Sequential monitoring of the tail behavior of dependent data

Yannick Hoga[*] Dominik Wied[†]

August 16, 2016

Abstract

We construct a sequential monitoring procedure for changes in the tail index and extreme quantiles of β-mixing random variables, which can be based on a large class of tail index estimators. The assumptions on the data are general enough to be satisfied in a wide range of applications. In a simulation study empirical sizes and power of the proposed tests are studied for linear and non-linear time series. Finally, we use our results to monitor Bank of America stock log-losses from 2007 to 2012 and detect changes in extreme quantiles without an accompanying detection of a tail index break.

Keywords: Sequential monitoring, change point, β-mixing, tail index, extreme quantiles, functional central limit theorem

JEL classification: C12 (Hypothesis Testing), C14 (Semiparametric and Nonparametric Methods), C22 (Time-Series Models)

1 Motivation

The tail index of a random variable is arguably one of the most important parameters of its distribution: It determines some fundamental properties like the existence of moments, tail asymptotics of the distribution and the asymptotic behavior of sums and maxima. As a measure of tail thickness, the tail index is used in fields where heavy tails are frequently encountered, such as (re)insurance, finance, and teletraffic engineering (cf. Resnick, 2007, Sec. 1.3, and the references cited therein). Particularly in finance, the closely related extreme quantiles play a prominent role as a risk measure called Value-at-Risk (VaR).

The use of the variance as a risk measure has a long tradition in finance. Under Gaussianity the variance completely determines the tails of the distribution, which is no longer the case with heavy-tailed data. Hence, in order to assess the tail behavior of a time series, practitioners often estimate the tail index or an extreme quantile, the implicit assumption being their constancy over time. There are several suggestions in the literature on how to test this crucial assumption: Quintos, Fan and

[*]Faculty of Economics and Business Administration, University of Duisburg-Essen, Universitätsstraße 12, D-45117 Essen, Germany, tel. +49 201 1834365, `yannick.hoga@vwl.uni-due.de`. The author gratefully acknowledges the support of DFG (HA 6766/2-1).

[†]University of Cologne and TU Dortmund, Germany, tel. +49 221 4704514, `wied@statistik.tu-dortmund.de`. Financial support by Deutsche Forschungsgemeinschaft (SFB 823) is gratefully acknowledged.

The authors would like to thank Stefan Fremdt for helpful discussion and Christoph Hanck for carefully reading an earlier version of this manuscript.

Phillips (2001) developed so called recursive, rolling and sequential tests for independent and GARCH data for tail index constancy based on the Hill (1975) estimator. Kim and Lee (2011) investigated their tests for more general β-mixing time series. Taking a likelihood approach for independent data, Dierckx and Teugels (2010) focus on breaks in the tail index for environmental data. Tests based on other estimators than the Hill (1975) estimator were first proposed by Einmahl, de Haan and Zhou (2016) for independent and Hoga (2016+) for dependent data. To the best of our knowledge the only paper dealing with changes in extreme quantiles is Hoga (2015). All these tests are of a retrospective nature.

We are not aware of any work on online surveillance methods for constancy of the tail index and extreme quantiles. This is important because, as noted in Chu, Stinchcombe and White (1996), '[b]reaks can occur at any point, and given the costs of failing to detect them, it is desirable to detect them as rapidly as possible. One-shot tests cannot be applied in the usual way each time new data arrive, because repeated application of such tests yields a procedure that rejects a true null hypothesis of no change with probability one as the number of applications grows.' This paper will fill this gap for closed-end procedures. To allow for sufficient flexibility in the use of tail index estimators, we will use the approach of Hoga (2016+).

Whether a monitoring procedure for a change in the tail index or an extreme quantile is of interest will largely be a matter of context. If interest centers on VaR, which is widely used in the banking industry and by financial regulators as a risk measure, the quantile monitoring procedure will be more relevant. If however interest centers on the mean excess function of the (log-transformed) data X, then, since $\mathrm{E}\left(\log X - \log t | X > t\right)$ converges to the extreme value index of X as $t \to \infty$, the tail index alternative seems more appropriate. Furthermore, the tail index per se could also be of interest as there are indications that it has predictive power for stock returns (Kelly and Jiang, 2014), where higher (lower) tail indices of returns indicate higher (lower) absolute returns.

The outline of this paper is as follows. The main results under the null and two alternatives are stated in Section 2, where an example of a time series satisfying our assumptions is also given. Simulations and an empirical application are presented in Sections 3 and 4 respectively. All proofs are collected in an appendix.

2 Main results

2.1 Preliminaries and assumptions

To introduce the required notation let $X_1, \ldots, X_n$ be a sequence of random variables defined on some probability space $(\Omega, \mathcal{A}, P)$ with survivor function $\bar{F}_i(x) := 1 - F_i(x) = P(X_i > x)$, that is regularly varying with parameter $-\alpha_i$ (written $\bar{F}_i \in RV_{-\alpha_i}$), i.e.,

$$\bar{F}_i(x) = x^{-\alpha_i} L_i(x), \quad x > 0, \tag{1}$$

where $L_i : (0, \infty) \to (0, \infty)$ is slowly varying, i.e.,

$$\lim_{x \to \infty} \frac{L_i(\lambda x)}{L_i(x)} = 1 \quad \forall\, \lambda > 0. \tag{2}$$

If X_i is Pareto distributed, then $L_i(x) \equiv c > 0$. Since slow variation of the function $L_i(x)$ means, loosely speaking, that it behaves like a constant function at infinity, we say that X_i with tails as in (1) has *Pareto-type tails.* In the context of extreme value theory, α_i is called the tail index and $\gamma_i := 1/\alpha_i$ the extreme value index (Resnick, 2007, Sec. 4.5.1).

Define

$$U_i(x) := F_i^{-1}\left(1 - \frac{1}{x}\right), \quad x > 1,$$

as the $(1 - 1/x)$-quantile, F_i^{-1} being the left-continuous inverse of F_i. Then, recall that (1) is equivalent to

$$\frac{U_i(\lambda x)}{U_i(x)} \underset{(x \to \infty)}{\longrightarrow} \lambda^{\gamma} \tag{3}$$

(e.g., Resnick, 2007, Prop. 2.6 (v)). Throughout, $k = k_n \in \mathbb{N}$ will denote a sequence satisfying $k \leq n - 1$,

$$k \underset{(n \to \infty)}{\longrightarrow} \infty \quad \text{and} \quad \frac{k}{n} \underset{(n \to \infty)}{\longrightarrow} 0, \tag{4}$$

controlling the number of upper order statistics used in the estimation of the tail index and $p = p_n \to 0$, $n \to \infty$, will denote a sequence of small probabilities, for which we want to test for a change in an appertaining extreme (right-tail) quantile $U_i(1/p)$. As is customary in extreme value theory, we will usually drop the subindex n and simply write k and p. For $t - s \geq 1/n$ and $y \in [0, 1]$ set

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

Under the assumption of strictly stationary X_i we write $\bar{F} = \bar{F}_i$ and $U = U_i$. Let

$$\widehat{\gamma}(s, t) := \widehat{\gamma}_n(s, t), \qquad 0 \leq s < t < \infty,\ t - s \geq 1/n,$$

denote a generic tail index estimator based on the $(\lfloor k(t - s) \rfloor + 1)$-largest order statistics of the subsample $X_{\lfloor ns \rfloor + 1}, \ldots, X_{\lfloor nt \rfloor}$. Then approximate with (3) for $x = 1/p$, $\lambda = pn/k$ with small $p > 0$ to get

$$\begin{aligned} U\left(\frac{1}{p}\right) &\approx U\left(\frac{n}{k}\right)\left(\frac{pn}{k}\right)^{-\gamma} \\ &\approx X_k(s, t, 1)\left(\frac{np}{k}\right)^{-\widehat{\gamma}(s,t)} =: \widehat{x}_p(s, t), \end{aligned} \tag{5}$$

which motivates and defines the so called Weissman (1978) estimator for the extreme quantile $U(1/p)$. Hence, the idea is to use the (within sample range) estimator $X_k(s, t, 1)$ of $U(n/k)$ to estimate the (possibly out of sample range) quantile $U(1/p)$ by exploiting the regular variation of the quantile function in (3). In view of this we will require $p \ll k/n$.

