on Wu [51, Proposition 1] (see also [53]), we can present the following result.

Proposition 3.11. *Let* $l_n\in\mathbb{N}$, $l_n\to\infty$ *as n increases with* $l_n = \mathcal{O}\left(n^{1/2-\chi}\right)$, $\chi > 0$. *If* Assumption 3.1(i), (ii) *holds, then*

$$P\big(\max\big|\widehat{\boldsymbol{\Gamma}}_{\mathbf{W}^{(d)}}-\boldsymbol{\Gamma}_{\mathbf{W}^{(d)}}\big|\ge n^{-\chi}\big) = \mathcal{O}\big(1\big),$$

where $\widehat{\boldsymbol{\Gamma}}_{\mathbf{W}^{(d)}}$ *is constructed via* (3.7). *A similar result holds if one uses Bartlett-based estimators.*

We now list a few examples of popular processes together with explicit conditions such that Assumption 3.1 is valid. In case of the Garch-type model (Example 3.13), we borrow heavily from the findings of Aue et al. [4].

Example 3.12 (*Linear Processes*). We first reconsider the example from the introduction, namely a linear d-dimensional process $\{\mathbf{Y}\}_{k\in\mathbb{Z}}$ Let $\{\boldsymbol{\epsilon}_k\}_{k\in\mathbb{Z}}$ be a d-dimensional I.I.D. sequence, where we denote the single elements with $\epsilon_{k,h}$, $1\le h\le d$. Define the d-dimensional linear process $\{\mathbf{Y}\}_{k\in\mathbb{Z}}$ as

$$\mathbf{Y}_k = \sum_{i=0}^{\infty}\mathbf{A}_i\boldsymbol{\epsilon}_{k-i}, \tag{3.8}$$

where $\mathbf{A}_i = \big\{a_{r,s}^{(i)}\big\}_{1\le r,s\le d}$ is a sequence of $d\times d$ dimensional matrices. We need to verify Assumption 3.1 for $X_{k,h}^{(1)}$, and $X_{k,h}^{(2)}$. It holds that

$$\max_{1\leq h\leq d}\left\|X_{k,h}^{(1)}-X_{k,h}^{(1),'}\right\|_p \leq \left\|\max\left|\mathbf{A}_k(\boldsymbol{\epsilon}_k-\boldsymbol{\epsilon}_k')\right|\right\|_p \leq 2\left\|\max\left|\mathbf{A}_k\boldsymbol{\epsilon}_k\right|\right\|_p$$
$$\leq 2\max_{1\leq r\leq d}\sum_{s=1}^{d}|a_{r,s}^{(k)}|\max_{1\leq h\leq d}\left\|\epsilon_{0,h}\right\|_p = 2\max\left|\mathbf{A}_k\right|\max_{1\leq h\leq d}\left\|\epsilon_{0,h}\right\|_p,$$

hence we thus require that

- $d=\mathcal{O}\big((\log n)^{\delta}\big)$, for some $\delta>0$,
- $\max_{1\leq h\leq d}\left\|\epsilon_{0,h}\right\|_p<\infty$ for $p>8$,
- $\max\left|\mathbf{A}_k\right|=\max_{1\leq r\leq d}\sum_{s=1}^{d}|a_{r,s}^{(k)}|=\mathcal{O}\big(k^{-\beta}\big)$, where $\beta>(3+\sqrt{3})(2\sqrt{3}-2)^{-1}\approx 3.232\cdots$,

and conditions (i)–(iii) of Assumption 3.1 are satisfied. As mentioned earlier in Section 2, verifying Assumption 3.1(iv) is very difficult, and indeed providing a general, simple verifiable condition for the matrix sequence $\mathbf{A}_i$ seems to be impossible. We now turn to $X_{k,h}^{(2)}$. Suppose that $(i,j)\in\mathcal{I}_{\mathfrak{d}}$ corresponds to h. Then the Cauchy–Schwarz inequality implies

$$\left\|X_{k,h}^{(2)}-X_{k,h}^{(2),'}\right\|_p \leq \left\|X_{k,j}^{(1)}(X_{k,i}^{(1)}-X_{k,i}^{(1),'})\right\|_p+\left\|X_{k,i}^{(1)}(X_{k,j}^{(1)}-X_{k,j}^{(1),'})\right\|_p$$
$$\leq 2\max_{1\leq h\leq d}\left\|X_{k,h}^{(1)}\right\|_{2p}\max_{1\leq h\leq d}\left\|X_{k,h}^{(1)}-X_{k,h}^{(1),'}\right\|_{2p}.$$

Hence the conditions remain the same, except that we require $\max_{1\leq h\leq d}\left\|\epsilon_{0,h}\right\|_p<\infty$ for $p>16$. Using the general theory of multivariate ARMA models (cf. [13,1]), one readily provides conditions based on those for linear processes.

Example 3.13 (*GARCH Model*)**.** Arch and Garch Models have been introduced by Engle [20] and Bollerslev [11], and have had a major impact on economic time series analysis. Bollerslev [12] suggested the following constant conditional correlation (CCC) model as a generalization to the multivariate case. Denote with $\circ$ the Hadamard product, and let $\{\boldsymbol{\epsilon}_k\}_{k\in\mathbb{Z}}$ be a d-dimensional I.I.D. sequence, where we denote the single elements with $\epsilon_{k,h}$, $1\leq h\leq d$. Define the process $\{\mathbf{Y}\}_{k\in\mathbb{Z}}$ as

$$\mathbf{Y}_k=\boldsymbol{\sigma}_k\circ\boldsymbol{\epsilon}_k, \tag{3.9}$$

$$\boldsymbol{\sigma}_k\circ\boldsymbol{\sigma}_k=\boldsymbol{\mu}+\sum_{j=1}^{p^*}\boldsymbol{\alpha}_j\circ\boldsymbol{\alpha}_{k-j}\circ\boldsymbol{\sigma}_{k-j}+\sum_{j=1}^{q^*}\boldsymbol{\beta}_j\circ\mathbf{Y}_{k-j}\circ\mathbf{Y}_{k-j}, \tag{3.10}$$

where $\boldsymbol{\mu}$ is coordinate wise strictly positive, $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_{p^*}$ and $\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_{q^*}$ are coordinate wise nonnegative d-dimensional vectors, and $p^*,q^*>0$. The process $\{\mathbf{Y}_k\}_{k\in\mathbb{Z}}$ reflects the structure of an univariate GARCH(p^*,q^*) time series, and in fact each coordinate represents a one-dimensional GARCH equation, whose orders are at most (p^*,q^*). Let $\boldsymbol{\alpha}_i=\big(\alpha_{i,1},\ldots,\alpha_{i,d}\big)^t$ and $\boldsymbol{\beta}_i=\big(\beta_{i,1},\ldots,\beta_{i,d}\big)^t$. Proceeding as in the proof of Theorem 4.2 in [4] combined with the computations in Example 3.12, one derives the conditions

- $d=\mathcal{O}\big((\log n)^{\delta}\big)$, for some $\delta>0$,
- $\max_{1\leq h\leq d}\left\|\epsilon_{0,h}\right\|_p<\infty$, $p>8$.
- $\max_{1\leq h\leq d}\sum_{j=1}^{\max\{p^*,q^*\}}\left\|\alpha_{j,h}+\beta_{j,h}\epsilon_{0,h}^2\right\|_{p/2}<1$ for $p>8$,

where $\boldsymbol{\alpha}_i=\big(\alpha_{i,1},\ldots,\alpha_{i,d}\big)^t$ and $\boldsymbol{\beta}_i=\big(\beta_{i,1},\ldots,\beta_{i,d}\big)^t$, and possible undefined spots are filled up with zeros. These conditions ensure the validity of Assumption 3.1(i)–(iii) in case of $X_{k,h}^{(1)}$. As in the previous Example 3.12, one only needs to replace the condition $p>8$ with $p>16$ to deal with $X_{k,h}^{(2)}$, and similarly it seems to be impossible to give general, easy verifiable conditions to ensure the validity of Assumption 3.1(iv).

An extension of the CCC-GARCH model was introduced by Jeantheau [30] by replacing the vectors $\boldsymbol{\alpha}_i$, $\boldsymbol{\beta}_i$ with matrices $\mathbf{A}_i$, $\mathbf{B}_i$. As before one can derive sufficient conditions by following the proof of Theorem 4.3 in [4] combined with the computations in Example 3.12.

Example 3.14 (*Iterative Random Function Model*)**.** Elton and Diaconis and Freedman introduced the following generalization of the AR(1) process. Let $\{\boldsymbol{\epsilon}_k\}_{k\in\mathbb{Z}}$ be a d-dimensional I.I.D. sequence, where we denote the single elements with $\epsilon_{k,h}$, $1\leq h\leq d$, and $R_h\big(,\big)$ be a collection of functions from $\mathbb{R}\times\mathbb{R}^{\mathbb{Z}}\mapsto\mathbb{R}$. We define the iterated random process as

$$Y_{k,i}=R_h\big(Y_{k-1,i},\boldsymbol{\epsilon}_k\big),\quad k\in\mathbb{Z},1\leq h\leq d. \tag{3.11}$$

Denote with

$$\mathcal{L}_{\boldsymbol{\epsilon},h}=\sup_{x\neq y}\frac{\left|R_h\big(x,\boldsymbol{\epsilon}\big)-R_h\big(y,\boldsymbol{\epsilon}\big)\right|}{\left|x-y\right|}$$

the Lipschitz coefficient, and suppose that

- $\max_{1\le h\le d}\left\| Y_{0,h}^{(1)}\right\|_p < \infty$, for $p > 8$,
- $\max_{1\le h\le d}\left\| \mathcal{L}_{\boldsymbol{\epsilon},h}\right\|_p \le r < 1$.

Since for $k > 0$

$$\left\|X_{k,h}^{(1)} - X_{k,h}^{(1),'}\right\|_p \le \left\|\mathcal{L}_{\boldsymbol{\epsilon}_k,h}\right\|_p \left\|X_{k-1,h}^{(1)} - X_{k-1,h}^{(1),'}\right\|_p \le \cdots \le \prod_{j=0}^{k-1}\left\|\mathcal{L}_{\boldsymbol{\epsilon}_{k-j},h}\right\|_p \left\|X_{0,h}^{(1)} - X_{0,h}^{(1),'}\right\|_p \le Cr^k \left\|Y_{0,h}^{(1)}\right\|_p,$$

it follows that conditions (i)–(iii) of Assumption 3.1 are satisfied. As in the previous examples, we cannot give simple yet general conditions that ensure the validity of Assumption 3.1(iv). As before one only needs to replace the condition $p > 8$ with $p > 16$ to deal with $X_{k,h}^{(2)}$.

Example 3.15 (*Another Linear Model*)**.** Finally, we present a model where Assumption 3.1(iv) is verifiable via a simple condition. Consider the following multivariate MA(∞) process: Let

$$X_{k,h} = \sum_{i=0}^{\infty} a_i L_{k-i,h} + \mu_h, \tag{3.12}$$

where $\mu_h \in \mathbb{R}$, $\left\{L_{k,h}\right\}_{k\in\mathbb{Z}}$ is an $\mathbb{R}^\infty$ valued I.I.D. sequence, and $\left\{L_{k,h}\right\}_{h\in\mathbb{Z}}$ is an AR(d) process with parameter $\boldsymbol{\zeta} = \left(1, \zeta_1, \ldots, \zeta_d\right)$ for every fixed k, i.e.

$$L_{k,h} = \zeta_1 L_{k,h-1} + \cdots + \zeta_d L_{k,h-d} + \epsilon_{k,h}, \tag{3.13}$$

where $\left\{\epsilon_{k,h}\right\}_{h\in\mathbb{Z}}$ is a zero mean white noise sequence, i.e. it holds that $\mathbb{E}\left(\epsilon_{k,i}\epsilon_{k,j}\right) = 0$ for $i \ne j$. Clearly, for fixed $k \in \mathbb{Z}$, the single components $X_{k,i}, X_{k,j}$ are dependent, and we have

$$n^{-1}\mathrm{Cov}\left(S_i^{(n)}, S_j^{(n)}\right) = n^{-1}\sum_{k=1}^{n}\sum_{\substack{r,s=0,\\ 1\le k-r+s\le n}}^{\infty} a_r a_s \mathbb{E}\left(L_{0,i}L_{0,j}\right),$$

where $S_h^{(n)} = \sum_{k=1}^n X_{k,h}$. Suppose in addition that $\mathcal{L} = \lim_{n\to\infty} n^{-1}\sum_{k=1}^n \sum_{\substack{r,s=0,\\ 1\le k-r+s\le n}}^{\infty} a_r a_s \ne 0$ and $\sum_{j=1}^q |\zeta_j| \le \vartheta < 1$. Then we obtain from [13, Theorem 4.4.2] that the spectral density function $f_L(\lambda)$ is given as

$$f_L(\lambda) = \frac{\sigma^2}{2\pi\left|\underline{\zeta}\left(e^{-i\lambda}\right)\right|^2},$$

where $\underline{\zeta}(s) = 1 - \sum_{j=1}^q \zeta_j s^j$. Since $\underline{\zeta}\left(e^{-i\lambda}\right) = 1 - \sum_{j=1}^q \zeta_i e^{-i\lambda j}$, it holds that

$$0 < \left|1 - \sum_{j=1}^q |\zeta_j|\right|^2 \le \left|\zeta\left(e^{-i\lambda}\right)\right|^2 \le \left|1 + \sum_{j=1}^q |\zeta_j|\right|^2 < \infty,$$

hence we obtain from Lemma 3.2 that the eigenvalues of the covariance matrix $\boldsymbol{\Gamma}_{\mathbf{W}^{(d)}}$ are uniformly positively bounded from below and above. Hence if we assume that

- $d = \mathcal{O}\left((\log n)^\delta\right)$, for some $\delta > 0$,
- $\max_{1\le h\le d}\left\|\epsilon_{0,h}\right\|_p < \infty$, $p > 8$.
- $\mathcal{L} \ne 0$, $\left|a_k\right| = \mathcal{O}\left(k^{-\beta}\right)$, where $\beta > (3+\sqrt{3})(2\sqrt{3}-2)^{-1} \approx 3.232\cdots$,
- $\sum_{j=1}^q |\zeta_j| < 1$,

then Assumption 3.1(i)–(iv) is valid in case of $X_{k,h}^{(1)}$. As before, we only need to replace condition $p > 8$ with $p > 16$ to deal with $X_{k,h}^{(2)}$.

