*Bernoulli* **26**(3), 2020, 1891–1911
https://doi.org/10.3150/19-BEJ1174

# Functional weak limit theorem for a local empirical process of non-stationary time series and its application

ULRIKE MAYER[1,*], HENRYK ZÄHLE[1,†] and ZHOU ZHOU[2]

[1] *Department of Mathematics, Saarland University, 66123 Saarbrücken, Germany.*
*E-mail:* [*] *mayer@math.uni-sb.de;* [†] *zaehle@math.uni-sb.de*
[2] *Department of Statistics, 100 St. George St., Toronto, ON, M5S 3G3 Canada.*
*E-mail: zhou@utstat.toronto.edu*

We derive a functional weak limit theorem for a local empirical process of a wide class of piece-wise locally stationary (PLS) time series. The latter result is applied to derive the asymptotics of weighted empirical quantiles and weighted V-statistics of non-stationary time series. The class of admissible underlying time series is illustrated by means of PLS linear processes and PLS ARCH processes.

*Keywords:* local empirical process; piece-wise locally stationary time series; weak convergence; weighted empirical quantile; weighted V-statistic

## 1. Introduction

Non-stationary time series analysis has attracted much attention in the statistics community over the last two decades. Among others, see Dahlhaus [8] for a seminal work which proposed a class of locally stationary time series from a time-varying spectral representation point of view. Nason et al. [19] investigated locally stationary time series via the time-varying wavelet spectrum. Zhou and Wu [30] formulated locally stationary time series from a time-varying physical system perspective. In Zhou [28], the framework in Zhou and Wu [30] was extended to a class of piece-wise locally stationary models which allows both smooth and abrupt changes in the physical system. Recently Dahlhaus et al. [10] studied a general class of locally stationary time series using both stationary approximations and the physical system representation. We also refer to Dahlhaus [9] for a more comprehensive review and additional references.

The purpose of this paper is to provide a systematic asymptotic theory for the local empirical processes of a wide class of piece-wise locally stationary (PLS) time series in the sense of [28]. As empirical processes are fundamental tools for many statistical problems, the results of this paper serve as a unified theoretical basis for a wide range of nonparametric statistical problems in non-stationary time series analysis.

To define our time series model, we fix a finite partition $0 = p_0 < p_1 < \cdots < p_\ell < p_{\ell+1} = 1$ of the unit interval $[0, 1]$. For every $j = 0, \dots, \ell$, we let $G_j : (p_j, p_{j+1}] \times \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ be any

1350-7265 

$(\mathcal{B}((p_j, p_{j+1}]) \otimes \mathcal{B}(\mathbb{R})^{\otimes \mathbb{N}}, \mathcal{B}(\mathbb{R}))$-measurable map. For every $n \in \mathbb{N}$, we define by

$$\mathfrak{G}_n\big(i, (x_k)_{k\in\mathbb{N}}\big) := \sum_{j=0}^{\ell} G_j\big(i/n, (x_k)_{k\in\mathbb{N}}\big)\mathbb{1}_{(p_j, p_{j+1}]}(i/n) \tag{1}$$

a time dependent filter $\mathfrak{G}_n : \{1, \ldots, n\} \times \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$. Then, given a two-sided sequence $\boldsymbol{\epsilon} = (\varepsilon_k)_{k\in\mathbb{Z}}$ of i.i.d. real-valued random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we can define a non-stationary time series $(X_{n,i})_{i=1}^n$ on $(\Omega, \mathcal{F}, \mathbb{P})$ by

$$X_{n,i} := \mathfrak{G}_n(i, \boldsymbol{\epsilon}_i) = \sum_{j=0}^{\ell} G_j(i/n, \boldsymbol{\epsilon}_i)\mathbb{1}_{(p_j, p_{j+1}]}(i/n), \quad i = 1, \ldots, n, \tag{2}$$

where $\boldsymbol{\epsilon}_i := (\varepsilon_i, \varepsilon_{i-1}, \varepsilon_{i-2}, \ldots)$. For every $j = 1, \ldots, \ell$, this time series is subject to a structural break at the smallest time point $i$ with $i > np_j$. Note that the number of observations between any two adjacent structural break points increases linearly in $n$.

Time series of the form (2) were introduced first in [28]. Under suitable assumptions on $G_0, \ldots, G_\ell$ and $\mathbb{P}_{\varepsilon_0}$ such time series are approximatively stationary in every small (relative to $n$) time range in between adjacent structural break points. Meanwhile the series can experience abrupt changes in its data generating mechanism at break points $p_1, \ldots, p_\ell$. Hence, the above PLS framework allows for a very flexible modeling of complexly time-varying temporal dynamics with both smooth and abrupt changes. We refer to [25,28] for more discussions and examples of the PLS time series models.

Suppose that we are interested in (a characteristic derived from) the distribution of $X_{n,i_{p,n}}$ for $i_{p,n} := \lfloor pn \rfloor$ for some fixed $p \in (0, 1)$. For our mathematical results, we will assume that $p \notin \{p_1, \ldots, p_\ell\}$. Let us use $F_{p,n}$ to denote the distribution function of $X_{n,i_{p,n}}$. Under some assumptions $F_{p,n}$ stabilizes as $n \to \infty$. Indeed, if $F_p$ denotes the distribution function of $\xi_p := \sum_{j=0}^{\ell} G_j(p, \boldsymbol{\epsilon}_0)\mathbb{1}_{(p_j, p_{j+1}]}(p)$, we obtain under some assumptions that $F_{p,n} \to F_p$ in some (nonuniform) sup-norm; for details and the rate of convergence see Lemma 3.1 below. Thus, under suitable conditions it can be reasonable to use

$$\widehat{F}_{p,n} := c_n \sum_{i=1}^{n} \kappa\left(\frac{i/n - i_{p,n}/n}{b_n}\right)\mathbb{1}_{[X_{n,i},\infty)} = c_n \sum_{i=1}^{n} \kappa\left(\frac{i - i_{p,n}}{nb_n}\right)\mathbb{1}_{[X_{n,i},\infty)} \tag{3}$$

as an estimator for $F_{p,n}$, where $\kappa : \mathbb{R} \to \mathbb{R}_+$ is a suitable (kernel) function, $b_n \in \mathbb{R}_{++}$ is a bandwidth, and $c_n := 1/\sum_{i=1}^{n} \kappa((i/n - i_{p,n}/n)/b_n)$ is a normalizing constant.

Our main result (Theorem 2.4 in conjunction with Remark 2.5) shows that under suitable assumptions

$$\mathcal{E}_{p,n}(\cdot) := \sqrt{nb_n}\big(\widehat{F}_{p,n}(\cdot) - F_{p,n}(\cdot)\big) \rightsquigarrow^* B_p \tag{4}$$

(with respect to a nonuniform sup-norm) for a non-degenerate Gaussian process $B_p$, where $\rightsquigarrow^*$ means convergence in distribution in the Hoffmann-Jørgensen sense [17] (see Remark 2.5 for the definition). In fact, we will show that under suitable assumptions

$$\widetilde{\mathcal{E}}_{p,n}(\cdot) := \sqrt{nb_n}\big(\widehat{F}_{p,n}(\cdot) - \mathbb{E}\big[\widehat{F}_{p,n}(\cdot)\big]\big) \rightsquigarrow^* B_p, \tag{5}$$

and we will discuss additional assumptions under which $\sqrt{nb_n}(F_{p,n}(\cdot) - \mathbb{E}[\widehat{F}_{p,n}(\cdot)]) \to 0$ (with respect to a nonuniform sup-norm). The convergence in (5) can be seen as the analogue of Theorem 1 in [27] where Wu proved a similar result for stationary time series (and with $\widehat{F}_{p,n}$ replaced by the classical empirical distribution function). For convergence in distribution of the finite-dimensional distributions, Wu employed Hannan's [16] central limit theorem for stationary time series. In our setting, where the underlying time series is non-stationary, we have to argue differently; see Section 4.1.

On the one hand, (4) yields consistency and the rate of convergence of the function-valued estimator $\widehat{F}_{p,n}(\cdot)$ for the distribution function $F_{p,n}(\cdot)$. On the other hand, in view of tools as the (extended) continuous mapping theorem and the functional delta-method, it can also be seen as a building stone for deriving the asymptotic distribution of the empirical plug-in estimator $\mathcal{T}(\widehat{F}_{p,n})$ for some characteristic $\mathcal{T}(F_{p,n})$ derived from $F_{p,n}$. Two specific examples will be discussed in Section 3.

The rest of the article is organized as follows. In Section 2, we present our main result, Theorem 2.4. At first glance, the imposed assumptions might look somewhat cumbersome. However, they are in line with the assumptions imposed by Wu [27] in the stationary case, and we will demonstrate that they are satisfied by two relevant PLS time series models, namely PLS linear processes and PLS ARCH processes. In Section 3, the functional weak limit theorem of Theorem 2.4 is applied to derive the asymptotic distribution of point estimators for quantiles and von Mises-characteristics of $F_{p,n}$. The proof of Theorem 2.4 is carried out in Section 4 (and Section 1 of the supplemental article [18]). All the others results will be proven in Section 2 of the supplemental article [18].

# 2. Main result

## 2.1. Physical dependence measure revisited

Before presenting our main result, we recall the definition of the physical dependence measure introduced by Wu [26] and extended by Zhou and Wu [30]. The dependence measure (more precisely the objects introduced in (6) and (7) below) will appear in assumptions (A5) and (A8) in Section 2.2. Let $\varepsilon^*$ be a real-valued random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{P}_{\varepsilon_0} = \mathbb{P}_{\varepsilon^*}$ and being independent of $\boldsymbol{\epsilon} = (\varepsilon_k)_{k\in\mathbb{Z}}$. If necessary, consider an enlargement of $(\Omega, \mathcal{F}, \mathbb{P})$. For every $i \in \mathbb{Z}$ and $r \in \mathbb{N}$, let

$$\boldsymbol{\epsilon}^*_{i,i-r} := \big(\varepsilon_i, \varepsilon_{i-1}, \dots, \varepsilon_{i-r+1}, \varepsilon^*, \varepsilon_{i-r-1}, \dots\big).$$

Note that $\boldsymbol{\epsilon}^*_{i,i-r}$ is a coupled version of $\boldsymbol{\epsilon}_i$ with $\varepsilon_{i-r}$ replaced by the i.i.d. copy $\varepsilon^*$. Let $I \subseteq \mathbb{R}$ be an interval, and $H : I \times \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ be any $(\mathcal{B}(I) \otimes \mathcal{B}(\mathbb{R})^{\otimes\mathbb{N}}, \mathcal{B}(\mathbb{R}))$-measurable map. For any $r \in \mathbb{N}$, $q > 0$, and $t \in I$, the *physical dependence measure* (associated with $H(t, \cdot)$ and $\boldsymbol{\epsilon}$) is defined by

$$\delta_{\boldsymbol{\epsilon},r;q}(H;t) := \big\| H(t, \boldsymbol{\epsilon}_0) - H\big(t, \boldsymbol{\epsilon}^*_{0,-r}\big)\big\|_q, \tag{6}$$

where $\|\cdot\|_q := \mathbb{E}[|\cdot|^q]^{1/q}$. Moreover, for any $r \in \mathbb{N}$ and $q > 0$, the *physical dependence measure* (associated with $H$ and $\boldsymbol{\epsilon}$) is defined by

$$\delta_{\boldsymbol{\epsilon},r;q}(H) := \sup_{t \in I} \delta_{\boldsymbol{\epsilon},r;q}(H;t). \tag{7}$$

Note that $\delta_{\boldsymbol{\epsilon},r;q}(H;t)$ and $\delta_{\boldsymbol{\epsilon},r;q}(H)$ will not change if in (6) $\boldsymbol{\epsilon}_0$ and $\boldsymbol{\epsilon}^*_{0,-r}$ are replaced by $\boldsymbol{\epsilon}_k$ and $\boldsymbol{\epsilon}^*_{k,k-r}$, respectively, for any $k \in \mathbb{Z} \setminus \{0\}$. The following Example 2.1 was already discussed on page 6 in [28].