For concreteness we will focus in the following on the Hill (1975) estimator of γ given by

$$\widehat{\gamma}_H(0,1) := \frac{1}{k}\sum_{i=0}^{k-1}\log\left(\frac{X_{n-i:n}}{X_{n-k:n}}\right), \tag{6}$$

where $X_{n:n} \geq X_{n-1:n} \geq \ldots \geq X_{1:n}$ denote the order statistics of the sample $X_1,\ldots,X_n$. But the main results in Theorems 1 and 2 below hold without modification for the moments-ratio estimator of Daníelsson *et al.* (1996) and the class of estimators introduced by Csörgő and Viharos (1998); see also the proof of Theorem 1 below.

The dependence concept we will use in the following is that of β-mixing, that is for a sequence of random variables $\{X_i\}_{i\in\mathbb{N}}$ the β-mixing coefficients $\beta(l)$ converge to zero:

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

where $\mathcal{F}_m^{\infty} := \sigma(X_m, X_{m+1},\ldots)$ and $\mathcal{F}_l^m := \sigma(X_l,\ldots,X_m)$. For more on this mixing concept we refer to Bradley (1986).

Write $D[a,b]$ for the space of cadlag functions on $[a,b]$ $(0\leq a<b<\infty)$ endowed with the Skorohod topology and $D(I)$, $I\subset\mathbb{R}^2$ compact, for the two-parameter extension.

In order to construct a monitoring procedure for a tail index change, we have to assume tail index (or, equivalently, extreme value index) constancy over some historical period (also called training period) of suitable length n:

(A1) $\gamma_1 = \ldots = \gamma_n, \qquad n\in\mathbb{N}.$

This assumption can of course be tested by any of the retrospective change point tests proposed in, e.g., Hoga (2016+, 2015).

As soon as a period $X_1,\ldots,X_n$ of tail index or extreme quantile stability is identified and more observations $X_{n+1}, X_{n+2},\ldots$ become available, we are interested in an online surveillance method testing

$$\begin{aligned} \mathcal{H}_{0,\gamma}:&\quad \gamma_1 = \ldots = \gamma_n = \gamma_{n+1} = \ldots \quad \text{vs.}\\ \mathcal{H}_{1,\gamma}^{\lesseqgtr}:&\quad \gamma_1 = \ldots = \gamma_n = \ldots = \gamma_{\lfloor nt^*\rfloor} \lesseqgtr \gamma_{\lfloor nt^*\rfloor+1} = \gamma_{\lfloor nt^*\rfloor+2} = \ldots, \end{aligned} \tag{7}$$

and

$$\begin{aligned} \mathcal{H}_{0,U}:&\quad U_1\left(\frac{1}{p}\right) = \ldots = U_n\left(\frac{1}{p}\right) = U_{n+1}\left(\frac{1}{p}\right) = \ldots \quad \text{vs.}\\ \mathcal{H}_{1,U}^{\lesseqgtr}:&\quad U_1\left(\frac{1}{p}\right) = \ldots = U_n\left(\frac{1}{p}\right) = \ldots = U_{\lfloor nt^*\rfloor}\left(\frac{1}{p}\right) \lesseqgtr U_{\lfloor nt^*\rfloor+1}\left(\frac{1}{p}\right) = U_{\lfloor nt^*\rfloor+2}\left(\frac{1}{p}\right) = \ldots \end{aligned} \tag{8}$$

for some $t^*\geq 1$ denoting the unknown change point. We use $\mathcal{H}_0$ or $\mathcal{H}_1^{\lesseqgtr}$ as shorthand notation for both of $\mathcal{H}_{0,\gamma}$, $\mathcal{H}_{0,U}$ or $\mathcal{H}_{1,\gamma}^{\lesseqgtr}$, $\mathcal{H}_{1,U}^{\lesseqgtr}$.

We use the following detectors for (7)

$$V_n^{\widehat{\gamma}}(t) := \frac{\left[(t-1)\left(\widehat{\gamma}(1,t)-\widehat{\gamma}(0,1)\right)\right]^2}{\int_{t_0}^1 \left[s\left(\widehat{\gamma}(0,s)-\widehat{\gamma}(0,1)\right)\right]^2 \mathrm{d}s}, \qquad t \geq 1+t_0,$$

$$W_n^{\widehat{\gamma}}(t) := \frac{\left[t_0\left(\widehat{\gamma}(t-t_0,t)-\widehat{\gamma}(0,1)\right)\right]^2}{\int_{t_0}^1 \left[t_0\left(\widehat{\gamma}(s-t_0,s)-\widehat{\gamma}(0,1)\right)\right]^2 \mathrm{d}s}, \qquad t \geq 1+t_0,$$

and for (8)

$$V_n^{\widehat{x}_p}(t) := \frac{\left[(t-1)\log\left(\frac{\widehat{x}_p(1,t)}{\widehat{x}_p(0,1)}\right)\right]^2}{\int_{t_0}^1 \left[s\log\left(\frac{\widehat{x}_p(0,s)}{\widehat{x}_p(0,1)}\right)\right]^2 \mathrm{d}s}, \qquad t \geq 1+t_0,$$

$$W_n^{\widehat{x}_p}(t) := \frac{\left[t_0\log\left(\frac{\widehat{x}_p(t-t_0,t)}{\widehat{x}_p(0,1)}\right)\right]^2}{\int_{t_0}^1 \left[t_0\log\left(\frac{\widehat{x}_p(s-t_0,s)}{\widehat{x}_p(0,1)}\right)\right]^2 \mathrm{d}s}, \qquad t \geq 1+t_0,$$

where $t_0 > 0$ defines the (minimal) fraction of n upon which the tail index and extreme quantile estimators are based. To motivate our detectors consider $V_n^{\widehat{\gamma}}$, the others can be motivated similarly. In the numerator the training period estimate $\widehat{\gamma}(0,1)$ is compared with the current observation period estimate $\widehat{\gamma}(1,t)$. If the observation period length $(t-1)$ is large, that difference is weighted more heavily. The denominator 'self-normalizes' the numerator. While we could have chosen a wide range of functionals for this (e.g., the denominator of $W_n^{\widehat{\gamma}}$), it seemed more natural to incorporate the functional form of the numerator to do so. With this motivation in mind we are inclined to reject $\mathcal{H}_0$ if the following stopping times terminate (in the sense of being finite):

