Econometric Theory, **33**, 2017, 915–954.
doi:10.1017/S0266466616000189

CHANGE POINT TESTS FOR THE TAIL INDEX OF β-MIXING RANDOM VARIABLES

YANNICK HOGA
University of Duisburg-Essen

The tail index as a measure of tail thickness provides information that is not captured by standard volatility measures. It may however change over time. Currently available procedures for detecting those changes for dependent data (e.g., Quintos *et al.*, 2001) are all based on comparing Hill (1975) estimates from different subsamples. We derive tests for a wide class of other tail index estimators. The limiting distribution of the test statistics is shown not to depend on the particular choice of the estimator, while the assumptions on the dependence structure allow for sufficient generality in applications. A simulation study investigates empirical sizes and powers of the tests in finite samples.

1. MOTIVATION

The tail index of a distribution is of great importance in statistics, in particular in extreme value theory. It determines the limit distribution of the (suitably normalized) sample maximum and minimum. Also, the tail index determines the existence of higher-order moments and consequently is used as a measure for the thickness of the tail of a distribution. As such it is of interest in fields as diverse as finance, hydrology, and internet-traffic engineering, where heavy tails are frequently encountered in real data. Moreover, it is important to know if the tail index of a time series has changed at some point during the observation period, since ignoring such a change can have negative consequences. For example, being unaware of a change to thicker tails of financial returns may lead to avoidable losses due to inadequate risk management, or, if tails vary from thicker to thinner, foregone profits because too much capital is put aside as a cushion against extreme losses. Indeed, there is empirical evidence that such changes do occur for many time series (Quintos, Fan and Phillips, 2001; Galbraith and Zernov, 2004; Werner and Upper, 2004).

The tail index is superior to other volatility measures, like the variance, when measuring volatility in at least the following two respects: It only captures the

The author would like to thank seminar participants at TU Dortmund for helpful discussion. Furthermore, the author is indebted to two anonymous referees and Christoph Hanck for their detailed comments, that greatly improved the quality of the paper. Full responsibility is taken for all remaining errors. Support of DFG (HA 6766/2-1) is gratefully acknowledged. Address correspondence to Yannick Hoga, Faculty of Economics and Business Administration, University of Duisburg-Essen, Universitätsstraße 12, D-45117 Essen, Germany, tel. +49 201 1834365, e-mail: yannick.hoga@vwl.uni-due.de.

behavior of the distribution in the tail, upon which interest in, e.g., financial risk management frequently centers. This is by definition not the case for the variance, suggesting that there is information in the tail index that is not present in other volatility measures, which Werner and Upper (2004) also found indications for in empirical work. Secondly, the variance as a measure of tail behavior is only available if second moments exist. Well-known empirical results show that this is not always the case (e.g., Resnick, 2007, Figures 4.12 and 4.15). The tail index, in contrast, does not require the existence of p-th moments for some $p > 0$.

Much research has been devoted to tail index estimation, see, e.g., Drees (1998a,b) for some general results in the independent and identically distributed (i.i.d.) case and Drees (2000) for the dependent case. But estimating the tail index from a sample $X_1, \dots, X_n$ assumes (often implicitly) homogeneity in the tail index, which might not be warranted. A test of this assumption is useful for at least two reasons: First, in the case of an undetected break in the tail index from, say, α_1 to $\alpha_2 > \alpha_1$, where tails get lighter after the break, most tail index estimators will consistently estimate $1/\alpha_1$ (see Theorem 3), suggesting a heavier tail for the postbreak period. In the above example of financial returns, this would lead to excessive conservatism. Second, the tail behavior depends in a very sensitive way on the tail index; e.g., for a Student's t_2-distribution (where the tail index equals the degrees of freedom) the 99.9%-quantile is 22.3 and for a t_1-distribution the same quantity is 318. Combined with the first reason this suggests that an undetected break in the sample would lead to a wrong tail index estimate and hence a very misleading picture of the tail behavior.

Tests for a change in the tail index at a known breakpoint have been available for some time (Koedijk, Schafgans and de Vries, 1990). So called recursive, rolling and sequential tests for an unknown break point in the tail index were first proposed by Quintos *et al.* (2001) for i.i.d. and GARCH(1,1) data. These tests were subsequently investigated by Kim and Lee (2011) to cover strictly stationary, β-mixing random variables. All these tests are based on comparing Hill (1975) estimates of different subsamples. However, many other, possibly better (in a mean-squared error sense), estimators exist; see, e.g., Figure 1 in de Haan and Peng (1998) for the i.i.d. case and the simulation results in Wagner and Marsh (2004) for ARCH-type data. Very recently, under 'heteroscedastic extremes', Einmahl, de Haan and Zhou (2016) allowed for other estimators to be used, although they only focused on the Hill (1975) estimator. Our first main contribution is to show that a vast range of tail index estimators is covered under their and (equivalently) our scheme, while, unlike Einmahl *et al.* (2016), allowing for dependent data, which is crucial for most real-world applications. Previously, consistency of change point tests for the tail index has only been proved under independence (Quintos *et al.*, 2001; Kim and Lee, 2009) or not at all (Einmahl *et al.*, 2016; Kim and Lee, 2011). The second main contribution is to demonstrate consistency under dependence. Further, we show that if there is a single tail index break in the sample, tail index estimators will still converge weakly, though with a different limit distribution. This result might be of independent interest.

A simulation study investigates whether gains in power can be achieved in a change point context by using other tail index estimators than Hill's covered by our framework. For instance for ARCH data 'there is a tendency of the Hill estimator to overestimate small tail indices and to underestimate large tail indices' (Wagner and Marsh, 2004, p. 3), which may lead to poor power properties of the change point tests based on the Hill estimator, as confirmed by our simulations. Another problem with the application of change point tests for the tail index to ARCH-models is that to the best of our knowledge, all currently available tests (Quintos *et al.*, 2001) are derived using standard-normally distributed innovations, which in empirical work is often not credible (e.g., Aguilar and Hill, 2015, Figure 2). This issue is also addressed in this paper by allowing for, e.g., t-distributed innovations, which are also used in the simulations. Furthermore, our simulations reveal that the problem of nonmonotonic power emerges for some tail index estimators.

The generality of our results rests on the insight that a wide range of tail index estimators can be written as functionals of the (slightly adapted) sequential tail empirical process. Allowing for dependence requires consistent estimates of the asymptotic variance of the tail index estimator. With the exception of Drees (2003), such estimators have so far only been considered for very specific dependence structures and tail index estimators, e.g., in Quintos *et al.* (2001) (for the Hill estimator and GARCH(1,1) data) and Chan, Li, Peng and Zhang (2013) (for a moment-type estimator and AR(1) data with ARCH innovations). We propose a consistent variance estimator for the tail index estimators we consider under weak conditions on the dependence of the time series observations, which might be of independent interest.

The main results are stated in Section 2. Simulation evidence is presented in Section 3 and the proofs are relegated to Section 4.

2. MAIN RESULTS

This section is organized as follows: Subsection 2.1 introduces basic notation and the main assumptions that will be used throughout. It also gives examples of linear and nonlinear models for which these assumptions have been verified. Subsection 2.2 introduces some of the estimators that can be used under our scheme and states convergence results under the null. Results under a one-break alternative are stated in Subsection 2.3.

2.1. Preliminaries

Consider stationary random variables (r.v.s) $\{X_i\}_{i\in\mathbb{N}}$ defined on some probability space $(\Omega, \mathcal{A}, P)$. Let F be the distribution function (d.f.) of X_1, where $1-F$ is assumed to be regularly varying with parameter $-\alpha < 0$ (written $1-F \in RV_{-\alpha}$), i.e.,

$$\frac{1-F(ty)}{1-F(t)} \underset{(t\to\infty)}{\longrightarrow} y^{-\alpha} \qquad \forall\, y > 0, \tag{1}$$

where α is called the tail index of X_1. If we define

$$U(t) := F^{\leftarrow}\left(1-\frac{1}{t}\right), \qquad t > 1,$$

as the $(1-1/t)$-quantile, where $^{\leftarrow}$ denotes the left-continuous inverse, then (1) is equivalent to

$$\frac{U(ty)}{U(t)} \underset{(t\to\infty)}{\longrightarrow} y^{\gamma} \qquad \forall y > 0, \tag{2}$$

with $\gamma = 1/\alpha > 0$ the extreme value index (cf. de Haan and Ferreira, 2006). We will use both notations, γ and α.

Remark 1. If $\{X_i\}$ are i.i.d., then by the well-known Fisher–Tippett theorem the extreme value index γ determines (apart from a location and scale parameter) the possible limiting d.f.s of

$$\frac{\max(X_1,\ldots,X_n)-b_n}{a_n} \qquad (a_n > 0, b_n \in \mathbb{R}), \tag{3}$$

namely $G_\gamma(x) = \exp\left(-(1+\gamma x)^{-1/\gamma}\right)$, $1+\gamma x > 0$.

In the sequel, $k = k_n \in \mathbb{N}$ with $k \le n-1$ will be an intermediate sequence, i.e.,

$$k \underset{(n\to\infty)}{\longrightarrow} \infty \qquad \text{and} \qquad \frac{k}{n} \underset{(n\to\infty)}{\longrightarrow} 0,$$

controlling the number of "extremely large" observations used in the estimation of the tail index. For $t-s \ge 1/n$ and $y \in [0,1]$ set

$$X_k(s,t,y) := \left(\lfloor k(t-s)y\rfloor + 1\right)\text{-th largest value of } X_{\lfloor ns\rfloor+1},\ldots,X_{\lfloor nt\rfloor}. \tag{4}$$

For clarity of exposition we will sometimes write $X_{1:n} \le \cdots \le X_{n:n}$ for the order statistics of $X_1,\ldots,X_n$. The dependence concept used here is that of β-mixing. Recall that a sequence of random variables $\{X_i\}_{i\in\mathbb{N}}$ is β-mixing iff

$$\beta(l) := \sup_{m\in\mathbb{N}} \mathbb{E}\left[\sup_{A\in\mathcal{F}_{m+l+1}^{\infty}} \left|P\left(A\middle|\mathcal{F}_1^m\right) - P(A)\right|\right] \underset{(l\to\infty)}{\longrightarrow} 0,$$

where $\mathcal{F}_m^\infty := \sigma(X_m, X_{m+1},\ldots)$ and $\mathcal{F}_l^m := \sigma(X_l,\ldots,X_m)$ are the σ-algebras generated by the respective r.v.s.

If it is in doubt whether all $X_1,\ldots,X_n$ have the same extreme value index $\gamma_1 = \cdots = \gamma_n$, it is important for reasons detailed in the motivation to test the following hypothesis:

$$\begin{aligned} \mathcal{H}_0 &: \quad \gamma_1 = \cdots = \gamma_n \qquad \text{versus} \\ \mathcal{H}_1 &: \quad \text{Not } \mathcal{H}_0. \end{aligned} \tag{5}$$

We now state our main assumptions that will be maintained throughout.

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

(C1) $\{X_i\}_{i\in\mathbb{N}}$ is a strictly stationary β-mixing process with continuous marginals and mixing coefficients $\beta(\cdot)$, such that

$$\lim_{n\to\infty}\frac{n}{r_n}\beta(l_n)+\frac{r_n}{\sqrt{k}}\log^2(k)=0 \tag{6}$$

for sequences $\{l_n\}_{n\in\mathbb{N}}\subset\mathbb{N}$, $\{r_n\}_{n\in\mathbb{N}}\subset\mathbb{N}$ tending to infinity with $l_n=o(r_n)$, $r_n=o(n)$.

(C2) There exists a function $r(\cdot,\cdot)$, s.t. for all $x, y\in[0,y_0+\delta]$ $(\delta>0)$

$$\lim_{n\to\infty}\frac{n}{r_n k}Cov\left(\sum_{i=1}^{r_n}I_{\{X_i>U(\frac{n}{kx})\}},\sum_{j=1}^{r_n}I_{\left\{X_j>U\left(\frac{n}{ky}\right)\right\}}\right)=r(x,y). \tag{7}$$

(C3) For some constant $C>0$

$$\frac{n}{r_n k}\mathbb{E}\left[\sum_{i=1}^{r_n}I_{\left\{U\left(\frac{n}{ky}\right)<X_i\le U\left(\frac{n}{kx}\right)\right\}}\right]^4\le C(y-x)\qquad\forall\, 0\le x<y\le y_0+\delta,\ n\in\mathbb{N}.$$

(C4) There exist $\rho<0$ and a function $A(\cdot)$, eventually positive or negative, $\lim_{t\to\infty}A(t)=0$, s.t.

$$\lim_{t\to\infty}\frac{\frac{U(ty)}{U(t)}-y^\gamma}{A(t)}=y^\gamma\frac{y^\rho-1}{\rho}\qquad\forall\, y>0,$$

where $\sqrt{k}A(n/k)\to 0$ as $n\to\infty$.

Remark 2. (a) Conditions **(C1)**, **(C2)**, and **(C3)** are discussed in some detail in Drees (2000, 2003) and Rootzén (2009). They are almost identical to conditions $(\widetilde{C1})$, $(\widetilde{C2})$, and $(\widetilde{C3^*})$ in Drees (2000). **(C4)** is a standard second-order condition (used in, e.g., Einmahl *et al.*, 2016) that controls the speed of convergence in (2). It is slightly stronger than Drees' (2000) corresponding condition (3.5), which can be seen from de Haan and Ferreira (2006, Theorem 2.3.9).

(b) If **(C2)** holds for k_n and $k_{n,\lambda_j}\sim\lambda_j k_n$ $(\lambda_j\in(0,1))$, $j=1,2$, then (cf. Drees, 2003, p. 629)

$$r(tx,ty)=tr(x,y), \tag{8}$$

which will simplify the expressions for the asymptotic variances of the estimators we consider.

Conditions **(C1)**–**(C4)** relax some assumptions of previous tests. For instance, our scheme covers a wide range of short-memory processes (see Rootzén, 2009, Section 4, for an overview) in contrast to Einmahl *et al.* (2016), where only independent r.v.s are considered. Allowing for dependence is essential as, under

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

dependence, the limit distribution of the test statistic considered in Einmahl *et al.* (2016, Corollary 2) will be scaled by some (dependence-structure dependent) factor. Note that the presence of 'heteroscedastic extremes' introduced in Einmahl *et al.* (2016) does not influence the limit behavior of their test statistic. Next, heavy-tailed innovations for ARCH(1)-processes are allowed under our conditions (and not under those of Quintos *et al.*, 2001); see also Remark 8.

The next two examples, taken from Drees (2000, Section 4) and Drees (2003, Subsections 3.1 & 3.2), give specific models where **(C1)–(C3)** have been verified. While the first-order condition in (2) is satisfied for both examples, the second-order condition **(C4)** has not yet been verified.

Example 1 (Linear model)
Consider stationary $\{X_i\}_{i\in\mathbb{N}}$ with representation

$$X_i = \sum_{j=0}^{\infty} \Psi_j Z_{i-j}, \quad i \in \mathbb{N},$$

where $\{Z_i\}_{i\in\mathbb{Z}}$ are i.i.d., $\Psi_0 = 1$ without loss of generality (w.l.o.g.) and $\left|\Psi_j\right| = \mathcal{O}\left(\tau^j\right)$, $j \to \infty$, for some $\tau \in (0,1)$. If for F_Z the d.f. of Z_1 we have $1 - F_Z \in RV_{-\alpha}$ $(\alpha > 0)$ and some further conditions hold, then **(C1)–(C3)** hold for sequences $k = k_n$ satisfying

$$\log^2(n)\log^4(\log n) = o(k) \quad \text{and} \quad k = o\left(n/\log(n)\right). \tag{9}$$

In that case, the X_i also have tail index α.

Example 2 (Nonlinear model)
Consider a squared ARCH(1)-process $\left\{X_i^2\right\}_{i\in\mathbb{N}}$

$$X_i^2 = \left(\alpha_0 + \alpha_1 X_{i-1}^2\right) Z_i^2, \quad i \in \mathbb{N},$$

where $\alpha_0, \alpha_1 > 0$ and $\{Z_i\}_{i\in\mathbb{N}} \overset{\text{i.i.d.}}{\sim} (0,1)$. If Z_1 satisfies the following moment conditions

$$\begin{aligned} &\exists\, \kappa, \xi > 0 \;:\; \mathbb{E}\log\left(\alpha_1 Z_1^2\right) < 0, \quad \mathbb{E}\left(\alpha_1 Z_1^2\right)^{\kappa} = 1, \\ &\mathbb{E}\left(\alpha_1 Z_1^2\right)^{\kappa+\xi} < \infty, \quad \mathbb{E}\left(\alpha_0 Z_1^2\right)^{\kappa+\xi} < \infty, \end{aligned} \tag{10}$$

then conditions **(C1)–(C3)** were shown to hold for sequences $k = k_n$ satisfying

$$\log^2(n)\log^4(\log n) = o(k) \quad \text{and} \quad k = o\left(n^{2\rho/(2\rho+1)}\right) \quad \text{for some } \rho > 0.$$

The tail index $\alpha = \kappa > 0$ of (the strictly stationary) X_i^2 is determined by the moment condition $\mathbb{E}\left(\alpha_1 Z_i^2\right)^{\alpha} = 1$. Hence, α can be changed either by varying α_1 or the distribution of Z_i. Note that light-tailed Z_i, e.g., $Z_i \sim \mathcal{N}(0,1)$, lead to heavy tails in X_i, which is not true for Example 1.

2.2. Results Under the Null

The generality of our approach rests on a (weighted) weak convergence result for

$$F_n(s,t,y) := \frac{1}{\lfloor k(t-s)\rfloor} \sum_{i=\lfloor ns\rfloor+1}^{\lfloor nt\rfloor} I_{\{X_i > y X_k(s,t,1)\}}.$$

A wide range of tail index estimators can be written as functionals of $F_n(s,t,y)$. For example, the Hill estimator $\widehat{\gamma}_H(0,1)$ based on the full sample $X_1,\dots,X_n$ can be written as

$$\widehat{\gamma}_H(0,1) := \frac{1}{k}\sum_{i=0}^{k}\log\left(\frac{X_{n-i:n}}{X_{n-k:n}}\right) = \int_1^\infty F_n(0,1,y)\frac{\mathrm{d}y}{y}; \tag{11}$$

see also Examples 3–5. In the first step, we will establish weighted convergence of the *sequential tail empirical process*

$$\sqrt{k}\left\{\frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor} I_{\{X_i > yU(n/k)\}} - y^{-1/\gamma}t\right\}.$$

Then a variant of this result for $F_n(s,t,y)$ will be used to investigate weak convergence of (suitably normalized) generic tail index estimators

$\widehat{\gamma}(s,t)$ that can be written as functionals of $F_n(s,t,y)$,

i.e., estimators based on subsamples $X_{\lfloor ns\rfloor+1},\dots,X_{\lfloor nt\rfloor}$.