**Example 2.1.** In the setting of Section 1, assume that specifically $G_j(\pi, (x_k)_{k\in\mathbb{N}}) := \sum_{s=0}^{\infty} a_{j,s}(\pi) x_{i+s}$ for some arbitrary functions $a_{j,s} : (p_j, p_{j+1}] \to \mathbb{R}$, $s \in \mathbb{N}_0$. Then

$$\delta_{\boldsymbol{\epsilon},r;q}(G_j) = \sup_{\pi \in (p_j, p_{j+1}]} \big\| a_{j,r}(\pi)\big(\varepsilon^*_{-r} - \varepsilon_{-r}\big)\big\|_q \le 2^{\max\{1,1/q\}} \|\varepsilon_0\|_q \sup_{\pi \in (p_j, p_{j+1}]} \big|a_{j,r}(\pi)\big|$$

for every $r \in \mathbb{N}$ and $q > 0$.

## 2.2. Assumptions and main result

As already mentioned in Section 1, our main result (Theorem 2.4 below) is a variant of Theorem 1 in [27]. In the latter theorem, Wu studied the case of stationary time series (i.e. $\ell = 0$ and $G_0$ independent of the first argument), where the role of $\widehat{F}_{p,n}$ was played by the classical empirical distribution function. For our result, we will impose nine assumptions, (A1)–(A9). Assumptions (A7) and (A8) are the analogues of Wu's assumptions (6) and (7), respectively. Assumption (A3) is the analogue of a moment condition on the marginal distribution of the time series in [27], and the analogue of (A6) was tacitly assumed in [27]. The additional assumptions (A1), (A2), (A4), and (A9) are due to the non-stationarity of our underlying time series model, and the additional assumption (A5) is a short range dependence condition which we need for the proof of convergence in distribution of the finite-dimensional distributions in (5) (recall that we will argue differently than Wu at this point). The assumption of exponential decaying physical dependence measures in (A5) can be relaxed to a sufficiently fast polynomial rate. However, such relaxation will require a much more tedious writing and it does not reflect the essence and main contributions of the article. Hence, we will assume exponentially decaying dependence throughout this paper.

In Theorem 2.4 below, we will assume that the following conditions (A3), (A7), and (A8) hold for a common $\lambda \ge 0$. Thus let $\lambda \ge 0$ be arbitrary but fixed. We will frequently use the function $\phi_s : \mathbb{R} \to [1, \infty)$ defined by $\phi_s(x) := (1 + |x|)^s$ for different $s \in \mathbb{R}$. We will also use the corresponding nonuniform sup-norm $\|\cdot\|_{(s)}$ defined by $\|v\|_{(s)} := \|v\phi_s\|_\infty$ with $\|v\|_\infty := \sup_{x\in\mathbb{R}} |v(x)|$. Please do not confuse the nonuniform sup-norm $\|\cdot\|_{(s)}$ for real-valued functions on $\mathbb{R}$ with the $L^q$-norm $\|\cdot\|_q := \mathbb{E}[|\cdot|^q]^{1/q}$ for random variables on $(\Omega, \mathcal{F}, \mathbb{P})$.

Regarding the kernel and the bandwidth we make the following assumptions.

(A1) The kernel function $\kappa$ is twice continuously differentiable on $\mathbb{R}$ with support $[-1, 1]$ and (without loss of generality) $\int_{\mathbb{R}} \kappa(u)\, du = 1$.

(A2) $\lim_{n\to\infty} nb_n = \infty$ and $\lim_{n\to\infty} b_n = 0$.

Let $j_p$ be the unique index $j$ with $p \in (p_j, p_{j+1})$. Then we have for $n$ sufficiently large (depending only on $p_{j_p}$ and $p_{j_p+1}$) that $i_{p,n}/n \in (p_{j_p}, p_{j_p+1})$. For every $n \in \mathbb{N}$ we use $I_{n;p}$ to denote the set of all $i \in \{1,\dots,n\}$ with $i/n \in (p_{j_p}, p_{j_p+1})$. We make the following assumptions, where the constant $q'$ in (A4) might differ from the constant $q'$ in (A5).

(A3) The distribution of $X_{n,i}$ has a Lebesgue density $f_{n,i}$ for any $i = 1,\dots,n$ and $n \in \mathbb{N}$, and $\sup_{n\in\mathbb{N}} \max_{1\le i\le n} \|f_{n,i}\|_{(\gamma)} < \infty$ for some $\gamma \in (2\lambda+1, \infty)$.
(A4) $\|G_{j_p}(\pi, \boldsymbol{\epsilon}_0) - G_{j_p}(\pi', \boldsymbol{\epsilon}_0)\|_{q'} \le C_{p,q'}|\pi - \pi'|$ for all $\pi, \pi' \in (p_{j_p}, p_{j_p+1}]$, and for some constants $q' \in (0,1]$ and $C_{p,q'} > 0$.
(A5) $\delta_{\boldsymbol{\epsilon},r;q'}(G_{j_p}) = \mathcal{O}(a^r)$ in $r \in \mathbb{N}$, for some constants $a \in [0,1)$ and $q' \in (0,1]$.

Here $\delta_{\boldsymbol{\epsilon},r;q'}$ refers to the physical dependence measure as defined in (7). Thus assertion (A5) means that $\delta_{\boldsymbol{\epsilon},r;q'}(G_{j_p};\pi)$ decays exponentially in $r$ uniformly in $\pi \in (p_{j_p}, p_{j_p+1}]$.

Now, denote by $\mathbb{P}_{X_{n,i}\|\boldsymbol{\epsilon}_{i-1}}$ a factorized regular version of the conditional distribution of $X_{n,i}$ (w.r.t. $\mathbb{P}$) given $\boldsymbol{\epsilon}_{i-1}$, that is, a probability kernel satisfying $\mathbb{P}_{X_{n,i}\|\boldsymbol{\epsilon}_{i-1}}(\boldsymbol{x}, B) = \mathbb{P}[X_{n,i} \in B \| \boldsymbol{\epsilon}_{i-1} = \boldsymbol{x}]$ for $\mathbb{P}_{\boldsymbol{\epsilon}_{i-1}}$-a.e. $\boldsymbol{x} \in \mathbb{R}^{\mathbb{N}}$, for all $B \in \mathcal{B}(\mathbb{R})$. Define a map $\mathfrak{F}_{n,i} : \mathbb{R}\times\mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ by

$$\mathfrak{F}_{n,i}(x, \boldsymbol{x}) := \mathbb{P}_{X_{n,i}\|\boldsymbol{\epsilon}_{i-1}}\big(\boldsymbol{x}, (-\infty, x]\big) \quad \big(= \mathbb{E}\big[\mathbb{1}_{(-\infty,x]}(X_{n,i}) \,\|\, \boldsymbol{\epsilon}_{i-1} = \boldsymbol{x}\big]\big),$$

which we refer to as *factorized conditional distribution function* of $X_{n,i}$ given $\boldsymbol{\epsilon}_{i-1}$. If $x \mapsto \mathfrak{F}_{n,i}(x, \boldsymbol{x})$ is twice differentiable for $\mathbb{P}_{\boldsymbol{\epsilon}_{i-1}}$-a.e. $\boldsymbol{x} \in \mathbb{R}^{\mathbb{N}}$, then we may define maps $\mathfrak{f}_{n,i} : \mathbb{R}\times\mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ and $\mathfrak{f}'_{n,i} : \mathbb{R}\times\mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ by

$$\mathfrak{f}_{n,i}(x, \boldsymbol{x}) := \begin{cases} \dfrac{\partial}{\partial x}\mathfrak{F}_{n,i}(x, \boldsymbol{x}), & \boldsymbol{x} \notin N_{i-1}, \\ 0, & \boldsymbol{x} \in N_{i-1} \end{cases} \quad \text{and} \quad \mathfrak{f}'_{n,i}(x, \boldsymbol{x}) := \frac{\partial}{\partial x}\mathfrak{f}_{n,i}(x, \boldsymbol{x}),$$

respectively, where $N_{i-1} \in \mathcal{B}(\mathbb{R})^{\otimes\mathbb{N}}$ is the respective $\mathbb{P}_{\boldsymbol{\epsilon}_{i-1}}$-null set. In this case, we refer to $\mathfrak{f}_{n,i}$ as *factorized conditional density* of $X_{n,i}$ given $\boldsymbol{\epsilon}_{i-1}$, and to $\mathfrak{f}'_{n,i}$ as its derivative. We make the following assumptions, where $\delta_{\boldsymbol{\epsilon},r;2}$ is defined as in (6).

(A6) For any $n \in \mathbb{N}$ and $i \in I_{n;p}$, the factorized conditional distribution function $x \mapsto \mathfrak{F}_{n,i}(x, \boldsymbol{x})$ is twice continuously differentiable for $\mathbb{P}_{\boldsymbol{\epsilon}_{i-1}}$-a.e. $\boldsymbol{x} \in \mathbb{R}^{\mathbb{N}}$.
(A7) For some $q \in (2,\infty)$ we have $\lim_{w\to\infty} M_q(\mathbb{R}\setminus(-w,w)) = 0$ and $M_q(\mathbb{R}) < \infty$, where

$$M_q(J) := \sup_{n\in\mathbb{N}} \int_J \max_{i\in I_{n;p}} \big\|\mathfrak{f}_{n,i}(x, \boldsymbol{\epsilon}_{i-1})\big\|_{q/2}^{q/2} \phi_{q\lambda-1+q/2}(x)\,dx.$$

(A8) For some $\alpha \in [0,1]$ and $\beta \in (0,\infty)$ we have $\lim_{w\to\infty} M_{i,\alpha}(\mathbb{R}\setminus(-w,w)) = 0$ and $M_{i,\alpha}(\mathbb{R}) < \infty$ for $i = 1, 2$ as well as $M_\beta(\mathbb{R}) < \infty$, where

$$M_{1,\alpha}(J) := \sup_{n\in\mathbb{N}} \sum_{r=1}^{\infty} \left\{ \max_{i\in I_{n;p}} \int_J \delta^2_{\boldsymbol{\epsilon},r-1;2}(\mathfrak{F}_{n,i}; x)\phi_{2\lambda-\alpha}(x)\,dx \right\}^{1/2},$$

$$M_{2,\alpha}(J) := \sup_{n\in\mathbb{N}} \sum_{r=1}^{\infty} \left\{ \max_{i\in I_{n;p}} \int_J \delta^2_{\boldsymbol{\epsilon},r-1;2}(\mathfrak{f}_{n,i};x)\phi_{2\lambda+\alpha}(x)\,dx \right\}^{1/2},$$

$$M_{\beta}(\mathbb{R}) := \sup_{n\in\mathbb{N}} \sum_{r=1}^{\infty} \left\{ \max_{i\in I_{n;p}} \int_{\mathbb{R}} \delta^2_{\boldsymbol{\epsilon},r-1;2}\big(\mathfrak{f}'_{n,i};x\big)\phi_{-\beta}(x)\,dx \right\}^{1/2}.$$

(A9) The distribution of $\xi_p := G_{j_p}(p, \boldsymbol{\epsilon}_0)$ has a bounded Lebesgue density $f_p$.