$$\tau_n^{V^{\widehat{\gamma}}} := \inf\left\{t \in [1+t_0, T] \ : \ V_n^{\widehat{\gamma}}(t) > c\right\},$$
$$\tau_n^{W^{\widehat{\gamma}}} := \inf\left\{t \in [1+t_0, T] \ : \ W_n^{\widehat{\gamma}}(t) > c\right\},$$

and

$$\tau_n^{V^{\widehat{x}_p}} := \inf\left\{t \in [1+t_0, T] \ : \ V_n^{\widehat{x}_p}(t) > c\right\},$$
$$\tau_n^{W^{\widehat{x}_p}} := \inf\left\{t \in [1+t_0, T] \ : \ W_n^{\widehat{x}_p}(t) > c\right\},$$

where from now on $\inf\emptyset := \infty$ and $c > 0$ is chosen, such that under $\mathcal{H}_0$, $\lim_{n\to\infty} P(\tau_n < \infty) = \alpha$ for some prespecified significance level $\alpha \in (0,1)$ (see Theorem 1 below). Here $T > 1$ denotes the arbitrarily large closed end of the procedure, i.e., the method terminates after observations $X_{n+1}, \ldots, X_{\lfloor nT \rfloor}$. Closed-end procedures are quite common, e.g., Aue *et al.* (2012) consider breaks in portfolio betas, Wied and Galeano (2013) breaks in cross-correlations, Zeileis *et al.* (2005) and Aschersleben *et al.* (2015) breaks in regression and cointegrating relationships respectively.

Remark 1. (a) The detector $V_n^{\widehat{\gamma}}$ comes closer in spirit to many of the detectors in the online monitoring literature, where an estimate of some parameter based on the historical period is compared to that based on the (ever longer) monitoring period; see the references in the paragraph before. However, the procedure based on $V_n^{\widehat{\gamma}}$ is not consistent against $\mathcal{H}_{1,\gamma}^{>}$, cf. Theorem 2 below, which is the reason for introducing the method based on $W_n^{\widehat{\gamma}}$.

(b) We could have based our procedure equally well on detectors of the type

$$V_n^{\widehat{\gamma}}(t) := \frac{1}{\sigma^2_{\widehat{\gamma},\gamma}}\left[(t-1)\sqrt{k}\left(\widehat{\gamma}(1,t)-\widehat{\gamma}(0,1)\right)\right]^2,$$

where $\sigma^2_{\widehat{\gamma},\gamma}$ is a consistent estimator of the asymptotic variance of $\sqrt{k}(\widehat{\gamma}(0,1)-\gamma)$ based on the observations $X_1,\ldots,X_n$ in the observation period (e.g., the one in Theorem 2 of Hoga (2016+)). It turned out however, that in simulations values of even $n=500$ for the training period were not sufficient to deliver accurate variance estimates for a wide range of model parameters for the models we investigated, which lead to severe size distortions of our surveillance methods. This is why we opted for the self-normalized approach advocated in Shao and Zhang (2010) in our sequential setting. To the best of our knowledge we are the first to do so. Shao and Zhang (2010) found that for retrospective change point tests self-normalized test statistics delivers far superior size in simulations. However, the price to be paid for using a self-normalization approach vs. a variance estimation approach is slightly lower power.

Under $\mathcal{H}_0$ we will assume beyond **(A1)** that:

(A2) $\{X_i\}_{i\in\mathbb{N}}$ is a strictly stationary β-mixing process with continuous marginals and mixing coefficients $\beta(\cdot)$, such that

$$\lim_{n\to\infty}\frac{n}{r_n}\beta(l_n)+\frac{r_n}{\sqrt{k}}\log^2(k)=0 \tag{9}$$

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

(A3) There exists a function $r=r(x,y)$, such that for all $x,y\in[0,y_0+\varepsilon]$ $(y_0\geq 1,\varepsilon>0)$

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

(A4) For some constant $C>0$

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

(A5) There exist $\rho<0$ and a function $A(\cdot)$ with eventually positive or negative sign, $\lim_{t\to\infty}A(t)=0$, such that for any $y>0$

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

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

For the detectors for changes in extreme quantiles we need the following further assumptions:

(A6) $\lim_{n\to\infty} \frac{np}{k} = 0$, $\lim_{n\to\infty} k^{-1/2} \log(np) = 0$.

(A7) The sequence k satisfies

$$\frac{U(1/p)}{U(n/k)} \left(\frac{np}{k}\right)^{\gamma} - 1 = o\left(\frac{1}{\sqrt{k}}\right).$$

The conditions **(A2)**-**(A4)** correspond (almost) exactly to conditions $(\widetilde{\text{C1}})$, $(\widetilde{\text{C2}})$ and $(\widetilde{\text{C3*}})$ in Drees (2000). Condition **(A5)**, which is a widely used second-order strengthening of (3) (e.g., Kim and Lee, 2011; Einmahl *et al.*, 2016), is stronger than Drees's (2000) condition (3.5). **(A2)** ensures a standard 'big block - small block' approach may be applied to deduce weak convergence of what Einmahl *et al.* (2016, p. 42) termed the *simple sequential tail empirical process* in (A.1) below. The limit process has a well-defined covariance structure by virtue of **(A3)**. **(A6)** provides a range for p: $\lim_{n\to\infty} \frac{np}{k} = 0$ provides an upper bound for the decay of p (indicating the limitations of the extreme value theory approach towards the center of the distribution) while $\lim_{n\to\infty} k^{-1/2} \log\left(np/k\right) = \lim_{n\to\infty} k^{-1/2} \log(np) = 0$ provides a lower bound, beyond which estimation is no longer feasible. Note that if the d.f. obeys the quite general expansion

$$1 - F(x) = Cx^{-\alpha}(1 + \mathcal{O}(x^{-\beta})) \quad \text{as} \quad x \to \infty; \quad C, \alpha, \beta > 0,$$

then by inversion $U(x) = C^{1/\alpha} x^{1/\alpha}(1 + \mathcal{O}(x^{-\beta/\alpha}))$, whence **(A7)** does not impose an additional constraint on the choice of k.

2.2 Results under the null and the alternative

We are now ready to state the main results under the null, describing the asymptotic behavior of our monitoring procedures based on the stopping rules τ_n.

Theorem 1. *Suppose* **(A1)**-**(A5)** *hold for* $y_0 = 1$. *Then for any* $t_0 > 0$, $T > 1 + t_0$ *and*

$$\begin{aligned} V_{t_0,T} &:= \frac{\sup_{t\in[1+t_0,T]} \left[W(t) - tW(1)\right]^2}{\int_{t_0}^{1} \left[W(s) - sW(1)\right]^2 \mathrm{d}s}, \\ W_{t_0,T} &:= \frac{\sup_{t\in[1+t_0,T]} \left[W(t) - W(t-t_0) - t_0 W(1)\right]^2}{\int_{t_0}^{1} \left[W(s) - W(s-t_0) - t_0 W(1)\right]^2 \mathrm{d}s}, \end{aligned} \tag{10}$$

with $\left\{W(t)\right\}_{t\in[0,T]}$ *a standard Brownian motion,*

(i) under $\mathcal{H}_{0,\gamma}$

$$\lim_{n\to\infty} P\left(\tau_n^{V^{\hat\gamma}} < \infty\right) = P\left(V_{t_0,T} > c\right),$$
$$\lim_{n\to\infty} P\left(\tau_n^{W^{\hat\gamma}} < \infty\right) = P\left(W_{t_0,T} > c\right),$$