Under **(C1)**–**(C4)** it will be possible to derive the limiting distributions of the test statistics (where $\widehat{\sigma}^2_{\widehat{\gamma},\gamma} \in \left\{\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\mathrm{nor}}, \widehat{\sigma}^2_{\widehat{\gamma},\gamma,\mathrm{rev}}\right\}$ is defined in Theorem 2)

$$\begin{aligned}
Q_{\mathrm{rec}} &:= \frac{1}{\widehat{\sigma}^2_{\widehat{\gamma},\gamma}} \sup_{t\in[t_0,1-t_0]} \left\{t\sqrt{k}\left[\widehat{\gamma}(0,t)-\widehat{\gamma}(0,1)\right]\right\}^2;\\
Q^{\leftarrow}_{\mathrm{rec}} &:= \frac{1}{\widehat{\sigma}^2_{\widehat{\gamma},\gamma}} \sup_{t\in[t_0,1-t_0]} \left\{(1-t)\sqrt{k}\left[\widehat{\gamma}(t,1)-\widehat{\gamma}(0,1)\right]\right\}^2;\\
Q_{\mathrm{seq}} &:= \frac{1}{\widehat{\sigma}^2_{\widehat{\gamma},\gamma}} \sup_{t\in[t_0,1-t_0]} \left\{t(1-t)\sqrt{k}\left[\widehat{\gamma}(0,t)-\widehat{\gamma}(t,1)\right]\right\}^2;\\
Q_{\mathrm{rol}} &:= \frac{1}{\widehat{\sigma}^2_{\widehat{\gamma},\gamma}} \sup_{t\in[t_0,1-t_0]} \left\{t_0\sqrt{k}\left[\widehat{\gamma}(t,t+t_0)-\widehat{\gamma}(0,1)\right]\right\}^2;
\end{aligned} \tag{12}$$

for the testing problem (5), namely (see Corollary 2)

$$\begin{aligned}
Q^{(\leftarrow)}_{\mathrm{rec}} &\xrightarrow[(n\to\infty)]{\mathcal{D}} \sup_{t\in[t_0,1-t_0]} \{W(t)-tW(1)\}^2,\\
Q_{\mathrm{seq}} &\xrightarrow[(n\to\infty)]{\mathcal{D}} \sup_{t\in[t_0,1-t_0]} \{W(t)-tW(1)\}^2,\\
Q_{\mathrm{rol}} &\xrightarrow[(n\to\infty)]{\mathcal{D}} \sup_{t\in[t_0,1-t_0]} \{[W(t+t_0)-W(t)]-t_0W(1)\}^2,
\end{aligned} \tag{13}$$

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

where $W(\cdot)$ denotes a standard Brownian motion. The general form of the test statistics in (12) is taken from Quintos *et al.* (2001). We have modified Q_{seq} slightly by including the factor $(1-t)$. Without it Q_{seq}, by construction, would be more likely to detect a change in the tail index at the end of the observation period, where t is large, than towards the beginning, which may not be desirable.

We assume throughout that $t_0 \in (0, 1/2)$. In our framework t_0 and $(1-t_0)$ denote the time before and after which the change is not allowed to occur. Since all tests allow to take t_0 arbitrarily close to zero, this does not impose a serious restriction. Further, by choosing t_0 closer to $1/2$ one can incorporate prior knowledge of the change point location in the tests, which, as unreported simulations for $Q_{\text{rec}}^{(\leftarrow)}$ show, leads to higher power. It is easy to verify that asymmetric intervals à la $[t_0, t_1]$, $t_1 \in (t_0, 1)$ over which the supremum is taken in (12) and (13) are also possible, lending more flexibility to the incorporation of prior beliefs.

The weighted convergence result stated in the next theorem is fundamental to our approach (see also Remark 3 (b)). For this, define non-negative, continuous weight functions $q(\cdot)$, similarly as in Drees (2000, Eq. (1.3)), as functions satisfying

$$\inf_{y>\vartheta} q(y) > 0 \quad \forall \vartheta > 0 \qquad \text{and} \qquad y^{\nu} |\log y|^{\mu} = \mathcal{O}(q(y)), \quad y \downarrow 0, \tag{14}$$

for some $\nu \in [0, 1/2)$, $\mu \in \mathbb{R}$ or $\nu = 1/2$, $\mu > 1/2$. Then we may prove

THEOREM 1. *Suppose* **(C1)–(C4)** *hold and* $q(\cdot)$ *satisfies (14). Then for some* $\widetilde{\delta} > 0$*, under a Skorohod construction,*

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\geq y_0^{-\gamma}-\widetilde{\delta}}} \frac{1}{q\left(y^{-1/\gamma}\right)} \left| \sqrt{k} \begin{pmatrix} \frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor} I_{\{X_i>yU(n/k)\}} - y^{-1/\gamma} t \\ \frac{1}{k}\sum_{i=\lfloor nt\rfloor+1}^{n} I_{\{X_i>yU(n/k)\}} - y^{-1/\gamma}(1-t) \\ \frac{1}{k}\sum_{i=\lfloor nt\rfloor+1}^{\lfloor n(t+t_0)\rfloor} I_{\{X_i>yU(n/k)\}} - y^{-1/\gamma} t_0 \end{pmatrix} - \begin{pmatrix} W(t, y^{-1/\gamma}) \\ W(1, y^{-1/\gamma}) - W(t, y^{-1/\gamma}) \\ W(t+t_0, y^{-1/\gamma}) - W(t, y^{-1/\gamma}) \end{pmatrix} \right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0, \tag{15}$$

where $\{W(t,y)\}$ *is a continuous zero-mean Gaussian process with covariance function*

$$Cov\,(W(t_1, y_1), W(t_2, y_2)) = \min(t_1, t_2)\, r(y_1, y_2).$$

A slightly modified version of (15), where $U(n/k)$ is replaced by an appropriate empirical counterpart, will be more convenient for our purposes. This results in a change of the limiting processes.

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

COROLLARY 1. *Suppose* **(C1)–(C4)** *hold for some* $y_0 \geq 1$ *and* $q(\cdot)$ *satisfies (14). Then, under a Skorohod construction,*

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\geq y_0^{-\gamma}}} \frac{1}{q\left(y^{-1/\gamma}\right)} \left| \sqrt{k} \begin{pmatrix} t\left(F_n(0,t,y)-y^{-1/\gamma}\right) \\ (1-t)\left(F_n(t,1,y)-y^{-1/\gamma}\right) \\ t_0\left(F_n(t,t+t_0,y)-y^{-1/\gamma}\right) \end{pmatrix} - \begin{pmatrix} W(t,y^{-1/\gamma})-y^{-1/\gamma}W(t,1) \\ W(1,y^{-1/\gamma})-W(t,y^{-1/\gamma})-y^{-1/\gamma}\left[W(1,1)-W(t,1)\right] \\ W(t+t_0,y^{-1/\gamma})-W(t,y^{-1/\gamma})-y^{-1/\gamma}\left[W(t+t_0,1)-W(t,1)\right] \end{pmatrix} \right| \xrightarrow[(n\to\infty)]{\text{a.s.}} 0, \quad \mathbf{(16)}$$

where $\{W(t,y)\}$ *is as in Theorem 1.*

Since the above asymptotic results become stronger the smaller $q(\cdot)$ is, it would in fact suffice to only consider the case $\nu = 1/2$ in (14).

In the following three examples, we will demonstrate how the convergence result of Corollary 1 can be used to establish joint convergence of $\sqrt{k}(\widehat{\gamma}(0,t) - \gamma, \widehat{\gamma}(t,1)-\gamma, \widehat{\gamma}(t,t+t_0)-\gamma)^T$.

Example 3 (WLS estimator)
Consider the class of weighted least squares (WLS) estimators of the tail index

$$\widehat{\gamma}_{WLS}(0,1) := \sum_{j=1}^{k} \int_{(j-1)/k}^{j/k} J(s)\mathrm{d}s \log(X_{n+1-j:n})$$

and with a finite-sample correction

$$\widehat{\gamma}_{\widetilde{WLS}}(0,1) := \frac{\sum_{j=1}^{k} \int_{(j-1)/k}^{j/k} J(s)\mathrm{d}s \log(X_{n+1-j:n})}{\sum_{j=1}^{k} \int_{(j-1)/k}^{j/k} J(s)\mathrm{d}s \log(k/j)}$$

discussed in Csörgő and Viharos (1998), where the weighting function $J(\cdot)$ satisfies

(W1) $\int_0^1 J(s)\mathrm{d}s = 0$,
(W2) $J(\cdot)$ is nonincreasing and continuous on $[0,1]$,
(W3) $-\int_0^1 \log(s)J(s)\mathrm{d}s = 1$.

PROPOSITION 1. *Suppose* **(C1)–(C4)** *hold for* $y_0 = 1$. *Then for* $J(\cdot)$ *satisfying* **(W1)–(W3)**

$$\sqrt{k}\begin{pmatrix} \widehat{\gamma}_{WLS}(0,t)-\gamma \\ \widehat{\gamma}_{WLS}(t,1)-\gamma \\ \widehat{\gamma}_{WLS}(t,t+t_0)-\gamma \end{pmatrix} \xrightarrow[(n\to\infty)]{\mathcal{D}} \sigma_{\widehat{\gamma}_{WLS},\gamma} \begin{pmatrix} W(t)/t \\ (W(1)-W(t))/(1-t) \\ (W(t+t_0)-W(t))/t_0 \end{pmatrix}$$

in $D^3[t_0, 1-t_0]$, **(17)**

where $W(\cdot)$ *is a standard Brownian motion and*

$$\sigma^2_{\widehat{\gamma}_{WLS},\gamma} = \gamma^2 \int_0^1 \int_0^1 \frac{r(x,y)}{xy} J(x)J(y)\mathrm{d}x\mathrm{d}y. \quad \mathbf{(18)}$$

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

Specifically, Csörgő and Viharos (1998) consider weight functions fulfilling **(W1)–(W3)**

$$J_\theta(s) := \frac{\theta+1}{\theta} - \frac{(\theta+1)^2}{\theta} s^\theta, \qquad s \in [0,1], \quad \theta > 0,$$

yielding estimators denoted by $\widehat{\gamma}_{CV_\theta}$, that possess certain optimality properties in a mean-squared error sense (cf. Csörgő and Viharos, 1998, Weight Theorem (ii)). Then, under (8), (18) simplifies to

$$\gamma^2 \int_0^1 \int_0^1 \frac{r(z,1)}{z} J_\theta(zy) J_\theta(y) \mathrm{d}y \mathrm{d}z = 2\frac{(\theta+1)^2}{2\theta+1}\gamma^2 \int_0^1 \frac{r(x,1)}{x} x^\theta \mathrm{d}x. \tag{19}$$

Remark 3. (a) Inclusion of the finite-sample correction does not change the convergence result in (17) (cf. Csörgő and Viharos, 1998, p. 18).

(b) The need for a weighted convergence result as in Corollary 1 can be seen most clearly from (62) in the proof of Proposition 1, where without weighting (i.e., $q \equiv 1$) the integral in that expression would not generally be finite.

Example 4 (Hill estimator)
As in the proof of Proposition 1, we will only derive weak convergence of $\widehat{\gamma}_H(0,t)$ from the first component of (16). Joint convergence as in (17) can again be obtained from (16). Check that similarly as in (11) we have $\widehat{\gamma}_H(0,t) = \int_1^\infty F_n(0,t,y)\frac{\mathrm{d}y}{y}$, such that by Corollary 1

$$\begin{aligned}\sqrt{k}\left(\widehat{\gamma}_H(0,t)-\gamma\right) &= \sqrt{k}\int_1^\infty \left(F_n(0,t,y) - y^{-1/\gamma}\right)\frac{\mathrm{d}y}{y} \\ &\underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} \frac{1}{t}\int_1^\infty \left[W(t,y^{-1/\gamma}) - y^{-1/\gamma}W(t,1)\right]\frac{\mathrm{d}y}{y} \\ &= \frac{\gamma}{t}\int_0^1 \left[W(t,u) - uW(t,1)\right]\frac{\mathrm{d}u}{u}.\end{aligned}$$

Calculate covariances to obtain that the right-hand side is distributed as $\sigma_{\widehat{\gamma}_H,\gamma} W(t)/t$, where $W(\cdot)$ denotes a standard Brownian motion and

$$\sigma^2_{\widehat{\gamma}_H,\gamma} = \gamma^2 \int_0^1 \int_0^1 \left\{\frac{r(x,y)}{xy} - \frac{r(x,1)}{x} - \frac{r(1,y)}{y} + r(x,y)\right\} \mathrm{d}x\mathrm{d}y \overset{(8)}{=} \gamma^2 r(1,1). \tag{20}$$

Example 5 (Moments ratio estimator)
We consider convergence of the moments ratio estimator based on the subsample $X_1,\ldots,X_{\lfloor nt\rfloor}$. Define, for $j = 1,2$,

$$M_j(t) := \frac{1}{\lfloor kt\rfloor}\sum_{i=1}^{\lfloor kt\rfloor}\left(\log(X_{\lfloor nt\rfloor-i+1:\lfloor nt\rfloor}) - \log(X_{\lfloor nt\rfloor-\lfloor kt\rfloor:\lfloor nt\rfloor})\right)^j.$$

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

Then

$$\widehat{\gamma}_{MR}(0,t) := \frac{1}{2}\frac{M_2(t)}{M_1(t)} = \frac{1}{2}\frac{M_2(t)}{\widehat{\gamma}_H(0,t)}$$

is the so called moments ratio (MR) estimator of the tail index introduced by Daníelsson, Jansen and de Vries (1996). One may verify that (cf. also the proof of Proposition 1)

$$M_1(t) = \int_1^\infty F_n(0,t,y)\frac{\mathrm{d}y}{y} \qquad \text{and} \qquad M_2(t) = \int_1^\infty F_n(0,t,y)2\log(y)\frac{\mathrm{d}y}{y}. \quad \textbf{(21)}$$

Then, under **(C1)–(C4)**, for $y_0 = 1$

$$\begin{aligned}
\widehat{\gamma}_H(0,t)\cdot\sqrt{k}\,(\widehat{\gamma}_{MR}(0,t)-\gamma) &= \sqrt{k}\left(\frac{1}{2}\int_1^\infty F_n(0,t,y)2\log(y)\frac{\mathrm{d}y}{y} - \gamma\int_1^\infty F_n(0,t,y)\frac{\mathrm{d}y}{y}\right)\\
&= \sqrt{k}\int_1^\infty F_n(0,t,y)\left[\log(y)-\gamma\right]\frac{\mathrm{d}y}{y}\\
&= \int_1^\infty \sqrt{k}\left[F_n(0,t,y)-y^{-1/\gamma}\right]\left[\log(y)-\gamma\right]\frac{\mathrm{d}y}{y}\\
&\quad + \sqrt{k}\int_1^\infty \left[\log(y)-\gamma\right]y^{-(1/\gamma+1)}\mathrm{d}y\\
\xrightarrow[(n\to\infty)]{\mathcal{D}} \frac{1}{t}\int_1^\infty &\left[W(t,y^{-1/\gamma}) - y^{-1/\gamma}W(t,1)\right]\left[\log(y)-\gamma\right]\frac{\mathrm{d}y}{y} + 0\\
&= -\frac{\gamma^2}{t}\int_0^1 \left[W(t,u)/u - W(t,1)\right]\left[\log(u)+1\right]\mathrm{d}u\\
&= -\frac{\gamma^2}{t}\int_0^1 W(t,u)/u\left[\log(u)+1\right]\mathrm{d}u.
\end{aligned}$$

Use $\widehat{\gamma}_H(0,t) = \gamma + o_P(1)$ uniformly in $t \in [t_0, 1-t_0]$ (from Example 4) and calculate covariances to obtain

$$t\sqrt{k}\,(\widehat{\gamma}_{MR}(0,t)-\gamma) \xrightarrow[(n\to\infty)]{\mathcal{D}} \sigma_{\widehat{\gamma}_{MR},\gamma}W(t) \quad \text{in } D[t_0, 1-t_0],$$

where $W(\cdot)$ is a standard Brownian motion and

$$\begin{aligned}
\sigma^2_{\widehat{\gamma}_{MR},\gamma} &= \gamma^2\int_0^1\int_0^1 \frac{r(x,y)}{xy}\left[\log(x)+1\right]\left[\log(y)+1\right]\mathrm{d}x\,\mathrm{d}y\\
&\overset{(8)}{=} 2\gamma^2\int_0^1 \frac{r(x,1)}{x}\mathrm{d}x. \qquad \textbf{(22)}
\end{aligned}$$

Again, joint convergence as in (17) can be obtained by virtue of the joint convergence in (16).