Before stating our main result (Theorem 2.4), we present two lemmas which are needed for (the statement of) the main result.

**Lemma 2.2.** *Let* $\kappa_2 := \int \kappa(x)^2\,dx$ *and assume that* (A1)–(A5) *and* (A9) *hold. Then*

$$\gamma_p(x, y) := \kappa_2 \sum_{k=-\infty}^{\infty} \mathbb{C}\mathrm{ov}\big(\mathbb{1}_{(-\infty,x]}\big(G_{j_p}(p, \boldsymbol{\epsilon}_k)\big), \mathbb{1}_{(-\infty,y]}\big(G_{j_p}(p, \boldsymbol{\epsilon}_0)\big)\big) \tag{8}$$

*is well-defined for any* $x, y \in \mathbb{R}$, *and the mapping* $(x, y) \mapsto \gamma_p(x, y)$ *is symmetric and positive semi-definite. Moreover we have* $\lim_{n\to\infty} \mathbb{E}[\widetilde{\mathcal{E}}_{p,n}(x)\widetilde{\mathcal{E}}_{p,n}(y)] = \gamma_p(x, y)$ *for any* $x, y \in \mathbb{R}$.

As a consequence of Lemma 2.2 there exists a centered Gaussian process with covariance function $\gamma_p$. This Gaussian process (respectively a suitable modification of it) will play the role of the limiting process in Theorem 2.4 below.

For any $\lambda \in \mathbb{R}_+$, let $\boldsymbol{D}_{(\lambda)}$ be the set of all bounded càdlàg functions $v : \mathbb{R} \to \mathbb{R}$ with $\lim_{x\to\pm\infty} v(x) = 0$ and $\|v\|_{(\lambda)}$ $(= \sup_{x\in\mathbb{R}} |v(x)|\phi_\lambda(x)) < \infty$. We equip $\boldsymbol{D}_{(\lambda)}$ with the nonuniform sup-norm $\|\cdot\|_{(\lambda)}$ and with the $\sigma$-algebra $\mathcal{D}_{(\lambda)}$ generated by the open balls w.r.t. $\|\cdot\|_{(\lambda)}$, often referred to as *open-ball $\sigma$-algebra*. Convergence in distribution will take place in $(\boldsymbol{D}_{(\lambda)}, \|\cdot\|_{(\lambda)})$. Empirical processes are known not to be Borel measurable w.r.t. sup-norms but only measurable w.r.t. the open-ball $\sigma$-algebra. In our setting the Borel $\sigma$-algebra is indeed strictly larger than the open-ball $\sigma$-algebra $\mathcal{D}_{(\lambda)}$ (which is possible due to the non-separability of càdlàg spaces w.r.t. sup-norms), and the $\mathcal{D}_{(\lambda)}$-measurability of (empirical) processes is a simple consequence of Lemma 4.1 in [4] (which states that $\mathcal{D}_{(\lambda)}$ coincides with the $\sigma$-algebra generated by the one-dimensional coordinate projections). Since continuous functions $f : \boldsymbol{D}_{(\lambda)} \to \mathbb{R}$ can be non-measurable w.r.t. $\mathcal{D}_{(\lambda)}$, we will say that a sequence $(\Xi_n)_{n\in\mathbb{N}}$ of $\mathcal{D}_{(\lambda)}$-measurable maps $\Xi_n : \Omega \to \boldsymbol{D}_{(\lambda)}$ converges in distribution *w.r.t. the open-ball $\sigma$-algebra* to some separable $(\boldsymbol{D}_{(\lambda)}, \mathcal{D}_{(\lambda)})$-valued random variable $\Xi$ (in symbols $\Xi_n \rightsquigarrow^\circ \Xi$) if $\int f\,d\mathbb{P}\circ\Xi_n^{-1} \to \int f\,d\,\mathrm{law}(\Xi)$ for all bounded, $\|\cdot\|_{(\lambda)}$-continuous *and $\mathcal{D}_{(\lambda)}$-measurable* functions $f : \boldsymbol{D}_{(\lambda)} \to \mathbb{R}$. This type of convergence was used in [20,21]; see also [6], Section 1.6, and the Appendices of [4,5] for further details. When the limiting process $\Xi$ is continuous, $\Xi_n \rightsquigarrow^\circ \Xi$ is equivalent to convergence in distribution of $\Xi_n$ to $\Xi$ in the Hoffmann-Jørgensen sense [17]; see Remark 2.5 below.

**Lemma 2.3.** *Assume that assumptions* (A1)–(A5) *and* (A9) *hold and let* $\gamma_p$ *be defined as in* (8). *Then any centered Gaussian process with covariance function* $\gamma_p$ *possesses a modification whose paths all lie in the set* $\mathbf{C}_{(\lambda)}$ *of all continuous elements of* $\boldsymbol{D}_{(\lambda)}$.

Lemma 2.3 ensures that we may and do assume that the Gaussian limiting process in the following theorem takes values only in a separable and measurable subset of $\boldsymbol{D}_{(\lambda)}$. This is crucial for the claim of the theorem. The processes $\mathcal{E}_{p,n}$ and $\widetilde{\mathcal{E}}_{p,n}$ were defined in (4) and (5), respectively.

**Theorem 2.4.** *If conditions* (A1)–(A9) *hold true for some common* $\lambda \geq 0$, *then*

$$\widetilde{\mathcal{E}}_{p,n}(\cdot) \rightsquigarrow^{\circ} B_p \quad \text{in } \big(\boldsymbol{D}_{(\lambda)}, \mathcal{D}_{(\lambda)}, \|\cdot\|_{(\lambda)}\big) \tag{9}$$

*for a continuous centered Gaussian process $B_p$ with covariance function $\gamma_p$ as defined in* (8). *In particular, if we assume in addition* $\sqrt{nb_n}\|F_{p,n}(\cdot) - \mathbb{E}[\widehat{F}_{p,n}(\cdot)]\|_{(\lambda)} \to 0$,

$$\mathcal{E}_{p,n}(\cdot) \rightsquigarrow^{\circ} B_p \quad \text{in } \big(\boldsymbol{D}_{(\lambda)}, \mathcal{D}_{(\lambda)}, \|\cdot\|_{(\lambda)}\big). \tag{10}$$

**Remark 2.5.** As the limiting process $B_p$ in (9) and (10) is continuous, we may replace in either case $\rightsquigarrow^{\circ}$ by convergence in distribution in the Hoffmann-Jørgensen sense [17] (usually denoted by $\rightsquigarrow^{*}$). This is ensured by part (i) of Theorem 1.7.2 in [24].

Recall that by definition a sequence $(\Xi_n)_{n\in\mathbb{N}}$ of maps $\Xi_n : \Omega \to \boldsymbol{D}_{(\lambda)}$ (possibly not being Borel-measurable w.r.t. $\|\cdot\|_{(\lambda)}$) is said to converge in distribution *in the Hoffmann-Jørgensen sense* to some Borel-measurable $(\boldsymbol{D}_{(\lambda)}, \|\cdot\|_{(\lambda)})$-valued random variable $\Xi$ if $\int^* f\, d\mathbb{P}\circ\Xi_n^{-1} \to \int f\, d\,\mathrm{law}(\Xi)$ for all bounded and $\|\cdot\|_{(\lambda)}$-continuous functions $f : \boldsymbol{D}_{(\lambda)} \to \mathbb{R}$. Here $\int^*$ refers to the outer integral (as introduced in, e.g., Section 1.2 in [24]).

The following Lemma 2.6 provides sufficient conditions for the additional condition in the second part of Theorem 2.4 to hold. It involves the following two conditions.

(B2) $\lim_{n\to\infty} nb_n^{(3q+1)/(q+1)} = 0$.
(B4) $\|G_{j_p}(\pi, \boldsymbol{\epsilon}_0) - G_{j_p}(\pi', \boldsymbol{\epsilon}_0)\|_q \leq C_{p,q}|\pi - \pi'|$ for all $\pi, \pi' \in (p_{j_p}, p_{j_p+1}]$.

Note that conditions (A2) and (B2) on the bandwidth $b_n$ are simultaneously fulfilled if, for instance, $b_n = n^{-\beta}$ for some $\beta \in (\frac{q+1}{3q+1}, 1)$.

**Lemma 2.6.** *If* (B2), (A3), (B4) *hold true for some* $\lambda \in [0,\infty)$, $q \in [\lambda,\infty)\cap(0,\infty)$, $C_{p,q} \in [0,\infty)$, *then* $\lim_{n\to\infty}\sqrt{nb_n}\|\mathbb{E}[\widehat{F}_{p,n}] - F_{p,n}\|_{(\lambda)} = 0$.

The proofs of Theorem 2.4 and Lemmas 2.2, 2.3, 2.6 will be carried out in Section 4 and in Sections 1–2 of the supplemental article [18]. There we will frequently use that under (A1) and (A2)

$$c_n = \mathcal{O}\big((nb_n)^{-1}\big) \quad \big(\text{in particular } c_n\sqrt{nb_n} = \mathcal{O}\big((nb_n)^{-1/2}\big)\big), \tag{11}$$

which follows from $\sum_{i=1}^n \kappa(\frac{i-i_{p,n}}{nb_n}) = nb_n\int_{-1}^{+1}\kappa(u)\,du + \mathcal{O}(1)$ under (A1) and (A2).

## 2.3. Some comments

It is worth mentioning that the structural change points $p_1,\dots,p_\ell$ in the time series model (2) do not have any impact on the mathematical result of Theorem 2.4. On the one hand, the number of observations $X_{n,i}$ that the estimator $\widehat{F}_{p,n}$ effectively employs is proportional to $nb_n$, or more specifically the time index $i$ of any effectively employed observation satisfies $|i - i_{p,n}| \leq nb_n$. On the other hand, $i_{p,n}$ lies in the interval $(np_{j_p}, np_{j_p+1})$ (for sufficiently large $n$) and the number of observations with time index $i$ in between $np_{j_p}$ and $i_{p,n}$ and the number of observations with time index $i$ in between $i_{p,n}$ and $np_{j_p+1}$ are proportional to $n$. Since $nb_n$ increases at a lower rate than $n$ (recall assumption (A2)), it follows that for sufficiently large $n$ the time indices $i$ of the observations $X_{n,i}$ that are effectively used by the estimator $\widehat{F}_{p,n}$ all lie in the (same) interval $(np_{j_p}, np_{j_p+1})$ or, equivalently, the corresponding relative indices $i/n$ all lie in the (same) interval $(p_{j_p}, p_{j_p+1})$. So we could have basically restricted ourselves to the case $\ell = 0$ (no break point).