Example 3.16 (*Additional Examples*)**.** Mimicking the proofs in [4], one readily provides additional examples such as Dynamic factor models and Multivariate exponential Garch models.

4. Numerical examples

In this section, we present a short simulation study to some of the examples discussed before. The focus of this study is to investigate the relation of the dimension d_n and the power of Λ_n, Ω_n. We do not present results comparing weighted

Table 1
Simulation of a d-dimensional CCC GARCH(1,1) with 1000 repetitions.

$\boldsymbol{\mu}_k*$	FULL $n = 200$								
	$d = 10$			$d = 20$			$d = 50$		
	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$
0	7.7	4.6	0.7	8.1	4.4	0.8	7.2	4.2	1
0.01**e**	12.1	8.6	0.9	16.2	10.1	0.9	5.8	3.6	1.2
0.02**e**	13.8	8.8	1.1	18.3	12.1	0.8	4.3	1.6	3
0.03**e**	22.9	14.7	1.2	29.7	16.7	1.4	5.5	2.7	2.9

Table 2
Simulation of a d-dimensional CCC GARCH(1,1) process with 1000 repetitions.

$\boldsymbol{\mu}_k*$	FULL $n = 500$								
	$d = 10$			$d = 20$			$d = 50$		
	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$
0	9.6	4.8	0.6	8.3	4.9	0.7	9.2	4.2	0.8
0.01**e**	11.8	6.3	0.78	17.1	5.8	0.9	18.5	4.0	1.4
0.02**e**	29.4	18.1	4.7	35.2	19.4	5	40.5	21.6	7.1
0.03**e**	54.5	48.3	20.7	63.7	55.8	23.8	80.7	59.3	27.8

statistics to non-weighted ones, as the literature provides numerous examples (cf. [16,14,7]) on this subject. We will therefore always set $w(t) \equiv 1$, and only consider the case where the time of change is the 'middle', i.e; $\tau = 1/2$. Similarly, we do not provide examples to illustrate the rate of convergence in Corollaries 3.6 and 3.8. We just mention that the rate significantly depends on the choice of λ_n, for details we again refer to [16,14,7] and the references therein.

Due to the weak dependence of the underlying processes, we use a Bartlett-estimator for the long-run covariances matrix $\boldsymbol{\Gamma}_{\mathbf{W}^{(d)}}$ with window length $l_n = \lceil \log n \rceil$. We do not use any parametric or semi-parametric methods. The critical values for the corresponding statistics are empirical quantiles, which were obtained by sampling from a d-dimensional sequence of independent Gaussian zero mean random vectors $\mathbf{Z}_k$, where $\mathrm{Cov}\big(\mathbf{Z}_k, \mathbf{Z}_k\big) = \mathbf{I}$, where $\mathbf{I}$ denotes the identical $d \times d$ matrix.

As our first example, we consider the CCC GARCH model of Example 3.13, with specifications $p^* = q^* = 1, \alpha_{1,h} = \beta_{1,h} = 0.1$. We let $\boldsymbol{\epsilon}_k \sim \mathcal{N}(0, \sum)$ be Gaussian, where

$$\sum := \begin{pmatrix} 1 & 0.5 & 0 & \cdots & \cdots \\ 0.5 & 1 & 0.5 & \cdots & \cdots \\ 0 & 0.5 & 1 & 0.5 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix},$$

which thus imposes a slight dependence among the component-wise GARCH-processes. To assess the power, we introduce several alternatives. At the change point $\tau = 1/2$, the drift $\boldsymbol{\mu}_k^* = \delta\,\mathbf{e}$ is added, where $\delta \in \{0, 0.01, 0.02, 0.03\}$ and $\mathbf{e} = \big(1, \ldots, 1\big)^t$. The simulations are carried out for the dimensions $d \in \{10, 20, 50\}$, with a sample size of $n = 200$ or $n = 500$. For each scenario, the experiment was repeated a thousand times, and a burn-in phase of 500 iterations was used. The results are shown in Tables 1 and 2, where we only display the results of the statistic Ω_n. The results of Λ_n are quite similar, slight differences can only be observed in case of $d = 50$, where Λ_n performs a little better. This difference is worked out more thoroughly in our next experiment. As can be seen from the results in Tables 1 and 2, we observe an increase in power as the dimension and sample size increase, which is in accordance with the theory. A notable exception is the case $d = 50, n = 200$, where the statistic Λ_n performs rather poor. This can be explained by the relatively large dimension $d = 50$, compared to the sample size $n = 200$ and the changes $\boldsymbol{\mu}_k^*$.

In Tables 3 and 4, the corresponding results are displayed in case of $\mathbf{e} = \big(1, \ldots, 1, 0, \ldots, 0\big)^t$, where only the first $\lceil \log n \rceil$ elements are non-zero. This clearly leads to a decrease in power as the dimension d increases.

For the second experiment, we consider a d-dimensional AR(1) process

$$\mathbf{Y}_k = \mathbf{A}\mathbf{Y}_{k-1} + \boldsymbol{\epsilon}_k, \quad k = 1, \ldots, n,$$

where $\mathbf{A} = 0.1\mathbf{I}$, $\boldsymbol{\epsilon}_k \sim \mathcal{N}\big(0, \mathbf{I}\big)$, where $\mathbf{I}$ denotes the $d \times d$ identity matrix. Again, we introduce several alternatives. At the change point $\tau = 1/2$, the matrix $\mathbf{A}$ changes to $\mathbf{A}^* = 0.1\mathbf{I} + \delta\mathbf{E}$, where δ takes on values which depend on the dimension d, and $\mathbf{E}$ denotes the $d \times d$ matrix for which all entries are equal to 1, thus introducing a correlation structure. This example is also considered in Aue et al. [4] for $d = 4$. As outlined in Section 3.2, testing for changes in the cross-sectional covariance structure for a d-dimensional process results in a $\mathfrak{d} = d(d+1)/2$ dimensional statistic. We consider the cases $d \in \{4, 6, 10\}$ which results in $\mathfrak{d} \in \{10, 21, 55\}$, and a sample size of $n = 200$. The results are displayed in Tables 5 and 6. As can be seen, we clearly have an increase in power as the dimension d increases. Note that δ varies for each d, and becomes smaller as the dimension d increases. This is necessary since otherwise the process $\{\mathbf{Y}_k\}_{k\in\mathbb{Z}}$ is no longer stationary under the alternative matrix $\mathbf{A}^*$.

Table 3
Simulation of a d-dimensional CCC GARCH(1,1) process 1000 repetitions.

$\boldsymbol{\mu}_k*$	SPARSE $n = 200$								
	$d = 10$			$d = 20$			$d = 50$		
	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$
0	7.7	4.6	0.7	8.1	4.4	0.8	7.2	4.2	1
0.03**e**	21.4	10.1	1.2	13.8	6.2	1.1	6.4	3.6	1.8
0.04**e**	24.1	14.2	3.4	15.5	7.8	3.5	6.5	4.2	2.6
0.05**e**	31.6	22.7	6.1	20	9.3	3.9	7.1	3.8	2.8

Table 4
Simulation of a d-dimensional CCC GARCH(1,1) process with 1000 repetitions.

$\boldsymbol{\mu}_k*$	SPARSE $n = 500$								
	$d = 10$			$d = 20$			$d = 50$		
	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$
0	9.6	4.8	0.6	8.3	4.9	0.7	9.2	4.2	0.8
0.03**e**	36.2	25.1	9.3	31.6	22.8	4.1	15.4	6.6	2.8
0.04**e**	57.4	48.5	24.9	53.1	36.5	11.3	28.5	17.4	3.7
0.05**e**	84.7	75.3	48.8	72.7	55.2	29.9	34.2	25.8	8.2

Table 5
Simulation of a d-dimensional AR(1) process with 1000 repetitions.

Ω_n FULL $n = 200$											
$d = 4$				$d = 6$				$d = 10$			
$\delta \setminus \alpha$	10%	5%	1%	$\delta \setminus \alpha$	10%	5%	1%	$\delta \setminus \alpha$	10%	5%	1%
0	8	3.5	0.8	0	9	4.8	0.9	0	8.7	4.6	0.45
0.1	10.7	5.1	2.2	0.1	23.4	13.8	1.3	0.03	10.7	5.9	3.4
0.15	33.5	24.3	9.1	0.125	37.4	27.7	10	0.05	13.9	8.6	5.1
0.2	64	52.2	34.6	0.135	50.1	36.6	15.7	0.07	13.8	9.6	4.9
0.22	82.2	78.3	58.5	0.15	74.8	66.3	55.1	0.09	39.7	9.2	6.3

Table 6
Simulation of a d-dimensional AR(1) process with 1000 repetitions.

Λ_n FULL $n = 200$											
$d = 4$				$d = 6$				$d = 10$			
$\delta \setminus \alpha$	10%	5%	1%	$\delta \setminus \alpha$	10%	5%	1%	$\delta \setminus \alpha$	10%	5%	1%
0	9.5	4.5	0.6	0	9.5	4.9	0.7	0	9.2	4.7	0.8
0.1	14.6	7.2	2.1	0.1	29.9	12.3	2.4	0.03	12.8	7.5	1.3
0.15	34.8	23.3	9.2	0.125	48.4	32.7	14	0.05	14.3	10.7	1.6
0.2	63.4	52.7	35.2	0.135	57.3	40.7	26.6	0.07	20.1	9.3	2.8
0.22	79.6	72.1	54.8	0.15	76.4	64.3	59.1	0.09	54.9	41.7	3.1

It is interesting to note that for $d = 10$, the test based on Λ_n significantly outperforms the one based on Ω_n. Looking at the critical values, one finds that the tails of the distribution function $P\big(\Omega_n \leq x\big)$ are much heavier than those of $P\big(\Lambda_n \leq x\big)$. This can also be deduced from Remark 2.1 in [4]. Also note that the power is largest for $d = 10$ ($\eth = 55$), which is in contrast to the results of Table 1. This is because the changes in the cross-sectional covariance structure in this experiment have a much larger numerical and structural impact on the process $\{\mathbf{Y}_k\}_{k\in\mathbb{Z}}$. For example, for the pairs ($d = 4, \delta = 0.22$), ($d = 6$, $\delta = 0.15$), and ($d = 10, \delta = 0.09$), the process $\{\mathbf{Y}_k\}_{k\in\mathbb{Z}}$ is 'close' to non-stationarity under the alternative matrix $\mathbf{A}^*$.

In Tables 7 and 8, we display the corresponding results under a 'sparse' alternative matrix $\mathbf{A}^* = 0.1\mathbf{I} + \delta\mathbf{E}^*$, where $\delta \in \{0, 0.3, 0.35, 0.4, 0.45\}$, and $\mathbf{E}^*$ denotes the $d \times d$ matrix

$$\mathbf{E}^* := \begin{pmatrix} 1 & 1 & 0 & 0 & \cdots \\ 0 & 1 & 1 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix}.$$

In case of Ω_n, we observe a decrease of power in Table 7 as d increases. However, in case of Λ_n (Table 8) we first observe a slight increase in power from $\eth = 10$ to $\eth = 21$, before it drops below both previous levels for $\eth = 55$. In addition, its performance is significantly better than the test based on Ω_n, particularly in the case $\eth = 55$. Again, this is due to the fact of the heavier tails of the distribution function $P\big(\Omega_n \leq x\big)$.

Table 7
Simulation of a d-dimensional AR(1) process with 1000 repetitions.

Ω_n SPARSE $n = 200$											
$d = 4$				$d = 6$				$d = 10$			
$\delta \setminus \alpha$	10%	5%	1%	$\delta \setminus \alpha$	10%	5%	1%	$\delta \setminus \alpha$	10%	5%	1%
0	8	3.5	0.8	0	9	4.8	0.9	0	8.7	4.6	0.45
0.3	40.5	23.4	6.6	0.3	31.8	17.9	6.1	0.3	12.3	4.3	1
0.35	59.3	43.5	19.1	0.35	52.2	33	11.7	0.35	13.7	4.6	1.7
0.4	66.2	50.9	28.3	0.4	76.9	54.3	23.4	0.4	14	4.5	1.6
0.45	87.8	75.7	48.6	0.45	86.3	65.6	39.8	0.45	18.3	5.1	1.9

Table 8
Simulation of a d-dimensional AR(1) process with 1000 repetitions.