Clearly, we have to consistently estimate $\sigma^2_{\widehat{\gamma},\gamma}$, the asymptotic variance of $\widehat{\gamma}(0,1)$. To that end we propose the following method. The basic idea is as follows: With only one sample $X_1,\ldots,X_n$ we can only estimate γ once with $\widehat{\gamma}(0,1)$

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

and infer nothing on the variance of the estimate. To get more estimates we calculate $\widehat{\gamma}(0,t)$ for $t \in (0,1]$. Calculating suitably normalized sample variances of all these estimates yields a consistent estimate of the variance of $\widehat{\gamma}(0,1)$, as shown in

THEOREM 2. *Let $W(\cdot)$ denote a standard Brownian motion.*

(a) *If for any $t_0 > 0$*

$$t\sqrt{k}\left(\widehat{\gamma}(0,t)-\gamma\right) \xrightarrow[(n\to\infty)]{\mathcal{D}} \sigma_{\widehat{\gamma},\gamma} W(t) \quad \text{in } D[t_0,1], \tag{23}$$

then for all sequences $t_n \downarrow 0$ tending to 0 not too fast,

$$\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\text{nor}} := \frac{1}{\log(n/(\lfloor nt_n\rfloor+1))}\frac{k}{n}\sum_{i=\lfloor nt_n\rfloor+1}^{n}\left[\widehat{\gamma}\left(0,\frac{i}{n}\right)-\widehat{\gamma}(0,1)\right]^2 \xrightarrow[(n\to\infty)]{P} \sigma^2_{\widehat{\gamma},\gamma}.$$

(b) *If for any $t_0 > 0$*

$$(1-t)\sqrt{k}\left(\widehat{\gamma}(t,1)-\gamma\right) \xrightarrow[(n\to\infty)]{\mathcal{D}} \sigma_{\widehat{\gamma},\gamma}\left(W(1)-W(t)\right) \quad \text{in } D[0,1-t_0], \tag{24}$$

then for all sequences $t_n \downarrow 0$ tending to 0 not too fast,

$$\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\text{rev}} := \frac{1}{\log(n/(\lfloor nt_n\rfloor+1))}\frac{k}{n}\sum_{i=0}^{n-(\lfloor nt_n\rfloor+2)}\left[\widehat{\gamma}\left(\frac{i}{n},1\right)-\widehat{\gamma}(0,1)\right]^2 \xrightarrow[(n\to\infty)]{P} \sigma^2_{\widehat{\gamma},\gamma}.$$

Remark 4. (a) See the proof of Theorem 2 for why t_n must not approach 0 too fast.

(b) If **(C1)–(C4)** hold for $y_0 = 1$, then the convergences in (23) and (24) hold for any $t_0 > 0$ for the estimators given in Examples 3–5.

(c) In simulations we choose t_n as small as possible such that $\widehat{\gamma}\left(0,\frac{\lfloor nt_n\rfloor+1}{n}\right)$ (or $\widehat{\gamma}\left(\frac{n-(\lfloor nt_n\rfloor+2)}{n},1\right)$) is still well-defined for all choices of k. In fact, unreported simulations show that the estimates are quite robust with respect to the choice of t_n.

(d) Note that in the case of (e.g.) (a) in the above theorem

$$\frac{1}{\sigma^2_{\widehat{\gamma},\gamma}}\frac{k}{n}\sum_{i=\lfloor nt_n\rfloor+1}^{n}\left[\widehat{\gamma}\left(0,\frac{i}{n}\right)-\widehat{\gamma}(0,1)\right]^2 \overset{\mathcal{D}}{\approx} \frac{1}{n}\sum_{i=\lfloor nt_n\rfloor+1}^{n}\left[\frac{W(i/n)}{(i/n)}-W(1)\right]^2,$$

where the expectation of the right-hand side is (approximately)

$$\left[\left(\sum_{i=\lfloor nt_n\rfloor+1}^{n}\frac{1}{i}\right)-(1-t_n)\right] \sim \log(n/(\lfloor nt_n\rfloor+1)). \tag{25}$$

Hence, we use the left-hand side of (25) as a finite-sample correction for $\log(n/(\lfloor nt_n\rfloor+1))$ in the estimator from Theorem 2 (a). A similar argument reveals that the finite-sample correction is also sensible for $\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\text{rev}}$.

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

(e) The result of the above theorem may be of independent interest and could be adapted to a wide range of estimators investigated in a change-point context, where limit results as (23) with $\sqrt{k}$ replaced by some other sequence tending to infinity and $\widehat{\gamma}(0,t)$ replaced by some other estimator (based on observations $X_1,\ldots,X_{\lfloor nt \rfloor}$) of an unknown parameter.

The joint convergences (as in (17)) established in the above examples and the continuous mapping theorem now allow us to easily derive the null distributions of our test statistics from (12).

COROLLARY 2. *For the estimators from Examples 3–5 and* $\widehat{\sigma}^2_{\widehat{\gamma},\gamma} \in \{\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\mathrm{nor}}, \widehat{\sigma}^2_{\widehat{\gamma},\gamma,\mathrm{rev}}\}$ *the convergences in (13) hold under the conditions of Corollary 1 with* $y_0 = 1$.

2.3. Results Under the Alternative

We will explore the behavior of our tests under two specific one-break alternatives:

$$\mathcal{H}_1^{\lessgtr}: \quad \gamma_1 = \cdots = \gamma_{\lfloor nt^* \rfloor} \lessgtr \gamma_{\lfloor nt^* \rfloor + 1} = \cdots = \gamma_n, \tag{26}$$

for some $t^* \in (t_0, 1-t_0)$ and $\gamma_i > 0$ for all $i = 1,\ldots,n$. To avoid repetition in the following theorem we state conditions that must hold under $\mathcal{H}_1^{<}$ and that differ from the ones under $\mathcal{H}_1^{>}$ in parentheses (e.g., (28)).

THEOREM 3. *Under* $\mathcal{H}_1^{>}$ $\left(\mathcal{H}_1^{<}\right)$ *let the triangular array* $\left\{X_{i,n}\right\}_{i=1,\ldots,n;n\in\mathbb{N}}$ *be given by*

$$X_{i,n} := \begin{cases} Y_i^{\mathrm{pre}}, & i \in I_{\mathrm{pre}} := \{1,\ldots,\lfloor nt^* \rfloor\}, \\ Y_i^{\mathrm{post}}, & i \in I_{\mathrm{post}} := \{\lfloor nt^* \rfloor + 1,\ldots,n\}, \end{cases}$$

where $\left\{Y_i^{\mathrm{pre}}\right\}_{i\in\mathbb{N}}$ *and* $\left\{Y_i^{\mathrm{post}}\right\}_{i\in\mathbb{N}}$ *both satisfy conditions* **(C1)–(C4)** *with*

$$k_{\mathrm{pre}},\ \gamma_{\mathrm{pre}},\ U_{\mathrm{pre}}(\cdot),\ r_{\mathrm{pre}}(\cdot,\cdot),\ y_{0,\mathrm{pre}} = \frac{1-t_0}{t_0} \left(y_{0,\mathrm{pre}} = \left(\frac{1-t_0}{t_0}\right)^{\gamma_{\mathrm{post}}/\gamma_{\mathrm{pre}}} \right) \quad \text{and}$$

$$k_{\mathrm{post}},\ \gamma_{\mathrm{post}},\ U_{\mathrm{post}}(\cdot),\ r_{\mathrm{post}}(\cdot,\cdot),\ y_{0,\mathrm{post}} = \left(\frac{1-t_0}{t_0}\right)^{\gamma_{\mathrm{pre}}/\gamma_{\mathrm{post}}} \left(y_{0,\mathrm{post}} = \frac{1-t_0}{t_0} \right),$$

respectively. Suppose further that $q(\cdot)$ *satisfies (14), and*

$$k_{\mathrm{post}} = \mathcal{O}\left(k_{\mathrm{pre}}\right),\ s.t.\ k_{\mathrm{post}} \left(\frac{U_{\mathrm{post}}\left(\frac{n}{k_{\mathrm{post}}}\right)}{U_{\mathrm{pre}}\left(\frac{n}{k_{\mathrm{pre}}}\right)} \right)^{1/\gamma_{\mathrm{post}}} \underset{(n\to\infty)}{\longrightarrow} 0 \tag{27}$$

$$\left(k_{\mathrm{pre}} = \mathcal{O}\left(k_{\mathrm{post}}\right),\ s.t.\ k_{\mathrm{pre}} \left(\frac{U_{\mathrm{pre}}\left(\frac{n}{k_{\mathrm{pre}}}\right)}{U_{\mathrm{post}}\left(\frac{n}{k_{\mathrm{post}}}\right)} \right)^{1/\gamma_{\mathrm{pre}}} \underset{(n\to\infty)}{\longrightarrow} 0 \right). \tag{28}$$

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

Then, for the estimators from Examples 3–5, under $\mathcal{H}_1^{>}$

$$\sqrt{k_{\text{pre}}}\left(\widehat{\gamma}(0,t)-\gamma_{\text{pre}}\right)\underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} B_{\text{pre}}(t)/t \qquad \text{in } D[t_0,1], \tag{29}$$

where

$$B_{\text{pre}}(t) := \begin{cases} \gamma_{\text{pre}}\int_0^1 W_{\text{pre}}\left(t_{\min}, u\frac{t}{t_{\min}}\right) J(u)\frac{\mathrm{d}u}{u}, & \text{for } \widehat{\gamma}_{WLS}, \\ \gamma_{\text{pre}}\int_0^1 \left[W_{\text{pre}}\left(t_{\min}, u\frac{t}{t_{\min}}\right) - u W_{\text{pre}}\left(t_{\min}, \frac{t}{t_{\min}}\right)\right]\frac{\mathrm{d}u}{u}, & \text{for } \widehat{\gamma}_{H}, \\ -\gamma_{\text{pre}}\int_0^1 W_{\text{pre}}\left(t_{\min}, u\frac{t}{t_{\min}}\right)\left[\log(u)+1\right]\frac{\mathrm{d}u}{u}, & \text{for } \widehat{\gamma}_{MR}, \end{cases}$$

with $t_{\min} := \min(t,t^*)$ *and* $W_{\text{pre}}(\cdot,\cdot)$ *as in Theorem 1 with* $r(\cdot,\cdot)$ *replaced by* $r_{\text{pre}}(\cdot,\cdot)$*, and under* $\mathcal{H}_1^{<}$

$$\sqrt{k_{\text{post}}}\left(\widehat{\gamma}(t,1)-\gamma_{\text{post}}\right)\underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} B_{\text{post}}(t)/(1-t) \qquad \text{in } D[t_0,1],$$

where

$$B_{\text{post}}(t) := \begin{cases} \gamma_{\text{post}}\int_0^1 \widetilde{W}_{\text{post}}(t,u) J(u)\frac{\mathrm{d}u}{u}, & \text{for } \widehat{\gamma}_{WLS}, \\ \gamma_{\text{post}}\int_0^1 \left[\widetilde{W}_{\text{post}}(t,u) - u\widetilde{W}_{\text{post}}(t,1)\right]\frac{\mathrm{d}u}{u}, & \text{for } \widehat{\gamma}_{H}, \\ -\gamma_{\text{post}}\int_0^1 \widetilde{W}_{\text{post}}(t,u)\left[\log(u)+1\right]\frac{\mathrm{d}u}{u}, & \text{for } \widehat{\gamma}_{MR}, \end{cases}$$

with $\widetilde{W}_{\text{post}}(t,u) := W_{\text{post}}\left(1, u\frac{1-t}{1-t_{\max}}\right) - W_{\text{post}}\left(t_{\max}, u\frac{1-t}{1-t_{\max}}\right)$*,* $t_{\max} := \max(t,t^*)$*, and* $W_{\text{post}}(\cdot,\cdot)$ *as in Theorem 1 with* $r(\cdot,\cdot)$ *replaced by* $r_{\text{post}}(\cdot,\cdot)$*.*

Remark 5. (a) Quintos *et al.* (2001, Theorem 3) show that the Hill estimator applied to an i.i.d. sample with one break in the tail index as in (26) converges in probability to $\max(\gamma_{\text{pre}},\gamma_{\text{post}})$. Theorem 3 obviously substantially generalizes this result.

(b) Theorem 3 does not make any assumption on the dependence between Y_i^{pre} and Y_i^{post}.

(c) For the time series models from Examples 1 and 2 conditions **(C1)–(C3)** were satisfied for sequences k with lower and upper bound

$$\log^2(n)\log^4(\log n) = o(k) \qquad \text{and} \qquad k = o(n^{\xi}), \quad \xi \in (0,1)$$

(recall (2) and (9)). Hence for a sample with a break in the tail index it does not seem to be overly restrictive to assume $k = k_{\text{pre}} = k_{\text{post}}$, which is what we do in the following.

(d) Under $\mathcal{H}_1^{>}$ ($\mathcal{H}_1^{<}$) condition (27) ((28)) ensures that the part of the sequential tail empirical process appertaining to the post- (pre-) break period is asymptotically negligible (see the proof of Theorem 3).

How stringent is (27)? (A similar argument also holds for (28).) For any $\varepsilon > 0$ and n sufficiently large, note that

$$\frac{U_{\text{post}}\left(\frac{n}{k_{\text{post}}}\right)}{U_{\text{pre}}\left(\frac{n}{k_{\text{pre}}}\right)} = \frac{\left(\frac{n}{k_{\text{post}}}\right)^{\gamma_{\text{post}}+\varepsilon}\left(\frac{n}{k_{\text{post}}}\right)^{-\varepsilon} L_{\text{post}}\left(\frac{n}{k_{\text{post}}}\right)}{\left(\frac{n}{k_{\text{pre}}}\right)^{\gamma_{\text{pre}}-\varepsilon}\left(\frac{n}{k_{\text{pre}}}\right)^{\varepsilon} L_{\text{pre}}\left(\frac{n}{k_{\text{pre}}}\right)}$$
$$< \left(\frac{n}{k_{\text{post}}}\right)^{\gamma_{\text{post}}+\varepsilon}\left(\frac{n}{k_{\text{pre}}}\right)^{-\gamma_{\text{pre}}+\varepsilon},$$

where $L_{\text{pre(post)}}(x) = x^{-\gamma_{\text{pre(post)}}} U_{\text{pre(post)}}(x) \in RV_0$ and Bingham, Goldie and Teugels' (1987) Proposition 1.3.6 (v) was used for the inequality. If $k = k_{\text{pre}} = k_{\text{post}} = n^{\xi}$ for some $\xi \in (0,1)$, then (27) is satisfied for $\xi < 1 - \gamma_{\text{post}}/\gamma_{\text{pre}}$. That is, for small breaks, i.e., $\gamma_{\text{pre}} - \gamma_{\text{post}}$ close to 0, k must be rather small relative to n. Similar arguments apply to (28).

(In-)Consistency results can now easily be proved:

COROLLARY 3. *Under the conditions of Theorem 3 with $k = k_{\text{pre}} = k_{\text{post}}$, we have for the estimators from Examples 3–5 and for all sequences $t_n \downarrow 0$ tending to zero not too fast:*

(a) $\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\text{nor}} \underset{(n\to\infty)}{\overset{P}{\longrightarrow}} \sigma^2_{\widehat{\gamma},\gamma_{\text{pre}}}$ *under* $\mathcal{H}_1^{>}$,

(b) $\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\text{rev}} \underset{(n\to\infty)}{\overset{P}{\longrightarrow}} \sigma^2_{\widehat{\gamma},\gamma_{\text{post}}}$ *under* $\mathcal{H}_1^{<}$.

Using $\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\text{nor}}$ *or* $\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\text{rev}}$ *according as* $\mathcal{H}_1^{>}$ *or* $\mathcal{H}_1^{<}$*, we further have:*

(c) *The tests based on Q_{seq} and Q_{rol} are consistent under $\mathcal{H}_1^{\lessgtr}$, where for Q_{rol} the additional assumption $t^* \in (t_0, 1-2t_0)$ $(t^* \in (2t_0, 1-t_0))$ has to hold under $\mathcal{H}_1^{>}$ $\left(\mathcal{H}_1^{<}\right)$.*

(d) *The test based on Q_{rec} $\left(Q_{\text{rec}}^{\leftarrow}\right)$ is consistent under $\mathcal{H}_1^{<}$ $\left(\mathcal{H}_1^{>}\right)$, whereas under $\mathcal{H}_1^{>}$ $\left(\mathcal{H}_1^{<}\right)$ we have*

$$Q_{\text{rec}}^{(\leftarrow)} \underset{(n\to\infty)}{=} \mathcal{O}_P(1).$$

Remark 6. For an estimator of the change point t^* that is consistent under weak conditions we refer to Kim and Lee (2009, Theorem 3).

3. SIMULATIONS

This section investigates the finite-sample properties of our tests for specific models from Examples 1 and 2. We do so only for Q_{rec} and $Q_{\text{rec}}^{\leftarrow}$, because we want to explore the differences between the tail index estimators and not between the different test statistics in (12). The latter has already been done in the literature (Quintos *et al.*, 2001; Kim and Lee, 2011). We just remark that the qualitative

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

conclusions from the other studies hold here as well. The $Q_{\text{rec}}^{(\leftarrow)}$- and Q_{seq}-test have slightly better size than the Q_{rol}-test, presumably because the estimates $\widehat{\gamma}(t,t+t_0)$ are always based on relatively small rolling windows. Under the alternative, when the tests based on $Q_{\text{rec}}^{(\leftarrow)}$ are consistent, they have the highest power of all alternatives, which is why we focus on these tests here.

We use $t_0 = 0.2$, sample sizes of $n = 500$ and $n = 2000$, and $t_n = 50/n$. With this choice of t_n all estimates $\widehat{\gamma}\big(0, \frac{\lfloor nt_n \rfloor + 1}{n}\big)$ and $\widehat{\gamma}\big(\frac{n-(\lfloor nt_n \rfloor+2)}{n}, 1\big)$ remained well-defined and t_n is reasonably small, as required by Theorem 2. Table 1 shows critical values, obtained by 100,000 realizations of the approximations to the limit distribution $\sup_{t\in[t_0,1-t_0]}\{W(t) - tW(1)\}^2$ from (13) (where the Brownian motion itself was generated from 100,000 independent normally distributed r.v.s). We use the estimators $\widehat{\gamma}_H$, $\widehat{\gamma}_{MR}$, and $\widehat{\gamma}_{CV_\theta}$ for $\theta = 1$. The corresponding tests will be denoted H, MR and CV_1.