(ii) under $\mathcal{H}_{0,U}$ and additionally **(A6)**-**(A7)**

$$\lim_{n\to\infty} P\left(\tau_n^{V^{\widehat{x}_p}} < \infty\right) = P\left(V_{t_0,T} > c\right),$$
$$\lim_{n\to\infty} P\left(\tau_n^{W^{\widehat{x}_p}} < \infty\right) = P\left(W_{t_0,T} > c\right).$$

Next, we investigate the behavior of our procedures under the 'one-sided' alternatives $\mathcal{H}_1^{\lessgtr}$. To prove our results the observations will be denoted by the triangular array of random variables $X_{n,i}$, $n \in \mathbb{N}$, $i = 1, \ldots, n$, which have a common marginal survivor function $\overline{F}_{\text{pre}} \in RV_{-\alpha_{\text{pre}}}$ ($\overline{F}_{\text{post}} \in RV_{-\alpha_{\text{post}}}$) for $i \in I_{\text{pre}} := \{1, \ldots, \lfloor nt^* \rfloor\}$ ($i \in I_{\text{post}} := \{\lfloor nt^* \rfloor + 1, \ldots, \lfloor nT \rfloor\}$). Set

$$U_{\text{pre}}(x) = F_{\text{pre}}^{-1}\left(1 - \frac{1}{x}\right) \quad \text{and} \quad U_{\text{post}}(x) = F_{\text{post}}^{-1}\left(1 - \frac{1}{x}\right).$$

Theorem 2. *Let the triangular array of observations $\{X_{n,i}\}_{n\in\mathbb{N}, i=1,\ldots,n}$ be given by*

$$X_{n,i} := \begin{cases} Y_n, & i \in I_{\text{pre}}, \\ Z_n, & i \in I_{\text{post}}, \end{cases}$$

where $\{Y_n\}_{n\in\mathbb{N}}$ and $\{Z_n\}_{n\in\mathbb{N}}$ both satisfy conditions **(A2)**-**(A5)** *with $y_0 = T$ and*

$$k, \gamma_{\text{pre}}, U_{\text{pre}}(\cdot), r_{\text{pre}}(\cdot,\cdot) \qquad \text{and} \qquad k, \gamma_{\text{post}}, U_{\text{post}}(\cdot), r_{\text{post}}(\cdot,\cdot)$$

respectively. Then

(i)

$$\lim_{n\to\infty} P\left(\tau_n^{V^{\widehat{\gamma}}} < \infty\right) = 1 \qquad \text{under } \mathcal{H}_{1,\gamma}^{<},$$
$$\lim_{n\to\infty} P\left(\tau_n^{V^{\widehat{\gamma}}} < \infty\right) < 1 \qquad \text{under } \mathcal{H}_{1,\gamma}^{>},$$

(ii)

$$\lim_{n\to\infty} P\left(\tau_n^{W^{\widehat{\gamma}}} < \infty\right) = 1 \qquad \text{under } \mathcal{H}_{1,\gamma}^{\lessgtr},$$

where under $\mathcal{H}_1^{>}$ additionally $t^* \in [1, T - t_0]$ must hold.

Suppose that additionally **(A6)**-**(A7)** *hold for $\{Y_n\}$ and $\{Z_n\}$. Then*

(iii)

$$\lim_{n\to\infty} P\left(\tau_n^{V^{\widehat{x}_p}} < \infty\right) = 1 \qquad \text{under } \mathcal{H}_{1,\gamma}^{\lessgtr},$$

(iv) if $t^* \in [1, T - t_0]$ and $|\sqrt{k}/\log(k/(np))\log(U_{\text{pre}}(1/p)/U_{\text{post}}(1/p))| \underset{(n\to\infty)}{\longrightarrow} \infty$,

$$\lim_{n\to\infty} P\left(\tau_n^{W^{\widehat{x}_p}} < \infty\right) = 1 \qquad \text{under } \mathcal{H}_{1,U}^{\lessgtr}.$$

Remark 2. (a) Note that the sequence k in the pre- and post-break period must be the same. This is however not too restrictive as (see (11) below) one can frequently choose $k = n^{\nu}$ for some arbitrarily small $\nu > 0$. So if **(A2)**-**(A7)** are satisfied for $k = n^{\nu}$ with $\nu \in (0, \nu_{\text{pre}})$ for the pre-break period and with $\nu \in (0, \nu_{\text{post}})$ in the post-break period, then taking $\nu \in (0, \min(\nu_{\text{pre}}, \nu_{\text{post}}))$ leads to a sequence for which all assumptions are satisfied for the whole sample.

(b) If the $X_{n,i}$ are generated as in Theorem 2, the hypothesis $\mathcal{H}_{1,\gamma}^{\lessgtr}$ is a strict subset of the hypothesis $\mathcal{H}_{1,U}^{\lessgtr}$. E.g., taking $Z_n = aY_n$, $a \neq 1$, is covered under $\mathcal{H}_{1,U}^{\lessgtr}$, but not under $\mathcal{H}_{1,\gamma}^{\lessgtr}$, since scaling does not affect the tail index. If however (e.g.) $\mathcal{H}_{1,\gamma}^{>}$ is true, we have $U_{\text{pre}}/U_{\text{post}} \in RV_{\gamma_{\text{pre}}-\gamma_{\text{post}}>0}$ and hence

$$\frac{U_{\text{pre}}(1/p_n)}{U_{\text{post}}(1/p_n)} \underset{(n\to\infty)}{\longrightarrow} \infty,$$

s.t. $\mathcal{H}_{1,U}^{>}$ is true.

(c) Under $\mathcal{H}_1^{>}$ the procedure based on $\tau_n^{V\widehat{\gamma}}$ is not consistent, which motivated the study of $\tau_n^{W\widehat{\gamma}}$. The reason for the inconsistency is, simply speaking, that in a sample with one extreme value index break, extreme value index estimators will consistently estimate the larger extreme value index. Hence, if there is a break in t^* toward lighter tails in the observation period, then $\widehat{\gamma}(1,t)$, $t > t^*$, will still estimate the larger extreme value index, even though the last part of the sample upon which it is based possess a smaller extreme value index. Thus, the change goes unnoticed.

2.3 An Example

In this subsection we verify the conditions **(A1)**-**(A4)** for the following stochastic volatility model

$$X_i = \sigma_i Z_i, \qquad i \in \mathbb{Z},$$

where $\{Z_i\}$ are i.i.d. and independent from $\{\sigma_i\}$. Denote by $1 - F_{|Z|} \in RV_{-\alpha}$, $\alpha > 0$, the survivor function of $|Z_0|$. The volatility process is assumed to be generated according to

$$\sigma_i = \exp(Y_i), \qquad \text{where } Y_i = \sum_{j=0}^{\infty} \psi_j \epsilon_{i-j},$$

with (w.l.o.g.) $\psi_0 = 1$, $\epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ and geometrically decreasing coefficients $|\psi_j| \underset{(j\to\infty)}{=} \mathcal{O}(\eta^j)$, $\eta \in (0,1)$, covering all finite order ARMA(p,q)-models. Many popular stochastic volatility models use (zero-mean) Gaussian AR(1)-models for the volatility dynamics (Asai *et al.*, 2006).