Λ_n SPARSE $n = 200$											
$d = 4$				$d = 6$				$d = 10$			
$\delta \setminus \alpha$	10%	5%	1%	$\delta \setminus \alpha$	10%	5%	1%	$\delta \setminus \alpha$	10%	5%	1%
0	9.5	4.5	0.6	0	9.5	4.9	0.7	0	9.2	4.7	0.8
0.3	43.8	27.2	13.1	0.3	40	30.5	11.6	0.3	30.2	12.5	2.5
0.35	62.6	49.3	25.4	0.35	63.8	51	27.8	0.35	47.7	20.8	2.3
0.4	69.8	55.7	36.9	0.4	81.1	76.5	43.4	0.4	57.4	31.3	3.4
0.45	85.3	74.8	55.6	0.45	92.6	83.3	60.5	0.45	74.1	44.5	4.8

5. Proofs and ramifications

Throughout the proofs, C denotes a generic constant that may vary from one formula to another. The proofs are essentially based on Theorem 5.4 given below, whose proof is based on results given in Section 6. We make the following assumption.

Assumption 5.1. For $m = m_n = \mathcal{O}\left(n^{\theta}\right)$, $0 < \theta < 1$, $d = d_n = \mathcal{O}\left(n^{\delta}\right)$, $\delta > 0$ we suppose that

(i) $\sup_h \|X_{1,h}\|_p < \infty, \quad \text{for some } p > 8, \qquad \mathbb{E}\big(X_{k,h}\big) = \mu_h, \quad \text{for all } 1 \leq h \leq d_n,$

(ii) $\sup_h \sum_{j=0}^{\infty} j|\phi_{j,h}| < \infty,$

(iii) $\limsup_{n\to\infty} \max_{1\leq h\leq d_n} \max_{1\leq l\leq n} \left\| \sum_{j=1}^{l} Y_{j,h}^{(>m_n)} \right\|_p = \mathcal{O}\left(1\right), \quad p > 8.$

Remark 5.2. Note that Assumption 5.1(ii) implies that

$$\psi_h^2 = \lim_n n^{-1}\mathrm{Var}\left(\sum_{1\leq k\leq n} X_{k,h}\right) < \infty, \tag{5.1}$$

and in particular $\sup_h \psi_h^2 < \infty$.

The link between Assumptions 3.1 and 5.1 is provided by the following proposition.

Proposition 5.3. *Suppose that* $\max_{1\leq h\leq d_n}\left\|X_{k,h} - X'_{k,h}\right\|_p = \mathcal{O}\left(k^{-\beta}\right)$, *with* $\theta \geq \frac{2}{2\beta-1}$, $\theta \geq \delta$, $p > 8$, *where* $\beta > 5/2$. *Then Assumption* 5.1(ii), (iii) *are valid. If* $d_n = \mathcal{O}\big((\log n)^{\delta}\big)$, *we do not require that* $\theta \geq \delta$.

We can now present the main approximation result.

Theorem 5.4. *Suppose that Assumption* 5.1(i)–(iii) *holds, and let* $\boldsymbol{\Gamma}_n$ *be a sequence of regular matrices such that* $\max|\boldsymbol{\Gamma}_n^{-1}| \leq L_n$ *for some sequence* L_n. *Then on a possible larger probability space, we have that*

(i) $\left| \sup_{\lambda_n/n\leq t\leq 1-\lambda_n/n} \left|w(t)^{-1}\big(\mathbf{M}_t^{(n)}\big)^t \boldsymbol{\Gamma}_n^{-1}\mathbf{M}_t^{(n)}\right| - \sup_{\lambda_n/n\leq t\leq 1-\lambda_n/n} \left|w(t)^{-1}\mathbf{B}_t^t \boldsymbol{\Gamma}_n^{-1}\mathbf{B}_t\right| \right| = \mathcal{O}_p\big(\sqrt{d_n}\big),$

(ii) $\sup_{\lambda_n/n\leq t\leq 1-\lambda_n/n} \left|w(t)^{-1}\big(\mathbf{M}_t^{(n)}\big)^t \boldsymbol{\Gamma}_n^{-1}\mathbf{M}_t^{(n)} - w(t)^{-1}\mathbf{B}_t^t \boldsymbol{\Gamma}_n^{-1}\mathbf{B}_t\right| = \mathcal{O}_p\big(\sqrt{d_n}\big).$

The dimension $d_n = \mathcal{O}(n^\delta)$ must satisfy the relation

$$\theta \leq \delta < \min\left\{\frac{4(p-2\nu)}{(-4+3p)\nu}, \frac{2+p-2(1+\theta)\nu}{(2+4\theta+p(4+\theta))\nu}\right\}, \tag{5.2}$$

where we require $p > 4\nu$ and $L_n d_n^{1/2} = \mathcal{O}\left((\log n)^{-\kappa}\right)$, for some $\kappa > 0$. Moreover, it holds that $\max\left|\boldsymbol{\Gamma}_{\mathbf{S}^{(n)}} - \boldsymbol{\Gamma}_{\mathbf{W}^{(n)}}\right| = \mathcal{O}(n^{-\gamma})$, for some $\gamma > 0$. Alternatively, if one sets $d_n = \mathcal{O}\left((\log n)^\delta\right)$, for arbitrary $\delta > 0$, then we require

$$2 < \min\left\{\frac{4+2p}{4+4\theta+p\theta+4\theta^2+p\theta^2}, \frac{2}{1-2\theta}, p/4\right\}, \tag{5.3}$$

and $L_n d_n^{1/2} = \mathcal{O}\left((\log\log n)^{-\kappa}\right)$, for some $\kappa > 0$.

Remark 5.5. Conditions (i) and (iii) of Assumption 5.1 can in fact be weakened to $p > 4$. This, however, leads to a less tractable bound for δ.

Before proving the results of Section 3, we will show the validity of Proposition 5.3.

Proof of Proposition 5.3. First, note that condition $\theta \geq \delta$ is only needed to ensure that $X_{k,d}$ is $\mathcal{F}_{k+m}$-measurable. Clearly this is no longer necessary if $d_n = \mathcal{O}\left((\log n)^\delta\right)$, since then $X_{k,d}$ is $\mathcal{F}_{k+m}$-measurable for large enough n. Let $\mathcal{F}_k' = \sigma(\epsilon_k', \epsilon_{k-1}', \ldots)$. Then for any $p \geq 1$ we have by Jensen's and the triangular inequality

$$\begin{aligned}\left\|Y_{k,h}^{(>m_n)}\right\|_p &\leq \left\|X_{k,h} - X_{k,h}^{(m_n+1,*)}\right\|_p + \left\|\mathbb{E}\left(X_{k,h} - X_{k,h}^{(m_n+1,*)} \mid \sigma\left(\mathcal{F}_{k-m_n}^{k+m_n} \cup \mathcal{F}_{k-m_n-1}'\right)\right)\right\|_p \\ &\leq 2\left\|X_{k,h} - X_{k,h}^{(m_n+1,*)}\right\|_p,\end{aligned}$$

where we also used the fact that

$$X_{k,h}^{(m_n+1,*)} - \mathbb{E}\left(X_{k,h}^{(m_n+1,*)} \mid \mathcal{F}_{k-m_n}^{k+m_n}\right) = \mathbb{E}\left(X_{k,h}^{(m_n+1,*)} - X_{k,h} \mid \sigma\left(\mathcal{F}_{k-m_n}^{k+m_n} \cup \mathcal{F}_{k-m_n-1}'\right)\right).$$

By [54, Theorem 1(iii)] we have for $p \geq 2$

$$\left\|X_{1,h} - X_{1,h}^{(m_n+1,*)}\right\|_p^2 \leq C \sum_{i=-\infty}^{0} \left\|X_{m_n+1-i,h} - X_{m_n+1-i,h}'\right\|_p^2 = \mathcal{O}\left(m_n^{1-2\beta}\right), \tag{5.4}$$

which leads to

$$\left\|Y_{k,h}^{(>m_n)}\right\|_p = \mathcal{O}\left(m_n^{1/2-\beta}\right). \tag{5.5}$$

Consequently, using the triangular inequality and the above, we obtain that

$$\max_{1\leq l\leq n}\left\|\sum_{j=1}^{l} Y_{j,h}^{(>m_n)}\right\|_p \leq \sum_{k=1}^{n}\left\|Y_{k,h}^{(>m_n)}\right\|_p = \mathcal{O}\left(n\, m_n^{1/2-\beta}\right) = \mathcal{O}(1),$$

which proves Assumption 5.1(iii). In order to show (ii), note that the Cauchy–Schwarz inequality implies

$$\begin{aligned}\left|\mathbb{E}\left((X_{k,h}-\mu_h)(X_{0,h}-\mu_0)\right)\right| &= \left|\mathbb{E}(X_{k,0}-\mu_0)\mathbb{E}\left((X_{k,h}-\mu_h) \mid \mathcal{F}_d\right)\right| \leq 2\left\|X_{0,h}\right\|_2\left\|\mathbb{E}\left((X_{k,h}-\mu_h) \mid \mathcal{F}_d\right)\right\|_2 \\ &\leq 2\left\|X_{0,h}\right\|_2\left\|X_{k,h} - X_{k,h}^{(k,*)}\right\|_2,\end{aligned}$$

and it follows from (5.4) that

$$\sum_{j=0}^{\infty} j|\phi_{j,h}| \leq C\sum_{j=0}^{\infty} j\left\|X_{j,h} - X_{j,h}^{(j,*)}\right\|_2 = \mathcal{O}\left(\sum_{j=1}^{\infty} j^{3/2-\beta}\right) = \mathcal{O}(1). \quad \square$$

5.1. Proofs of Section 3

The proof of Theorem 3.3 will be developed in a series of Lemmas. The difficulty mainly consists of controlling the error of $\max\left|\widehat{\boldsymbol{\Gamma}}_{\mathbf{S}^{(n)}}^{-1} - \boldsymbol{\Gamma}_{\mathbf{W}^{(n)}}^{-1}\right|$. To this end, we require the following Lemma.

Lemma 5.6. *Let $\mathbf{A}$, $\mathbf{B}$ be two regular $d \times d$ dimensional matrices, such that*

- $\max\left|\mathbf{A} - \mathbf{B}\right| = \mathcal{O}\left(d^{-\gamma}\right)$, *for some $\gamma > 0$,*

- *The smallest eigenvalue $\sigma_{\min}(\mathbf{A})$ of $\mathbf{A}$ is positive and satisfies $\sigma_{\min}(\mathbf{A})^{-1} = \mathcal{O}(d^{\kappa})$, $\kappa > 0$, such that $d^{\kappa+1/2} = \mathcal{O}(d^{\gamma})$. Then for large enough d, it holds that*

(i) $\max|\mathbf{A}| = \mathcal{O}(d^{\kappa+1/2})$,
(ii) $\max|\mathbf{A}^{-1} - \mathbf{B}^{-1}| = \mathcal{O}(d^{2\kappa+1-\gamma})$.

We are now ready to prove Theorem 3.3.

Proof of Theorem 3.3. First note that by Proposition 5.3, the assumptions of Theorem 5.4 are validated. For a function $f(t)$, we denote with $\lfloor f \rfloor_n^* := \sup_{\lambda/n \le t \le 1-\lambda/n} w(t)^{-1}|f(t)|$, where $w(t) = t(1-t)$. Define the random vector

$$\mathbf{B}_{w,n} = \left(B_1^{(\lambda/n)}, B_2^{(\lambda/n)}, \ldots, B_{d_n}^{(\lambda/n)}\right)^t,$$

where $B_h^{(\lambda/n)} = \lfloor B_{t,h} \rfloor_n^*$.