In order for tests to be consistent, we estimate $\sigma^2_{\widehat{\gamma},\gamma}$ using $\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\text{nor}}$ for $Q_{\text{rec}}^{\leftarrow}$ (under $\mathcal{H}_0$ and $\mathcal{H}_1^{>}$) and $\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\text{rev}}$ for Q_{rec} (under $\mathcal{H}_1^{<}$). We modify $\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\text{nor}}$ and $\widehat{\sigma}^2_{\widehat{\gamma},\gamma,\text{rev}}$ by requiring that they be at least as large as the (consistent) variance estimate in the independent case, i.e., $\widehat{\gamma}_H^2(0,1)$ for $\widehat{\gamma}_H(0,1)$, $2\widehat{\gamma}_{MR}^2(0,1)$ for $\widehat{\gamma}_{MR}(0,1)$ and $\frac{2\theta+2}{2\theta+1}\widehat{\gamma}^2_{CV_\theta}(0,1)$ for $\widehat{\gamma}_{CV_\theta}(0,1)$. This is warranted by the observation that our models from Examples 1 and 2 satisfy the conditions of Drees (2003, Prop. 2.1), whence $r(x,y) \geq \min(x,y)$. (Note that $r(x,y) = \min(x,y)$ under independence and **(C4)**.) Hence, the asymptotic variances given in (19), (20), and (22) cannot be lower under dependence than under independence, such that our modified estimators are still consistent for $\sigma^2_{\widehat{\gamma},\gamma}$ under independence *and* dependence.

Concretely, we simulate from the two data generating processes (DGPs)

$$X_i = 0.5 \cdot X_{i-1} + Z_i, \tag{AR}$$

$$X_i^2 = \left(\alpha_0 + \alpha_1 \cdot X_{i-1}^2\right) Z_i^2. \tag{ARCH}$$

Remark 7. (a) In the AR(1) case, it is also possible to use the change point test proposed in Kim and Lee (2012), which is based on AR(p)-residuals. However, in the context of extreme quantile estimation Drees (2008, Section 2) cautions against using residual-based tail index estimators for AR(p)-models, since they can be very sensitive to ever so slight misspecifications. We therefore advocate using nonmodel-based estimators in a change point context as well.

(b) For the (G)ARCH-model with innovations that have finite $(4+\delta)$-th moments, there exist more precise estimators of the tail index (e.g., in Berkes, Horváth and Kokoszka, 2003, and Chan *et al.*, 2013) in the sense of

TABLE 1. Quantiles of $\sup_{t\in[t_0,1-t_0]}\{W(t) - tW(1)\}^2$ for $t_0 = 0.2$

α_q	0.50	0.60	0.70	0.80	0.90	0.95	0.99
α_q-quantile	0.650	0.767	0.918	1.128	1.478	1.821	2.653

being $\sqrt{n}$-consistent instead of the slower $\sqrt{k}$. Using these estimators in a change point test could potentially result in more powerful tests. However, as in part (a) of this remark, slight departures from the model could then lead to severe mis-estimation of tail parameters.

For the ARCH-model one often uses $Z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ or the normalized $\sqrt{\frac{\nu-2}{\nu}}t_\nu$ $(\nu > 2)$ with unit variance. For standardized t_ν-innovations the first moment condition in (10) implies that $\alpha_1 \in (0, \exp\{\psi(\nu/2) - \psi(1/2)\}/\{\nu-2\})$, where $\psi(z) = \Gamma'(z)/\Gamma(z)$ denotes the digamma function, since

$$0 > \mathbb{E}\log(\alpha_1 Z_i^2) = \log(\alpha_1) + \log\left(\frac{\nu-2}{\nu}\right) + \log(\nu) + \psi(1/2) - \psi(\nu/2), \tag{30}$$

where $\log(t_\nu^2)/2 \sim \log(F_{1,\nu})/2$ follows Fisher's z-distribution with mean $\left[\log(\nu) + \psi(1/2) - \psi(\nu/2)\right]/2$.

Remark 8. The only existing change point test known to the author for ARCH data in Quintos *et al.* (2001) relies on standard-normally distributed innovations, whereas our tests permit, e.g., $\sqrt{\frac{\nu-2}{\nu}}t_\nu$-distributed innovations for $\nu > 2$. If only standard-normally distributed innovations are permitted, tail index break tests degenerate to tests for parameter constancy for α_1 (recall from Example 2 that the tail index of an ARCH(1) can only be changed by varying α_1 or the distribution of Z_1). We venture to claim that tests for parameter constancy for GARCH(p,q) models as proposed in Berkes, Horváth and Kokoszka (2004) then perform better, as more observations are effectively used in the estimation of α_1.

For the results under the null in Table 2, we choose $Z_i \sim \sqrt{\frac{\nu-2}{\nu}}t_\nu$ with $\nu = 5/2$ for the ARCH(1)-model along with $\alpha_0 = 0.01$ and $\alpha_1 = 0.95$, i.e., tail index equal to 1.01 determined from $\mathbb{E}(\alpha_1 Z_i^2)^\alpha = 1$. Note that by choosing $\nu = 5/2$ the innovations barely have existing second moments, which is required in ARCH-type models. Note further that by (30) the choice of $\nu = 5/2$ necessitates $\alpha_1 \in (0, 11.34...)$, which is of course satisfied for our particular choice of α_1. For the AR(1)-model we also use $Z_i \sim t_\nu$ with $\nu = 5/2$, i.e., tail index equal to 2.5. Hence, the process in (AR) does have finite second moments, while that in (ARCH) does not.

For both models the results show that, by and large, sizes only slightly decrease in k. This is encouraging since the choice of k can be a very sensitive issue in tail index estimation, see, e.g., Section 4.4.2 in Resnick (2007) for a Hill horror plot and some references on the topic. As a referee pointed out, this may be explained by the canceling of bias terms (that arise in tail index estimation for large k) in (12). For $n = 500$ most tests are conservative for both models. The convergence to the nominal level for $n = 2000$ is satisfactory for the H and the CV_1 test for a wide range of k's, while the MR test is still somewhat conservative.

TABLE 2. Empirical sizes of $Q_{\text{rec}}^{\leftarrow}$-tests in % for n realizations of (AR) and (ARCH)

Test	Size	DGP	$n = 500$					$n = 2000$				
			k/n									
			0.04	0.08	0.12	0.16	0.2	0.04	0.08	0.12	0.16	0.2
H	0.05	(AR)	6.3	4.5	4.8	4.2	4.1	5.7	5.6	4.8	4.5	4.1
	0.01		1.8	1.3	1.0	1.1	0.8	1.4	1.2	0.9	0.8	0.7
MR	0.05		2.1	2.9	2.5	2.0	1.5	3.9	3.4	3.2	2.7	2.3
	0.01		0.5	1.1	0.7	0.4	0.3	1.5	0.8	0.8	0.6	0.4
CV_1	0.05		2.4	2.0	1.9	1.9	1.7	3.8	3.8	3.8	3.4	3.2
	0.01		0.4	0.4	0.3	0.2	0.3	0.7	0.7	0.6	0.5	0.3
H	0.05	(ARCH)	7.1	5.7	5.5	4.4	4.4	7.8	6.9	6.1	5.4	5.2
	0.01		2.4	1.6	1.0	1.0	0.9	2.2	1.8	1.1	1.0	1.1
MR	0.05		2.0	3.0	2.8	2.4	1.9	5.0	4.6	3.8	3.2	3.0
	0.01		0.5	1.1	1.0	0.7	0.6	1.7	1.6	1.1	0.7	0.7
CV_1	0.05		4.0	2.8	2.1	1.8	1.5	4.5	4.5	4.4	4.6	4.4
	0.01		1.2	0.8	0.6	0.4	0.2	1.0	0.9	0.7	0.7	0.7

To examine power we simulate according to model (AR) again, only that now $Z_i := Z_{i,n}$ with

$$Z_{i,n} \sim \begin{cases} t_2, & i \le \lfloor nt^* \rfloor, \\ t_\nu, & i > \lfloor nt^* \rfloor, \end{cases} \tag{31}$$

where we choose $t^* = 0.25, 0.5$, and $\nu \in [2,4]$. Hence, there is a break in the tail index of $\{X_i\}$ from 2 to ν, i.e., lighter postbreak tails. To investigate power for a nonlinear model as well, we simulate from

$$X_{i,n}^2 := \begin{cases} (0.01 + 0.45 \cdot X_{i-1,n}^2) Z_i^2, & i \le \lfloor nt^* \rfloor, \\ (0.01 + \alpha_1 \cdot X_{i-1,n}^2) Z_i^2, & i > \lfloor nt^* \rfloor, \end{cases} \tag{32}$$

where again $t^* = 0.25, 0.5, \alpha_1 \in [0.45, 0.95]$, and $Z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. This ARCH(1)-model with a break in α_1 has tails varying from thinner ($\alpha = 2.67$ corresponding to $\alpha_1 = 0.45$) to thicker ($\alpha = 1.07$ corresponding to $\alpha_1 = 0.95$). Throughout we take $n = 2000$.

The simulation results for different values of ν in (31) and α_1 in (32) are displayed in Figure 1. We choose $k = 0.2 \cdot n$ for both models, because for these values of k the differences in size (i.e., $\nu = 2$ in the AR(1)-case and $\alpha_1 = 0.45$ in the ARCH(1)-case) are smallest, such that direct power comparisons are more meaningful. Figure 1 (a) displays the results for the AR(1)-model with innovations as in (31) using the $Q_{\text{rec}}^{\leftarrow}$-test, which is consistent. In the bottom part, where results

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

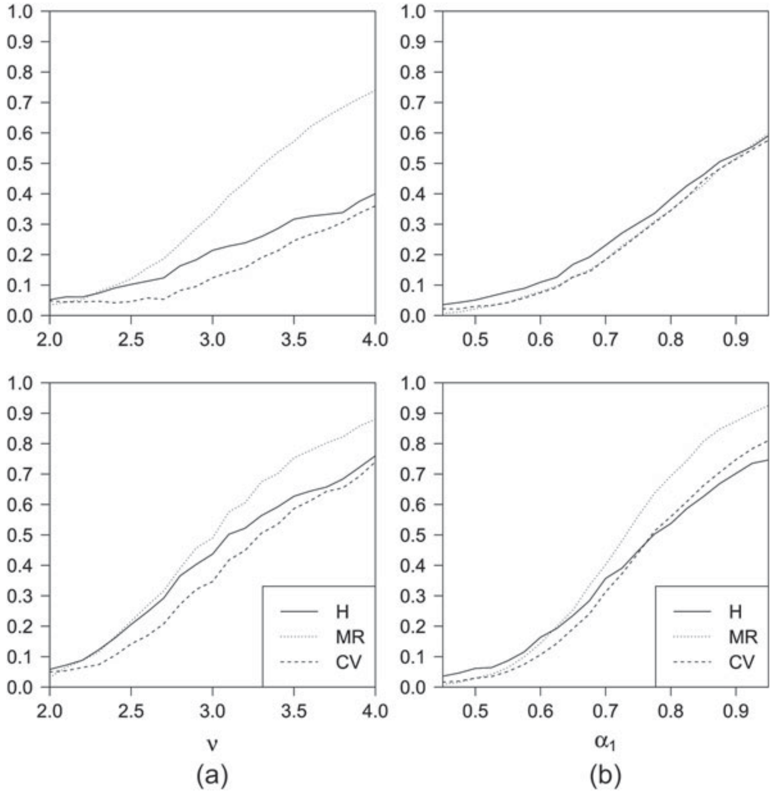

FIGURE 1. (a) Fraction of rejections for $Q_{\text{rec}}^{\leftarrow}$-test using (AR) with innovations (31) and (b) for Q_{rec}-test using (32) with $t^* = 0.25$ (top), $t^* = 0.5$ (bottom).

are shown for $t^* = 0.5$, tests have roughly comparable properties. It is notable that, despite being slightly conservative, the MR test offers higher power than the other two tests, the more so the larger ν. The difference for $\nu = 4$ is between 13 and 15 percentage points. This is even more apparent when $t^* = 0.25$, where the MR test performs only marginally worse than before, but the CV_1 test and the H test in particular lose sizable amounts of power. Here the biggest difference is as high as 38 percentage points.

Panel (b) shows results for the ARCH(1) with a break in α_1 using Q_{rec}. The top part for $t^* = 0.25$ shows that the MR test has comparable power as the other two tests despite being more conservative. When $t^* = 0.5$ (bottom part) it seems to gain more power than the other two offerings, so that power for the MR test is 18 percentage points higher than that for the H test for $\alpha_1 = 0.95$. In light of the simulation, evidence in Wagner and Marsh (2004) already mentioned in the motivation, this was to be expected.

In the upper left plot in Figure 1, the CV_1 test seems to suffer slightly from non-monotonic power—a well-known phenomenon in the literature on change point detection (Vogelsang, 1997, 1999)—i.e., it does not show increasing power in distance from the null in some ranges. In the context of mean-shift detection, Vogelsang (1999) identifies long-run variance estimates as one major source of nonmonotonicity. We also find indications for this here. For $\nu = 2$ the average estimate of $\widehat{\sigma}^2_{\widehat{\gamma}_{CV_1},\gamma}$ ($\widehat{\sigma}^2_{\widehat{\gamma}_{MR},\gamma}$) over all 5000 replications is 0.74 (0.82), while for $\nu = 2.7$ it is 0.95 (0.70). Hence, while for the MR test the denominator of $Q^{\leftarrow}_{\text{rec}}$ even decreased, it increased markedly for the CV_1 test. Apparently, it increased roughly proportionately to the numerator of $Q^{\leftarrow}_{\text{rec}}$ for CV_1, which could be a reason for the flat profile for $\nu \in [2, 2.7]$.

All in all, our simulations reveal reasonable size of our tests for quite a wide range of k's with some conservative tendencies. Under the alternative we find the MR test to clearly deliver the best results, with the H and CV_1 test performing similarly.

4. PROOFS

In the following K, K_1, K_2, and $\widetilde{\delta}$ denote large and small positive constants that may change from line to line. $D[t_0, 1]$ denotes the space of càdlàg functions equipped with the Skorohod metric and the Borel σ-field $\mathcal{D}[t_0, 1]$. For brevity put $D^2 := D([t_0, 1] \times [0, y_0 + \delta])$ ($\delta \geq 0$) for the space of two-parameter càdlàg functions on $[t_0, 1] \times [0, y_0 + \delta]$, which is equipped with the multiparameter extension of the Skorohod metric (cf. Bickel and Wichura, 1971, p. 1662) and the Borel σ-field $\mathcal{D}^2$. As usual, define $\|\cdot\|_2$ to be the Euclidean metric, $|\cdot|$ to be cardinality when applied to a set and $\sum_i^j := 0$ for $i > j$.

To derive weighted weak convergence results involving the weight function $q(\cdot)$, we may assume w.l.o.g., as in the proof of Drees (2000, Theorem 2.2), that for some $\vartheta > 0$ sufficiently small

$$q(y) = y^{\nu} |\log y|^{\mu}, \; y \in (0, \vartheta], \quad \text{s.t. } q \text{ is increasing and } q/Id \text{ decreasing on } (0, \vartheta], \tag{33}$$

where $Id(\cdot)$ denotes the identity function.

In the first step, we will consider uniformly distributed r.v.s $U_i \sim \mathcal{U}[0, 1]$ and then suitably apply this result to X_i satisfying **(C4)**. To this end we need the following analogs of conditions **(C1)–(C3)**:

(U1) $\{U_i\}_{i \in \mathbb{N}}$ is a strictly stationary β-mixing process with mixing coefficients $\beta(\cdot)$, such that

$$\lim_{n \to \infty} \frac{n}{r_n} \beta(l_n) + \frac{r_n}{\sqrt{k}} \log^2(k) = 0$$

for sequences $\{l_n\}_{n \in \mathbb{N}} \subset \mathbb{N}$, $\{r_n\}_{n \in \mathbb{N}} \subset \mathbb{N}$ tending to infinity with $l_n = o(r_n)$, $r_n = o(n)$.

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

(U2) There exists a function $r(x, y)$, such that for all $x, y \in [0, y_0 + \delta]$

$$\lim_{n\to\infty} \frac{n}{r_n k} Cov\left(\sum_{i=1}^{r_n} I_{\left\{U_i > 1 - \frac{k}{n}x\right\}}, \sum_{j=1}^{r_n} I_{\left\{U_j > 1 - \frac{k}{n}y\right\}}\right) = r(x, y).$$

(U3) For some constant $C > 0$,

$$\frac{n}{r_n k}\mathbb{E}\left[\sum_{i=1}^{r_n} I_{\left\{1-\frac{k}{n}y < U_i \le 1 - \frac{k}{n}x\right\}}\right]^4 \le C(y - x) \qquad \forall 0 \le x < y \le y_0 + \delta,\ n \in \mathbb{N}.$$

We start with a nonweighted weak convergence result for the sequential tail empirical process for the $\{U_i\}$:

THEOREM 4. *Suppose* **(U1)–(U3)** *hold. Then*

$$\sqrt{k}\left\{\frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor}\left[I_{\left\{U_i > 1-\frac{k}{n}y\right\}} - \frac{k}{n}y\right]\right\} \xrightarrow[(n\to\infty)]{\mathcal{D}} W(t, y) \quad \text{in } D^2, \tag{34}$$

where $\{W(t, y)\}$ *is a continuous zero-mean Gaussian process with covariance function*

$$Cov(W(t_1, y_1), W(t_2, y_2)) = \min(t_1, t_2)\, r(y_1, y_2). \tag{35}$$

Proof. For notational convenience and w.l.o.g. $\delta = 0$. We use a classical 'big block - small block' approach, where the small blocks are asymptotically negligible. For $t \in [0, 1]$ define

$$m_n(t) := \left\lfloor \frac{\lfloor nt\rfloor}{r_n + l_n}\right\rfloor$$

and for $j = 1, \ldots, m_n(1)$ define I_j (the big blocks) and J_j (the small blocks) to be consecutive blocks of integers of length $\left|I_j\right| = r_n$ and $\left|J_j\right| = l_n$, i.e.,

$$I_1 = \{1, \ldots, r_n\},\ J_1 = \{r_n + 1, \ldots, r_n + l_n\}, \text{ etc.}$$

Choose the length of $I_{m_n(t)+1}$ such that the integers $\{1, \ldots, \lfloor nt\rfloor\}$ are covered. Now decompose

$$\sqrt{k}\left\{\frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor}\left[I_{\left\{U_i > 1-\frac{k}{n}y\right\}} - \frac{k}{n}y\right]\right\} = \sum_{j=1}^{m_n(t)} Y_j^I(y) + \sum_{j=1}^{m_n(t)} Y_j^J(y) + R_n(t, y),$$