On the other hand, several real world dynamics are subject to structural break points and can be considered locally stationary only piecewise. See for instance the discussions and examples in [12,28]. To make explicitly clear that our approach is applicable in such "real world" examples, we decided to present the model in its full generality ($\ell \geq 1$). In this context, it should also be mentioned that the specific choice of the bandwidth $b_n$ for finite sample size $n$ is an important issue in practical applications. The bandwidth $b_n$ should obviously be chosen in such a way that all of the observations effectively used by $\widehat{F}_{p,n}$ lie in between the same two adjacent break points. That means that $b_n$ cannot be chosen independently of the break points. But this means at the same time that the estimation of the break points is a crucial issue to be addressed in future research.

Finally, we would like to point out that if our time series (2) is stationary in between any two adjacent break points (i.e., if the functions $G_j$ are independent of the first argument) and if the bandwidth $b_n$ satisfies condition (A2), then, in view of the first paragraph of this Section, Theorem 1 in [27] (with sample size "$n := nb_n$") provides basically the same result as Theorem 2.4 (with $K = \frac{1}{2}\mathbb{1}_{[-1,1]}$, respectively a smooth approximation of it).

## 2.4. Illustrating examples

### 2.4.1. *PLS linear processes*

As in Example 2.1, let for any $j = 0, \ldots, \ell$ specifically $G_j(\pi, (x_k)_{k\in\mathbb{N}}) := \sum_{s=0}^{\infty} a_{j,s}(\pi) x_{i+s}$ for some functions $a_{j,s} : (p_j, p_{j+1}] \to \mathbb{R}$, $s \in \mathbb{N}_0$. In this case the corresponding process $(X_{n,i})_{i=1}^n$ can be seen as a PLS linear process. Without loss of generality, we assume $a_{j,0} \equiv 1$.

**Corollary 2.7.** *Let assumptions* (A1) *and* (A2) *be fulfilled. Assume that $a_{j_p,s}$ is continuously differentiable on $(p_{j_p}, p_{j_p+1}]$ for any $s \in \mathbb{N}$, and that the distribution of $\varepsilon_0$ has a Lebesgue density $f_\varepsilon$ that is twice continuously differentiable. Moreover, assume that for some given $\lambda \in [0, \infty)$ the following assertions hold.*

(a) $\sum_{s=1}^{\infty} \sup_{\pi\in(p_j, p_{j+1}]} |a_{j,s}(\pi)| < \infty$, $j = 0, \ldots, \ell$, *and* $\sup_{\pi\in(p_{j_p}, p_{j_p+1}]} |a_{j_p,s}(\pi)| = \mathcal{O}(a^s)$ *for some* $a \in [0, 1)$.
(b) $\sum_{s=1}^{\infty} \sup_{\pi\in(p_{j_p}, p_{j_p+1}]} |a'_{j_p,s}(\pi)|^{q'} < \infty$ *for some* $q' \in (0, 1]$.
(c) $\|\varepsilon_0\|_\gamma < \infty$ *for some* $\gamma \in (2\lambda + 3, \infty)$.
(d) $\|f'_\varepsilon\|_{(\alpha)} < \infty$ *and* $\|f''_\varepsilon\|_{(\beta)} < \infty$ *for some* $\alpha \in (\lambda + 1/2, \infty)$ *and* $\beta \in (1/2 - \lambda, \infty)$.

*Then* (9) *holds true. Moreover, if in addition* (b) *is satisfied for* $q' := 1$ *and condition* (B2) *is satisfied for* $q := 2\lambda + 4$, *then also* (10) *holds true.*

In the proof of Corollary 2.7 in Section 2.5 of the supplemental article [18], we will show that the assumptions of the corollary imply (A3)–(A9) and (B4).

### 2.4.2. *PLS ARCH processes*

Recall that the filters $\mathfrak{G}_n$, $n \in \mathbb{N}$, introduced in (1) are generated by $G_0, \dots, G_\ell$, and that $\boldsymbol{\epsilon} = (\varepsilon_k)_{k\in\mathbb{Z}}$ is a two-sided sequence of i.i.d. real-valued random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Assume that $\varepsilon_k$, $k \in \mathbb{Z}$, are nonnegative and that

$$G_j(\pi, \boldsymbol{x}_i) = \left(a_{j,0}(\pi) + \sum_{s=1}^{\mathcal{P}} a_{j,s}(\pi) G_j(\pi, \boldsymbol{x}_{i-s})\right) x_i \qquad \text{for any } \pi \in (p_j, p_{j+1}], \quad \mathbb{P}_{\boldsymbol{\epsilon}}\text{-a.e. } \boldsymbol{x} \in \mathbb{R}^{\mathbb{Z}} \tag{12}$$

for any $j = 0, \dots, \ell$ and $i \in \mathbb{N}$. Here, $\mathcal{P} \in \mathbb{N}$ is fixed, $a_{j,s} : [p_j, p_{j+1}] \to \mathbb{R}_+$, $s = 0, \dots, \mathcal{P}$, are any functions, and $\boldsymbol{x} := (x_k)_{k\in\mathbb{Z}}$ as well as $\boldsymbol{x}_i := (x_i, x_{i-1}, x_{i-2}, \dots)$. The existence of such functions $G_0, \dots, G_\ell$ under certain restrictions on $a_{j,s}$ and $\varepsilon_0$ will be provided in Lemma 2.8 below. In this case, we have in particular

$$G_j(i/n, \boldsymbol{\epsilon}_i) = \rho_{n,i,j}\varepsilon_i \quad \mathbb{P}\text{-a.s., where } \rho_{n,i,j} := a_{j,0}(i/n) + \sum_{s=1}^{\mathcal{P}} a_{j,s}(i/n) G_j(i/n, \boldsymbol{\epsilon}_{i-s}) \tag{13}$$

for any $j = 0, \dots, \ell$, $n \in \mathbb{N}$, and $i = 1, \dots, n$ with $i/n \in (p_j, p_{j+1}]$. If no structural break is possible (i.e., $\ell = 0$), then (13) can be seen as a variant of the time-varying ARCH (tvARCH) model introduced by Dahlhaus and Subba Rao [11] (and developed further by Fryzlewicz et al. [13], Fryzlewicz and Subba Rao [14], and others). In the latter references the roles of $\rho_{n,i,0}$ and $G_0(i/n, \boldsymbol{\epsilon}_{i-s})$ are played by $\sigma_i^2$ and $X_{i-s}^2$ respectively (similarly as in [15], page 4, in the stationary case). However, we do not only allow for smooth but also for abrupt changes of the coefficients (i.e., $\ell \geq 1$).

As before let $X_{n,i}$ be defined by (2) (with $G_0, \dots, G_\ell$ defined by (14) below). In view of (13) and the preceding comments, we refer to the process $(X_{n,i})_{i=1}^n$ as PLS ARCH($\mathcal{P}$) process. With regard to applications one might think of $X_{n,i}$ for instance, as the absolute value or squared value of an asset return.

Let us give a criterion for (12) to be valid (see Lemma 2.8 below). To this end let $v_{(1)}$ refer to the first entry of a vector $v \in \mathbb{R}^{\mathcal{P}}$ and set

$$\overline{b}_j(\pi, x) := \begin{bmatrix} a_{j,0}(\pi)x \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

$$A_j(\pi,x) := \begin{bmatrix} a_{j,1}(\pi)x & a_{j,2}(\pi)x & \dots & a_{j,\mathcal{P}-1}(\pi)x & a_{j,\mathcal{P}}(\pi)x \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

for any $\pi \in [0,1]$ and $x \in \mathbb{R}$. Under the validity of assertion (i) of Lemma 2.8 below we may define a function $G_j : [p_j, p_{j+1}] \times \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ by

$$G_j\big(\pi, (x_k)_{k\in\mathbb{N}}\big) := \begin{cases} \overline{b}_j(\pi, x_1)_{(1)} + \left(\sum_{r=0}^{\infty}\left\{\prod_{t=0}^{r} A_j(\pi, x_{t+1})\right\}\overline{b}_j(\pi, x_{r+2})\right)_{(1)}, & (x_k)_{k\in\mathbb{N}} \notin N, \\ 0, & (x_k)_{k\in\mathbb{N}} \in N \end{cases} \tag{14}$$

for some suitable $\mathbb{P}_{\boldsymbol{\epsilon}}$-null set $N$. In this case, we have

$$G_j(\pi, \boldsymbol{\epsilon}_i) = \overline{b}_j(\pi, \varepsilon_i)_{(1)} + \sum_{r=0}^{\infty}\left\{\left[\prod_{s=0}^{r} A_j(\pi, \varepsilon_{i-s})\right]\overline{b}_j(\pi, \varepsilon_{i-r-1})\right\}_{(1)} \quad \text{for any } \pi \in (p_j, p_{j+1}], \quad \mathbb{P}\text{-a.s.} \tag{15}$$

for any $j = 0, \dots, \ell$ and $i \in \mathbb{N}$. Note that (15) is in line with the vector representation of ARCH and GARCH processes considered in [1,7,14,22] and others.

In the following lemma, we mean by solution to (12) a measurable map $G_j : [p_j, p_{j+1}] \times \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ for which (12) holds for any $i \in \mathbb{N}$. We say that two solutions $G_j$ and $H_j$ generate the same samples almost surely if $G_j(i/n, \boldsymbol{\epsilon}_i) = H_j(i/n, \boldsymbol{\epsilon}_i)$ for all $j = 0, \dots, \ell$, $n \in \mathbb{N}$ and $i = 1, \dots, n$ with $i/n \in [p_j, p_{j+1}]$ $\mathbb{P}$-a.s. The proof of the lemma can be found in Section 2.6 of the supplemental article [18].