Next, notice that

$$\left| \lfloor \mathbf{M}_t^{(n)} \boldsymbol{\Gamma}_{\mathbf{S}^{(n)}}^{-1} \mathbf{M}_t^{(n)} \rfloor_n^* - \lfloor \mathbf{M}_t^{(n)} \widehat{\boldsymbol{\Gamma}}_{\mathbf{S}^{(n)}}^{-1} \mathbf{M}_t^{(n)} \rfloor_n^* \right| \le \left| \lfloor \mathbf{M}_t^{(n)} (\boldsymbol{\Gamma}_{\mathbf{S}^{(n)}}^{-1} - \widehat{\boldsymbol{\Gamma}}_{\mathbf{S}^{(n)}}^{-1}) \mathbf{M}_t^{(n)} \rfloor_n^* \right| \le d_n \max\left| \boldsymbol{\Gamma}_{\mathbf{S}^{(n)}}^{-1} - \widehat{\boldsymbol{\Gamma}}_{\mathbf{S}^{(n)}}^{-1} \right| \max\left|\mathbf{Z}_n\right|^2. \tag{5.6}$$

We have $\max\left|\boldsymbol{\Gamma}_{\mathbf{S}^{(n)}} - \boldsymbol{\Gamma}_{\mathbf{W}^{(n)}}\right| = \mathcal{O}(n^{-\gamma})$, and combining this with Assumption 3.1(iii), (iv), we obtain from Lemma 5.6 (a slight adaption is necessary)

$$P\left(\max\left|\widehat{\boldsymbol{\Gamma}}_{\mathbf{S}^{(n)}}^{-1} - \boldsymbol{\Gamma}_{\mathbf{W}^{(n)}}^{-1}\right| \ge d_n^{2\kappa+1} n^{-(\gamma\wedge\chi)}\right) = \mathcal{O}(1). \tag{5.7}$$

Hence (5.6) implies that for $\epsilon > 0$ and large enough n

$$\begin{aligned} P\left(\left| \lfloor \mathbf{M}_t^{(n)} \widehat{\boldsymbol{\Gamma}}_{\mathbf{S}^{(n)}}^{-1} \mathbf{M}_t^{(n)} \rfloor_n^* - \lfloor \mathbf{M}_t^{(n)} \boldsymbol{\Gamma}_{\mathbf{W}^{(n)}}^{-1} \mathbf{M}_t^{(n)} \rfloor_n^* \right| > \sqrt{d_n}\epsilon\right) &\le \mathcal{O}(1) + P\left(\max\left|\mathbf{Z}_n\right|^2 \ge \epsilon d_n^{-2\kappa-3/2} n^{\gamma\wedge\delta}\right) \\ &\le \mathcal{O}(1) + d_n P\left(\left|Z_1^{(\lambda/n,n)}\right|^2 \ge \epsilon d_n^{-2\kappa-3/2} n^{\gamma\wedge\chi}\right). \end{aligned}$$

Moreover, since $d_n^{-2\kappa-3/2} n^{\gamma\wedge\chi} = \mathcal{O}(n^{\eta})$ for some $\eta > 0$, Theorem 5.4(i) implies that

$$P\left(\max\left|\mathbf{Z}_n\right|^2 \ge \epsilon d_n^{-2\kappa-3/2} n^{\gamma\wedge\chi}\right) \le \mathcal{O}_p\left(\sqrt{d_n}\right) + P\left(\max\left|\mathbf{B}_{w,n}\right|^2 \ge \epsilon d_n^{-2\kappa-3/2} n^{\gamma\wedge\chi}\right).$$

Let $u_n = n^2/\lambda_n^2 - n/\lambda_n + 1$. Then it holds that (cf. [15, A.3.19])

$$B_h^{(\lambda_n/n)} \stackrel{d}{=} \sup_{0 \le t \le \log u_n} \left|V_h(t)\right|, \quad 1 \le h \le d_n, \tag{5.8}$$

where $V_h(t)$ is a zero mean Ornstein–Uhlenbeck process with $\mathrm{Cov}\left(V_h(t), V_h(s)\right) = \exp(-|t-s|/2)$. Due to Theorem 1 in [41], it holds that for $p \ge 1$

$$\left\| \sup_{0 \le t \le \log u_n} \left|V_h(t)\right| (2 \log\log u_n)^{-1/2} \right\|_p \le C_p, \tag{5.9}$$

where C_p does not depend on n. Hence the Markov inequality and (5.9) imply that

$$\begin{aligned} P\left(\max\left|\mathbf{B}_{w,n}\right|^2 \ge \epsilon d_n^{-2\kappa-3/2} n^{\gamma\wedge\chi}\right) &\le d_n \max_{1 \le h \le d_n} P\left(\sup_{0 \le t \le \log u_n} \left|V_h(t)\right| \ge \epsilon^{1/2} d_n^{-\kappa-3/4} n^{(\gamma\wedge\chi)/2}\right) \\ &\le \epsilon^{-p/2} d_n \left\| \sup_{0 \le t \le \log u_n} \left|V_h(t)\right| (2\log\log u_n)^{-1/2} \right\|_p^p \left((2\log\log u_n)^{1/2} d_n^{\kappa+3/4} n^{-(\gamma\wedge\chi)/2}\right)^p \\ &= \mathcal{O}(1). \end{aligned}$$

Since $d_n^{-2\kappa-3/2} n^{\gamma\wedge\chi} = \mathcal{O}(n^{\eta})$ for some $\eta > 0$, we can thus choose a sequence ϵ_n that tends to zero such that

$$P\left(\left| \lfloor \mathbf{M}_t^{(n)} \widehat{\boldsymbol{\Gamma}}_{\mathbf{S}^{(n)}}^{-1} \mathbf{M}_t^{(n)} \rfloor_n^* - \lfloor \mathbf{M}_t^{(n)} \boldsymbol{\Gamma}_{\mathbf{W}^{(n)}}^{-1} \mathbf{M}_t^{(n)} \rfloor_n^* \right| > \epsilon_n \sqrt{d_n}\right) = \mathcal{O}(1). \tag{5.10}$$

Moreover, Theorem 5.4(i) implies that

$$\left| \sup_{\lambda_n/n \le t \le 1-\lambda_n/n} \left| w(t)^{-1} \left(\mathbf{M}_t^{(n)}\right)^t \boldsymbol{\Gamma}_{\mathbf{W}^{(n)}}^{-1} \mathbf{M}_t^{(n)} \right| - \sup_{\lambda_n/n \le t \le 1-\lambda_n/n} \left| w(t)^{-1} \sum_{h=1}^{d_n} \left(B_{t,h}^{(*)}\right)^2 \right| \right| = \mathcal{O}_p\left(\sqrt{d_n}\right), \tag{5.11}$$

provided that $(2\kappa + 1/2)\delta < (1/2 - 1/\nu)\lambda$. It thus remains to evaluate the bounds provided by Theorem 5.4. The condition $\theta < 1/2(\sqrt{3} - 1)$ implies that we may choose $\theta > (3 + \sqrt{3})(2\sqrt{3} - 2)^{-1}$, which completes the proof. □

The proof of Theorem 3.4 hinges on the following Lemma.

Lemma 5.7. *Put $w(t) = (t(1-t))$, and let $0 < l(n) < r(n) < 1$. Then*

$$\left\| \int_{l(n)}^{r(n)} \frac{B_t^2}{w(t)} dt - \frac{1}{n} \sum_{k=\lceil l(n)n \rceil}^{\lceil r(n)n \rceil} \frac{B_{k/n}^2}{w(k/n)} \right\|_1 = \mathcal{O}\left(n^{-1/2} (\log u(n))^2\right),$$

where $u(n) = \frac{1-l(n)r(n)}{l(n)(1-r(n))}$.

Proof of Theorem 3.4. Due to Theorem 5.4(ii), we have

$$\left| \frac{1}{n} \sum_{k=\lceil \lambda \rceil}^{\lceil n-\lambda \rceil} w(k/n)^{-1} \left(\mathbf{M}_{k/n}^{(n)}\right)^t \widehat{\boldsymbol{\Gamma}}_{\mathbf{S}^{(n)}}^{-1} \mathbf{M}_{k/n}^{(n)} - \sum_{h=1}^{d_n} \frac{1}{n} \sum_{k=\lceil \lambda \rceil}^{\lceil n-\lambda \rceil} w(k/n)^{-1} \left(B_{t,h}^{(*)}\right)^2 \right| = \mathcal{O}_p(1).$$

Setting $l(n) = \lambda/n$, $r(n) = 1 - \lambda/n$, Lemma 5.7 now yields the claim. □

Proof of Corollary 3.6. It holds that (cf. [14, A.3.19])

$$\sup_{\lambda \le t \le 1-\lambda} w(t)^{-1} \sum_{h=1}^{d_n} \left(B_{t,h}^{(*)}\right)^2 \overset{d}{=} \sup_{0 \le t \le \log u^*(\lambda)} \sum_{h=1}^{d_n} V_{t,h}^2, \quad u^*(\lambda) = \frac{1-\lambda^2}{\lambda(1-\lambda)},$$

where $\left\{V_{t,h}\right\}_{t \in \mathbb{R}}$ are independent zero mean Ornstein–Uhlenbeck process with $\mathrm{Cov}\left(V_{t,h}, V_{s,h}\right) = \exp(-|t-s|/2)$. Hence it suffices to establish that

$$(d_n 2)^{-1/2} \left(\sum_{h=1}^{d_n} V_{t,h}^2 - d_n \right) \overset{w}{\to} V(t)$$

on the space $\mathcal{C}[0, \log u^*(\lambda)]$. This, however, is provided by Ronzhin [41, Theorem 2]. □

Proof of Corollary 3.8. Let $u(n) = \frac{1-l(n)r(n)}{l(n)(1-r(n))}$. Due to Theorem 3.4, it suffices to establish that

$$(d_n \sigma^2 \log u_n)^{-1/2} \left(\sum_{h=1}^{d_n} \int_{\lambda/n}^{1-\lambda/n} w(t)^{-1} \left(B_{t,h}^{(*)}\right)^2 - \sqrt{\pi} d_n \right) \overset{w}{\to} \mathcal{N}(0, 1).$$

However, [15, Theorem A.3.5] implies that $X_{h,n} = \int_{\lambda/n}^{1-\lambda/n} w(t)^{-1} \left(B_{t,h}^{(*)}\right)^2 - \sqrt{\pi}$ is an array of I.I.D zero mean random variables with variance $\mathrm{Var}(X_{h,n}) = \sigma^2 \log u_n + \mathcal{O}\left(\log u_n\right)$. Hence the claim follows e.g. from [29, Theorem VII.2.35]. □

Proof of Lemma 5.6. For a matrix $\mathbf{A}$, denote with $\left|\mathbf{A}\right|_2$ the induced l_2-norm. It then holds that

$$\max\left|\mathbf{A}^{-1}\right| \le \sqrt{d} \left|\mathbf{A}^{-1}\right|_2 = \sqrt{d} \sigma_{\min}\left(\mathbf{A}\right)^{-1} = \mathcal{O}\left(d^{\kappa+1/2}\right),$$

which proves (i). Let $\mathbf{E} = \mathbf{B} - \mathbf{A}$, and note that for sufficiently large d, we have

$$\max\left|\mathbf{E}\right| \max\left|\mathbf{A}^{-1}\right| < 1.$$

Equation 1.5 in [45] then implies that

$$\max\left|\mathbf{A}^{-1} - \mathbf{B}^{-1}\right| = \max\left|\mathbf{A}^{-1} - \mathbf{E} + \mathbf{A}^{-1}\right| \le \frac{\max\left|\mathbf{E}\right| \max\left|\mathbf{A}^{-1}\right|^2}{1 - \max\left|\mathbf{E}\right| \max\left|\mathbf{A}^{-1}\right|} = \mathcal{O}\left(d^{2\kappa+1-\gamma}\right). \quad \square$$

Proof of Lemma 5.7. Put $u(n) = \frac{1-l(n)r(n)}{l(n)(1-r(n))}$. It holds that (cf. [15, A.3.19])

$$\left\{B_t^2 w(t)^{-1}\right\}_{l(n) \le t \le r(n)} \overset{d}{=} \left\{V(t)^2\right\}_{0 \le t \le u(n)}, \tag{5.12}$$

where $V(t)$ is a zero mean Ornstein–Uhlenbeck process with $\mathrm{Cov}\left(V(t), V(s)\right) = \exp(-|t-s|/2)$. Denote with $V_n(t) = V(k/n)$, $k \log u(n) \le t \le (k+1) \log u(n)$. Note that by using the properties of $V(t)$, one readily verifies that $\left\|V(t) - V_n(t)\right\|_2 = \mathcal{O}\left(n^{-1/2} \log u(n)\right)$. Using this together with the Cauchy–Schwarz inequality, one obtains

$$\begin{aligned} \left\| \int_0^{\log u(n)} V^2(t) - V_n^2(t) dt \right\|_1 &\le \int_0^{\log u(n)} \mathbb{E}\left|(V(t) - V_n(t))(V(t) + V_n(t))\right| dt \\ &\le C \int_0^{\log u(n)} \left\|V(t) - V_n(t)\right\|_2 dt = \mathcal{O}\left(n^{-1/2} (\log u(n))^2\right). \quad \square \end{aligned}$$

Proof of Theorem 3.9. As in the proof of Theorem 3.3 one derives that

$$\widehat{\boldsymbol{\Gamma}}_{\mathbf{S}^{(n)}}^{-1} = \tau \boldsymbol{\Gamma}_{\mathbf{S}^{(n,1)}}^{-1} + (1-\tau)\boldsymbol{\Gamma}_{\mathbf{S}^{(n,2)}}^{-1} + \mathcal{O}_p(1),$$

where the matrices $\boldsymbol{\Gamma}_{\mathbf{S}^{(n,1)}}, \boldsymbol{\Gamma}_{\mathbf{S}^{(n,2)}}$ denote the pre respectively post-change covariance matrices. Since $\limsup_{n\to\infty} \max_{1\leq i\leq d_n} \sum_{j=1}^{d_n} \big|\rho_{i,j}^{(n)}\big| < \infty$, we have that $\sigma_{\max}\big(\boldsymbol{\Gamma}_{\mathbf{S}^{(n,1)}}\big), \sigma_{\max}\big(\boldsymbol{\Gamma}_{\mathbf{S}^{(n,2)}}\big) < \infty$, uniformly in n. Since both $\boldsymbol{\Gamma}_{\mathbf{S}^{(n,1)}}, \boldsymbol{\Gamma}_{\mathbf{S}^{(n,2)}}$ are clearly symmetric, we have a corresponding system of orthogonal eigenvectors, and we thus obtain

$$\big(\boldsymbol{\mu}_{k^*+1}^{(n)}\big)^t \big(\tau \boldsymbol{\Gamma}_{\mathbf{S}^{(n,1)}}^{-1} + (1-\tau)\boldsymbol{\Gamma}_{\mathbf{S}^{(n,2)}}^{-1}\big)\boldsymbol{\mu}_{k^*+1}^{(n)} \geq \Big(\sigma_{\max}\big(\boldsymbol{\Gamma}_{\mathbf{S}^{(n,1)}}\big)^{-1}\tau + \sigma_{\max}\big(\boldsymbol{\Gamma}_{\mathbf{S}^{(n,2)}}\big)^{-1}(1-\tau)\Big)\left(\Big(\boldsymbol{\mu}_{k^*+1}^{(n)}\Big)^t \boldsymbol{\mu}_{k^*+1}^{(n)}\right).$$