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

where

$$Y_j^I(y) = \frac{1}{\sqrt{k}} \sum_{i \in I_j} \left[I_{\left\{U_i > 1 - \frac{k}{n} y\right\}} - \frac{k}{n} y \right],$$

$$Y_j^J(y) = \frac{1}{\sqrt{k}} \sum_{i \in J_j} \left[I_{\left\{U_i > 1 - \frac{k}{n} y\right\}} - \frac{k}{n} y \right],$$

$$R_n(t, y) = \frac{1}{\sqrt{k}} \sum_{i \in I_{m_n(t)+1}} \left[I_{\left\{U_i > 1 - \frac{k}{n} y\right\}} - \frac{k}{n} y \right].$$

We will consider these terms separately. First, noting that the cardinality $\left|I_{m_n(t)+1}\right| \leq r_n + l_n - 1$,

$$0 \leq \sup_{(t,y) \in [t_0, 1] \times [0, y_0]} |R_n(t, y)| \leq 2 \frac{r_n + l_n - 1}{\sqrt{k}} \underset{(n \to \infty)}{\overset{\textbf{(U1)}}{\longrightarrow}} 0. \tag{36}$$

Second, set $L_n(t, y) = \sum_{j=1}^{m_n(t)} Y_j^I(y)$ and define the measurable mapping

$$M_n : \left(D^{m_n}[0, y_0], \mathcal{D}^{m_n}[0, y_0]\right) \to \left(D^2, \mathcal{D}^2\right)$$

$$M_n\left(t, x_1(\cdot), \ldots, x_{m_n}(\cdot)\right) = \sum_{i=1}^{m_n(t)} x_i(y), \qquad (t, y) \in [t_0, 1] \times [0, y_0].$$

Then for $H \in \mathcal{D}^2$ using Lemma 2 of Eberlein (1984) and **(U1)**

$$\begin{aligned} P(L_n \in H) &= P\left(\left(Y_1^I(\cdot), \ldots, Y_{m_n}^I(\cdot)\right) \in M_n^{-1}(H)\right) \\ &= \widetilde{P}\left(\left(\widetilde{Y}_1^I(\cdot), \ldots, \widetilde{Y}_{m_n}^I(\cdot)\right) \in M_n^{-1}(H)\right) + \mathcal{O}\left(m_n \beta(l_n)\right) \\ &\overset{\textbf{(U1)}}{=} \widetilde{P}(\widetilde{L}_n \in H) + o(1), \end{aligned} \tag{37}$$

where $\widetilde{L}_n(t, y) = \sum_{j=1}^{m_n(t)} \widetilde{Y}_j^I(y)$ and the $\widetilde{Y}_j^I(\cdot)$ are i.i.d. copies of $Y_1^I(\cdot)$ defined on the product probability space $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P}) := \otimes_{i=1}^{\infty} \left(D[0, y_0], \mathcal{D}[0, y_0], P_{Y_i}\right)$ via

$$\tilde{Y}_i : (\tilde{\Omega}, \tilde{\mathcal{A}}) \to \left(D[0, y_0], \mathcal{D}[0, y_0]\right), \quad \tilde{Y}_i(\omega) := \pi_i(\omega) := \omega_i.$$

Now Corollary 3.3 of Hill (2009) implies

$$\widetilde{L}_n(t, y) \underset{(n \to \infty)}{\overset{\mathcal{D}}{\longrightarrow}} W(t, y) \quad \text{in } D^2, \tag{38}$$

where $\{W(t, y)\}$ is a zero-mean Gaussian process with continuous paths along t and y. For the covariance structure of the process consider weak convergence of the $\mathbb{R}^2$-valued random vector

$$\begin{pmatrix} \widetilde{L}_n(t_1, y_1) \\ \widetilde{L}_n(t_2, y_2) \end{pmatrix}$$

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

or, using the Cramér–Wold device, of

$$a\widetilde{L}_n(t_1,y_1)+b\widetilde{L}_n(t_2,y_2)=b\left[\widetilde{L}_n(t_2,y_2)-\widetilde{L}_n(t_1,y_2)\right]+\left[a\widetilde{L}_n(t_1,y_1)+b\widetilde{L}_n(t_1,y_2)\right]=:A_n+B_n,$$

for arbitrary $a,b\in\mathbb{R}$. Observe that A_n and B_n are independent for each n and hence it suffices to consider weak convergence of A_n and B_n separately. W.l.o.g. let $t_1\le t_2$.

For A_n we have

$$\widetilde{L}_n(t_2,y_2)-\widetilde{L}_n(t_1,y_2)=\sum_{j=m_n(t_1)+1}^{m_n(t_2)}\widetilde{Y}_j^I(y_2).$$

Then **(U2)** implies

$$s_n^2:=\sum_{j=m_n(t_1)+1}^{m_n(t_2)}Var\left(\widetilde{Y}_j^I(y_2)\right)=\frac{m_n(t_2)-m_n(t_1)}{k}Var\left(\sum_{i=1}^{r_n}I_{\left\{U_i>1-\frac{k}{n}y_2\right\}}\right)\underset{(n\to\infty)}{\longrightarrow}(t_2-t_1)\,r(y_2,y_2)=:\sigma_A^2. \tag{39}$$

The Lyapunov condition (cf., e.g., Billingsley, 1968, Theorem 7.3) is satisfied (for $\delta=2$), since

$$\frac{1}{s_n^4}\sum_{j=m_n(t_1)+1}^{m_n(t_2)}\mathbb{E}\left(\widetilde{Y}_j^I(y_2)\right)^4\le K\frac{1}{k}\frac{n}{r_nk}\mathbb{E}\left[\sum_{i=1}^{r_n}\left(I_{\left\{U_i>1-\frac{k}{n}y_2\right\}}-\frac{k}{n}y_2\right)\right]^4\underset{(n\to\infty)}{\overset{\textbf{(U3)}}{\longrightarrow}}0. \tag{40}$$

Using (39) we thus get

$$A_n\underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}}\mathcal{N}\left(0,b^2\sigma_A^2\right). \tag{41}$$

For B_n we have

$$a\widetilde{L}_n(t_1,y_1)+b\widetilde{L}_n(t_1,y_2)=\sum_{j=1}^{m_n(t_1)}\left(a\widetilde{Y}_j^I(y_1)+b\widetilde{Y}_j^I(y_2)\right).$$

Reasoning similarly as for the weak convergence of A_n (using Loève's c_r inequality for the analog of (40)) we get

$$B_n\underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}}\mathcal{N}\left(0,\sigma_B^2\right), \tag{42}$$

where $\sigma_B^2=a^2t_1r(y_1,y_1)+2abt_1r(y_1,y_2)+b^2t_1r(y_2,y_2)$. Combining (41) and (42) gives

$$A_n+B_n=a\widetilde{L}_n(t_1,y_1)+b\widetilde{L}_n(t_2,y_2)\underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}}\mathcal{N}\left(0,\sigma^2\right),$$

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

where $\sigma^2 = b^2\sigma_A^2 + \sigma_B^2 = a^2 t_1 r(y_1, y_1) + 2ab t_1 r(y_1, y_2) + b^2 t_2 r(y_2, y_2)$, i.e.,

$$\begin{pmatrix} \widetilde{L}_n(t_1, y_1) \\ \widetilde{L}_n(t_2, y_2) \end{pmatrix} \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} \mathcal{N}(\mathbf{0}, \Sigma) \overset{\mathcal{D}}{=} \begin{pmatrix} W(t_1, y_1) \\ W(t_2, y_2) \end{pmatrix},$$

where

$$\Sigma = \begin{pmatrix} t_1 r(y_1, y_1) & t_1 r(y_1, y_2) \\ t_1 r(y_1, y_2) & t_2 r(y_2, y_2) \end{pmatrix}.$$

Thus, $\{W(t, y)\}$ has the claimed covariance structure in (35). By (37) we also have

$$L_n(t, y) \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} W(t, y) \quad \text{in } D^2. \tag{43}$$

Third, in view of (36) and (43) it remains to prove

$$\sup_{\substack{t\in[t_0,1] \\ y\in[0,y_0]}} \left| \sum_{j=1}^{m_n(t)} Y_j^J(y) \right| \underset{(n\to\infty)}{=} o_P(1). \tag{44}$$

Set

$$S_m(y) := \sum_{j=1}^{m} \widetilde{Y}_j^J(y) \qquad \text{and} \qquad \|S_m\| := \sup_{y\in[0,y_0]} |S_m(y)|,$$

where $\widetilde{Y}_j^J(\cdot)$ are i.i.d. copies of $Y_j^J(\cdot)$ as above. Similarly as in (37), to show that $\sum_{j=1}^{m_n(t)} Y_j^J(y)$ is asymptotically negligible, it suffices to do so for $\sum_{j=1}^{m_n(t)} \widetilde{Y}_j^J(y)$. To this end the Ottaviani inequality (cf., e.g., Shorack and Wellner, 1996, Proposition A.1.1) yields for any $\varepsilon > 0$

$$P\left\{ \sup_{\substack{t\in[t_0,1] \\ y\in[0,y_0]}} \left| \sum_{j=1}^{m_n(t)} \widetilde{Y}_j^J(y) \right| > 2\varepsilon \right\} \le P\left\{ \max_{m\in\{1,\ldots,m_n(1)\}} \|S_m\| > 2\varepsilon \right\}$$
$$\le \frac{P\left\{ \|S_{m_n(1)}\| > \varepsilon \right\}}{1 - \max_{m\in\{1,\ldots,m_n(1)\}} P\{\|S_m\| > \varepsilon\}}.$$

We show that $P\{\|S_m\| > \varepsilon\} = o(1)$ uniformly in $m = 1, \ldots, m_n(1)$. For this let $\Delta = \Delta_n > 0$ be a sequence, s.t. $\Delta = \mathcal{O}(k^{-1/2})$ and $y_0/\Delta \in \mathbb{N}$. Observe that (because of $m \le n/r_n$) for all $y \in [(i-1)\Delta, i\Delta]$

$$S_m((i-1)\Delta) - \underbrace{\frac{l_n}{r_n}\sqrt{k}\Delta}_{\to 0} \le S_m(y) \le S_m(i\Delta) + \underbrace{\frac{l_n}{r_n}\sqrt{k}\Delta}_{\to 0},$$

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

from which we conclude via Markov's inequality

$$P\{\|S_m\| > \varepsilon\} \leq P\left\{\max_{i\in\{0,\dots,y_0/\Delta\}} |S_m(i\Delta)| > \varepsilon/2\right\}$$
$$\leq (\varepsilon/2)^{-4}\mathbb{E}\left[\max_{i\in\{0,\dots,y_0/\Delta\}} |S_m(i\Delta)|^4\right].$$

First we bound $\mathbb{E}\left[|S_m(i\Delta)|^4\right]$ by arguments similar to Rootzén (2009, p. 479). We have

$$\mathbb{E}\left[\sum_{i=1}^{r_n} I_{\left\{U_i>1-\frac{k}{n}y\right\}}\right]^4 \geq \mathbb{E}\left[\sum_{i=1}^{r_n} I_{\left\{U_i>1-\frac{k}{n}y\right\}}\right]^2, \tag{45}$$

because the sum of the indicators is $\mathbb{N}_0$-valued, and (also using strict stationarity) for $p = 2, 4$

$$\mathbb{E}\left[\sum_{i=1}^{r_n} I_{\left\{U_i>1-\frac{k}{n}y\right\}}\right]^p \geq \mathbb{E}\left[\sum_{w=1}^{\lfloor r_n/l_n\rfloor} \sum_{i=(w-1)l_n+1}^{wl_n} I_{\left\{U_i>1-\frac{k}{n}y\right\}}\right]^p$$
$$\geq \left\lfloor \frac{r_n}{l_n} \right\rfloor \mathbb{E}\left[\sum_{i=1}^{l_n} I_{\left\{U_i>1-\frac{k}{n}y\right\}}\right]^p,$$

whence $\mathbb{E}\left[\widetilde{Y}_j^J(y)\right]^p \leq K\frac{l_n}{n}k^{1-p/2}y$ with **(U3)**. Rosenthal's inequality now implies

$$\mathbb{E}\left[|S_m(i\Delta)|^4\right] \leq K\left\{m\mathbb{E}\left[\widetilde{Y}_j^J(i\Delta)\right]^4 + \left(m\mathbb{E}\left[\widetilde{Y}_j^J(i\Delta)\right]^2\right)^2\right\}$$
$$\leq K\left\{\frac{l_n}{r_n k} i\Delta + \left(\frac{l_n}{r_n}\right)^2 i^2\Delta^2\right\}$$

for K independent of m. Then, applying Móricz' (1982) Theorem in a similar way as for (5.2) in Drees (2000), we get

$$\mathbb{E}\left[\max_{i\in\{0,\dots,y_0/\Delta\}} |S_m(i\Delta)|^4\right] \leq K\left\{\frac{l_n}{r_n k}\log^4\left(\frac{1}{\Delta}\right) + \left(\frac{l_n}{r_n}\right)^2\right\} \underset{(n\to\infty)}{\longrightarrow} 0,$$

whence (44) follows, completing the proof. ■

Based upon the result of Theorem 4, we can derive a weighted version of the convergence in (34):

THEOREM 5. *Suppose* **(U1)–(U3)** *hold and* $q(\cdot)$ *satisfies (14). Then*

$$\frac{\sqrt{k}}{q(y)}\left\{\frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor}\left[I_{\left\{U_i>1-\frac{k}{n}y\right\}} - \frac{k}{n}y\right]\right\} \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} \frac{W(t,y)}{q(y)} \quad \text{in } D^2,$$

where $\{W(t,y)\}$ *is as in Theorem 4.*

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

Proof. For brevity put $e_n(t,y) := \frac{1}{\sqrt{k}}\sum_{i=1}^{\lfloor nt\rfloor}\left[I_{\left\{U_i>1-\frac{k}{n}y\right\}} - \frac{k}{n}y\right]$. In view of Theorem 4 and Billingsley (1968, Theorem 4.2), it suffices to prove that for all $\varepsilon > 0$

$$\lim_{\vartheta\downarrow 0}\limsup_{n\to\infty} P\left\{\sup_{\substack{t\in[t_0,1]\\ y\in(0,\vartheta]}}\left|\frac{e_n(t,y)}{q(y)}\right| > 3\varepsilon\right\} = 0, \tag{46}$$

$$\lim_{\vartheta\downarrow 0} P\left\{\sup_{\substack{t\in[t_0,1]\\ y\in(0,\vartheta]}}\left|\frac{W(t,y)}{q(y)}\right| > \varepsilon\right\} = 0. \tag{47}$$

We first show (46). For $s = 1,\dots,n$ define

$$S_s(y) := \frac{1}{q(y)}\frac{1}{\sqrt{k}}\sum_{i=1}^{s}\left(I_{\left\{U_i>1-\frac{k}{n}y\right\}} - \frac{k}{n}y\right) \qquad \text{and} \qquad \|S_s\| := \sup_{y\in(0,\vartheta]}|S_s(y)|,$$

such that $e_n(t,y)/q(y) = S_{\lfloor nt\rfloor}(y)$. Then

$$\begin{aligned} &P\left\{\sup_{\substack{t\in[t_0,1]\\ y\in(0,\vartheta]}}\left|\frac{e_n(t,y)}{q(y)}\right| > 3\varepsilon\right\} \\ &= P\left\{\max_{s\in\{\lfloor nt_0\rfloor,\dots,n\}}\|S_s\| > 3\varepsilon\right\} \\ &\le \frac{P\{\|S_n\|>\varepsilon\} + P\left\{\max_{\substack{r<s\in\{\lfloor nt_0\rfloor,\dots,n\}\\ s-r\le 2r_n}}\|S_s - S_r\| > \varepsilon\right\} + \frac{n}{r_n}\beta(r_n)}{1 - \max_{s\in\{\lfloor nt_0\rfloor,\dots,n\}} P\{\|S_n - S_s\| > \varepsilon\}}, \end{aligned} \tag{48}$$

where the last step follows from the Ottaviani-type inequality in Bücher (2015, Lemma 3) combined with the fact that α-mixing coefficients are bounded by β-mixing coefficients. Next, we show that the numerator tends to zero and the denominator tends to 1. First consider the three terms in the numerator.