**Lemma 2.8.** *Assume that* $\|\varepsilon_0\|_q \max_{j=0,\dots,\ell} \sum_{s=0}^{\mathcal{P}} \sup_{\pi\in[p_j,p_{j+1}]} a_{j,s}(\pi) < 1$ *for some* $q \in [1,\infty)$. *Then for any* $j = 0, \dots, \ell$ *the following assertions hold true.*

- (i) *For any fixed* $t \in \mathbb{N}$, $\| \sup_{\pi\in[p_j,p_{j+1}]} \sum_{r=0}^{\infty}\{[\prod_{s=0}^{r} A_j(\pi, \varepsilon_{t-s})]\overline{b}_j(\pi, \varepsilon_{t-r-1})\}_{(1)}\|_q < \infty$ *and, in particular,* $\mathbb{P}$*-a.s. the series* $\sum_{r=0}^{\infty}\{[\prod_{s=0}^{r} A_j(\pi, \varepsilon_{t-s})]\overline{b}_j(\pi, \varepsilon_{t-r-1})\}_{(1)}$ *converges for any* $\pi \in [p_j, p_{j+1}]$.
- (ii) *The function* $G_j$ *defined by* (14) *is a solution of* (12).
- (iii) *If another solution* $H_j$ *of* (12) *satisfies* $\|H_j(i/n, \boldsymbol{\epsilon}_0)\|_q < \infty$ *for all* $n \in \mathbb{N}$ *and* $i = 1, \dots, n$ *with* $i/n \in [p_j, p_{j+1}]$, *then* $H_j$ *and* $G_j$ *generate the same samples almost surely.*

**Corollary 2.9.** *Let assumptions* (A1) *and* (A2) *be fulfilled. Assume that* $a_{j_p,s}$ *is continuously differentiable on* $[p_{j_p}, p_{j_p+1}]$ *for any* $s = 0, \dots, \mathcal{P}$, *and that the distribution of* $\varepsilon_0$ *has a Lebesgue density* $f_\varepsilon$ *that is twice continuously differentiable. Moreover, assume that for some given* $\lambda \in [0,\infty)$ *the following assertions hold.*

(a) $\min_{j=0,\ldots,\ell} \inf_{\pi\in(p_{j_p},p_{j_p+1}]} a_{j,0}(\pi) > 0$, *and*

$$\|\varepsilon_0\|_q \max_{j=0,\ldots,\ell} \sum_{s=0}^{\mathcal{P}} \sup_{\pi\in(p_j,p_{j+1}]} a_{j,s}(\pi) < 1 \quad \textit{for some } q \in (4\lambda+2,\infty).$$

(b) $\|f_\varepsilon\|_{(\gamma)} + \|f_\varepsilon'\|_{(\gamma)} < \infty$ *for some* $\gamma \in (2\lambda+1,\infty)$.
(c) $\|f_\varepsilon''\|_{(0)} < \infty$.

*Then* (9) *holds true. Moreover, if in addition condition* (B2) *is satisfied for q from assumption* (a), *then also* (10) *holds true.*

In the proof of Corollary 2.9 in Section 2.7 of the supplemental article [18], we will show that the assumptions of the corollary imply (A3)–(A9) and (B4).

# 3. Applications

## 3.1. A preliminary result

Theorem 2.4 and Lemma 2.6 show that the convergence in (10) holds true if conditions (A1)–(A9) as well as (B2) and (B4) are satisfied. By the following Lemma 3.1 (and Slutsky's theorem in the form of Corollary A.2 in [5]) we can immediately conclude that under the same assumptions

$$\sqrt{nb_n}\big(\widehat{F}_{p,n}(\cdot) - F_p(\cdot)\big) \rightsquigarrow^\circ B_p \quad \text{in } \big(\boldsymbol{D}_{(\lambda)}, \mathcal{D}_{(\lambda)}, \|\cdot\|_{(\lambda)}\big), \tag{16}$$

because Lemma 3.1 ensures

$$\sqrt{nb_n}\|F_{p,n} - F_p\|_{(\lambda)} \to 0. \tag{17}$$

Here $F_p$ refers to the distribution function of $\xi_p$ introduced a few lines before (3). Lemma 3.1 involves the following condition.

(C2) $\lim_{n\to\infty} n^{(1-q)/(1+q)} b_n = 0$.

Note that (B2) implies (C2), and that (A2) implies (C2) if $q \geq 1$.

**Lemma 3.1.** *If* (C2), (A3), (B4) *hold true for some* $\lambda \in [0,\infty)$, $q \in [\lambda,\infty)\cap(0,\infty)$, $C_{p,q} \in [0,\infty)$, *then* (17) *holds.*

In the proof of Lemma 3.1 (see Section 2.8 of the supplemental article [18]), we will show that (A3) and (B4) imply $\|F_{p,n} - F_p\|_{(\lambda)} = \mathcal{O}(n^{-q/(q+1)})$; together with (C2) this ensures the claim of the lemma. Let us summarize our findings.

**Corollary 3.2.** *Assume that* (A1)–(A9) *hold for some common* $\lambda \in [0,\infty)$. *Moreover, assume that* (B2) *and* (B4) *hold for some* $q \in [\lambda,\infty)\cap(0,\infty)$ *with the same* $\lambda$. *Then* (16) *and* (17) *hold.*

## 3.2. Weighted empirical quantiles

The (lower) $\alpha$-quantile functional associated with some given level $\alpha \in (0, 1)$ is defined by

$$\mathcal{Q}_\alpha(F) := \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}$$

on the set of all distribution functions $F$ on the real line. Given the time series $X_{n,1}, \dots, X_{n,n}$, it can be reasonable to use $\mathcal{Q}_\alpha(\widehat{F}_{p,n})$ as an estimator for $\mathcal{Q}_\alpha(F_{p,n})$. Note that $\mathcal{Q}_\alpha(\widehat{F}_{p,n})$ can be seen as a weighted $\alpha$-quantile. The estimator $\widehat{F}_{p,n}$ is indeed supported by the finite set $\{X_{n,1}, \dots, X_{n,n}\}$, but the mass assigned to the individual points of this set is not uniform. More precisely, denoting by $X_{n,1(n)}, \dots, X_{n,n(n)}$ the order statistics of $X_{n,1}, \dots, X_{n,n}$, that is, using $i(n)$ to denote the time index of the $i$-th largest element of $X_{n,1}, \dots, X_{n,n}$ (where the specific ranking of elements with the same value is not relevant), we have

$$\mathcal{Q}_\alpha(\widehat{F}_{p,n}) = X_{n,k(n)} \quad \text{for the smallest } k \in \{1, \dots, n\} \text{ with } \sum_{i=1}^{k} w_n\big(i(n)\big) \geq \alpha,$$

where $w_n(i(n)) := c_n \kappa(\frac{i(n)-i_{p,n}}{nb_n})$ refers to the mass assigned to $X_{n,i(n)}$.

Given (16) and (17), we can use the functional delta-method to obtain

$$\sqrt{nb_n}\big(\mathcal{Q}_\alpha(\widehat{F}_{p,n}) - \mathcal{Q}_\alpha(F_{p,n})\big) \rightsquigarrow Z_\alpha \tag{18}$$

for some centered normally distributed random variable $Z_\alpha$ with variance

$$\mathbb{V}\mathrm{ar}[Z_\alpha] = \frac{\gamma_p(F_p^{-1}(\alpha), F_p^{-1}(\alpha))}{F_p'(F_p^{-1}(\alpha))^2} \tag{19}$$

under some assumption on $F_p$, where $\gamma_p$ is the covariance function defined by (8). In fact, we even get a process version of (18) for the corresponding quantile process, namely

$$\big(\sqrt{nb_n}\big(\mathcal{Q}_\alpha(\widehat{F}_{p,n}) - \mathcal{Q}_\alpha(F_{p,n})\big)\big)_{\alpha\in[\alpha_1,\alpha_2]} \rightsquigarrow^* (Z_\alpha)_{\alpha\in[\alpha_1,\alpha_2]} \quad \text{in } \ell^\infty\big([\alpha_1, \alpha_2]\big) \tag{20}$$

for a continuous centered Gaussian process $(Z_\alpha)_{\alpha\in[\alpha_1,\alpha_2]}$ with covariance function

$$\mathbb{C}\mathrm{ov}(Z_\alpha, Z_\beta) = \frac{\gamma_p(F_p^{-1}(\alpha), F_p^{-1}(\beta))}{F_p'(F_p^{-1}(\alpha))F_p'(F_p^{-1}(\beta))}, \tag{21}$$

where $0 < \alpha_2 < \alpha_2 < 1$ are arbitrary but fixed.

**Theorem 3.3.** *Assume that (16) and (17) hold for $\lambda = 0$. If $F_p$ is continuously differentiable on a closed interval $[a_1, a_2] \subseteq (0, 1)$ with strictly positive derivative, and $a_1 < F_p^{-1}(\alpha_1)$, $F_p^{-1}(\alpha_2) < a_2$, then (20) holds. In particular, if $F_p$ is continuously differentiable in a neighborhood of $F_p^{-1}(\alpha)$ with strictly positive derivative at $F_p^{-1}(\alpha)$, then (18) holds.*

**Proof.** In view of (16) and Remark 2.5, we obtain by Lemma 21.4 and Theorem 20.8 in [23] that $(\sqrt{nb_n}(\mathcal{Q}_\alpha(\widehat{F}_{p,n}) - \mathcal{Q}_\alpha(F_p)))_{\alpha\in[\alpha_1,\alpha_2]} \rightsquigarrow^* (Z_\alpha)_{\alpha\in[\alpha_1,\alpha_2]}$ in $\ell^\infty([\alpha_1,\alpha_2])$, noting that $(Z_\alpha)_{\alpha\in[\alpha_1,\alpha_2]}$ defined by $Z_\alpha := -B_p(F_p^{-1}(\alpha))/F_p'(F_p^{-1}(\alpha))$ is a continuous centered Gaussian process with covariance function as in (21). Moreover, in view of (17), we obtain by another application of Lemma 21.4 and Theorem 20.8 in [23] (to purely deterministic processes) that $(\sqrt{nb_n}(\mathcal{Q}_\alpha(F_{p,n}) - \mathcal{Q}_\alpha(F_p)))_{\alpha\in[\alpha_1,\alpha_2]} \to \mathbf{0}$. Along with Slutsky's theorem (in the form of Theorems 18.10(v) and 18.11(i) in [23]) this gives (20). This proves the first assertion. The second assertion is implied by the first one. □

## 3.3. Weighted V-statistics

The V-functional (von Mises functional) of degree two associated with some given measurable function $g : \mathbb{R}^2 \to \mathbb{R}$ (often referred to as kernel) is defined by

$$\mathcal{V}_g(F) := \iint g(x_1, x_2)\, dF(x_1)\, dF(x_2) \tag{22}$$

on the set $\mathbf{F}_g$ of all distribution functions $F$ on the real line for which the double integral in (22) exists. Given the time series $X_{n,1}, \dots, X_{n,n}$, it can be reasonable to use $\mathcal{V}_g(\widehat{F}_{p,n})$ as an estimator for $\mathcal{V}_g(F_{p,n})$. Note that $\mathcal{V}_g(\widehat{F}_{p,n})$ can be seen as a weighted V-statistic. It indeed admits the representation

$$\mathcal{V}_g(\widehat{F}_{p,n}) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_n(i,j) g(X_{n,i}, X_{n,j})$$

with $w_n(i,j) := c_n^2 \kappa(\frac{i-i_{p,n}}{nb_n})\kappa(\frac{j-i_{p,n}}{nb_n})$.

Given (16) and (17), we can follow the continuous mapping approach by Beutner and Zähle [3] to show that under some assumptions (see Theorem 3.4 below)

$$\sqrt{nb_n}\big(\mathcal{V}_g(\widehat{F}_{p,n}) - \mathcal{V}_g(F_{p,n})\big) \rightsquigarrow Z \tag{23}$$

for some centered normally distributed random variable $Z$ with variance

$$\mathbb{V}\mathrm{ar}[Z] = \iint \gamma_p(x_1, x_2)\, dg_{F_p}(x_1)\, dg_{F_p}(x_2), \tag{24}$$

where $\gamma_p$ is the covariance function defined by (8), and $g_{F_p} := g_{1,F_p} + g_{2,F_p}$ with $g_{1,F_p}(\cdot) := \int g(\cdot, x_2)\, dF_p(x_2)$ and $g_{2,F_p}(\cdot) := \int g(x_1, \cdot)\, dF_p(x_1)$.