Using Theorem 3.3 and Corollary 3.6, one thus obtains

$$\sup_{\lambda_n/n\leq t\leq 1-\lambda_n/n} \Big| w(t)^{-1}\big(\mathbf{M}_t^{(n)}\big)^t \widehat{\boldsymbol{\Gamma}}_{\mathbf{S}^{(n)}}^{-1}\mathbf{M}_t^{(n)}\Big| \geq \mathcal{O}_P\Big(n\big(\boldsymbol{\mu}_{k^*+1}^{(n)}\big)^t \boldsymbol{\mu}_{k^*+1}^{(n)}\Big),$$

hence the claim follows. □

Proof of Proposition 3.11. Using Boole's inequality, one may proceed exactly as in the proof of Proposition 1 in [51]. □

6. Gaussian approximation

Let $\{X_{k,h}\}_{k,h\geq 1}$ be a collection of random variables such that for each h_0, $\{X_{k,h_0}\}_{k\geq 1}$ is a zero mean stationary sequence. Throughout the proofs, C denotes a generic constant that may vary from one formula to another. Recall the notation

$$Y_{k,h}^{(\leq m)} = \mathbb{E}\big(X_{k,h} \mid \mathcal{F}_{k-m}^{k+m}\big), \tag{6.1}$$

$$Y_{k,h}^{(>m)} = X_{k,h} - Y_{k,h}^{(\leq m)} = X_{k,h} - \mathbb{E}\big(X_{k,h} \mid \mathcal{F}_{k-m}^{k+m}\big). \tag{6.2}$$

The Gaussian approximation is obtained under the following Assumption.

Assumption 6.1. For $m = m_n = n^{\theta}$, $0 < \theta < 1$, $d = d_n = n^{\delta}$, $0 < \delta$, $\psi_h > 0$ we suppose that

(i) $\limsup_{n\to\infty} \max_{1\leq h\leq d_n} \big\|Y_{k,h}^{(\leq m)}\big\|_p < \infty, \quad p > 8,$

(ii) $\limsup_{n\to\infty} \max_{1\leq h\leq d_n} \max_{1\leq l\leq n} \left\|\sum_{j=1}^{l} Y_{j,h}^{(>m_n)}\right\|_p = \mathcal{O}(1), \quad p > 8,$

(iii) $\limsup_{n\to\infty} \max_{1\leq h\leq d_n} \left|\mathrm{Var}\left(\sum_{k=1}^{n} X_{k,h}\right) - \psi_h n\right| < \infty.$

Remark 6.2. If the above assumptions hold for some $m = m_n = (\log n)^{\lambda}$, one can set $\theta = 0$ in all the conditions given below that involve θ.

Lemma 6.3. *Suppose that* $\sup_h \sum_{j=0}^{\infty} j|\phi_{j,h}| < \infty$. *Then Assumption* 6.1(iii) *holds, and* $\psi_h = \phi_{0,h} + 2\sum_{j=1}^{\infty}\phi_{j,h}$.

Proof of Lemma 6.3. We have

$$\begin{aligned}\mathrm{Var}\left(\sum_{i=1}^{n} X_{i,h}\right) &= \sum_{1\leq i,j\leq n} \phi_{|i-j|,h} = \sum_{i=1}^{n}\sum_{j=1-i}^{n-i} \phi_{|j|,h} = n\psi_h + \mathcal{O}\left(\sum_{i=0}^{\infty}\sum_{j=i}^{\infty} |\phi_{j,h}|\right)\\ &= n\psi_h + \mathcal{O}\left(\sup_h \sum_{j=0}^{\infty} j|\phi_{j,h}|\right) = n\psi_h + \mathcal{O}(1). \quad \square\end{aligned}$$

For a d_n-dimensional Brownian motion $\{\mathbf{W}_{\mathbf{t}}^{(\mathbf{n})}\}_{t\geq 0} = \{W_{t,h}^{(n)}\}_{\substack{t\geq 0,\\ 0\leq h\leq d_n}}$, we denote the covariance matrix with $\boldsymbol{\Gamma}_{\mathbf{W}}^{(n)}$, and similarly, we write $\boldsymbol{\Gamma}_{\mathbf{S}}^{(n)}$ for the covariance matrix of the vector $n^{-1/2}\mathbf{S}^{(n)}$. The main Theorem is formulated below.

Theorem 6.4. *Suppose that Assumption 6.1 is valid. Then for each n and* $\nu \geq 2$, *on a possible larger probability space, there exists a* d_n*-dimensional Brownian motion* $\{\mathbf{W}_{\mathbf{t}}^{(\mathbf{n})}\}_{t\geq 0} = \{W_{t,h}^{(n)}\}_{\substack{t\geq 0,\\ 0\leq h\leq d_n}}$ *such that for some* $q > 1$

$$P\left(\max_{1\leq h\leq d_n}\max_{1\leq i\leq n}\left|\sum_{k=1}^{i} X_{k,h} - \psi_h W_{i,h}^{(n)}\right| \geq n^{1/\nu}\right) = \mathcal{O}\left(n^{-q}\right),$$

where $0 < \psi_h^2 = \lim_n n^{-1}\mathrm{Var}\big(\sum_{1\leq i\leq n} X_{k,h}\big) < \infty$, *and*

$$\theta \leq \delta < \min\left\{\frac{4(p-2\nu)}{(-4+3p)\nu}, \frac{2+p-2(1+\theta)\nu}{(2+4\theta+p(4+\theta))\nu}\right\},$$

where we require $p > 4\nu$. *In addition, we have that* $\max\big|\boldsymbol{\Gamma}_{\mathbf{W}}^{(n)} - \boldsymbol{\Gamma}_{\mathbf{S}}^{(n)}\big| = \mathcal{O}\big(n^{-\gamma}\big)$, *for some* $\gamma > 0$. *Alternatively, if one sets* $d_n = \mathcal{O}\big((\log n)^{\delta}\big)$, *for arbitrary* $\delta > 0$, *then we require*

$$\nu < \min\left\{\frac{4+2p}{4+4\theta+p\theta+4\theta^2+p\theta^2}, \frac{2}{1-2\theta}, p/4\right\}.$$

Remark 6.5. Note that by setting $\theta = 0$ and letting $p \to \infty$, we obtain the upper bound $\delta < 1/9$. In addition, we point out that conditions (i), (ii) of Assumption 6.1 can be weakened to $p > 4$, which however leads to a less tractable bound for δ.

Based on this result, we can derive the following two Theorems.

Theorem 6.6. *Assume that the assumptions of Theorem 6.4 hold. Then, for each n, we can define two independent* d_n*-dimensional Brownian motions* $\{\mathbf{W}_{t,h}^{(1,n)}\}_{t\geq 0} = \{W_{t,h}^{(1,n)}\}_{\substack{t\geq 0,\\ 0\leq h\leq d_n}}$, $\{\mathbf{W}_{t,h}^{(2,n)}\}_{t\geq 0} = \{W_{t,h}^{(2,n)}\}_{\substack{t\geq 0,\\ 0\leq h\leq d_n}}$

$$P\left(\max_{1\leq h\leq d_n}\sup_{1\leq x\leq n/2}\left|\sum_{1\leq i\leq x} X_{i,h} - \psi_h W_{x,h}^{(n)}\right| \Big/ x^{1/\nu}\right) = \mathcal{O}(1),$$

and

$$P\left(\max_{1\leq h\leq d_n}\sup_{1\leq x\leq n/2}\left|\sum_{n-x\leq i\leq n} X_{i,h} - \psi_h W_{x,h}^{(n)}\right| \Big/ x^{1/\nu}\right) = \mathcal{O}(1),$$

with $\nu > 2$.

Theorem 6.7. *Assume that the assumptions of Theorem 6.4 hold, and let* $\lambda > 0$. *Then, for each n, we can define a* d_n*-dimensional Brownian Bridge* $\{\mathbf{B}_t^{(n)}\}_{t\geq 0} = \{B_{t,h}^{(n)}\}_{\substack{0\leq t\leq 1,\\ 0\leq h\leq d_n}}$ *such that*

$$\max_{1\leq h\leq d_n}\sup_{\lambda\leq t\, n\leq n-\lambda}\frac{\big|M_{t,h}^{(n)} - \psi_h B_{t,h}^{(n)}\big|}{\big(t(1-t)\big)^{1/2}} = \mathcal{O}_P\big(1\big),$$

for $\nu \geq 2$.

Remark 6.8. Note that in both Theorems 6.6 and 6.7 we still have the relation $\max\big|\boldsymbol{\Gamma}_{\mathbf{W}}^{(n)} - \boldsymbol{\Gamma}_{\mathbf{S}}^{(n)}\big| = \mathcal{O}\big(n^{-\gamma}\big)$, for some $\gamma > 0$, for the corresponding Brownian motion. Moreover, it follows immediately from the proof of Theorem 6.4 that one may replace the norm $\max_{0\leq h\leq d_n}\big|\cdot\big|$ with the l_2 norm $\big|\cdot\big|_2$ in Theorems 6.4, 6.6 and 6.7.

The proof of Theorem 6.4 follows [7, Theorem 4.1] in broad brushes, with some (essential) changes in the details. To this end, we require some preliminary results. The following coupling inequality is due to Berthet and Mason [10].

Lemma 6.9 (*Coupling Inequality*)**.** *Let* $X_1, \ldots, X_N$ *be independent, mean zero random vectors in* $\mathbb{R}^n$, $n \geq 1$, *such that for some* $B > 0$, $|X_i|_2 \leq B$, $i = 1, \ldots, N$. *If the probability space is rich enough, then for each* $\delta > 0$, *one can define independent normally distributed mean zero random vectors* $\xi_1, \ldots, \xi_N$ *with* ξ_i *and* X_i *having the same variance/covariance matrix for* $i = 1, \ldots, N$, *such that for universal constants* $C_1 > 0$ *and* $C_2 > 0$,

$$P\left\{\left|\sum_{i=1}^{N}(X_i - \xi_i)\right|_2 > \delta\right\} \leq C_1 n^2 \exp\left(-\frac{C_2\delta}{Bn^2}\right).$$

Lemma 6.10. *There is an absolute constant C such that*

$$\mathbb{E}\left|\sum_{l\leq i\leq k} Y_{k,h}^{(\leq m)}\right|^p \leq C\big((k-l+1)\lceil m+1\rceil\big)^{p/2}.$$

Proof of Lemma 6.10. Put $K = 2\lceil m+1\rceil$, and denote with $\|\cdot\|_p = (\mathbb{E}|\cdot|^p)^{1/p}$. Then per construction, we can rewrite

$$\sum_{l\leq i\leq k} Y_{k,h}^{(\leq m)} = R_1 + \cdots + R_K,$$

where R_i is a sum of independent random variables with at most $(k-l+1)/K$ terms. Minikowski's inequality gives us

$$\|R_1 + \cdots + R_K\|_p \le \|R_1\|_p + \cdots + \|R_K\|_p.$$

By Rosenthal's inequality and Assumption 6.1(i), we have

$$\mathbb{E}\big|R_i\big|^p \le C\big((k-l+1)/K\big)^{p/2} = C\left((k-l+1)/K\right)^{p/2},$$

hence

$$\Big\|\sum_{l\le i\le k} Y_{k,h}^{(\le m)}\Big\|_p^p \le C\big((k-l+1)K\big)^{p/2}. \quad \square$$

Proof of Theorem 6.4. The proof is based on a blocking and truncation argument, which requires us to have numbers $\beta, \delta, \kappa, \theta, p, q, \nu$ that satisfy the following conditions

(A) $\max\{\theta, \delta\} < \beta(\beta+1)^{-1}$,
(B) $\nu^{-1} - \beta(2+2\beta)^{-1} - \kappa - 3\delta > 0$,
(C) $\nu^{-1} - (1-\beta)(2+2\beta)^{-1} > 0$,
(D) $\nu^{-1} - (\beta/4 + 1/p)(1+\beta)^{-1} - \delta/4 - (1+\delta)/p > 0$,
(E) $1 < \nu^{-1} + p\kappa/2 - \delta - \theta\beta(1+\beta)^{-1}(p/4+1) - \theta$,
(F) $p > 4\nu$,
(G) $\beta(2(1+\beta))^{-1} = \gamma + \delta$.