First, because β-mixing coefficients are nonincreasing in the argument, we can bound

$$\frac{n}{r_n}\beta(r_n) \le \frac{n}{r_n}\beta(l_n) \underset{(n\to\infty)}{\overset{\textbf{(U1)}}{=}} o(1).$$

Second, because of (5.3) in the proof of Drees (2000, Theorem 2.2),

$$\lim_{\vartheta\downarrow 0}\limsup_{n\to\infty} P\{\|S_n\|>\varepsilon\} = \lim_{\vartheta\downarrow 0}\limsup_{n\to\infty} P\left\{\sup_{y\in(0,\vartheta]}\left|\frac{e_n(1,y)}{q(y)}\right| > \varepsilon\right\} = 0.$$

Now, for the numerator it remains to be shown that

$$\lim_{\vartheta\downarrow 0}\limsup_{n\to\infty} P\left\{\max_{\substack{r<s\in\{\lfloor nt_0\rfloor,\dots,n\}\\ s-r\le 2r_n}}\sup_{y\in(0,\vartheta]}\frac{1}{q(y)}\left|\frac{1}{\sqrt{k}}\sum_{i=r+1}^{s}\left(I_{\left\{U_i>1-\frac{k}{n}y\right\}}-\frac{k}{n}y\right)\right|>\varepsilon\right\}=0,$$

where

$$\begin{aligned}
&\max_{\substack{r<s\in\{\lfloor nt_0\rfloor,\dots,n\}\\ s-r\le 2r_n}}\sup_{y\in(0,\vartheta]}\frac{1}{q(y)}\left|\frac{1}{\sqrt{k}}\sum_{i=r+1}^{s}\left(I_{\left\{U_i>1-\frac{k}{n}y\right\}}-\frac{k}{n}y\right)\right|\\
&\quad\le \underbrace{\max_{\substack{r<s\in\{\lfloor nt_0\rfloor,\dots,n\}\\ s-r\le 2r_n}}\sup_{y\in(0,\vartheta]}\frac{1}{q(y)}\left|\frac{1}{\sqrt{k}}\sum_{i=r+1}^{s}I_{\left\{U_i>1-\frac{k}{n}y\right\}}\right|}_{=:A_n}+\underbrace{\sup_{y\in(0,\vartheta]}\frac{1}{q(y)}\left|\frac{2r_n}{\sqrt{k}}\frac{k}{n}y\right|}_{=:B_n}.
\end{aligned}$$

By condition **(U1)** B_n tends to zero. As for A_n

$$\begin{aligned}
&P\left\{\max_{\substack{r<s\in\{\lfloor nt_0\rfloor,\dots,n\}\\ s-r\le 2r_n}}\sup_{y\in(0,\vartheta]}\frac{1}{q(y)}\left|\frac{1}{\sqrt{k}}\sum_{i=r+1}^{s}I_{\left\{U_i>1-\frac{k}{n}y\right\}}\right|>\varepsilon/2\right\}\\
&\quad\le P\left\{\max_{m\in\{0,\dots,\lfloor n/(2r_n)\rfloor\}}\sup_{y\in(0,\vartheta]}\frac{1}{q(y)}\left|\frac{1}{\sqrt{k}}\sum_{i=\lfloor m2r_n\rfloor+1}^{\lfloor(m+2)2r_n\rfloor}I_{\left\{U_i>1-\frac{k}{n}y\right\}}\right|>\varepsilon/2\right\}\\
&\quad\le \left\lfloor\frac{n}{2r_n}\right\rfloor P\left\{\sup_{y\in(0,\vartheta]}\frac{1}{q(y)}\left|\frac{1}{\sqrt{k}}\sum_{i=1}^{4r_n}I_{\left\{U_i>1-\frac{k}{n}y\right\}}\right|>\varepsilon/2\right\}\\
&\quad\le \left\lfloor\frac{n}{2r_n}\right\rfloor\sum_{j=0}^{\infty}P\left\{\sup_{y\in\left(\vartheta e^{-(j+1)},\vartheta e^{-j}\right]}\frac{1}{q(y)}\left|\frac{1}{\sqrt{k}}\sum_{i=1}^{4r_n}I_{\left\{U_i>1-\frac{k}{n}y\right\}}\right|>\varepsilon/2\right\}\\
&\quad\overset{(33)}{\le} \left\lfloor\frac{n}{2r_n}\right\rfloor\sum_{j=0}^{\infty}P\left\{\frac{1}{\sqrt{k}}\sum_{i=1}^{4r_n}I_{\left\{U_i>1-\frac{k}{n}\vartheta e^{-j}\right\}}>\varepsilon/2q(\vartheta e^{-(j+1)})\right\}\\
&\quad\le \left\lfloor\frac{n}{2r_n}\right\rfloor\sum_{j=0}^{\infty}\mathbb{E}\left[\frac{1}{\sqrt{k}}\sum_{i=1}^{4r_n}I_{\left\{U_i>1-\frac{k}{n}\vartheta e^{-j}\right\}}\right]^2(\varepsilon/2)^{-2}q^{-2}\left(\vartheta e^{-(j+1)}\right)\\
&\quad\le \left\lfloor\frac{n}{2r_n}\right\rfloor\sum_{j=0}^{\infty}K\frac{r_nk}{n}\frac{1}{k}\vartheta e^{-j}(\varepsilon/2)^{-2}q^{-2}\left(\vartheta e^{-(j+1)}\right)\\
&\quad\le K\sum_{j=0}^{\infty}\vartheta e^{-j}q^{-2}\left(\vartheta e^{-(j+1)}\right),
\end{aligned}$$

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

where the second inequality follows from strict stationarity and the second to last one from Loève's c_r inequality combined with (45) and **(U3)**. Using (33) the last term can be bounded by

$$K\sum_{j=0}^{\infty}\left[\vartheta e^{-(j+1)}\right]^{1-2\nu}\left|\log\left(\vartheta e^{-(j+1)}\right)\right|^{-2\mu} \le K\int_0^{\infty}\left(\vartheta e^{-t}\right)^{1-2\nu}\left|\log\left(\vartheta e^{-t}\right)\right|^{-2\mu}\mathrm{d}t$$
$$= K\int_{-\log(\vartheta)}^{\infty} e^{-z(1-2\nu)}z^{-2\mu}\mathrm{d}z,$$

which tends to 0 as $\vartheta \downarrow 0$, if and only if $\nu < \frac{1}{2}$ or $\nu = \frac{1}{2}$ and $\mu > \frac{1}{2}$. All in all the numerator tends to zero as $n \to \infty$ followed by $\vartheta \downarrow 0$.

Now consider the denominator of (48): by strict stationarity we can write

$$\max_{s\in\{\lfloor nt_0\rfloor,\ldots,n\}} P\{\|S_n - S_s\| > \varepsilon\}$$
$$= \max_{s\in\{\lfloor nt_0\rfloor,\ldots,n\}} P\{\|S_m\| > \varepsilon\}$$
$$= \max_{s\in\{\lfloor nt_0\rfloor,\ldots,n\}} P\left\{\sup_{y\in(0,\vartheta]}\left|\frac{1}{q(y)}\frac{1}{\sqrt{k}}\sum_{i=1}^{m}\left(I_{\left\{U_i>1-\frac{k}{n}y\right\}} - \frac{k}{n}y\right)\right| > \varepsilon\right\},$$

where

$$\frac{1}{q(y)}\frac{1}{\sqrt{k}}\sum_{i=1}^{m}\left(I_{\left\{U_i>1-\frac{k}{n}y\right\}} - \frac{k}{n}y\right) = \underbrace{\frac{1}{q(y)}\sum_{w=0}^{\lfloor m/r_n\rfloor-1}\frac{1}{\sqrt{k}}\sum_{i=wr_n+1}^{(w+1)r_n}\left(I_{\left\{U_i>1-\frac{k}{n}y\right\}} - \frac{k}{n}y\right)}_{=:C_{m,n}}$$
$$+\underbrace{\frac{1}{q(y)}\frac{1}{\sqrt{k}}\sum_{i=\lfloor m/r_n\rfloor r_n+1}^{m}\left(I_{\left\{U_i>1-\frac{k}{n}y\right\}} - \frac{k}{n}y\right)}_{=:D_{m,n}}.$$

We have

$$D_{m,n} = \underbrace{\frac{1}{q(y)}\frac{1}{\sqrt{k}}\sum_{i=\lfloor m/r_n\rfloor r_n+1}^{m} I_{\left\{U_i>1-\frac{k}{n}y\right\}}}_{=:\widetilde{A}_{m,n}} - \underbrace{\frac{m-\lfloor m/r_n\rfloor r_n}{\sqrt{k}}\frac{k}{n}\frac{y}{q(y)}}_{=:\widetilde{B}_{m,n}}.$$

Because there are at most r_n terms in the sum in $\widetilde{A}_{m,n}$, that $\sup_{y\in(0,\vartheta]}\widetilde{A}_{m,n} = o_P(1)$ uniformly in m can be seen as for A_n. The convergence of $\sup_{y\in(0,\vartheta]}\widetilde{B}_{m,n}$ (uniformly in m) can also be seen as the one for B_n. It remains to investigate $C_{m,n}$. To this end consider the proof of Theorem 2.2 in Drees (2000). Replacing his $m_n = \left\lfloor \frac{n}{2r_n}\right\rfloor$ by $m_n = \left\lfloor \frac{m}{2r_n}\right\rfloor$ in the proof it is easy to see that

$$\lim_{\vartheta\downarrow 0}\lim_{n\to\infty} P\left\{\sup_{y\in\left(0,\vartheta e^{-j_n}\right]}\left|C_{m,n}\right| > \varepsilon\right\} = 0 \quad \text{uniformly in } m \in \{1,\ldots,n\},$$

where $j_n := \min\{j \in \mathbb{N} : \sqrt{k} \le \eta \frac{q}{Id}(\vartheta e^{-(j+1)})\}$ for some small $\eta > 0$. The uniformity is due to the fact that for all $m \in \{1,\ldots,n\}$ in Drees' (2000) notation

$$\mathbb{E}\left(\widetilde{S}_n(\vartheta e^{-j})\right) \le \sqrt{k}\vartheta e^{-j}$$

in the step leading to his (5.6). Using assumption **(U3)** and again replacing $m_n = \left\lfloor \frac{n}{2r_n} \right\rfloor$ by $m_n = \left\lfloor \frac{m}{2r_n} \right\rfloor$ in the proof of Drees (2000, Theorem 2.3), retracing the proof again yields

$$\lim_{\vartheta \downarrow 0} \lim_{n\to\infty} P\left\{ \sup_{y\in\left(\vartheta e^{-j_n},\vartheta\right]} \left|C_{m,n}\right| > \varepsilon \right\} = 0 \quad \text{uniformly in } m \in \{1,\ldots,n\}.$$

The uniformity is due to the uniformity of his moment inequality (5.14) derived there by an application of Burkholder's inequality.

Next we will prove (47) via Lin and Choi (1999, Lemma 2.1). Use the fact that **(U3)** implies

$$\frac{n}{r_n k} Var\left(\sum_{i=1}^{r_n} I_{\left\{1-\frac{k}{n}y_2 < U_i \le 1-\frac{k}{n}y_1\right\}}\right) \le \frac{n}{r_n k}\mathbb{E}\left[\sum_{i=1}^{r_n} I_{\left\{1-\frac{k}{n}y_2 < U_i \le 1-\frac{k}{n}y_1\right\}}\right]^4 \le C(y_2 - y_1).$$

(Recall again for the first inequality that the sum of the indicators is $\mathbb{N}_0$-valued.) By **(U2)** the left-hand side converges to $r(y_2,y_2) - 2r(y_1,y_2) + r(y_1,y_1)$ as $n \to \infty$. Hence,

$$r(y_2,y_2) - 2r(y_1,y_2) + r(y_1,y_1) \le C\,|y_1 - y_2|. \tag{49}$$

Assume w.l.o.g. $t_1 > t_2$ and use the Cauchy–Schwarz inequality in the last inequality to obtain

$$\begin{aligned}
\mathbb{E}[W(t_1,y_1) - W(t_2,y_2)]^2 &= Var(W(t_1,y_1)) + Var(W(t_2,y_2)) - 2Cov(W(t_1,y_1),W(t_2,y_2)) \\
&= t_1 r(y_1,y_1) + t_2 r(y_2,y_2) - 2\min(t_1,t_2) r(y_1,y_2) \\
&= (t_1 - t_2) r(y_1,y_1) + t_2\{r(y_2,y_2) - 2r(y_1,y_2) + r(y_1,y_1)\} \\
&\le C\{|t_1 - t_2| + |y_1 - y_2|\} \\
&\le \sqrt{2}C\left\{|t_1 - t_2|^2 + |y_1 - y_2|^2\right\}^{1/2} \\
&=: \varphi^2\left(\left\|\begin{pmatrix} t_1 \\ y_1 \end{pmatrix} - \begin{pmatrix} t_2 \\ y_2 \end{pmatrix}\right\|_2\right),
\end{aligned}$$

i.e., $\varphi(r) = \sqrt{2}C\sqrt{r}$. Next, define

$$\mathbb{D}_j := [t_0, 1] \times \left[\vartheta e^{-(j+1)}, \vartheta e^{-j}\right];$$

$$\Gamma_j^2 := \sup_{(t,y)\in\mathbb{D}_j} \mathbb{E}[W(t,y)]^2 = \sup_{y\in\left[\vartheta e^{-(j+1)},\vartheta e^{-j}\right]} r(y,y) \overset{(49)}{\le} C\vartheta e^{-(j+1)},$$

$$\lambda_j := \vartheta e^{-(j+1)}[e-1],$$

so that

$$\int_0^\infty \varphi(\sqrt{2}\lambda_j 2^{-x^2})\mathrm{d}x \le K\sqrt{\lambda_j}\int_0^\infty 2^{-x^2/2}\mathrm{d}x = \mathcal{O}\left(\vartheta^{1/2}e^{-\frac{1}{2}(j+1)}\right),$$

and apply Lemma 2.1 of Lin and Choi (1999) to get, using $\nu < \frac{1}{2}$ or $\nu = \frac{1}{2}$ and $\mu > 1/2$ from (14),

$$\begin{aligned}
&P\left\{\sup_{\substack{t\in[t_0,1]\\ y\in(0,\vartheta]}}\left|\frac{W(t,y)}{q(y)}\right| > \varepsilon\right\} \\
&\le \sum_{j=0}^{\infty} P\left\{\sup_{\substack{t\in[t_0,1]\\ y\in\left(\vartheta e^{-(j+1)},\vartheta e^{-j}\right]}} |W(t,y)| > \varepsilon\left(\vartheta e^{-(j+1)}\right)^{\nu}\left|\log(\vartheta e^{-(j+1)})\right|^{\mu}\right\} \\
&\le K\sum_{j=0}^{\infty}\exp\left\{-\frac{1}{2}\left(\frac{\varepsilon\left(\vartheta e^{-(j+1)}\right)^{\nu}\left|\log(\vartheta e^{-(j+1)})\right|^{\mu}}{\Gamma_j + (2\sqrt{2}+2)K_1\int_0^\infty \varphi(\sqrt{2}\lambda_j 2^{-x^2})\mathrm{d}x}\right)^2\right\} \\
&\le K\sum_{j=0}^{\infty}\exp\left\{-K_1\left(\vartheta e^{-(j+1)}\right)^{2\nu-1}\left|\log(\vartheta e^{-(j+1)})\right|^{2\mu}\right\} \\
&\le K\vartheta^{K_2}\sum_{j=0}^{\infty}\exp\{-K_2\}^{j+1} \underset{(\vartheta\downarrow 0)}{\longrightarrow} 0,
\end{aligned} \tag{50}$$

by boundedness of the sum. ■

PROPOSITION 2. *Suppose $q(\cdot)$ and $W(\cdot,\cdot)$ are as in Theorem 5. Then*

$$\lim_{\vartheta\downarrow 0}\sup_{\substack{t\in[t_0,1]\\ y\in(0,\vartheta]}}\left|\frac{W(t,y)}{q(y)}\right| \overset{\text{a.s.}}{=} 0.$$

Proof. The proof of (50) reveals that by choosing $\vartheta = \vartheta_n = n^{-2/K_2}$ one obtains

$$\sum_{n=1}^{\infty} P\left\{\sup_{\substack{t\in[t_0,1]\\ y\in(0,\vartheta_n]}}\left|\frac{W(t,y)}{q(y)}\right| > \varepsilon\right\} < \infty,$$

which implies by the Borel–Cantelli lemma that, a.s.,

$$\lim_{\vartheta\downarrow 0}\sup_{\substack{t\in[t_0,1]\\ y\in(0,\vartheta]}}\left|\frac{W(t,y)}{q(y)}\right| = \limsup_{\vartheta\downarrow 0}\sup_{\substack{t\in[t_0,1]\\ y\in(0,\vartheta]}}\left|\frac{W(t,y)}{q(y)}\right| = 0,$$

because $\varepsilon > 0$ was arbitrary. ■

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

Proof of Theorem 1. By the proof of Theorem 3.1 in Drees (2000) **(C1)–(C4)** imply **(U1)–(U3)** for $U_i := F(X_i) \sim \mathcal{U}[0,1]$. Hence, noting that

$$X_i > U\left(\frac{n}{ky}\right) \quad \Longleftrightarrow \quad U_i > 1 - \frac{k}{n}y,$$

we obtain from Theorem 5 that

$$\frac{\sqrt{k}}{q(y)}\left(\frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor} I_{\left\{X_i > U\left(\frac{n}{ky}\right)\right\}} - ty\right) \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} \frac{1}{q(y)} W(t,y) \quad \text{in } D^2. \tag{51}$$

Applying the continuous mapping theorem to (51), we get

$$\frac{\sqrt{k}}{q(y)}\begin{pmatrix} \frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor} I_{\left\{X_i > U\left(\frac{n}{ky}\right)\right\}} - ty \\ \frac{1}{k}\sum_{i=\lfloor nt\rfloor+1}^{n} I_{\left\{X_i > U\left(\frac{n}{ky}\right)\right\}} - (1-t)y \\ \frac{1}{k}\sum_{i=\lfloor nt\rfloor+1}^{\lfloor n(t+t_0)\rfloor} I_{\left\{X_i > U\left(\frac{n}{ky}\right)\right\}} - t_0 y \end{pmatrix} \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} \frac{1}{q(y)}\begin{pmatrix} W(t,y) \\ W(1,y)-W(t,y) \\ W(t+t_0,y)-W(t,y)\end{pmatrix}$$

in $D^3([t_0, 1-t_0]\times[0, y_0+\delta])$. By Skorohod's representation theorem (cf., e.g., Wichura, 1970, Theorem 1), we can pretend that this convergence holds almost surely on a suitable probability space:

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\in(0,y_0+\delta]}} \frac{1}{q(y)}\left|\sqrt{k}\begin{pmatrix} \frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor} I_{\left\{X_i > U\left(\frac{n}{ky}\right)\right\}} - ty \\ \frac{1}{k}\sum_{i=\lfloor nt\rfloor+1}^{n} I_{\left\{X_i > U\left(\frac{n}{ky}\right)\right\}} - (1-t)y \\ \frac{1}{k}\sum_{i=\lfloor nt\rfloor+1}^{\lfloor n(t+t_0)\rfloor} I_{\left\{X_i > U\left(\frac{n}{ky}\right)\right\}} - t_0 y \end{pmatrix} - \begin{pmatrix} W(t,y) \\ W(1,y)-W(t,y) \\ W(t+t_0,y)-W(t,y)\end{pmatrix}\right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0. \tag{52}$$

(Note that the limits are continuous, hence convergence is uniform.) It remains to show that $U\left(\frac{n}{ky}\right)$ can be replaced by $y^{-\gamma}U\left(\frac{n}{k}\right)$ in (52). For brevity we carry out the steps for the first component of (52) only (the others being dealt with similarly). Similarly as in the proof of Corollary 3 in Einmahl *et al.* (2016) we set

$$y_n := \frac{n}{k}\left[1 - F\left(y^{-\gamma}U\left(\frac{n}{k}\right)\right)\right], \qquad y \in (0, y_0+\delta],$$

so that by **(C4)** (cf. Einmahl *et al.*, 2016, p. 46)

$$\sup_{y\in(0,y_0+\delta]}\left|\frac{y_n - y}{A(n/k)y}\right| \underset{(n\to\infty)}{=} \mathcal{O}(1). \tag{53}$$