We now verify our conditions for $|X_i| = \sigma_i |Z_i|$. For **(A2)** strict stationarity is immediate (which also implies **(A1)**). By Bradley (1986, Ex. 6.1) the Y_i are geometrically β-mixing, whence, by Gaussianity of the Y_i, they are also geometrically ρ-mixing (Bradley, 1986, Eq. (1.7) & Thm. 5.1 and the comments below it), i.e., the ρ-mixing coefficients

$$\rho(j) = \sup_{\substack{U \in \mathcal{L}_2(\mathcal{B}_{-\infty}^0) \\ V \in \mathcal{L}_2(\mathcal{B}_j^{\infty})}} \left|\text{corr}(U,V)\right| \underset{(j\to\infty)}{\longrightarrow} 0$$

decay to zero geometrically fast. Here, $\mathcal{B}_{-\infty}^0 = \sigma(\ldots, Y_{-1}, Y_0)$, $\mathcal{B}_j^\infty = \sigma(Y_j, Y_{j+1}, \ldots)$ and $\mathcal{L}_2(\mathcal{A})$ is the space of square-integrable, $\mathcal{A}$-measurable real-valued functions. By Bradley (1986, p. 170) this implies geometric ρ-mixing of $\sigma_i = \exp(Y_i)$. This in turn implies geometric ρ-mixing of $X_i = \sigma_i Z_i$ since (by independence of the Z_i) the ρ-mixing coefficients of X_i are bounded by those of σ_i, such that geometric ρ-mixing is inherited from the volatility process (see Bradley, 2007, Thm. 6.6). We conclude that $\rho(j) \leq K\eta^j$ for some $\eta \in (0,1)$ for the mixing coefficients $\rho(\cdot)$ appertaining to the $|X_i|$. By Bradley (2007, Thm. 6.6) again, independence of the Z_i and recalling that the Y_i (and hence the σ_i) are geometrically β-mixing, the same holds true for the β-mixing coefficients of the $|X_i|$. Thus, **(A2)** is satisfied for the following choices

$$k = n^\nu \text{ for } \nu \in (0, 3/4), \quad l_n = -2\frac{\log n}{\log \eta}, \quad r_n = n^{\nu/3}. \tag{11}$$

We check conditions (C2) and (C3) of Drees (2003, Prop. 2.1), which implies **(A3)** because with the above choices $r_n k/n = o(1)$. For (C2) we get from Hill (2011, Thm. 2.1) that as $n \to \infty$

$$\begin{aligned} P(|X_1| > xU(n/k), |X_{1+m}| > yU(n/k)) &\sim \frac{\mathrm{E}[\sigma_1^\alpha \sigma_{1+m}^\alpha]}{\mathrm{E}[\sigma_1^\alpha]\,\mathrm{E}[\sigma_{1+m}^\alpha]} P(|X_1| > xU(n/k)) P(|X_{1+m}| > yU(n/k)) \\ &\sim \frac{\mathrm{E}[\sigma_1^\alpha \sigma_{1+m}^\alpha]}{\mathrm{E}[\sigma_1^\alpha]\,\mathrm{E}[\sigma_{1+m}^\alpha]} \frac{k}{n} x^{-\alpha} \frac{k}{n} y^{-\alpha}, \end{aligned}$$

where the last line follows from (12) below and (e.g.) Resnick (2007, Sec. 2.2.1). (C3) is satisfied due to geometric ρ-mixing of $|X_i|$ (see also Drees, 2003, Rem. 2.2). Assumption **(A4)** is again a consequence of Drees (2003, Prop. 2.1 & Rem. 2.3).

The $|X_i|$ inherit their heavy tails from the Z_i as, by Breiman's (1965) lemma,

$$1 - F(x) = P(|X_0| > x) \underset{(x \to \infty)}{\sim} \mathrm{E}[\sigma_1^\alpha] P(|Z_0| > x). \tag{12}$$

Hence, the $|X_i|$ also have tail index α and (1) is satisfied. Of course, (12) only gives $1 - F(x) = cx^{-\alpha}(1 + o(1))$. This is weaker than

$$\lim_{t \to \infty} \frac{\frac{1-F(ty)}{1-F(t)} - y^{-\alpha}}{A(1/[1 - F(t)])} = y^{-\alpha} \frac{y^{\rho\alpha} - 1}{\rho/\alpha}, \tag{13}$$

with $\sqrt{k}A(n/k) \to 0$ as $n \to \infty$, which is equivalent to **(A5)** (cf. de Haan and Ferreira, 2006, Thm. 2.3.9). The currently sharpest result on the second-order behavior of the d.f.s of stochastic volatility models seems to be Kulik and Soulier (2011, Prop. 2.8). Assume

$$1 - F_{|Z|}(x) = c_z x^{-\alpha} \exp\left(\int_1^x \frac{\eta(s)}{s} \mathrm{d}s\right), \qquad x > 0,\ c > 0$$

for some $\eta(s) = \mathcal{O}(s^{\alpha\rho})$, $\rho < 0$. For instance Fréchet-, $|t_\nu|$- and generalized Pareto distributions have such tails (cf. Beirlant *et al.*, 2004, Sec. 2.3.4). Then the aforementioned proposition implies

$$1 - F(x) = cx^{-\alpha}(1 + \mathcal{O}(x^{\alpha\rho})), \qquad x > 0,\ c > 0, \tag{14}$$

which is stronger than $1 - F(x) = cx^{-\alpha}(1 + o(1))$ from (12), but not quite sufficient for (13). However, if the $\mathcal{O}$-term in (14) satisfied an expansion $c_1 x^{\alpha\rho}(1 + o(1))$, then (13) would be satisfied for $k = o(n^{2\rho/(2\rho-1)})$ (cf. de Haan and Ferreira, 2006, p. 77). Recall from the discussion of conditions **(A2)**-**(A7)** that (14) implies **(A7)**.

3 Simulations

In this section we investigate the finite-sample behavior of the monitoring procedures based on the stopping times $\tau_n^{W^{\widehat{\gamma}/\widehat{x}_p}}$ and $\tau_n^{V^{\widehat{\gamma}/\widehat{x}_p}}$. Throughout we simulate 10,000 time series with historical periods of length $n = 125, 500$ and $T = 4$, such that the total length is $\lfloor nT \rfloor = 500, 2000$. Furthermore we use $t_0 = 0.2$ and $k/n = 0.2$ and the estimator $\widehat{\gamma} = \widehat{\gamma}_H$ of the extreme value index we employ is the Hill (1975) estimator. Simulation results were quite robust to the particular choice of k/n and are available from the authors upon request. The quantiles of the distributions of $V_{t_0,T}$ and $W_{t_0,T}$ from (10) are tabulated in Table 1 for $t_0 = 0.2$ and $T = 4$. To simulate them we used 100,000 realizations of Brownian motions on the interval $[0, 4]$, which themselves were generated using 400,000 normally distributed random variables.