Which we will use as reference, and therefore they are not completely simplified. If we fix γ, θ, p, ν, and suppose that $\theta \le \delta$, then using the above inequalities we obtain

$$\frac{\nu-2}{\nu+2} < \beta < \min\left\{\frac{-2(2+p-2\nu+2\gamma\nu+3\gamma p\nu-2\theta\nu)}{4+2p-6\nu+4\gamma\nu-4p\nu+6\gamma p\nu-8\theta\nu-p\theta\nu}, \frac{-16\nu+8\gamma\nu+8p+2\gamma\nu p}{12\nu-8\gamma\nu+8p-3\nu p+2\gamma\nu p} \vee \infty\right\},$$

where $x \vee y = \min(x, y)$ if $x, y \ge 0$, and $x \vee y = y$ if $x < 0$. Using relation (G) one thus obtains a bound for δ. Note that if we just require $\gamma > 0$, then the above simplifies to

$$\frac{\nu-2}{\nu+2} < \beta < \min\left\{\frac{-2(2+p-2\nu-2\theta\nu)}{4+2p-6\nu-4p\nu-8\theta\nu-p\theta\nu}, \frac{-16\nu+8p}{12\nu+8p-3\nu p} \vee 0\right\}.$$

Alternatively, if we set $d_n = \mathcal{O}\big((\log n)^\delta\big)$, then we may set $\delta = 0$ in (A)–(G), and an evaluation amounts to

$$\nu < \min\left\{\frac{4+2p}{4+4\theta+p\theta+4\theta^2+p\theta^2}, \frac{2}{1-2\theta}, p/4\right\}.$$

We will now construct an approximation for the random variables $R_i^{(h)}$. To this end, we first divide the set of integers $\{1, 2, \ldots\}$ into consecutive blocks $\mathcal{H}_1, \mathcal{I}_1, \mathcal{H}_2, \mathcal{I}_2 \cdots$. The blocks are defined by recursion. Fix $\beta > 0$. If the largest element of $\mathcal{I}_{i-1}$ is k_{i-1}, then $\mathcal{H}_i = \{k_{i-1}+1, \ldots, k_{i-1}+i^\beta\}$ and $\mathcal{I}_i = \{k_{i-1}+i^\beta+1, \ldots, k_i\}$, where $k_i = \min\{l : l - (\Upsilon d_n) \vee m_n - 1 \ge k_{i-1} + i^\beta\}$, for some constant $\Upsilon > 0$, where $x \vee y = \max(x, y)$ for $x, y \in \mathbb{R}$. Let $|\cdot|$ denote the cardinality of a set. It follows from the definition of $\mathcal{H}_i, \mathcal{I}_i$ that $|\mathcal{H}_i| = i^\beta$ and $|\mathcal{I}_i| \ge d_n + 1$. Note that the total number of blocks is approximately $c_n = n^{1/(1+\beta)}$, due to (A). For $1 \le h \le d_n$, let

$$U_{k,h}^{(m,1)} = \sum_{i\in\mathcal{H}_k} Y_{i,h}^{(\le m)} \quad \text{and} \quad U_{k,h}^{(m,2)} = \sum_{i\in\mathcal{I}_k} Y_{i,h}^{(\le m)},$$
$$V_{k,h}^{(m,1)} = \sum_{i\in\mathcal{H}_k} Y_{i,h}^{(> m)} \quad \text{and} \quad V_{k,h}^{(m,2)} = \sum_{i\in\mathcal{I}_k} Y_{i,h}^{(> m)},$$

and define the vectors

$$\mathbf{U}_k^{(m,i)} = \big(U_{k,1}^{(m,i)}, U_{k,2}^{(m,i)}, \ldots, U_{k,d_n}^{(m,i)}\big)^T,$$
$$\mathbf{V}_k^{(m,i)} = \big(V_{k,1}^{(m,i)}, V_{k,2}^{(m,i)}, \ldots, V_{k,d_n}^{(m,i)}\big)^T, \quad i \in \{1, 2\}.$$

Throughout this proof, we will always assume that $m = m_n = n^\theta$. For a random variable X, let $\mathbf{I}_B^X = \mathbf{1}(X)_{\{|X|\le B\}}$ for $B > 0$, and similarly, $\mathbf{I}_{B^c}^X = 1 - \mathbf{I}_B^X = \mathbf{1}(X)_{\{|X|>B\}}$. In addition, we put $E_B^X = \mathbb{E}\big(X\mathbf{I}_B^X\big)$. Let

$$\xi_{k,h}^{(m)} = U_{k,h}^{(m,1)}\mathbf{I}_{B_n}^{U_{k,h}^{(m,1)}} - E_{B_n}^{U_{k,h}^{(m,1)}}, \qquad \eta_{k,h}^{(m)} = U_{k,h}^{(m,2)}\mathbf{I}_{B_n}^{U_{k,h}^{(m,2)}} - E_{B_n}^{U_{k,h}^{(m,2)}},$$

and define the random vectors

$$\boldsymbol{\xi}_j^{(m)} = \big(\xi_{j,1}^{(m)}, \xi_{j,2}^{(m)}, \ldots, \xi_{j,d_n}^{(m)}\big)^t, \qquad \boldsymbol{\eta}_j^{(m)} = \big(\eta_{j,1}^{(m)}, \eta_{j,2}^{(m)}, \ldots, \eta_{j,d_n}^{(m)}\big)^t.$$

As a first step, we will show that the truncation error is negligible, more precisely, we will show that

$$P\left(\max_{1\leq h\leq d_n}\max_{1\leq i\leq c_n}\left|\sum_{j=1}^{i} U_{j,h}^{(m,1)} + U_{j,h}^{(m,2)} - \xi_{j,h}^{(m)} - \eta_{j,h}^{(m)}\right| \geq n^{1/\nu}\right) = \mathcal{O}\left(n^{-q}\right). \tag{6.3}$$

To this end, let $x > 0$. Then the Markov and Lévy's maximal inequality imply that

$$P\left(\max_{1\leq h\leq d_n}\max_{1\leq i\leq c_n}\left|\sum_{j=1}^{i} U_{j,h}^{(m,1)} - \xi_{j,h}^{(m)}\right| \geq x\right) \leq Cx^{-2} d_n \max_{1\leq h\leq d_n}\sum_{i=1}^{c_n}\left\|U_{i,h}^{(m,1)} - \xi_{i,h}^{(m)}\right\|_2^2.$$

Using the Cauchy–Schwarz inequality, we obtain

$$\max_{1\leq h\leq d_n}\left\|U_{i,h}^{(m,1)} - \xi_{i,h}^{(m)}\right\|_2^2 \leq \left\|U_{i,h}^{(m,1)}\right\|_4^2 \left\|\mathbf{I}_B^{U_{i,h}^{(m,1)}}\right\|_4^2 \leq \left\|U_{i,h}^{(m,1)}\right\|_4^2 \left\|U_{i,h}^{(m,1)}\right\|_p^{p/2} B^{-p/2},$$

which, by Lemma 6.10, is of the magnitude $\mathcal{O}\left((m\, i^{\beta})^{p/4+1}B^{-p/2}\right)$. We thus obtain that

$$\begin{aligned}P\left(\max_{1\leq h\leq d_n}\max_{1\leq i\leq c_n}\left|\sum_{j=1}^{i} U_{j,h}^{(m,1)} - \xi_{j,h}^{(m)}\right| \geq x\right) &\leq Cx^{-1} d_n\, m^{p/4+1} B^{-p/2}\sum_{i=1}^{c_n}(i^{\beta})^{p/4+1}\\ &= \mathcal{O}\left(x^{-1} d_n\, c_n^{\beta(p/4+4)+1} B^{-p/2}\right).\end{aligned}$$

Setting $x = 2n^{1/\nu}$ and $B = B_n = n^{\kappa}$, we find that relation (E) establishes $\mathcal{O}\left(x^{-1} d_n\, c_n^{\beta(p/4+4)+1} B^{-p/2}\right) = \mathcal{O}\left(n^{-q}\right)$.

By the same argument, one also establishes that

$$P\left(\max_{1\leq h\leq d_n}\max_{1\leq i\leq c_n}\left|\sum_{j=1}^{i} U_{j,h}^{(m,2)} - \eta_{j,h}^{(m)}\right| \geq x\right) = \mathcal{O}\left(n^{-q}\right), \tag{6.4}$$

which together with the previous result gives us (6.3).

Note that per construction and relation (A), choosing the constant Υ big enough, we have that $\left\{\boldsymbol{\xi}_j^{(m)}\right\}_{j\in\mathbb{N}}$ and $\left\{\boldsymbol{\eta}_j^{(m)}\right\}_{j\in\mathbb{N}}$ are sequences of independent random vectors. In addition, we have the bound

$$\left|\boldsymbol{\xi}_j^{(m)}\right|_{d_n} \leq d_n B_n \qquad \left|\boldsymbol{\eta}_j^{(m)}\right|_{d_n} \leq d_n B_n. \tag{6.5}$$

Hence, by Lemma 6.9, we can define a sequence of independent normal random vectors $\boldsymbol{\xi}_j^{(m,*)} = \left(\xi_{j,1}^{(m,*)}, \xi_{j,2}^{(m,*)}, \ldots, \xi_{j,d_n}^{(m,*)}\right)^t$, such that for $x > 0$

$$\begin{aligned}P\left(\max_{1\leq h\leq d_n}\max_{1\leq i\leq c_n}\left|\sum_{j=1}^{i}\left(\xi_{j,h}^{(m)} - \xi_{j,h}^{(m,*)}\right)\right| \geq x\right) &\leq \sum_{h=1}^{d_n}\sum_{i=1}^{c_n} P\left(\left|\sum_{j=1}^{i}\left(\xi_{j,h}^{(m)} - \xi_{j,h}^{(m,*)}\right)\right| \geq x\right)\\ &= \sum_{h=1}^{d_n}\sum_{i=1}^{c_n} P\left(\left|\sum_{j=1}^{i}\left(\xi_{j,h}^{(m)} - \xi_{j,h}^{(m,*)}\right)\right|_2 \geq x\right)\\ &\leq \sum_{h=1}^{d_n}\sum_{i=1}^{c_n} P\left(\left|\sum_{j=1}^{i}\left(\boldsymbol{\xi}_{j,h}^{(m)} - \boldsymbol{\xi}_{j,h}^{(m,*)}\right)\right|_2 \geq x\right)\\ &\leq C_1\sum_{i=1}^{c_n} d_n^3 \exp\left(-\frac{C_2 x}{2d_n^3 B_n}\right)\\ &\leq C_1 c_n d_n^3 \exp\left(-\frac{C_2 x}{2d_n^3 B_n}\right).\end{aligned}$$

Hence due to (B), we obtain

$$P\left(\max_{1\leq h\leq d_n}\max_{1\leq i\leq c_n}\left|\sum_{j=1}^{i}\left(\xi_{j,h}^{(m)} - \xi_{j,h}^{(m,*)}\right)\right| \geq n^{1/\nu}\right) = \mathcal{O}\left(n^{-q}\right), \tag{6.6}$$

for $q > 1$. Similar arguments show that under the same conditions as above, there exists a sequence of independent normal random vectors $\boldsymbol{\eta}_j^{(m,*)} = \left(\eta_{j,1}^{(m,*)}, \eta_{j,2}^{(m,*)}, \ldots, \eta_{j,d_n}^{(m,*)}\right)^t$, such that

$$P\left(\max_{1\leq h\leq d_n}\max_{1\leq i\leq c_n}\left|\sum_{j=1}^{i}\left(\eta_{j,h}^{(m)} - \eta_{j,h}^{(m,*)}\right)\right| \geq n^{1/\nu}\right) = \mathcal{O}\left(n^{-q}\right),$$

for $q > 1$. By Lévy's maximal inequality, we have

$$P\left(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\left|\sum_{j=1}^{i}\eta_{j,h}^{(m,*)}\right|\ge n^{1/\nu}\right)\le 2\sum_{h=1}^{d_n}P\left(\left|\sum_{j=1}^{c_n}\eta_{j,h}^{(m,*)}\right|\ge n^{1/\nu}\right).$$

By Lemma 6.10, we have that $\mathrm{Var}\big(\eta_{j,h}^{(m,*)}\big)\le Cd_n^2$ for all $j\le c_n$, $h\le d_n$. Hence if (D) holds, by known properties of the tails of a normal cdf, we obtain that

$$P\left(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\left|\sum_{j=1}^{i}\eta_{j,h}^{(m,*)}\right|\ge n^{1/\nu}\right)=\mathcal{O}\left(n^{-q}\right), \tag{6.7}$$

for $q > 1$. This yields

$$P\left(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\left|\sum_{j=1}^{i}\big(\xi_{j,h}^{(m)}+\eta_{j,h}^{(m)}-\xi_{j,h}^{(m,*)}\big)\right|\ge n^{1/\nu}\right)=\mathcal{O}\left(n^{-q}\right), \tag{6.8}$$

for $q > 1$.