Inserting y_n for y in the first component of (52) gives

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\in(0,y_0+\delta]}} \frac{1}{q(y_n)}\left|\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor} I_{\{X_i > y^{-\gamma}U(\frac{n}{k})\}} - ty_n\right) - W(t,y_n)\right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0. \tag{54}$$

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

Now we have to show that y_n can be replaced with y at the three occurences in (54). For $q(\cdot)$, by the first property in (14), it suffices to note that (using $y_n \underset{(n\to\infty)}{=} y(1+o(1))$ uniformly in y from (53))

$$\sup_{y\in(0,\vartheta/2]}\left|\frac{q(y_n)}{q(y)}\right| \overset{(33)}{=} \sup_{y\in(0,\vartheta/2]} \frac{y_n^{\nu}|\log y_n|^{\mu}}{y^{\nu}|\log y|^{\mu}}$$
$$= (1+o(1))^{\nu}\sup_{y\in(0,\vartheta/2]}\left|\frac{\log(y)+\log(1+o(1))}{\log(y)}\right|^{\mu} \underset{(n\to\infty)}{=} 1+o(1). \quad \textbf{(55)}$$

Combine (53) with $\sqrt{k}A(n/k)\to 0$ to see that ty_n may be replaced with ty. Finally, by a simple (uniform) continuity argument we have that for all $\vartheta>0$

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\in[\vartheta,y_0+\delta]}} \frac{1}{q(y)}|W(t,y_n)-W(t,y)| \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0. \quad \textbf{(56)}$$

Further, by Proposition 2, (53) and (55)

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\in(0,\vartheta]}} \frac{1}{q(y)}|W(t,y_n)-W(t,y)| \le \sup_{\substack{t\in[t_0,1-t_0]\\ y\in(0,\vartheta]}} \frac{q(y_n)}{q(y)}\left|\frac{W(t,y_n)}{q(y_n)}\right|$$
$$+\sup_{\substack{t\in[t_0,1-t_0]\\ y\in(0,\vartheta]}}\left|\frac{W(t,y)}{q(y)}\right| \overset{\text{a.s.}}{\longrightarrow} 0,$$

as $\vartheta\downarrow 0$ and $n\to\infty$, justifying the replacement in $W(\cdot,\cdot)$. ■

Proof of Corollary 1. We will only prove the convergence of the first component in (16), the others being proved similarly. Theorem 1 implies (because $y_0\ge 1$)

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\in[1/2,1+\delta]}}\left|\sqrt{k}\left\{\frac{1}{kt}\sum_{i=1}^{\lfloor nt\rfloor} I_{\{X_i>y^{-\gamma}U(n/k)\}}-y\right\}-\frac{W(t,y)}{t}\right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0.$$

It follows exactly as in the proof of Einmahl *et al.* (2016, Theorem 3) using a generalized Vervaat lemma (cf. Einmahl, Gantner and Sawitzki, 2010, Lemma 5) that

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\in[1/2,1]}}\left|\sqrt{k}\left\{\left(\frac{X_k(0,t,y)}{U(n/k)}\right)^{-1/\gamma}-y\right\}+\frac{W(t,y)}{t}\right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0, \quad \textbf{(57)}$$

so in particular

$$\sup_{t\in[t_0,1-t_0]}\left|\sqrt{k}\left\{\left(\frac{X_k(0,t,1)}{U(n/k)}\right)^{-1/\gamma}-1\right\}+\frac{W(t,1)}{t}\right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0. \quad \textbf{(58)}$$

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

Replacing y with $y_n := yX_k(0,t,1)/U(n/k)$ in Theorem 1 implies, using (58),

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\geq y_0^{-\gamma}}} \frac{1}{q\left(y_n^{-1/\gamma}\right)} \left| \sqrt{k}\left\{ \frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor} I_{\{X_i > yX_k(0,t,1)\}} - y_n^{-1/\gamma} t \right\} - W\left(t, y_n^{-1/\gamma}\right) \right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0. \tag{59}$$

Now we show that $y_n^{-1/\gamma}$ can be replaced by $y^{-1/\gamma}$ at the three occurrences in (59). By the first property in (14), we need only justify the replacement in $q(\cdot)$ for large y. Since by (58) $y_n^{-1/\gamma} \underset{(n\to\infty)}{\overset{\text{a.s.}}{=}} y^{-1/\gamma}(1+o(1))$ uniformly in t and y, it follows as in (55) that

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\geq(\vartheta/2)^{-\gamma}}} \left| \frac{q(y_n^{-1/\gamma})}{q\left(y^{-1/\gamma}\right)} \right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{=}} 1+o(1).$$

As for $W(\cdot,\cdot)$, the same arguments as in (56) and below apply. Last, by (58) uniformly in t (and y)

$$\frac{1}{q\left(y^{-1/\gamma}\right)} t\sqrt{k}\left(y_n^{-1/\gamma} - y^{-1/\gamma}\right) \underset{(n\to\infty)}{\overset{\text{a.s.}}{=}} -\frac{1}{q\left(y^{-1/\gamma}\right)} y^{-1/\gamma} W(t,1) + o(1).$$

Making the replacements in (59) and multiplying through with $k/\lfloor kt\rfloor$ yields

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\geq y_0^{-\gamma}}} \frac{1}{q\left(y^{-1/\gamma}\right)} \left| \sqrt{k}\left(F_n(0,t,y) - \frac{kt}{\lfloor kt\rfloor} y^{-1/\gamma}\right) - \frac{k}{\lfloor kt\rfloor}\left(W\left(t, y^{-1/\gamma}\right) - y^{-1/\gamma} W(t,1)\right) \right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0.$$

Using $\frac{k}{\lfloor kt\rfloor} = 1/t + \mathcal{O}(1/k)$ uniformly in t and Proposition 2, the conclusion follows. ■

Proof of Proposition 1. We will derive convergence of $t\sqrt{k}(\widehat{\gamma}_{WLS}(0,t)-\gamma)$ from the first component of (16), the convergences of $(1-t)\sqrt{k}(\widehat{\gamma}_{WLS}(t,1)-\gamma)$, $t_0\sqrt{k}(\widehat{\gamma}_{WLS}(t,t+t_0)-\gamma)$ following similarly from the other components. The required joint convergence then follows from the joint convergence in (16).

Noting that for $i = 0,1,\ldots,\lfloor kt\rfloor - 1$

$$F_n(0,t,y) = \frac{\lfloor kt\rfloor - i}{\lfloor kt\rfloor} \quad \text{constant on } y \in \left[\frac{X_{\lfloor nt\rfloor-\lfloor kt\rfloor+i:\lfloor nt\rfloor}}{X_{\lfloor nt\rfloor-\lfloor kt\rfloor:\lfloor nt\rfloor}}, \frac{X_{\lfloor nt\rfloor-\lfloor kt\rfloor+i+1:\lfloor nt\rfloor}}{X_{\lfloor nt\rfloor-\lfloor kt\rfloor:\lfloor nt\rfloor}}\right) \quad \text{and}$$

$$F_n(0,t,y) = 0 \qquad \text{for } y \geq \frac{X_{\lfloor nt\rfloor:\lfloor nt\rfloor}}{X_{\lfloor nt\rfloor-\lfloor kt\rfloor:\lfloor nt\rfloor}},$$

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

it is straightforward to check with **(W1)** that

$$\int_1^\infty \left\{ \int_0^{F_n(0,t,y)} J(s)\mathrm{d}s \right\} \frac{\mathrm{d}y}{y} = \sum_{i=0}^{\lfloor kt\rfloor -1} \left\{ \int_0^{\frac{\lfloor kt\rfloor - i}{\lfloor kt\rfloor}} J(s)\mathrm{d}s \right\} \times \log\left(\frac{X_{\lfloor nt\rfloor - \lfloor kt\rfloor + i+1:\lfloor nt\rfloor}}{X_{\lfloor nt\rfloor - \lfloor kt\rfloor + i:\lfloor nt\rfloor}} \right) = \widehat{\gamma}_{WLS}(0,t).$$

Using **(W2)**, **(W3)** and partial integration it is further easy to establish that

$$\int_1^\infty \left\{ \int_0^{y^{-1/\gamma}} J(s)\mathrm{d}s \right\} \frac{\mathrm{d}y}{y} = \gamma \int_0^1 \left\{ \int_0^z J(s)\mathrm{d}s \right\} \frac{\mathrm{d}z}{z} = \gamma .$$

Combine these two facts to obtain

$$\begin{aligned}\sqrt{k}(\widehat{\gamma}_{WLS}(0,t) - \gamma) &= \sqrt{k} \int_1^\infty \left\{ \int_0^{F_n(0,t,y)} J(s)\mathrm{d}s - \int_0^{y^{-1/\gamma}} J(s)\mathrm{d}s \right\} \frac{\mathrm{d}y}{y} \\ &= \gamma \sqrt{k} \int_0^1 \left\{ \int_0^{F_n(0,t,u^{-\gamma})} J(s)\mathrm{d}s - \int_0^u J(s)\mathrm{d}s \right\} \frac{\mathrm{d}u}{u}. \qquad \textbf{(60)}\end{aligned}$$

Write the result of Corollary 1 as

$$\sup_{\substack{t\in[t_0,1-t_0]\\ u\in(0,1]}} \frac{1}{q(u)} \left| \sqrt{k}\left(F_n(0,t,u^{-\gamma}) - u\right) - \frac{1}{t}[W(t,u) - uW(t,1)] \right| \xrightarrow[(n\to\infty)]{\text{a.s.}} 0. \qquad \textbf{(61)}$$

Using the mean value theorem for the function $x \mapsto \int_0^x J(s)\mathrm{d}s$, we get for some $\xi = \xi_u \in (0,1)$

$$\int_0^{F_n(0,t,u^{-\gamma})} J(s)\mathrm{d}s = \int_0^u J(s)\mathrm{d}s + \left(F_n(0,t,u^{-\gamma}) - u\right) J\left(u + \xi\left(F_n(0,t,u^{-\gamma}) - u\right)\right).$$

Thus, from (60) and (61), uniformly in $t \in [t_0, 1-t_0]$

$$\begin{aligned}\sqrt{k}(\widehat{\gamma}_{WLS}(0,t) - \gamma) &= \gamma \sqrt{k} \int_0^1 \left(F_n(0,t,u^{-\gamma}) - u\right) J\left(u + \xi\left(F_n(0,t,u^{-\gamma}) - u\right)\right) \frac{\mathrm{d}u}{u} \\ &\overset{\text{a.s.}}{=} \frac{\gamma}{t} \int_0^1 (W(t,u) - uW(t,1) + o(1)q(u)) \\ &\quad \times J\left(u + \xi\left(F_n(0,t,u^{-\gamma}) - u\right)\right) \frac{\mathrm{d}u}{u} \qquad \textbf{(62)} \\ &\xrightarrow[(n\to\infty)]{} \frac{\gamma}{t} \int_0^1 (W(t,u) - uW(t,1)) J(u) \frac{\mathrm{d}u}{u} \\ &\overset{\textbf{(W1)}}{=} \frac{\gamma}{t} \int_0^1 W(t,u) J(u) \frac{\mathrm{d}u}{u}.\end{aligned}$$

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

Note that the integral in (62) is well-defined because of Proposition 2.

By calculating covariances of $\gamma \int_0^1 W(t,u) J(u) \frac{\mathrm{d}u}{u}$, we conclude (going back to the original probability space)

$$t\sqrt{k}(\widehat{\gamma}_{WLS}(0,t)-\gamma) \xrightarrow[(n\to\infty)]{\mathcal{D}} \sigma_{\widehat{\gamma}_{WLS},\gamma} W(t) \quad \text{in } D[t_0, 1-t_0],$$

where $W(\cdot)$ is a standard Brownian motion and

$$\sigma^2_{\widehat{\gamma}_{WLS},\gamma} = \gamma^2 \int_0^1 \int_0^1 \frac{r(x,y)}{xy} J(x)J(y)\mathrm{d}x\mathrm{d}y.$$

■

Proof of Theorem 2. We only prove part (a), the proof of (b) being similar. Because t_0 can be chosen arbitrarily close to 0 in (23), similarly as in the proof of Drees (2003, Theorem 2.3), one obtains using a diagonal argument and continuity of $W(\cdot)$ that in probability

$$\sup_{t\in[t_n,1]} \left| \sqrt{k}\,(\widehat{\gamma}(0,t)-\gamma) - \sigma_{\widehat{\gamma},\gamma} \frac{W(t)}{t} \right| \xrightarrow[(n\to\infty)]{} 0$$

for some sequence $t_n \downarrow 0$ tending to zero not too fast, whence with $\widetilde{t}_n := \frac{\lfloor nt_n \rfloor + 1}{n} (\geq t_n)$

$$\begin{aligned} \frac{k}{n} \sum_{i=\lfloor nt_n\rfloor+1}^{n} [\widehat{\gamma}(0,i/n) - \widehat{\gamma}(0,1)]^2 &= \int_{\widetilde{t}_n}^1 k\,(\widehat{\gamma}(0,t) - \widehat{\gamma}(0,1))^2 \,\mathrm{d}t \\ &= \int_{\widetilde{t}_n}^1 \left[\sqrt{k}\,(\widehat{\gamma}(0,t)-\gamma) - \sqrt{k}\,(\widehat{\gamma}(0,1)-\gamma)\right]^2 \mathrm{d}t \\ &= \sigma^2_{\widehat{\gamma},\gamma} \int_{\widetilde{t}_n}^1 \left(\frac{W(t)}{t} - W(1)\right)^2 \mathrm{d}t\,(1+o_P(1)) \\ &= \sigma^2_{\widehat{\gamma},\gamma} \int_{\log(\widetilde{t}_n)}^0 \left(\frac{W(e^y)}{e^{y/2}} - e^{y/2} W(1)\right)^2 \mathrm{d}y(1+o_P(1)), \quad \textbf{(63)} \end{aligned}$$

using the substitution $t = e^y$ ($\mathrm{d}t = e^y \mathrm{d}y$) in the fourth equality.

Now observe that $W(e^{\cdot})/e^{\cdot/2}$ is a zero-mean Gaussian process with covariance function

$$\mathbb{E}\left[\frac{W(e^x)}{e^{x/2}} \frac{W(e^y)}{e^{y/2}}\right] = e^{-(x+y)/2} \min(e^x, e^y) = e^{\min(x-y,y-x)/2} = e^{-|x-y|/2}$$

only depending on $x-y$, which implies (cf. Cramér and Leadbetter, 1967, p. 122) strict stationarity. By an application of the Birkhoff–Khintchine ergodic theorem (cf. Cramér and Leadbetter, 1967, p. 151) we obtain

$$\frac{1}{\log(1/\widetilde{t}_n)} \sigma^2_{\widehat{\gamma},\gamma} \int_{\log(\widetilde{t}_n)}^0 \left(\frac{W(e^y)}{e^{y/2}}\right)^2 \mathrm{d}y \xrightarrow[(n\to\infty)]{\text{a.s.}} \sigma^2_{\widehat{\gamma},\gamma} \mathbb{E}\left(\frac{W(e^y)}{e^{y/2}}\right)^2 = \sigma^2_{\widehat{\gamma},\gamma}. \quad \textbf{(64)}$$

Noting that $\int_{\log(\widetilde{t}_n)}^0 \left(e^{y/2} W(1)\right)^2 \mathrm{d}y = \mathcal{O}(1)$, the conclusion follows from (63) and (64). ■

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

Proof of Theorem 3. We will focus on the result under $\mathcal{H}_1^{>}$ (i.e., $\gamma_{\text{pre}} > \gamma_{\text{post}}$, meaning heavier prebreak tails) as the other can be established similarly. Write $X_i = X_{i,n}$ and $y_0 = y_{0,\text{pre}}$ for brevity. Then Theorem 1 implies for some $\widetilde{\delta} > 0$ that may change from line to line in this proof

$$\frac{1}{q(y)}\sqrt{k_{\text{pre}}}\left(\frac{1}{k_{\text{pre}}}\sum_{i=1}^{\lfloor nt\rfloor} I_{\{X_i > y^{-\gamma_{\text{pre}}}U_{\text{pre}}(n/k_{\text{pre}})\}} - yt\right) \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} \frac{W_{\text{pre}}(t,y)}{q(y)}, \tag{65}$$

in $D([t_0, t^*] \times [0, y_0 + \widetilde{\delta}])$ and for the postbreak r.v.s

$$\sup_{y \geq y_0^{-\gamma_{\text{pre}}} - \widetilde{\delta}} \frac{1}{q\left(y^{-1/\gamma_{\text{post}}}\right)}\left|\sqrt{k_{\text{post}}}\left(\frac{1}{k_{\text{post}}}\sum_{i=\lfloor nt^*\rfloor+1}^{n} I_{\{X_i > yU_{\text{post}}(n/k_{\text{post}})\}} - y^{-1/\gamma_{\text{post}}}(1-t^*)\right)\right.$$
$$\left. - \left(W_{\text{post}}\left(1, y^{-1/\gamma_{\text{post}}}\right) - W_{\text{post}}\left(t^*, y^{-1/\gamma_{\text{post}}}\right)\right)\right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0. \tag{66}$$

Inserting $y_n := yU_{\text{pre}}(n/k_{\text{pre}})/U_{\text{post}}(n/k_{\text{post}})$ for y in (66) and recalling Proposition 2 gives

$$\sup_{y \geq y_0^{-\gamma_{\text{pre}}} - \widetilde{\delta}} \frac{1}{q\left(y_n^{-1/\gamma_{\text{post}}}\right)}$$
$$\times \left|\sqrt{k_{\text{post}}}\left(\frac{1}{k_{\text{post}}}\sum_{i=\lfloor nt^*\rfloor+1}^{n} I_{\{X_i > yU_{\text{pre}}(n/k_{\text{pre}})\}} - y_n^{-1/\gamma_{\text{post}}}(1-t^*)\right)\right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0. \tag{67}$$