Let us collect the assumptions we need for (23). Assume $F_p \in \mathbf{F}_g$ and $F_{p,n} \in \mathbf{F}_g$ as well as $\iint |g(x_1,x_2)|\, dF_{p,n}(x_1)\, dF_p(x_2) < \infty$ and $\iint |g(x_1,x_2)|\, dF_p(x_1)\, dF_{p,n}(x_2) < \infty$ for any $n \in \mathbb{N}$. Assume that $g_{1,F_p}$ and $g_{2,F_p}$ are right-continuous and locally of bounded variation, that $g$ is upper right-continuous and locally of bounded bivariation, and that $g_{x_1}(\cdot) := g(x_1, \cdot)$ and $g_{x_2}(\cdot) := g(\cdot, x_2)$ are locally of bounded variation for every fixed real $x_1$ and $x_2$, respectively.

Under some weak additional assumptions (see, e.g., Remark 3.5) making the tail behavior of $g_{F_p}$ and $F_p$ and of $g$ and $F_p$ compatible, one can derive from (22) the decomposition

$$\begin{aligned}\mathcal{V}_g(\widehat{F}_{p,n}) - \mathcal{V}_g(F_p) = &-\int (\widehat{F}_{p,n} - F_p)(x-)\,dg_{F_p}(x) \\ &+ \iint (\widehat{F}_{p,n} - F_p)(x_1-)(\widehat{F}_{p,n} - F_p)(x_2-)\,dg(x_1, x_2)\end{aligned} \tag{25}$$

and its analogue with $\widehat{F}_{p,n}$ replaced by $F_{p,n}$. Then, under (16) and (17), the continuous mapping theorem (in the form of Theorem 6.4 of [6]) and Slutsky's theorem (in the form of Corollary A.2 in [5]) imply the following theorem, where one should note that $Z := -\int B_p(x-)\,dg_{F_p}(x)$ is normally distributed with variance as in (24).

**Theorem 3.4.** *Assume that (16) and (17) hold for some $\lambda \in [0, \infty)$ and that $F_p$ is continuous. Moreover, assume that (25) and its analogue with $\widehat{F}_{p,n}$ replaced by $F_{p,n}$ hold for any $n \in \mathbb{N}$, and that $\int \phi_{-\lambda}(x)|dg_{F_p}|(x) < \infty$ and $\iint \phi_{-\lambda}(x_1)\phi_{-\lambda}(x_2)|dg|(x_1, x_2) < \infty$. Then (23) holds.*

**Remark 3.5.** The conditions in Lemmas 3.4 and 3.6 in [3] (with $\widehat{F}_n$, $F$ replaced by $\widehat{F}_{p,n}$, $F_p$) provide simple (but lengthy) conditions for (25). The analogous assumptions with $\widehat{F}_{p,n}$ replaced by $F_{p,n}$ ensure (25) with $\widehat{F}_{p,n}$ replaced by $F_{p,n}$.

In some situations the function $g_{F_p}$ introduced subsequent to (24) is constant so that the first summand on the right-hand side of (25) vanishes. In this case, the weighted V-statistic is said to be *degenerate w.r.t. $F_p$*. For degenerate (weighted) V-statistics the variance in (24) is obviously equal to 0. Nevertheless for many degenerate (weighted) V-statistics one can derive a non-degenerate limiting distribution. Replacing the normalizing factor $\sqrt{nb_n}$ by $nb_n$, one can derive the following variant of (23):

$$nb_n\big(\mathcal{V}_g(\widehat{F}_{p,n}) - \mathcal{V}_g(F_{p,n})\big) \rightsquigarrow Y \tag{26}$$

with $Y := \iint B_p(x_1-)B_p(x_2-)\,dg(x_1, x_2)$. Indeed, in view of (25) an application of the continuous mapping theorem (in the form of Theorem 6.4 of [6]) along with Slutsky's theorem (in the form of Corollary A.2 in [5]) yields the following result.

**Theorem 3.6.** *Assume that the assumptions of Theorem 3.4 hold and that the weighted V-statistic is degenerate w.r.t. $F_p$. Then (26) holds.*

As elaborated in Section 3.2 of [3], the set of kernels $g$ that satisfy the conditions mentioned in Remark 3.5 (and thus admit the representation (25)) include the kernels corresponding to the variance, to Gini's mean difference, to the Cramér–von Mises goodness-of-fit test statistic, and to the Arcones–Giné test statistic for symmetry. The latter two kernels lead to degenerate (weighted) V-statistics; see Examples 3.13 and 3.14 in [3]. Let us detail out two of the four mentioned examples.

**Example 3.7 (Gini's mean difference).** Let specifically $g(x_1, x_2) := |x_1 - x_2|$. If $F$ is a distribution function on the real line that has a finite first absolute moment, then $\mathcal{V}_g(F)$ equals Gini's mean difference $\mathbb{E}[|X_1 - X_2|]$ of two i.i.d. random variables $X_1$ and $X_2$ distributed according to $F$. Assume that (16) and (17) hold for some $\lambda \in (1, \infty)$ and that $F_p$ is continuous and satisfies $\|F_p - \mathbb{1}_{[0,\infty)}\|_{(\lambda)} < \infty$. It can be easily seen from the elaborations in Example 3.10 of [3] that the assumptions of Theorem 3.4 hold true for this choice of $\lambda$ and $F_p$. Thus, Theorem 3.4 implies (23) where $dg_{1,F_p}(x) = dg_{2,F_p}(x) = (2F_p(x) - 1)\,dx$ (see Example 3.1 of [2]) for the specific choice of $g$.

**Example 3.8 (Cramér–von Mises test statistic).** Assume that $F_{p,n}$ approximates $F_p$ as $n \to \infty$ in a suitable sense (below we will specifically require (17) with $\lambda = 0$). Let $F_p^0$ be a distribution function on the real line. For large $n$, the Cramér–von Mises test statistic

$$\widehat{T}_n^0 := \int \big(\widehat{F}_{p,n}(x) - F_p^0(x)\big)^2\, dF_p^0(x)$$

based on the estimator $\widehat{F}_{p,n}$ is a candidate for testing the null hypothesis $F_p = F_p^0$ (or, a little more informally, $F_{p,n} \sim F_p^0$). It can be expressed as V-statistic $\mathcal{V}_g(\widehat{F}_{p,n})$ with kernel

$$g(x_1, x_2) := \int \big(\mathbb{1}_{[x_1,\infty)}(x) - F_p^0(x)\big)\big(\mathbb{1}_{[x_2,\infty)}(x) - F_p^0(x)\big)\, dF_p^0(x). \tag{27}$$

Under the null $F_p = F_p^0$ we have $g_{1,F_p} \equiv g_{2,F_p} \equiv 0$ (i.e. $\mathcal{V}_g(\widehat{F}_{p,n})$ is degenerate w.r.t. $F_p$). This implies that under the null the expression in (24) is equal to zero, and therefore Theorem 3.4 is not profitable in the present case. Though we can employ Theorem 3.6.

Assume that (16) and (17) hold for $\lambda = 0$ and that $F_p$ is continuous and $F_p = F_p^0$. It can be easily seen from the elaborations in Example 3.13 of [3] that the assumptions of Theorem 3.4 hold true for $\lambda = 0$. Thus, Theorem 3.6 implies (26) where $dg(x_1, x_2) = \mathcal{H}^1_{F_p^0}(d(x_1, x_2))$ with $\mathcal{H}^1_{F_p^0}(A) := \int \delta_{(x,x)}(A)\, dF_0^p(x)$, $A \in \mathcal{B}(\mathbb{R}^2)$, for the specific choice of $g$ in (27).

In Corollary 1 and Example 2 in [29], Zhou presents the analogue of (23) with $\mathcal{V}_g(F_{p,n})$ replaced by $\mathcal{V}_g(\mathbb{E}[\widehat{F}_{p,n}])$. More precisely, he proves that the standardized V-statistic $(\mathcal{V}_g(\widehat{F}_{p,n}) - \mathbb{E}[\mathcal{V}_g(\widehat{F}_{p,n})])/\mathbb{V}\mathrm{ar}[\mathcal{V}_g(\widehat{F}_{p,n})]^{1/2}$ is asymptotically standard normal under similar assumptions. In Theorem 3 and Corollary 4 in [29], he also presents an analogue of Theorem 3.6, where he establishes that a class of degenerated weighted V-statistics will converge to a mixture of i.i.d. centered $\chi^2(1)$ random variables.

## 4. Proof of Theorem 2.4

In the following, we will only show that (9) holds true, because (10) is a trivial consequence of (9) and Slutsky's theorem (in the form of Corollary A.2 in [5]). For (9) it suffices to show

$$\phi_\lambda \widetilde{\mathcal{E}}_{p,n} \rightsquigarrow^\circ \phi_\lambda B_p \quad \text{in } \big(\boldsymbol{D}_{(0)}, \mathcal{D}_{(0)}, \|\cdot\|_{(0)}\big) \tag{28}$$

(note that $\|\cdot\|_{(0)} = \|\cdot\|_\infty$). Indeed, the continuous mapping theorem (in the form of Theorem 6.4 of [6]) and the continuity of the mapping $v \mapsto v/\phi_\lambda$ from $(\boldsymbol{D}_{(0)}, \|\cdot\|_{(0)})$ to $(\boldsymbol{D}_{(\lambda)}, \|\cdot\|_{(\lambda)})$ together ensure that (28) implies (9). To show (28) we will verify in Sections 4.1 and 4.2 that conditions (a) and (b) of the following Theorem 4.1 are satisfied, if $\phi_\lambda \widetilde{\mathcal{E}}_{p,n}$ and $\phi_\lambda B_p$ play the roles of $\xi_n$ and $\xi$, respectively. Thus, (28) is a consequence of Theorem 4.1. Theorem 4.1 is a straightforward generalization of Theorem V.1.3 in [20]; we omit its proof.

**Theorem 4.1.** *Let $\xi_n$ be a $(\boldsymbol{D}_{(0)}, \mathcal{D}_{(0)})$-valued random variable on some probability space $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ for every $n \in \mathbb{N}$. Let $\boldsymbol{C}_{(0)} \in \mathcal{D}_{(0)}$ be separable, and $\xi$ be a $(\boldsymbol{D}_{(0)}, \mathcal{D}_{(0)})$-valued random variable on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\xi \in \boldsymbol{C}_{(0)}$ $\mathbb{P}$-a.s. Assume that the following two conditions hold.*

(a) *The finite-dimensional distributions of $\xi_n$ converge in distribution to those of $\xi$.*
(b) *For every $\epsilon > 0$ and $\delta > 0$ there exist $k \in \mathbb{N}$ and a partition $-\infty = x_0 < x_1 < \cdots < x_k < x_{k+1} = \infty$ such that*

$$\limsup_{n\to\infty} \mathbb{P}_n\Big[\max_{i=0,\dots,k} \sup_{x\in[x_i,x_{i+1})} \big|\xi_n(x) - \xi_n(x_i)\big| \ge \delta\Big] \le \epsilon.$$

*Then $\xi_n \rightsquigarrow^\circ \xi$ in $(\boldsymbol{D}_{(0)}, \mathcal{D}_{(0)}, \|\cdot\|_{(0)})$.*

Note that $\phi_\lambda \widetilde{\mathcal{E}}_{p,n}$ and $\phi_\lambda B_p$ take values in $\boldsymbol{D}_{(0)}$ and $\boldsymbol{C}_{(0)}$, respectively. This is ensured by Lemma 1.1 in the supplemental article [18] and Lemma 2.3, respectively.