α_q	0.50	0.60	0.70	0.80	0.90	0.95	0.99
α_q-quantile of $V_{t_0,T}$	78.35	113.0	166.2	257.8	464.5	723.4	1557
α_q-quantile of $W_{t_0,T}$	15.57	18.42	22.10	27.39	36.79	46.87	72.89

Table 1: Quantiles of $V_{t_0,T}$ and $W_{t_0,T}$ ($t_0 = 0.2$, $T = 4$)

X	DGP	Size	$\lfloor nT \rfloor = 500$				$\lfloor nT \rfloor = 2000$			
			$X^{\widehat{\gamma}}$	$X^{\widehat{x}_p}$			$X^{\widehat{\gamma}}$	$X^{\widehat{x}_p}$		
				$p = 0.1$	0.01	0.001		$p = 0.1$	0.01	0.001
V	(ARMA)	10	5.4	11	7.7	6.0	7.1	10	8.7	7.9
		5	2.7	6.2	4.0	3.0	3.6	5.5	4.5	3.7
	(ARCH)	10	5.9	12	9.0	6.9	7.5	11	9.0	8.0
		5	2.8	6.6	4.6	3.5	3.8	5.5	4.5	4.1
	(SV)	10	8.0	11	10	9.2	8.5	10	10	9.3
		5	4.3	5.9	6.0	5.3	4.3	5.5	5.1	4.6
W	(ARMA)	10	15	8.4	11	14	10	8.6	8.8	9.7
		5	8.8	4.4	6.7	8.3	5.5	4.4	4.7	4.9
	(ARCH)	10	13	7.9	12	13	11	8.5	11	12
		5	7.7	4.4	7.5	8.2	6.1	4.6	6.5	6.8
	(SV)	10	16	8.7	15	16	10	8.2	10	11
		5	11	5.2	10	11	5.8	4.0	5.7	5.9

Table 2: Empirical sizes in % of monitoring procedures based on $X^{\widehat{\gamma}}$ and $X^{\widehat{x}_p}$ ($X \in \{V, W\}$) for $\lfloor nT \rfloor$ realizations of (ARMA), (ARCH) and (SV)

We investigate size using data from a linear ARMA(1,1) and non-linear ARCH(1) and SV models.

X	DGP	t^*	Level	$\lfloor nT \rfloor = 500$				$\lfloor nT \rfloor = 2000$			
				$X^{\widehat{\gamma}}$	$X^{\widehat{x}_p}$			$X^{\widehat{\gamma}}$	$X^{\widehat{x}_p}$		
					$p = 0.1$	0.01	0.001		$p = 0.1$	0.01	0.001
V	(ARMA)	1.15	10	7.4	96	66	33	9.7	100	97	75
			5	3.4	92	53	23	4.5	100	94	62
		2.5	10	18	81	57	38	40	100	95	82
			5	11	70	45	27	27	98	89	71
	(ARCH)	1.15	10	17	43	37	28	37	77	71	59
			5	11	33	27	19	26	65	59	47
		2.5	10	12	26	23	18	22	44	42	34
			5	6.7	18	16	11	14	31	31	23
	(SV)	1.15	10	27	73	57	45	83	99	97	94
			5	15	58	40	29	69	97	91	85
		2.5	10	8.5	28	15	11	22	72	46	36
			5	3.6	17	7.4	5.0	11	57	32	22
W	(ARMA)	1.15	10	20	54	41	32	22	98	80	58
			5	13	41	31	23	14	95	68	45
		2.5	10	18	45	31	25	19	97	70	47
			5	11	33	22	17	12	92	57	34
	(ARCH)	1.15	10	32	30	43	40	50	60	69	64
			5	23	22	33	30	38	48	58	52
		2.5	10	25	22	31	30	37	47	54	49
			5	17	15	23	21	27	35	43	38
	(SV)	1.15	10	0.9	10	1.5	0.8	6.1	54	26	16
			5	0.4	5.3	0.6	0.0	2.2	38	14	7.8
		2.5	10	10	11	9.7	10	10	48	25	18
			5	6.2	5.6	5.9	6.1	4.9	33	14	9.2

Table 3: Empirical power in % of monitoring procedures based on $X^{\widehat{\gamma}}$ and $X^{\widehat{x}_p}$ ($X \in \{V, W\}$) for $\lfloor nT \rfloor$ realizations of (ARMA), (ARCH) and (SV)

Specifically, we simulate $\{X_i\}_{i=1,\ldots,\lfloor nT \rfloor}$ from the three data generating processes (DGPs)

$$X_i = 0.3 \cdot X_{i-1} + Z_i + 0.7 \cdot Z_{i-1}, \qquad Z_i \overset{\text{i.i.d.}}{\sim} t_{10}, \tag{ARMA}$$

$$X_i^2 = \left(0.01 + 0.3125 \cdot X_{i-1}^2\right) Z_i^2, \qquad Z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \tag{ARCH}$$

$$|X_i| = \sigma_i |Z_i|, \qquad Z_i \overset{\text{i.i.d.}}{\sim} t_{0.5}, \tag{SV}$$

where t_ν denotes a Student's t-distribution with ν degrees of freedom (i.e., tail index equal to ν) and $\sigma_i = 0.5\sigma_{i-1} + \varepsilon_i$, $\varepsilon \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, a Gaussian AR(1)-process. For the verification of the conditions **(A2)-(A7)** for the first two models we refer to Drees (2003, Secs. 3.1 & 3.2). The $|X_i|$ generated from the ARMA(1,1)-model have tail index 10 because of Lemma 5.2 in Datta and McCormick (1998). The tail index of the X_i^2 from the ARCH(1)-model is given by $8/2 = 4$ (cf. Davis and Mikosch, 1998, Table 1), while that of $|X_i|$ from (SV) is 0.5 (see Section 2.3). The parameters are chosen to demonstrate that our procedure works well for tails ranging from rather light (in the (ARMA) case) to very heavy with non-existent first moment (in the (SV) case).

The conclusions that can be drawn from Table 2 are quite similar for all models. When $\lfloor nT \rfloor = 500$ size varies around the nominal level quite a bit for different choices of p. This is no longer the case for the longer period, where size is always very close to the expected level. Although we cannot offer any intuition, it is interesting to observe that when the procedures based on the detectors V are oversized, those based on W are undersized and vice versa.

To assess the power of our tests we generate data from the following models, where the historical data in all three cases are generated according to the models already investigated under the null:

$$X_{i,n} = \begin{cases} 0.3 \cdot X_{i-1} + Z_i + 0.7 \cdot Z_{i-1}, & i = 1, \dots, \lfloor nt^* \rfloor, \\ 0.8 \cdot X_{i-1} + Z_i + 0.7 \cdot Z_{i-1}, & i = \lfloor nt^* \rfloor + 1, \dots, \lfloor nT \rfloor, \end{cases} \qquad Z_i \overset{\text{i.i.d.}}{\sim} t_{10}, \tag{15}$$

$$X_{i,n}^2 = \begin{cases} \left(0.01 + 0.3125 \cdot X_{i-1}^2\right) Z_i^2, & i = 1, \dots, \lfloor nt^* \rfloor, \\ \left(0.01 + 0.5773 \cdot X_{i-1}^2\right) Z_i^2, & i = \lfloor nt^* \rfloor + 1, \dots, \lfloor nT \rfloor, \end{cases} \qquad Z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \tag{16}$$

$$|X_i| = \sigma_i |Z_{i,n}|, \quad \text{where } Z_{i,n} \sim \begin{cases} t_{0.5}, & i = 1, \dots, \lfloor nt^* \rfloor, \\ t_1, & i = \lfloor nt^* \rfloor + 1, \dots, \lfloor nT \rfloor, \end{cases} \tag{17}$$

where $n = 125, 500$ and $T = 4$ as before, $t^* = 1.15, 2.5$, corresponding to breaks after 5% and 50% of the observation period, and the $Z_{i,n}$ are independent. Here, i.ni.d. stands for independent, non-identically distributed. In the ARMA(1,1)-model with the break in the AR-parameter from 0.3 to 0.8 there is no break in the tail index, but a break in the variance from 0.92 to 1.81, i.e., a more volatile distribution after the break. In the ARCH case the parameter shift induces a tail index change from $8/2 = 4$ to $4/2 = 2$ (cf. Davis and Mikosch, 1998, Table 1), i.e., heavier tails after the break. At the same time the variance is finite pre-break and (hairline) infinite post-break. For the stochastic volatility model with the break in the error distribution the break is in the opposite direction with a change in the tail index from 0.5 to 1.