Let $\boldsymbol{\eta}_j^{(m,**)}=\big(\eta_{j,1}^{(m,**)},\eta_{j,2}^{(m,**)},\ldots,\eta_{j,d_n}^{(m,**)}\big)^t$ be an independent copy of $\boldsymbol{\eta}_j^{(m,*)}$ such that $\boldsymbol{\eta}_j^{(m,*)}$ and $\boldsymbol{\xi}_j^{(m,**)}$ are independent. Proceeding as in the proof of [7, Theorem 4.1], by enlarging the probability space if necessary, there exists a d_n-dimensional Brownian motion $\{\mathbf{W}_t\}_{t\ge 0}=\{W_t^{(h)}\}_{\substack{t\ge 0,\\ 0\le h\le d_n}}$, such that

$$W_{k_i}^{(h)}=\sum_{1\le j\le i}d_j^{(h)}\big(\xi_{j,h}^{(m,*)}+\eta_{j,h}^{(m,**)}\big),$$

where $d_j^{(h)}$ is chosen such that $\big\|d_j^{(h)}(\xi_{j,h}^{(m,*)}+\eta_{j,h}^{(m,**)})\big\|_2^2=|H_j|+|J_j|$.

We will now establish that

$$\big(d_j^{(h)}\big)^2=1/\psi_h(1+\mathcal{O}(j^{-\beta})). \tag{6.9}$$

To this end, note that per construction

$$\big\|\xi_{j,h}^{(m,*)}+\eta_{j,h}^{(m,**)}\big\|_2^2=\big\|\xi_{j,h}^{(m,*)}\big\|_2^2+\big\|\eta_{j,h}^{(m,**)}\big\|_2^2=\big\|\xi_{j,h}^{(m)}\big\|_2^2+\big\|\eta_{j,h}^{(m)}\big\|_2^2.$$

In addition, Assumption 6.1(ii) implies that

$$\begin{aligned}\big|\big\|U_{j,h}^{(m,1)}+V_{j,h}^{(m,1)}\big\|_2-\big\|\xi_{j,h}^{(m)}\big\|_2\big| &\le \big\|\xi_{j,h}^{(m)}-U_{j,h}^{(m,1)}\big\|_2+\big|\big\|V_{j,h}^{(m,1)}\big\|_2\big|\\ &=\big\|\xi_{j,h}^{(m)}-U_{j,h}^{(m,1)}\big\|_2+\mathcal{O}\big(1\big).\end{aligned}$$

Moreover, it follows from computations performed when establishing (6.3) that $\big\|\xi_{j,h}^{(m)}-U_{j,h}^{(m,1)}\big\|_2=\mathcal{O}\big(1\big)$, hence

$$\big|\big\|U_{j,h}^{(m,1)}+V_{j,h}^{(m,1)}\big\|_2-\big\|\xi_{j,h}^{(m)}\big\|_2\big|=\mathcal{O}\big(1\big).$$

Similarly, one gets that

$$\big|\big\|U_{j,h}^{(m,2)}+V_{j,h}^{(m,2)}\big\|_2-\big\|\eta_{j,h}^{(m)}\big\|_2\big|=\mathcal{O}\big(1\big).$$

Hence Assumption 6.1(iii) implies that

$$\big\|(\xi_{j,h}^{(m,*)}+\eta_{j,h}^{(m,**)})\big\|_2^2=\psi_H\big(|H_j|+|J_j|\big)+\mathcal{O}\big(1\big),$$

and thus (6.9) follows. Relation (6.9) implies that

$$\mathrm{Var}\left(\sum_{j=1}^{c_n}(1-\psi_h d_j^{(h)})(\xi_{j,h}^{(m,*)}+\eta_{j,h}^{(m,**)})\right)=\mathcal{O}\left(\sum_{j=1}^{c_n}j^{-\beta}\right)=\mathcal{O}\left(n^{(1-\beta)/(1+\beta)}\right).$$

Hence by Lévy's maximal inequality, it follows from (C) that

$$\begin{aligned}&P\left(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\left|\sum_{j=1}^{i}\big(\xi_{j,h}^{(m,*)}+\eta_{j,h}^{(m,**)}\big)-\psi_h W_{k_i}^{(h)}\right|\ge n^{1/\nu}\right)\\ &\quad\le 2\sum_{h=1}^{d_n}P\left(\left|\sum_{j=1}^{c_n}\big(\xi_{j,h}^{(m,*)}+\eta_{j,h}^{(m,**)}\big)(1-\psi_h d_j^{(h)})\right|\ge n^{1/\nu}\right)\\ &\quad\le Cd_nP\big(Z_n\ge n^{1/\nu}\big)=\mathcal{O}\left(n^{-q}\right)\end{aligned}$$

for $q > 1$, where Z_n is a mean zero Gaussian random variable with $\mathrm{Var}(Z_n) = \mathcal{O}\big(n^{(1-\beta)/(1+\beta)}\big)$. Next, it is shown that.

$$P\left(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\max_{k_i\le l\le k_{i+1}}\left|\sum_{j=k_i+1}^{l} Y_{j,h}^{(\le m)}\right| > n^{1/\nu}\right) = \mathcal{O}\left(n^{-q}\right), \tag{6.10}$$

and

$$P\left(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\sup_{k_i\le s\le k_{i+1}}\left|W_s^{(h)} - W_{k_i}^{(h)}\right| > n^{1/\nu}\right) = \mathcal{O}\left(n^{-q}\right). \tag{6.11}$$

To this end, note that by Lemma 6.10 and Moricz et al. [38, Theorem 3.1], it holds that

$$\mathbb{E}\left(\max_{k_i\le l\le k_{i+1}}\left|\sum_{j=k_i+1}^{l} Y_{j,h}^{(\le m)}\right|^{p/2}\right) \le C\big(k_{i+1}-k_i\big)^{p/4}\big(m+1\big)^{p/4} \tag{6.12}$$

$$= \mathcal{O}\left(\big(i^{\beta}m\big)^{p/4}\right). \tag{6.13}$$

Using the Markov inequality, we thus obtain

$$\begin{aligned}
P\left(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\max_{k_i\le l\le k_{i+1}}\left|\sum_{j=k_i+1}^{l} Y_{j,h}^{(\le m)}\right| > n^{1/\nu}\right) &\le \sum_{h=1}^{d_n}\sum_{i=1}^{c_n} P\left(\max_{k_i\le l\le k_{i+1}}\left|\sum_{j=k_i+1}^{l} Y_{j,h}^{(\le m)}\right| > n^{1/\nu}\right)\\
&\le Cn^{-p/\nu}\sum_{h=1}^{d_n}\sum_{i=1}^{c_n}\big(k_{i+1}-k_i\big)^{p/4}\big(d_n+1\big)^{p/4}\\
&\le C\,n^{-p/\nu}\,d_n^{(p+4)/4}\sum_{i=1}^{c_n} i^{\beta p/4} = \mathcal{O}\left(n^{-p/\nu}c_n^{(\beta p+4)/4}d_n^{(p+4)/4}\right)\\
&= \mathcal{O}\left(n^{-p/\nu+(\beta p+4)/(4+4\beta)+\delta(p+4)/4}\right),
\end{aligned}$$

which proves (6.10) due to relation (D). The same argument also of applies to (6.11), by replacing the maximal inequality of Moricz et al. [38] by the corresponding results for the increments of the Wiener process in Csörgo and Horváth [14]. Piecing everything together, we obtain that

$$P\left(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\sup_{k_i\le s\le k_{i+1}}\left|\psi_h W_s^{(h)} - \sum_{j=1}^{i} U_{j,h}^{(m,1)} + U_{j,h}^{(m,2)}\right| > n^{1/\nu}\right) = \mathcal{O}\left(n^{-q}\right). \tag{6.14}$$

Suppose now that

$$P\left(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\max_{k_i\le l\le k_{i+1}}\left|\sum_{j=0}^{l} Y_{j,h}^{(>m)}\right| > n^{1/\nu}\right) = \mathcal{O}\left(n^{-q}\right). \tag{6.15}$$

This together with (6.10) yields

$$P\left(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\max_{k_i\le l\le k_{i+1}}\left|\sum_{r=0}^{l} X_{r,h} - \sum_{j=1}^{i} U_{j,h}^{(m,1)} + U_{j,h}^{(m,2)}\right| > n^{1/\nu}\right) = \mathcal{O}\left(n^{-q}\right), \tag{6.16}$$

which together with (6.14) gives the desired approximation result. Hence we need to verify (6.15). To this end, note that Assumption 6.1(ii) implies that

$$P\left(\max_{1\le h\le d_n}\max_{1\le i\le c_n}\max_{k_i\le l\le k_{i+1}}\left|\sum_{j=0}^{l} Y_{j,h}^{(>m)}\right| > n^{1/\nu}\right) \le Cd_nc_nn^{-p/\nu+1}\Lambda_{n,p}, \tag{6.17}$$

where $\Lambda_{n,p} = \max_{1\le l\le n}\big\|\sum_{j=1}^{l} Y_{j,h}^{(>m)}\big\|_p^p$. We thus require $1/\nu - 2/p\delta - p^{-1}(\beta+1)^{-1} - (q+1)/p > 0$. Using condition (A), this is true if $1/\nu - (2\beta+1)(p\beta+p)^{-1} - (1+q)/p > 0$. Since by (F) $p > 4\nu$, we can choose a $q > 1$ such that $p > (3+q)\nu$, and it follows that

$$\frac{(2+q)\nu - p}{p-(3+q)\nu} < 0 < \beta, \tag{6.18}$$

hence this imposes no additional restriction, and (6.15) thus holds.

Regarding the covariance structure of $\{\mathbf{W}_t\}_{t\geq 0}$, note that the blocking and truncation argument has slightly changed the covariance structure. In order to quantify the error, note first that by stationarity, the covariance structure within the vectors $\boldsymbol{\xi}_j^{(m)}$ and $\boldsymbol{\eta}_j^{(m)}$ is the same for all j, and hence this is also true for the approximations $\boldsymbol{\xi}_j^{(m,*)}$ and $\boldsymbol{\eta}_j^{(m,**)}$. Put $I_{k,i} := \sum_{r\in\mathcal{H}_k} X_{r,i}$ and $II_{k,i} := \sum_{r\in\mathcal{I}_k} X_{r,i}$, and define $I_{k,i}^{(*)}, II_{k,i}^{(*)}$ in the same manner. Then the Cauchy–Schwarz inequality implies

$$\begin{aligned}
&\left|\mathbb{E}\left(\sum_{k=1}^{l}(I_{k,i}+II_{k,i})\sum_{k=1}^{l}(I_{k,j}+II_{k,j})\right)-\mathbb{E}\left(\sum_{k=1}^{l}(I_{k,i}^{(*)}+II_{k,i}^{(*)})\sum_{k=1}^{l}(I_{k,j}^{(*)}+II_{k,j}^{(*)})\right)\right|\\
&\quad\leq\sqrt{\operatorname{Var}\left(\sum_{k=1}^{l}I_{k,i}\right)\operatorname{Var}\left(\sum_{k=1}^{l}II_{k,j}\right)}+\sqrt{\operatorname{Var}\left(\sum_{k=1}^{l}I_{k,j}\right)\operatorname{Var}\left(\sum_{k=1}^{l}II_{k,i}\right)}\\
&\quad=\mathcal{O}\left(\sqrt{\sum_{k=1}^{l}|\mathcal{I}_k|\sum_{k=1}^{l}|\mathcal{H}_k|}\right)=\mathcal{O}\left(l^{\beta/2+1}\sqrt{m_n\vee d_n}\right)\\
&\quad=\mathcal{O}\left(n^{\frac{2+\beta}{2(1+\beta)}}\sqrt{m_n\vee d_n}\right),
\end{aligned}$$

which gives us an upper bound for the error which stems from the blocking argument. Using the bound given in (6.3), a similar argument shows that the error which arises from the truncation of the random vectors $\mathbf{U}_{k,h}^{(m,i)}, \mathbf{V}_{k,h}^{(m,i)}$, $i\in\{1,2\}$ is of the magnitude $\mathcal{O}\big(n^{\frac{2+\beta}{2(1+\beta)}-1/4}\sqrt{m_n\vee d_n}\big)$. Finally, the error which comes from the conditioning argument is of the order $\mathcal{O}\big(n^{\frac{2+\beta}{2(1+\beta)}}\big)$, this follows again by a similar argument as before, using Assumption 6.1(ii). Combining all bounds, we obtain that the total error is of the magnitude $\mathcal{O}\big(n^{\frac{2+\beta}{2(1+\beta)}}\sqrt{m_n\vee d_n}\big)$. Using relation (G) completes the proof. □

Proof of Theorem 6.6. Using Theorem 6.4, one can proceed exactly as in the proof of [7, Theorem 4.2]. □

Proof of Theorem 6.7. Using Theorem 6.4, one can proceed exactly as in the proofs of Theorems 4.3 and 4.4 in [7]. □