## 4.1. Verification of condition (a) of Theorem 4.1

Let $\boldsymbol{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ be arbitrary but fixed, and assume that $x_1 < \cdots < x_d$. Here we show that (under assumptions (A1)–(A5) and (A9)) we have

$$\big(\phi_\lambda(x_1)\widetilde{\mathcal{E}}_{p,n}(x_1), \dots, \phi_\lambda(x_d)\widetilde{\mathcal{E}}_{p,n}(x_d)\big)' \rightsquigarrow \big(\phi_\lambda(x_1)B_p(x_1), \dots, \phi_\lambda(x_d)B_p(x_d)\big)'.$$

By the continuous mapping theorem (in the form of Theorem 6.4 of [6]) it suffices to show that

$$\big(\widetilde{\mathcal{E}}_{p,n}(x_1), \dots, \widetilde{\mathcal{E}}_{p,n}(x_d)\big)' \rightsquigarrow \big(B_p(x_1), \dots, B_p(x_d)\big)'.$$

Due to the Cramér–Wold theorem it even suffices to show that

$$\sum_{k=1}^d \lambda_k \widetilde{\mathcal{E}}_{p,n}(x_k) \rightsquigarrow \sum_{k=1}^d \lambda_k B_p(x_k) \tag{29}$$

for every $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d) \in \mathbb{R}^d$. For the proof of (29), we borrow arguments from the proof of Theorem 1 in [29]. Setting

$$Y_{n,i}(\boldsymbol{x}, \boldsymbol{\lambda}) := \sum_{k=1}^d \lambda_k \kappa\left(\frac{i - i_{p,n}}{nb_n}\right)\big(\mathbb{1}_{[X_{n,i},\infty)}(x_k) - \mathbb{E}\big[\mathbb{1}_{[X_{n,i},\infty)}(x_k)\big]\big),$$

$$Y_{n,i}^{\{m_n\}}(\boldsymbol{x}, \boldsymbol{\lambda}) := \mathbb{E}\big[Y_{n,i}(\boldsymbol{x}, \boldsymbol{\lambda})\big|\boldsymbol{\epsilon}_{i:i-m_n}\big]$$

with $\boldsymbol{\epsilon}_{i:i-m_n} := (\varepsilon_i, \varepsilon_{i-1}, \dots, \varepsilon_{i-m_n+1})$ and $m_n := \lceil \log(n) \rceil$, the left-hand side of (29) can be written as

$$\begin{aligned}
&\sum_{k=1}^{d} \lambda_k \widetilde{\mathcal{E}}_{p,n}(x_k) \\
&\quad = c_n \sqrt{nb_n} \sum_{i=1}^{n} Y_{n,i}(\boldsymbol{x}, \boldsymbol{\lambda}) \\
&\quad = c_n \sqrt{nb_n} \left( \sum_{i=1}^{n} Y_{n,i}(\boldsymbol{x}, \boldsymbol{\lambda}) - \sum_{i=1}^{n} Y_{n,i}^{\{m_n\}}(\boldsymbol{x}, \boldsymbol{\lambda}) \right) + c_n \sqrt{nb_n} \sum_{i=1}^{n} Y_{n,i}^{\{m_n\}}(\boldsymbol{x}, \boldsymbol{\lambda}) \\
&\quad =: S_{n,1}(\boldsymbol{x}, \boldsymbol{\lambda}) + S_{n,2}(\boldsymbol{x}, \boldsymbol{\lambda}). \qquad (30)
\end{aligned}$$

The summand $S_{n,1}(\boldsymbol{x}, \boldsymbol{\lambda})$ converges in probability to 0 by Lemma 1.4 in the supplemental article [18] and (11). We will now prove that the summand $S_{n,2}(\boldsymbol{x}, \boldsymbol{\lambda})$ converges in distribution to the right-hand side in (29), which is a centered normally distributed random variable with variance

$$\mathbb{V}\mathrm{ar}\left[\sum_{k=1}^{d} \lambda_k B_p(x_k)\right] = \sum_{k=1}^{d} \sum_{l=1}^{d} \lambda_k \lambda_l \mathbb{E}\big[B_p(x_k) B_p(x_l)\big] = \sum_{k=1}^{d} \sum_{l=1}^{d} \lambda_k \lambda_l \gamma_p(x_k, x_l). \qquad (31)$$

Along with Slutsky's theorem, this gives (29).

If the expression in (31) vanishes, then $\sum_{k=1}^{d} \lambda_k B_p(x_k) = 0$ $\mathbb{P}$-a.s. and $\lim_{n\to\infty} \mathbb{V}\mathrm{ar}[\sum_{k=1}^{d} \lambda_k \widetilde{\mathcal{E}}_{p,n}(x_k)] = 0$ by Lemma 2.2. The latter convergence implies $\lim_{n\to\infty} \| \sum_{k=1}^{d} \lambda_k \widetilde{\mathcal{E}}_{p,n}(x_k) \|_2 = 0$, i.e. $\lim_{n\to\infty} \| \sum_{k=1}^{d} \lambda_k \widetilde{\mathcal{E}}_{p,n}(x_k) - \sum_{k=1}^{d} \lambda_k B_p(x_k) \|_2 = 0$. Thus $\sum_{k=1}^{d} \lambda_k \widetilde{\mathcal{E}}_{p,n}(x_k)$ converges in distribution to $\sum_{k=1}^{d} \lambda_k B_p(x_k)$, i.e. (29) holds.

Now assume that the expression in (31) is strictly greater than 0. Then it suffices to show that

$$\frac{S_{n,2}(\boldsymbol{x}, \boldsymbol{\lambda})}{\sqrt{\mathbb{V}\mathrm{ar}[\sum_{k=1}^{d} \lambda_k B_p(x_k)]}} \rightsquigarrow Z$$

for a standard normally distributed random variable $Z$. By Slutsky's theorem and Lemma 1.7(iv) in the supplemental article [18] this is equivalent to

$$\frac{S_{n,2}(\boldsymbol{x}, \boldsymbol{\lambda})}{\sqrt{\mathbb{V}\mathrm{ar}[S_{n,2}(\boldsymbol{x}, \boldsymbol{\lambda})]}} \rightsquigarrow Z. \qquad (32)$$

To verify (32), we split $S_{n,2}(\boldsymbol{x}, \boldsymbol{\lambda})$ into sums

$$S_{n,2}(\boldsymbol{x}, \boldsymbol{\lambda}) = c_n \sqrt{nb_n} \sum_{j=1}^{\lceil n/s_n \rceil} R_{n,j}(\boldsymbol{x}, \boldsymbol{\lambda}) + c_n \sqrt{nb_n} \sum_{j=1}^{\lceil n/s_n \rceil} r_{n,j}(\boldsymbol{x}, \boldsymbol{\lambda}) \qquad (33)$$

of $\lceil n/s_n \rceil$ many big blocks

$$R_{n,j}(\boldsymbol{x},\boldsymbol{\lambda}) := \sum_{i=1}^{l_n} Y_{n,(j-1)s_n+i}^{\{m_n\}}(\boldsymbol{x},\boldsymbol{\lambda}), \qquad j = 1, 2, \ldots, \lceil n/s_n \rceil, \tag{34}$$

and $\lceil n/s_n \rceil$ many small blocks

$$r_{n,j}(\boldsymbol{x},\boldsymbol{\lambda}) := \sum_{i=l_n+1}^{s_n} Y_{n,(j-1)s_n+i}^{\{m_n\}}(\boldsymbol{x},\boldsymbol{\lambda}), \qquad j = 1, 2, \ldots, \lceil n/s_n \rceil, \tag{35}$$

where $l_n := \lceil \sqrt{nb_n} \rceil$ and $s_n := l_n + \lceil (\log n)^2 \rceil$. Recall $m_n = \lceil \log(n) \rceil$, and note that the big blocks are independent since $s_n - l_n > m_n - 2$, and that the small blocks are independent since $m_n < l_n + 2$.

Now, $c_n\sqrt{nb_n}\sum_{j=1}^{\lceil n/s_n \rceil} r_{n,j}(\boldsymbol{x},\boldsymbol{\lambda})$ converges in probability to 0 by (23) in the supplemental article [18] and (11). Moreover, $\lim_{n\to\infty} \mathbb{V}\mathrm{ar}[S_{n,2}(\boldsymbol{x},\boldsymbol{\lambda})]/\mathbb{V}\mathrm{ar}[c_n\sqrt{nb_n}\sum_{j=1}^{\lceil n/s_n \rceil} R_{n,j}(\boldsymbol{x},\boldsymbol{\lambda})] = 1$ by Lemma 1.7(ii) in the supplemental article [18] and (11). Thus, in view of (33) and Slutsky's theorem, for (32) it suffices to show

$$\frac{c_n\sqrt{nb_n}\sum_{j=1}^{\lceil n/s_n \rceil} R_{n,j}(\boldsymbol{x},\boldsymbol{\lambda})}{(\mathbb{V}\mathrm{ar}[c_n\sqrt{nb_n}\sum_{j=1}^{\lceil n/s_n \rceil} R_{n,j}(\boldsymbol{x},\boldsymbol{\lambda})])^{1/2}} = \frac{\sum_{j=1}^{\lceil n/s_n \rceil} R_{n,j}(\boldsymbol{x},\boldsymbol{\lambda})}{(\mathbb{V}\mathrm{ar}[\sum_{j=1}^{\lceil n/s_n \rceil} R_{n,j}(\boldsymbol{x},\boldsymbol{\lambda})])^{1/2}} \rightsquigarrow Z. \tag{36}$$

The big blocks, that is, the random variables in (34), are independent and centered. Thus, in view of Lyapunov's central limit theorem, for (36) it suffices to verify that the Lyapunov condition holds for $\sum_{j=1}^{\lceil n/s_n \rceil} R_{n,j}(\boldsymbol{x},\boldsymbol{\lambda})$. For $q \in (2,\infty)$ as in (A5) and sufficiently large $n$ we have

$$\begin{aligned}\frac{\sum_{j=1}^{\lceil n/s_n \rceil} \|R_{n,j}(\boldsymbol{x},\boldsymbol{\lambda})\|_q^q}{(\mathbb{V}\mathrm{ar}[\sum_j R_{n,j}(\boldsymbol{x},\boldsymbol{\lambda})])^{q/2}} &\le \frac{1}{c_{\boldsymbol{\lambda}}^{q/2}} \frac{\sum_{j=1}^{\lceil n/s_n \rceil} \|R_{n,j}(\boldsymbol{x},\boldsymbol{\lambda})\|_q^q}{(nb_n)^{q/2}} \\ &\le \frac{1}{c_{\boldsymbol{\lambda}}^{q/2}} C_{\boldsymbol{\lambda},q}^{q/2} \left(\frac{2nb_n - \sqrt{nb_n}}{\sqrt{nb_n} + \log^2(n)}\right) \frac{(\|\kappa\|_\infty^2 \sqrt{nb_n} + \mathcal{O}(1))^{q/2}}{(nb_n)^{q/2}},\end{aligned}$$

where we used Lemma 1.8 in the supplemental article [18] for the first step and Lemma 1.5(ii) in the supplemental article [18] for the second step. The latter bound converges to 0 by (A2) (and $q > 2$). This shows that the Lyapunov condition indeed holds.