Note that for the ARMA(1,1) model in (15) the null hypothesis $\mathcal{H}_{0,\gamma}$ is true. However, as in finite samples an upward break in the variance may not be clearly discerned from one in the tail index by our procedure, we should expect more rejections for $\tau_n^{V/W^{\widehat{\gamma}}}$ than in Table 2. This is generally confirmed by the results in Table 3. Furthermore, the variance change is most frequently detected using $\tau_n^{V/W^{\widehat{x}_p}}$ for $p = 0.1$. This may be explained by the higher variance of the estimates $\widehat{x}_p$ for smaller values of p, which makes detection of a quantile break very difficult, if the quantiles do not lie sufficiently far apart, as is plausible here, where a mere variance change occurred.

If there is a break in the tail index and the variance as in the ARCH- and SV-case, one can see from Table 3 that the procedures based on the Weissman (1978) estimator clearly perform better than those based solely on the Hill (1975) estimator. Heuristically, this may be explained by the fact that the Weissman (1978) estimator also takes differences in scale before and after the break into account

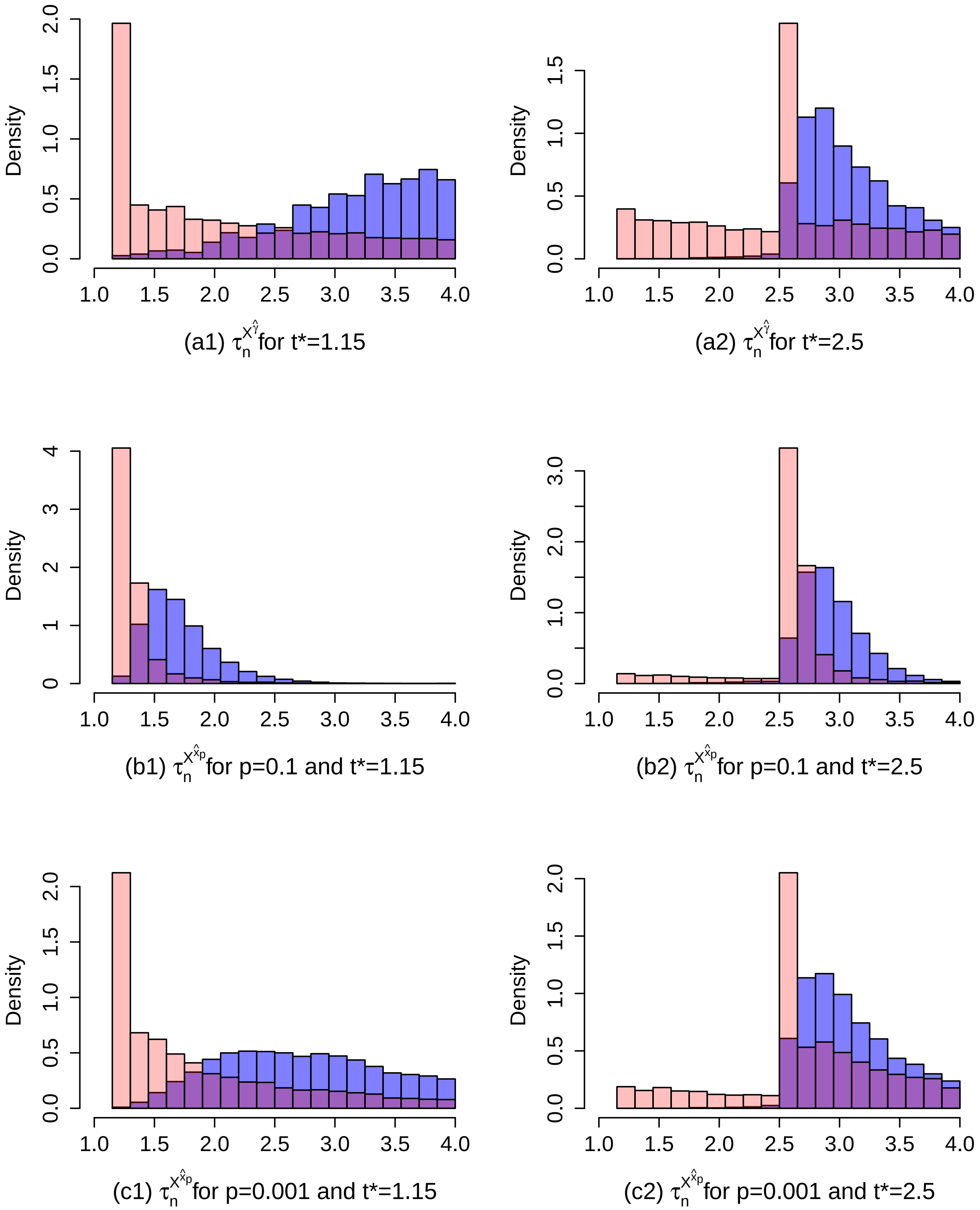

Figure 1: Histograms of detection times $\tau_n^{X^{\widehat{\gamma}}}$, $\tau_n^{X^{\widehat{x}_p}}$ for $X = V$ (blue) and for $X = W$ (bright red) for $p = 0.1, 0.001$ for (15) and $t^* = 1.15$ ((a1), (b1), (c1)), $t^* = 2.5$ ((a2), (b2), (c2)) for $\lfloor nT \rfloor = 2000$