Proof of Theorem 5.4. First note that without loss of generality, we may assume that $\mu_h = 0$, $1\leq h\leq d$, since the means μ_h always cancel per construction in $\mathbf{M}_t^{(n)}$. Using the same notation as in the proof of Theorem 3.3, recall that $|f|_n^* := \sup_{\lambda_n/n\leq t\leq 1-\lambda/n} w(t)^{-1}\big|f(t)\big|$, where $w(t)=t(1-t)$, and

$$\mathbf{B}_{w,n}=\left(B_1^{(\lambda_n/n)},B_2^{(\lambda_n/n)},\ldots,B_{d_n}^{(\lambda_n/n)}\right)^t,$$

where $B_h^{(\lambda_n/n)} = \big|B_{t,h}\big|_n^*$. We will only show (i) in case of $d_n = \mathcal{O}(n^\delta)$, since the other cases (including (ii)) follow in an analogue manner. We have that

$$\begin{aligned}
\Big||\mathbf{M}_t^{(n)}\boldsymbol{\Gamma}_n^{-1}\mathbf{M}_t^{(n)}|_n^* - |\mathbf{B}_t\boldsymbol{\Gamma}_n^{-1}\mathbf{B}_t|_n^*\Big| &\leq \Big||(\mathbf{M}_t^{(n)}-\mathbf{B}_t)^t\boldsymbol{\Gamma}_n^{-1}\mathbf{B}_t|_n^*\Big| + \Big||\mathbf{B}_t^t\boldsymbol{\Gamma}_n^{-1}(\mathbf{M}_t^{(n)}-\mathbf{B}_t)|_n^*\Big|\\
&\quad+\Big||(\mathbf{M}_t^{(n)}-\mathbf{B}_t)^t\boldsymbol{\Gamma}_n^{-1}(\mathbf{M}_t^{(n)}-\mathbf{B}_t)|_n^*\Big|.
\end{aligned}$$

Using that $\max|\boldsymbol{\Gamma}_n^{-1}|\leq L_n$, this is further smaller than

$$\Big||\mathbf{M}_t^{(n)}\boldsymbol{\Gamma}_n^{-1}\mathbf{M}_t^{(n)}|_n^* - |\mathbf{B}_t\boldsymbol{\Gamma}_n^{-1}\mathbf{B}_t|_n^*\Big| \leq CL_nd_n\left(\max\big|\mathbf{Z}_n-\mathbf{B}_{w,n}\big|\right)\left(\max\big|\mathbf{B}_{w,n}\big|\right)+CL_nd_n\left(\max\big|\mathbf{Z}_n-\mathbf{B}_{w,n}\big|\right)^2.$$

By Theorem 6.7, we have that $\max\big|\mathbf{Z}_n-\mathbf{B}_{w,n}\big| = \mathcal{O}_p(1)$. Let $\epsilon > 0$. Since we have $L_nd_n^{1/2} = \mathcal{O}\left((\log n)^{-\kappa}\right)$, we obtain the bound

$$\begin{aligned}
P\Big(\Big||\mathbf{M}_t^{(n)}\boldsymbol{\Gamma}_n^{-1}\mathbf{M}_t^{(n)}|_n^* - |\mathbf{B}_t\boldsymbol{\Gamma}_n^{-1}\mathbf{B}_t|_n^*\Big|\geq\epsilon\sqrt{d_n}\Big) &\leq \mathcal{O}\big(1\big)+2P\Big(\max\big|\mathbf{B}_{w,n}\big|\geq C(\log n)^\kappa\Big)\\
&\leq \mathcal{O}\big(1\big)+d_nP\Big(B_1^{(\lambda_n/n)}\geq C(\log n)^\kappa\Big),
\end{aligned}$$

for $C,\kappa > 0$. Arguing as in the proof of Theorem 3.3, one obtains

$$d_nP\Big(B_1^{(\lambda_n/n)}\geq C(\log n)^\kappa\Big)=\mathcal{O}(1).$$

Hence we can choose a sequence ϵ_n that tends to zero as n increases, such that

$$P\Big(\Big||\mathbf{M}_t^{(n)}\boldsymbol{\Gamma}_n^{-1}\mathbf{M}_t^{(n)}|_n^* - |\mathbf{B}_t\boldsymbol{\Gamma}_n^{-1}\mathbf{B}_t|_n^*\Big|\geq\epsilon_n\sqrt{d_n}\Big)=\mathcal{O}\big(1\big),\tag{6.19}$$

hence the claim follows. □

Acknowledgments

The author thanks the anonymous referees for their valuable remarks and comments that considerably helped to improve the results of this paper.

References

[1] T.W. Anderson, The Statistical Analysis of Time Series, John Wiley & Sons Inc., New York, 1971.
[2] A. Aue, R. Gabrys, L. Horváth, P. Kokoszka, Estimation of a change-point in the mean function of functional data, J. Multivariate Anal. 100 (10) (2009) 2254–2269.
[3] A. Aue, L. Horváth, M. Hušková, P. Kokoszka, Change-point monitoring in linear models, Econom. J. 9 (3) (2006) 373–403.
[4] A. Aue, S. Hörmann, L. Horváth, M. Reimherr, Break detection in the covariance structure of multivariate time series models, Ann. Statist. 37 (6B) (2009) 4046–4087.
[5] A. Banerjee, G. Urga, Modelling structural breaks, long memory and stock market volatility: an overview, J. Econometrics 129 (1–2) (2005) 1–34.
[6] I. Berkes, R. Gabrys, L. Horváth, P. Kokoszka, Detecting changes in the mean of functional observations, J. R. Stat. Soc. Ser. B 71 (5) (2009) 927–946.
[7] I. Berkes, E. Gombay, L. Horváth, Testing for changes in the covariance structure of linear processes, J. Statist. Plann. Inference 139 (6) (2009) 2044–2063.
[8] I. Berkes, S. Hörmann, J. Schauer, Asymptotic results for the empirical process of stationary sequences, Stochastic Process. Appl. 119 (4) (2009) 1298–1324.
[9] I. Berkes, W. Philipp, Approximation theorems for independent and weakly dependent random vectors, Ann. Probab. 7 (1) (1979) 29–54.
[10] P. Berthet, D.M. Mason, Revisiting two strong approximation results of Dudley and Philipp, in: High Dimensional Probability, in: IMS Lecture Notes Monogr. Ser., vol. 51, Inst. Math. Statist., Beachwood, OH, 2006, pp. 155–172.
[11] T. Bollerslev, Generalized autoregressive conditional heteroscedasticity, J. Econometrics 31 (3) (1986) 307–327.
[12] Tim Bollerslev, Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model, Rev. Econ. Statist. 72 (3) (1990) 498–505.
[13] P.J. Brockwel, R.A. Davis, Time Series: Theory and Methods, second ed. in: Springer Series in Statistics, Springer-Verlag, New York, 1991.
[14] M. Csörgo, L. Horváth, Limit Theorems in Change-Point Analysis, 1997.
[15] M. Csörgö, L. Horváth, Limit Theorems in Change-Point Analysis, John Wiley & Sons, 1997.
[16] M. Csörgő, L. Horváth, Weighted Approximations in Probability and Statistics, in: Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Ltd., Chichester, 1993, With a foreword by David Kendall.
[17] J. Dedecker, P. Doukhan, A new covariance inequality and applications, Stochastic Process. Appl. 106 (1) (2003) 63–80.
[18] D.M. DeLong, Crossing probabilities for a square root boundary by a Bessel process, Comm. Statist. A—Theory Methods 10 (21) (1981) 2197–2213.
[19] P. Doukhan, O. Wintenberger, An invariance principle for weakly dependent stationary general models, Probab. Math. Statist. 27 (1) (2007) 45–73.
[20] R. Engle, Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroscedasticity models, J. Bus. Econom. Statist. 20 (3) (2002) 339–350.
[21] J. Gao, Semiparametric and nonparametric methods, in: Nonlinear Time Series, in: Monographs on Statistics and Applied Probability, vol. 108, Chapman & Hall/CRC, Boca Raton, FL, 2007.
[22] E. Gombay, Change detection in autoregressive time series, J. Multivariate Anal. 99 (3) (2008) 451–464.
[23] E. Gombay, L. Horváth, Sequential tests and change detection in the covariance structure of weakly stationary time series, Comm. Statist. Theory Methods 38 (16–17) (2009) 2872–2883.
[24] U. Grenander, G. Szego, Toeplitz Forms and Their Applications, 1958.
[25] E.J. Hannan, Multiple Time Series, John Wiley and Sons, Inc., New York–London–Sydney, 1970.
[26] L. Horváth, M. Hušková, P. Kokoszka, Testing the stability of the functional autoregressive process, J. Multivariate Anal. 101 (2) (2010) 352–367.
[27] M. Husková, C. Kirch, Z. Prásková, J. Steinebach, On the detection of changes in autoregressive time series, ii, resampling procedures, J. Statist. Plann. Inference 138 (6) (2008) 1697–1721.
[28] C. Inclán, G.C. Tiao, Use of cumulative sums of squares for retrospective detection of changes of variance, J. Amer. Statist. Assoc. 89 (427) (1994) 913–923.
[29] J. Jacod, A.N. Shiryaev, Limit Theorems for Stochastic Processes, second ed. in: Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 288, Springer-Verlag, Berlin, 2003.
[30] Thierry Jeantheau, Strong consistency of estimators for multivariate arch models, Econometric Theory 14 (01) (1998) 70–86.
[31] T. Jiang, The limiting distributions of eigenvalues of sample correlation matrices, Sankhyā 66 (1) (2004) 35–48.
[32] T. Jiang, The asymptotic distributions of the largest entries of sample correlation matrices, Ann. Appl. Probab. 14 (2) (2004) 865–880.
[33] J.M. Jirak, On the maximum of covariance estimators, J. Multivariate Anal. 102 (6) (2011) 1032–1046.
[34] J. Keilson, H.F. Ross, Passage time distributions for Gaussian Markov (Ornstein–Uhlenbeck) statistical processes, in: Selected Tables in Mathematical Statistics, Vol. III, Amer. Math. Soc., Providence, RI, 1975, pp. 233–327.
[35] J. Komlós, P. Major, G. Tusnády, Weak convergence and embedding, in: Limit Theorems of Probability Theory (Colloq., Keszthely, 1974), in: Colloq. Math. Soc. János Bolyai, vol. 11, North-Holland, Amsterdam, 1975, pp. 149–165.
[36] J. Komlós, P. Major, G. Tusnády, An approximation of partial sums of independent RV's, and the sample DF. II, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 34 (1) (1976) 33–58.
[37] S. Lee, J. Ha, O. Na, S. Na, The cusum test for parameter change in time series models, Scand. J. Stat. 30 (4) (2003) 781–796.
[38] F.A. Moricz, R.J. Serfling, W.F. Stout, Moment and probability bounds with quasi-superadditive structure for the maximum partial sum, Ann. Probab. 10 (4) (1982) 1032–1040.
[39] P. Perron, Dealing with structural breaks. Boston University – Department of Economics – Working Papers Series WP2005-017, Boston University—Department of Economics, April 2005.
[40] M.B. Priestley, Nonlinear and Nonstationary Time Series Analysis, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1988.
[41] A.F. Ronzhin, On the probability of exit from a square-root boundary by the norm of a Brownian motion, Theory Probab. Appl. 30 (1) (1986) 160–162.
[42] A.I. Sahanenko, Estimates of the rate of convergence in the invariance principle, Dokl. Akad. Nauk SSSR 219 (1974) 1076–1078.
[43] X. Shao, W.B. Wu, Asymptotic spectral theory for nonlinear time series, Ann. Statist. 35 (4) (2007) 1773–1801.
[44] X. Shao, W.B. Wu, A limit theorem for quadratic forms and its applications, Econometric Theory 23 (5) (2007) 930–951.
[45] G.W. Stewart, On the continuity of the generalized inverse, SIAM J. Appl. Math. 17 (1) (1969) 33–45.
[46] H. Tong, A dynamical system approach, with an appendix by K.S. Chan, in: Nonlinear Time Series, in: Oxford Statistical Science Series, vol. 6, The Clarendon Press Oxford University Press, Oxford Science Publications, New York, 1990.
[47] R.S. Tsay, Analysis of Financial Time Series, second ed. in: Wiley Series in Probability and Statistics, Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2005.
[48] L.Yu. Vostrikova, Detection of "discordance" of a Wiener process, Teor. Veroyatnost. i Primenen. 26 (2) (1981) 362–368.
[49] David S.D.S. Watkins, Fundamentals of Matrix Computations, third ed. in: Pure and Applied Mathematics (Hoboken), John Wiley & Sons Inc., Hoboken, NJ, 2010.
[50] W.B. Wu, Strong invariance principles for dependent random variables, Ann. Probab. 35 (6) (2007) 2294–2320.
[51] W.B. Wu, An asymptotic theory for sample covariances of Bernoulli shifts, Stochastic Process. Appl. 119 (2) (2009) 453–467.

[52] W.B. Wu, Nonlinear system theory : another look at dependence, Proc. Natl. Acad. Sci. USA 102 (2005) 14150–14154.
[53] W.B. Wu, Recursive estimation of time-average variance constants, Ann. Appl. Probab. 19 (4) (2009) 1529–1552.
[54] W.B. Wu, Nonlinear system theory: another look at dependence, Proc. Natl. Acad. Sci. USA 102 (40) (2005) 14150–14154 (electronic).
[55] W.B. Wu, X. Han, Asymptotic inference of autocovariances of stationary processes, 2011. arXiv:1105.3423.
[56] Han Xiao, Wei Biao Wu, Covariance matrix estimation for stationary time series, 2011.
[57] A. Zaitsev, Estimates for the rate of strong approximation in the multidimensional invariance principle, J. Math. Sci. 145 (2007) 4856–4865. 10.1007/s10958-007-0319-7.