## 4.2. Verification of condition (b) of Theorem 4.1

In this section, we will show that (under assumptions (A1)–(A3) and (A6)–(A8)) there exist for every $\epsilon > 0$ and $\delta > 0$ some $k \in \mathbb{N}$ and a partition $-\infty = x_0 < x_1 < \cdots < x_k < x_{k+1} = \infty$ such that

$$\limsup_{n\to\infty} \mathbb{P}\Big[\max_{i=0,\ldots,k} \sup_{x\in[x_i,x_{i+1})} \big|\widetilde{\mathcal{E}}_{p,n}(x)\phi_\lambda(x) - \widetilde{\mathcal{E}}_{p,n}(x_i)\phi_\lambda(x_i)\big| \ge 2\delta\Big] \le 2\epsilon.$$

For the proof, we use the same idea as in [27]. Since we can write

$$\widetilde{\mathcal{E}}_{p,n}(x) = c_n\sqrt{nb_n}\sum_{i=1}^{n}\kappa\left(\frac{i-i_{p,n}}{nb_n}\right)\big(\mathbb{1}_{[X_{n,i},\infty)}(x) - \mathbb{E}\big[\mathbb{1}_{[X_{n,i},\infty)}(x)\big]\big)$$

as $\widetilde{\mathcal{E}}_{p,n}(x) = H_{p,n}(x) + Q_{p,n}(x)$ with

$$\begin{aligned}
H_{p,n}(x) &:= c_n\sqrt{nb_n}\sum_{i=1}^{n}\kappa\left(\frac{i-i_{p,n}}{nb_n}\right)\big(\mathbb{1}_{[X_{n,i},\infty)}(x) - \mathbb{E}\big[\mathbb{1}_{[X_{n,i},\infty)}(x)|\boldsymbol{\epsilon}_{i-1}\big]\big) \\
&= c_n\sqrt{nb_n}\sum_{i=1}^{n}\kappa\left(\frac{i-i_{p,n}}{nb_n}\right)\big(\mathbb{1}_{[X_{n,i},\infty)}(x) - \mathfrak{F}_{n,i}(x,\boldsymbol{\epsilon}_{i-1})\big), \\
Q_{p,n}(x) &:= c_n\sqrt{nb_n}\sum_{i=1}^{n}\kappa\left(\frac{i-i_{p,n}}{nb_n}\right)\big(\mathbb{E}\big[\mathbb{1}_{[X_{n,i},\infty)}(x)|\boldsymbol{\epsilon}_{i-1}\big] - \mathbb{E}\big[\mathbb{1}_{[X_{n,i},\infty)}(x)\big]\big) \\
&= c_n\sqrt{nb_n}\sum_{i=1}^{n}\kappa\left(\frac{i-i_{p,n}}{nb_n}\right)\big(\mathfrak{F}_{n,i}(x,\boldsymbol{\epsilon}_{i-1}) - \mathbb{E}\big[\mathbb{1}_{[X_{n,i},\infty)}(x)\big]\big),
\end{aligned}$$

it suffices to prove that for every $\epsilon > 0$ and $\delta > 0$ there exist $k_1, k_2 \in \mathbb{N}$ and partitions $-\infty = x_0 < x_1 < \cdots < x_{k_1} < x_{k_1+1} = \infty$ and $-\infty = x_0 < x_1 < \cdots < x_{k_2} < x_{k_2+1} = \infty$ with

$$\limsup_{n\to\infty}\mathbb{P}\Big[\max_{i=0,\ldots,k_1}\sup_{x\in[x_i,x_{i+1})}\big|H_{p,n}(x)\phi_\lambda(x) - H_{p,n}(x_i)\phi_\lambda(x_i)\big| \geq \delta\Big] \leq \epsilon \tag{37}$$

and

$$\limsup_{n\to\infty}\mathbb{P}\Big[\max_{i=0,\ldots,k_2}\sup_{x\in[x_i,x_{i+1})}\big|Q_{p,n}(x)\phi_\lambda(x) - Q_{p,n}(x_i)\phi_\lambda(x_i)\big| \geq \delta\Big] \leq \epsilon. \tag{38}$$

Let $\epsilon > 0$ and $\delta > 0$ be arbitrary but fixed. By Lemma 1.12(ii) of the supplemental article [18] and Lemma 1.14(ii) of the supplemental article [18], we can find $w_{1,\epsilon} \geq 0$ and $w_{2,\epsilon} \geq 0$ such that

$$\sup_{n\in\mathbb{N}}\mathbb{P}\Big[\sup_{|x|\geq w_{1,\epsilon}}\big|H_{p,n}(x)\big|\phi_\lambda(x) \geq \delta\Big] \leq \epsilon, \qquad \sup_{n\in\mathbb{N}}\mathbb{P}\Big[\sup_{|x|\geq w_{2,\epsilon}}\big|Q_{p,n}(x)\big|\phi_\lambda(x) \geq \delta\Big] \leq \epsilon.$$

Then (37) and (38) follow directly from Lemma 1.12(iii) of the supplemental article [18] and Lemma 1.14(iii) of the supplemental article [18].

## Acknowledgements

The authors would like to thank two reviewers for their helpful suggestions and comments which led to an improvement of the paper.

# Supplementary Material

**Supplement to "Functional weak limit theorem for a local empirical process of non-stationary time series and its application"** (DOI: 10.3150/19-BEJ1174SUPP; .pdf). The supplement [18] contains the technical details of the proof of Theorem 2.4 as well as the proofs of all the other results.

# References


[1] Basrak, B., Davis, R.A. and Mikosch, T. (2002). Regular variation of GARCH processes. *Stochastic Process. Appl.* **99** 95–115. MR1894253 https://doi.org/10.1016/S0304-4149(01)00156-9

[2] Beutner, E. and Zähle, H. (2012). Deriving the asymptotic distribution of U- and V-statistics of dependent data using weighted empirical processes. *Bernoulli* **18** 803–822. MR2948902 https://doi.org/10.3150/11-BEJ358

[3] Beutner, E. and Zähle, H. (2014). Continuous mapping approach to the asymptotics of $U$- and $V$-statistics. *Bernoulli* **20** 846–877. MR3178520 https://doi.org/10.3150/13-BEJ508

[4] Beutner, E. and Zähle, H. (2016). Functional delta-method for the bootstrap of quasi-Hadamard differentiable functionals. *Electron. J. Stat.* **10** 1181–1222. MR3499525 https://doi.org/10.1214/16-EJS1140

[5] Beutner, E. and Zähle, H. (2018). Bootstrapping average value at risk of single and collective risks. *Risks* **6** 96.

[6] Billingsley, P. (1999). *Convergence of Probability Measures*. New York: Wiley. MR0233396

[7] Bougerol, P. and Picard, N. (1992). Stationarity of GARCH processes and of some nonnegative time series. *J. Econometrics* **52** 115–127. MR1165646 https://doi.org/10.1016/0304-4076(92)90067-2

[8] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37. MR1429916 https://doi.org/10.1214/aos/1034276620

[9] Dahlhaus, R. (2012). Locally stationary processes. In *Handbook of Statistics* **30** 351–413.

[10] Dahlhaus, R., Richter, S. and Wu, W.B. (2019). Towards a general theory for nonlinear locally stationary processes. *Bernoulli* **25** 1013–1044. MR3920364 https://doi.org/10.3150/17-bej1011

[11] Dahlhaus, R. and Subba Rao, S. (2006). Statistical inference for time-varying ARCH processes. *Ann. Statist.* **34** 1075–1114. MR2278352 https://doi.org/10.1214/009053606000000227

[12] Dette, H., Wu, W. and Zhou, Z. (2019). Change point analysis of correlation in non-stationary time series. *Statist. Sinica* **29** 611–643. MR3931381

[13] Fryzlewicz, P., Sapatinas, T. and Subba Rao, S. (2008). Normalized least-squares estimation in time-varying ARCH models. *Ann. Statist.* **36** 742–786. MR2396814 https://doi.org/10.1214/07-AOS510

[14] Fryzlewicz, P. and Subba Rao, S. (2011). Mixing properties of ARCH and time-varying ARCH processes. *Bernoulli* **17** 320–346. MR2797994 https://doi.org/10.3150/10-BEJ270

[15] Giraitis, L., Kokoszka, P. and Leipus, R. (2000). Stationary ARCH models: Dependence structure and central limit theorem. *Econometric Theory* **16** 3–22. MR1749017 https://doi.org/10.1017/S0266466600161018

[16] Hannan, E.J. (1973). Central limit theorems for time series regression. *Z. Wahrsch. Verw. Gebiete* **26** 157–170. MR0331683 https://doi.org/10.1007/BF00533484

[17] Hoffmann-Jørgensen, J. (1984). Stochastic processes in Polish spaces. Unpublished manuscript.

[18] Mayer, U., Zähle, H. and Zhou, Z. (2020). Supplement to "Functional weak limit theorem for a local empirical process of non-stationary time series and its application." https://doi.org/10.3150/19-BEJ1174SUPP.

[19] Nason, G.P., von Sachs, R. and Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 271–292. MR1749539 https://doi.org/10.1111/1467-9868.00231

[20] Pollard, D. (1984). *Convergence of Stochastic Processes. Springer Series in Statistics*. New York: Springer. MR0762984 https://doi.org/10.1007/978-1-4612-5254-2

[21] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. New York: Wiley. MR0838963

[22] Straumann, D. and Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *Ann. Statist.* **34** 2449–2495. MR2291507 https://doi.org/10.1214/009053606000000803

[23] van der Vaart, A.W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. MR1652247 https://doi.org/10.1017/CBO9780511802256

[24] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics. Springer Series in Statistics*. New York: Springer. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2

[25] Wu, W. and Zhou, Z. (2018). Gradient-based structural change detection for nonstationary time series M-estimation. *Ann. Statist.* **46** 1197–1224. MR3798001 https://doi.org/10.1214/17-AOS1582

[26] Wu, W.B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. MR2172215 https://doi.org/10.1073/pnas.0506715102

[27] Wu, W.B. (2008). Empirical processes of stationary sequences. *Statist. Sinica* **18** 313–333. MR2384990

[28] Zhou, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *J. Amer. Statist. Assoc.* **108** 726–740. MR3174655 https://doi.org/10.1080/01621459.2013.787184

[29] Zhou, Z. (2014). Inference of weighted $V$-statistics for nonstationary time series and its applications. *Ann. Statist.* **42** 87–114. MR3161462 https://doi.org/10.1214/13-AOS1184

[30] Zhou, Z. and Wu, W.B. (2009). Local linear quantile estimation for nonstationary time series. *Ann. Statist.* **37** 2696–2729. MR2541444 https://doi.org/10.1214/08-AOS636


*Received January 2019 and revised September 2019*