Further, $y_n^{-1/\gamma_{\text{post}}}(1-t^*)$ may be omitted in (67), since, by (27) and (33), for n large

$$\sqrt{k_{\text{post}}}\sup_{y \geq y_0^{-\gamma_{\text{pre}}} - \widetilde{\delta}}\left|\frac{y_n^{-1/\gamma_{\text{post}}}}{q\left(y_n^{-1/\gamma_{\text{post}}}\right)}\right| \leq \sqrt{k_{\text{post}}}\, y_n^{-1/(2\gamma_{\text{post}})} \underset{(n\to\infty)}{=} o(1).$$

Using $k_{\text{post}} = \mathcal{O}\left(k_{\text{pre}}\right)$ and, by (14) and (27), for n sufficiently large $q\left(y_n^{-1/\gamma_{\text{post}}}\right)/q\left(y^{-1/\gamma_{\text{pre}}}\right) \leq 1$, this yields

$$\sup_{y \geq y_0^{-\gamma_{\text{pre}}} - \widetilde{\delta}} \frac{1}{q\left(y^{-1/\gamma_{\text{pre}}}\right)}\left|\sqrt{k_{\text{pre}}}\left(\frac{1}{k_{\text{pre}}}\sum_{i=\lfloor nt^*\rfloor+1}^{n} I_{\{X_i > yU_{\text{pre}}(n/k_{\text{pre}})\}}\right)\right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0.$$

Going back to the original probability space we obtain by non-negativity of the indicators

$$\frac{\sqrt{k_{\text{pre}}}}{q(y)}\left(\frac{1}{k_{\text{pre}}}\sum_{i=\lfloor nt^*\rfloor+1}^{\lfloor nt\rfloor} I_{\{X_i > y^{-\gamma_{\text{pre}}}U_{\text{pre}}(n/k_{\text{pre}})\}}\right) \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} 0$$

in $D([t^*, 1-t_0] \times [0, y_0 + \widetilde{\delta}])$. **(68)**

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

Hence, letting $k = k_{\text{pre}}$ for brevity in the rest of the proof, we get from (65) and (68) that

$$\frac{\sqrt{k}}{q(y)}\left(\frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor} I_{\{X_i > y^{-\gamma_{\text{pre}}} U_{\text{pre}}(n/k)\}} - y^{-1/\gamma_{\text{pre}}}\min(t^*,t)\right) \xrightarrow[(n\to\infty)]{\mathcal{D}} \frac{W_{\text{pre}}\left(\min(t^*,t),y\right)}{q(y)}$$

in $D([t_0, 1-t_0]\times[0, y_0+\widetilde{\delta}])$ or, invoking a Skorohod construction again and putting $t_{\min} := \min(t,t^*)$,

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\in(0,y_0+\widetilde{\delta}]}} \frac{1}{q(y)}\left|\sqrt{k}\left(\frac{1}{kt_{\min}}\sum_{i=1}^{\lfloor nt\rfloor} I_{\{X_i > y^{-\gamma_{\text{pre}}} U_{\text{pre}}(n/k)\}} - y\right) - \frac{W_{\text{pre}}(t_{\min},y)}{t_{\min}}\right| \xrightarrow[(n\to\infty)]{\text{a.s.}} 0. \tag{69}$$

Then, similarly as for (57), it follows that

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\in[1/2,y_0]}} \left|\sqrt{k}\left(\left(\frac{X_k\left(0,t,\frac{t_{\min}}{t}y\right)}{U_{\text{pre}}(n/k)}\right)^{-1/\gamma_{\text{pre}}} - y\right) + \frac{W_{\text{pre}}(t_{\min},y)}{t_{\min}}\right| \xrightarrow[(n\to\infty)]{\text{a.s.}} 0,$$

which, for $y = t/t_{\min} \le (1-t_0)/t_0 = y_0$, implies

$$\sup_{t\in[t_0,1-t_0]} \left|\sqrt{k}\left(\left(\frac{X_k(0,t,1)}{U_{\text{pre}}(n/k)}\right)^{-1/\gamma_{\text{pre}}} - \frac{t}{t_{\min}}\right) + \frac{W_{\text{pre}}\left(t_{\min},\frac{t}{t_{\min}}\right)}{t_{\min}}\right| \xrightarrow[(n\to\infty)]{\text{a.s.}} 0. \tag{70}$$

Substituting $\left(\frac{X_k(0,t,1)}{U_{\text{pre}}(n/k)}y\right)^{-1/\gamma_{\text{pre}}}$ $(y\in[1-\widetilde{\delta},\infty))$ for y in (69) thus yields

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\ge 1-\widetilde{\delta}}} \frac{1}{q\left(\left(\frac{X_k(0,t,1)}{U_{\text{pre}}(n/k)}y\right)^{-1/\gamma_{\text{pre}}}\right)} \left|\sqrt{k}\left(\frac{1}{kt_{\min}}\sum_{i=1}^{\lfloor nt\rfloor} I_{\{X_i > yX_k(0,t,1)\}} - y^{-1/\gamma_{\text{pre}}}\left(\frac{X_k(0,t,1)}{U_{\text{pre}}(n/k)}\right)^{-1/\gamma_{\text{pre}}}\right)\right.$$
$$\left. - \frac{1}{t_{\min}} W_{\text{pre}}\left(t_{\min}, \left(\frac{X_k(0,t,1)}{U_{\text{pre}}(n/k)}y\right)^{-1/\gamma_{\text{pre}}}\right)\right| \xrightarrow[(n\to\infty)]{\text{a.s.}} 0. \tag{71}$$

As in the proof of Corollary 1 it follows from (70) and (71) that

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\ge 1}} \frac{1}{q\left(\left(y\frac{t}{t_{\min}}\right)^{-1/\gamma_{\text{pre}}}\right)} \left|\sqrt{k}\left(\frac{1}{kt_{\min}}\sum_{i=1}^{\lfloor nt\rfloor} I_{\{X_i > yX_k(0,t,1)\}} - y^{-1/\gamma_{\text{pre}}}\frac{t}{t_{\min}}\right)\right.$$
$$\left. - \frac{1}{t_{\min}}\left[W_{\text{pre}}\left(t_{\min}, y^{-1/\gamma_{\text{pre}}}\frac{t}{t_{\min}}\right) - y^{-1/\gamma_{\text{pre}}} W_{\text{pre}}\left(t_{\min},\frac{t}{t_{\min}}\right)\right]\right| \xrightarrow[(n\to\infty)]{\text{a.s.}} 0,$$

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

or, using $0 \le q\left(y^{-1/\gamma_{\text{pre}}}\right)/q\left(\left(y\frac{t}{t_{\min}}\right)^{-1/\gamma_{\text{pre}}}\right) \le q\left(y^{-1/\gamma_{\text{pre}}}\right)/q\left(\left(y\frac{1-t_0}{t_0}\right)^{-1/\gamma_{\text{pre}}}\right) \le K$ for y sufficiently large, because $y \mapsto q(y^{-1/\gamma_{\text{pre}}})$ is $RV_{-\nu/\gamma_{\text{pre}}}$ (recall (33)),

$$\sup_{\substack{t\in[t_0,1-t_0]\\ y\ge 1}} \frac{1}{q\left(y^{-1/\gamma_{\text{pre}}}\right)} \left| \sqrt{k}\left(\frac{1}{kt}\sum_{i=1}^{\lfloor nt\rfloor} I_{\{X_i > yX_k(0,t,1)\}} - y^{-1/\gamma_{\text{pre}}}\right) \right.$$
$$\left. - \frac{1}{t}\left[W_{\text{pre}}\left(t_{\min}, y^{-1/\gamma_{\text{pre}}}\frac{t}{t_{\min}}\right) - y^{-1/\gamma_{\text{pre}}} W_{\text{pre}}\left(t_{\min}, \frac{t}{t_{\min}}\right)\right]\right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0.$$

Using this result the convergence in (29) can be checked easily by following the derivations in Examples 3–5. ∎

Proof of Corollary 3. For (a) ((b) being proved similarly) Theorem 3 implies the convergence in (29), where by calculating covariances the following distributional equality holds

$$\left\{B_{\text{pre}}(t)\right\}_{t\in[0,t^*]} \stackrel{\mathcal{D}}{=} \left\{\sigma_{\widehat{\gamma},\gamma_{\text{pre}}} W(t)\right\}_{t\in[0,t^*]}, \tag{72}$$

where $\sigma^2_{\widehat{\gamma},\gamma_{\text{pre}}}$ is either (18), (20), or (22) with γ, $r(\cdot,\cdot)$ replaced by γ_{pre}, $r_{\text{pre}}(\cdot,\cdot)$ and $W(\cdot)$ is a standard Brownian motion. Write, using $\widetilde{t}_n = \frac{\lfloor nt_n\rfloor+1}{n}$ as defined in the proof of Theorem 2,

$$\int_{\widetilde{t}_n}^{1} k\left(\widehat{\gamma}(0,t) - \widehat{\gamma}(0,1)\right)^2 \mathrm{d}t$$
$$= \int_{\widetilde{t}_n}^{t^*} k\left(\widehat{\gamma}(0,t) - \widehat{\gamma}(0,1)\right)^2 \mathrm{d}t + \int_{t^*}^{1} k\left(\widehat{\gamma}(0,t) - \widehat{\gamma}(0,1)\right)^2 \mathrm{d}t =: A_n + B_n.$$

By following the steps leading to (63) we get from (29)

$$\frac{1}{\log(1/\widetilde{t}_n)} A_n = \frac{1}{\log(1/\widetilde{t}_n)} \int_{\widetilde{t}_n}^{t^*} \left(\frac{B_{\text{pre}}(t)}{t} - B_{\text{pre}}(1)\right)^2 \mathrm{d}t\,(1+o_P(1))$$
$$\stackrel{(72)}{=} \frac{1}{\log(1/\widetilde{t}_n)} \int_{\widetilde{t}_n}^{t^*} \left(\sigma_{\widehat{\gamma},\gamma_{\text{pre}}}\frac{W(t)}{t} - B_{\text{pre}}(1)\right)^2 \mathrm{d}t\,(1+o_P(1))$$
$$= \frac{1}{\log(1/\widetilde{t}_n)} \int_{\log(\widetilde{t}_n)}^{\log(t^*)} \left(\sigma_{\widehat{\gamma},\gamma_{\text{pre}}}\frac{W(e^y)}{e^{y/2}} - e^{y/2} B_{\text{pre}}(1)\right)^2 \mathrm{d}y\,(1+o_P(1)).$$

By slightly adapting the arguments in the proof of Theorem 2 this term converges in probability to $\sigma^2_{\widehat{\gamma},\gamma_{\text{pre}}}$. Furthermore $B_n/\log(1/\widetilde{t}_n) = \mathcal{O}_P(1)\log^{-1}(1/\widetilde{t}_n) = o_P(1)$ by (29). The result follows.

For the consistency results in (c) and (d) combine (a) and (b) with the conclusion of Theorem 3 to deduce for (e.g.) Q_{seq}

$$Q_{\text{seq}} \geq \frac{1}{\widehat{\sigma}^2_{\widehat{\gamma},\gamma}} \left\{ t^*(1-t^*)\sqrt{k}\left(\widehat{\gamma}\left(0,t^*\right) - \widehat{\gamma}\left(t^*,1\right)\right)\right\}^2$$
$$= \left(\frac{1}{\sigma^2_{\widehat{\gamma},\max(\gamma_{\text{pre}},\gamma_{\text{post}})}} + o_P(1) \right) k \Bigg\{ \underbrace{t^*(1-t^*)}_{>0} \Big(\underbrace{\widehat{\gamma}\left(0,t^*\right)}_{\xrightarrow{P} \gamma_{\text{pre}}} - \underbrace{\widehat{\gamma}\left(t^*,1\right)}_{\xrightarrow{P} \gamma_{\text{post}}} \Big) \Bigg\}^2 \xrightarrow[(n\to\infty)]{P} \infty.$$

Note that assumption $k = k_{\text{pre}} = k_{\text{post}}$ was needed to deduce via (the equivalents of) (17)

$$\widehat{\gamma}\left(0,t^*\right) \xrightarrow{P} \gamma_{\text{pre}} \quad \text{and} \quad \widehat{\gamma}\left(t^*,1\right) \xrightarrow{P} \gamma_{\text{post}},$$

where by definition of Q_{seq} both extreme value index estimators rely on the *same* sequence k.

For the inconsistency in (d) of the test based on Q_{rec} (the proof for $Q^{\leftarrow}_{\text{rec}}$ is similar and hence omitted) combine the result of Theorem 3 with the continuous mapping theorem and part (a) of Corollary 3 to deduce

$$Q_{\text{rec}} \xrightarrow[(n\to\infty)]{\mathcal{D}} \frac{1}{\sigma^2_{\widehat{\gamma},\gamma_{\text{pre}}}} \sup_{t\in[t_0,1-t_0]} \left\{ B_{\text{pre}}(t) - t B_{\text{pre}}(1) \right\}^2,$$

whence $Q_{\text{rec}} = \mathcal{O}_P(1)$, $n \to \infty$. ■

REFERENCES

Aguilar, M. & J. Hill (2015) Robust score and portmanteau tests of volatility spillover. *Journal of Econometrics* 184, 37–61.

Berkes, I., Horváth, L., & P. Kokoszka (2003) Estimation of the maximal moment exponent of a GARCH(1,1) sequence. *Econometric Theory* 19, 565–586.

Berkes, I., Horváth, L., & P. Kokoszka (2004) Testing for parameter constancy in GARCH(p,q) models. *Statistics & Probability Letters* 70, 263–273.

Bickel, P. & M. Wichura (1971) Convergence criteria for multiparameter stochastic processes and some applications. *The Annals of Statistics* 42, 1656–1670.

Billingsley, P. (1968) *Convergence of Probability Measures*. Wiley.

Bingham, N., Goldie, C., & J. Teugels (1987) *Regular Variation*. Cambridge Universtiy Press.

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

Chan, N., Li, D., Peng, L., & R. Zhang (2013) Tail index of an AR(1) model with ARCH(1) errors. *Econometric Theory* 29, 920–940.

Cramér, H. & M. Leadbetter (1967) *Stationary and Related Stochastic Processes*. Wiley.

Csörgő, S. & L. Viharos (1998) Estimating the tail index. In Szyszkowicz, B. (ed.), *Asymptotic Methods in Probability and Statistics*, pp. 833–881. Elsevier.

Daníelsson, J., Jansen, D., & C. de Vries (1996) The method of moments ratio estimator for the tail shape parameter. *Communications in Statistics - Theory and Methods* 25, 711–720.

de Haan, L. & A. Ferreira (2006) *Extreme Value Theory*. Springer.

https://doi.org/10.1017/S0266466616000189 Published online by Cambridge University Press

de Haan, L. & L. Peng (1998) Comparison of tail index estimators. *Statistica Neerlandica* 52, 60–70.

Drees, H. (1998a) A general class of estimators of the extreme value index. *Journal of Statistical Planning and Inference* 66, 95–112.

Drees, H. (1998b) On smooth statistical tail functionals. *Scandinavian Journal of Statistics* 24, 187–210.

Drees, H. (2000) Weighted approximations of tail processes for β-mixing random variables. *The Annals of Applied Probability* 10, 1274–1301.

Drees, H. (2003) Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli* 9, 617–657.

Drees, H. (2008) Some aspects of extreme value statistics under serial dependence. *Extremes* 11, 35–53.

Eberlein, E. (1984) Weak convergence of partial sums of absolutely regular sequences. *Statistics & Probability Letters* 2, 291–293.

Einmahl, J., de Haan, L., & C. Zhou (2016) Statistics of heteroscedastic extremes. *Journal of the Royal Statistical Society: Series B* 78, 31–51.

Einmahl, J., Gantner, M., & G. Sawitzki (2010) Asymptotics of the shorth plot. *Journal of Statistical Planning and Inference* 140, 3003–3012.

Galbraith, J. & S. Zernov (2004) Circuit breakers and the tail index of equity returns. *Journal of Financial Econometrics* 2, 109–129.

Hill, B. (1975) A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* 3, 1163–1174.

Hill, J. (2009) On functional central limit theorems for dependent, heterogeneous arrays with applications to tail index and tail dependence estimation. *Journal of Statistical Planning and Inference* 139, 2091–2110.

Kim, M. & S. Lee (2009) Test for tail index change in stationary time series with Pareto-type marginal distribution. *Bernoulli* 15, 325–356.

Kim, M. & S. Lee (2011) Change point test for tail index for dependent data. *Metrika* 74, 297–311.

Kim, M. & S. Lee (2012) Change point test of tail index for autoregressive processes. *Journal of the Korean Statistical Society* 41, 305–312.

Koedijk, K., Schafgans, M., & C. de Vries (1990) The tail index of exchange rate returns. *Journal of International Economics* 29, 93–108.

Lin, Z. & YK. Choi (1999) Some limit theorems for fractional Lévy Brownian fields. *Stochastic Processes and their Applications* 82, 229–244.

Móricz, F. (1982) A general moment inequality for the maximum of partial sums of single series. *Acta Scientiarum Mathematicarum* 44, 67–75.

Quintos, C., Fan, Z., & P. Phillips (2001) Structural change tests in tail behaviour and the Asian crisis. *Review of Economic Studies* 68, 633–663.

Resnick, S. (2007) *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer.

Rootzén, H. (2009) Weak convergence of the tail empirical process for dependent sequences. *Stochastic Processes and their Applications* 119, 468–490.

Shorack, G. & J. Wellner (1996) *Weak Convergence and Empirical Processes*. Springer.

Vogelsang, T. (1997) Wald-type tests for detecting breaks in the trend function of a dynamic time series. *Econometric Theory* 13, 818–849.

Vogelsang, T. (1999) Sources of nonmonotonic power when testing for a shift in mean of a dynamic time series. *Journal of Econometrics* 88, 283–299.

Wagner, N. & T. Marsh (2004) Tail index estimation in small samples: Simulation results for independent and ARCH-type financial return models. *Statistical Papers* 45, 545–562.

Werner, T. & C. Upper (2004) Time variation in the tail behavior of bund future returns. *Journal of Future Markets* 24, 387–398.

Wichura, M. (1970) On the construction of almost uniformly convergent random variables with given weakly convergent image laws. *The Annals of Mathematical Statistics* 41, 284–291.