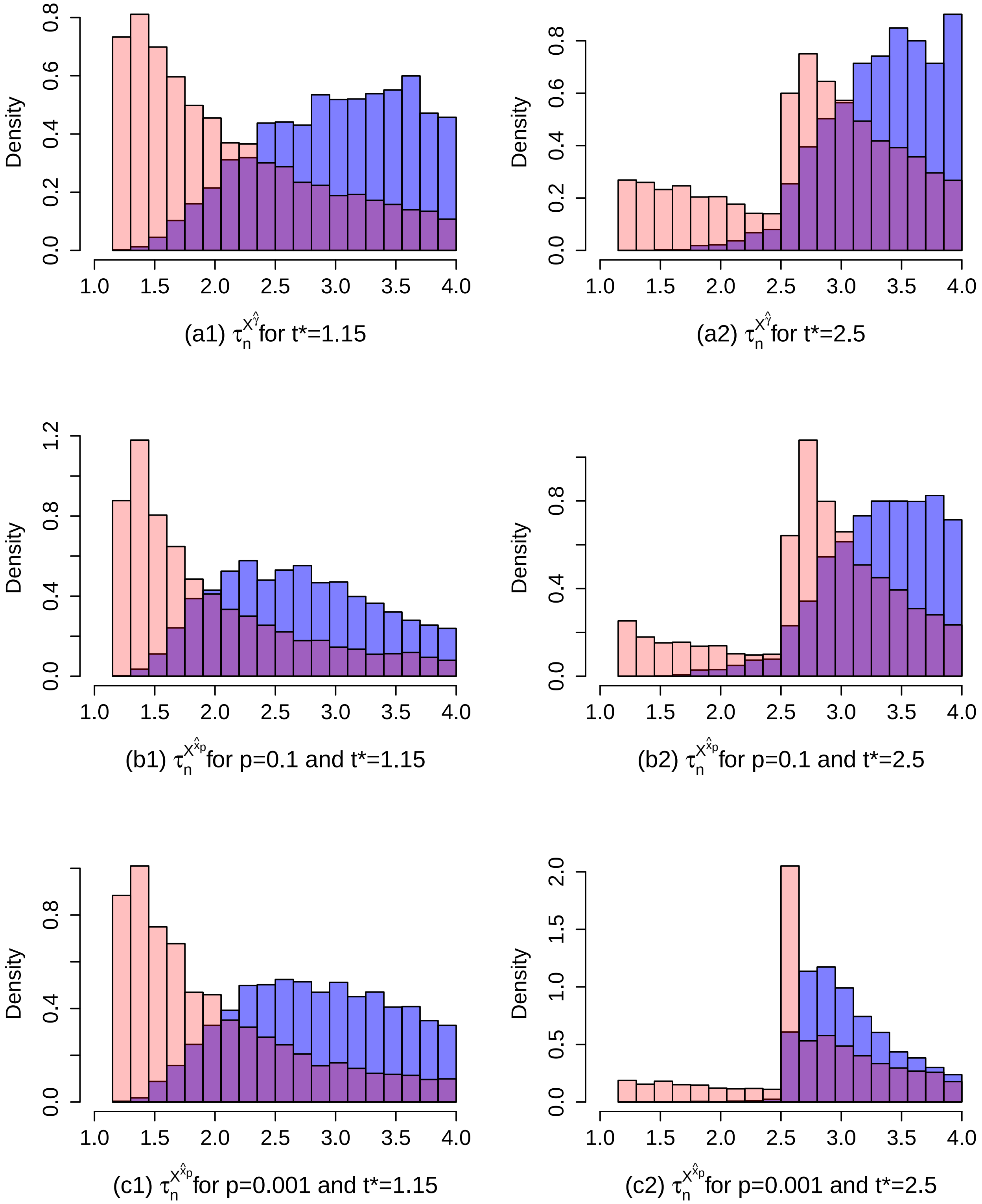

Figure 2: Histogram of detection times $\tau_n^{X^{\widehat{\gamma}}}$, $\tau_n^{X^{\widehat{x}_p}}$ for $X = V$ (blue) and for $X = W$ (bright red) for $p = 0.1, 0.001$ for (16) and $t^* = 1.15$ ((a1), (b1), (c1)), $t^* = 2.5$ ((a2), (b2), (c2)) for $\lfloor nT \rfloor = 2000$

(via $X_k(s,t,1)$; see (5)). Since in reality, changes in the tail index will most likely result in changes of scale, one should use the tests based on $V/W^{\widehat{x}_p}$. Further, the choice $\tau_n^{V^{\widehat{x}_p}}$ with $p = 0.1$ leads to the highest power, particularly for small sample lengths and the downward break in tail heaviness for model (17).

Recall that the procedure based on $V^{\widehat{\gamma}}$ is inconsistent under $\mathcal{H}_{1,\gamma}^{>}$. While this is not yet apparent for the early break ($t^* = 1.15$), it is for the late break ($t^* = 2.5$), where power is significantly lower as a larger portion of the sample upon which $\widehat{\gamma}(1,t)$ is based is 'contaminated' by very heavy tailed observations.

For sequential tests like ours, power is not the only criterion by which to judge a procedure, but also how promptly changes are detected. To look into this, Figures 1 and 2 show histograms of the (finite) realizations of $\tau_n^{V/W^{\widehat{\gamma}}}$ and $\tau_n^{V/W^{\widehat{x}_p}}$ (bright blue / red) at the 10% level for the ARMA and the ARCH models given in (15) and (16) respectively with $\lfloor nT \rfloor = 2000$. The histogram for (17) does not provide any additional insights and is omitted. There are 19 bars in all plots with breaks at $1 + l \cdot 0.15$ ($l = 1, \ldots, 20$). The value of t^* at which the changes are located are given by $1.15 = 1 + 1 \cdot 0.15$ ($l = 1$) and $2.5 = 1 + 10 \cdot 0.15$ ($l = 10$).

The results for the ARMA model are displayed in Figure 1. The high false detection rate for the tail index-based method using the detector $W_n^{\widehat{\gamma}}$ seems largely to be due to false detections just shortly after the break, as can be seen from the distinctive peaks in panels (a1) and (a2). The detections with $\tau_n^{W^{\widehat{x}_p}}$ for $p = 0.1$ in (b1) and (b2) indicate that a very large portion of detections occur within the time corresponding to the two bars right after the break. This holds to a lesser extent for the results shown in (c1) and (c2), where, however, detection rates were poor. Overall the detection speed is satisfactory but faster for larger values of p. For the ARCH model one can see slightly dissimilar detection patterns for all procedures based on W. The highest number of detections always occurs one or two bars after the break and that rate goes down only slowly so that detections (if they occur) take on average longer than in the ARMA case. This may be explained by the fact that ARCH models incorporate conditional heteroscedasticity, such that detection of changes in the variability of time series is inherently more difficult. We need to observe longer periods of higher volatility before one can reject the null here.

Comparing these results with those for the procedures based on V we see that for the latter detections take much longer. They never peak in the initial period, where the change occurs. This introduces a delicate trade-off for the detectors we introduced. The stopping times based on V terminate more often under the alternative than those based on W, but they take longer to do so. So if a swift detection is of the utmost importance, we recommend to use $\tau_n^{W^{\widehat{x}_p}}$ for $p = 0.01$. If it is more important that a break is detected at all, but speed is of lesser interest, then $\tau_n^{V^{\widehat{x}_p}}$ for $p = 0.01$ seems to be the wisest choice, unless a break towards lighter tails is expected in which case power was rather dismal.

4 Application

In this section we apply our tests to the lower tail of log-returns, i.e., log-losses, of Bank of America stocks covering the period of the financial crisis of 2007-2008, where short selling US financial stocks was halted until October 2, 2008 following an SEC order released on September 19, 2008. The return series we consider is displayed in the top part of Figure 4. Results for stock prices of other US financial companies (Morgan Stanley, Citigroup and Goldman Sachs) are very similar and can be obtained upon request. Since we try to detect changes in unconditional quantiles, our focus is on the long-term distributional changes in the time series, not on short-term changes in the conditional distribution. We set our (artificial) training and observation period to be the years from 2005 to 2012 corresponding to 2013 observations, $X_1, \ldots, X_{2012}$, the first quarter of which (roughly the years 2005 and 2006) make up the training period. The lengths of the training and observation period were chosen to correspond to the case $n = 500$ in the simulations, for which size and power proved to be very satisfying. Furthermore, we choose the training period to precede the onset of the financial crisis in 2007, so that we may analyze the performance of our procedures during these tumultuous years.

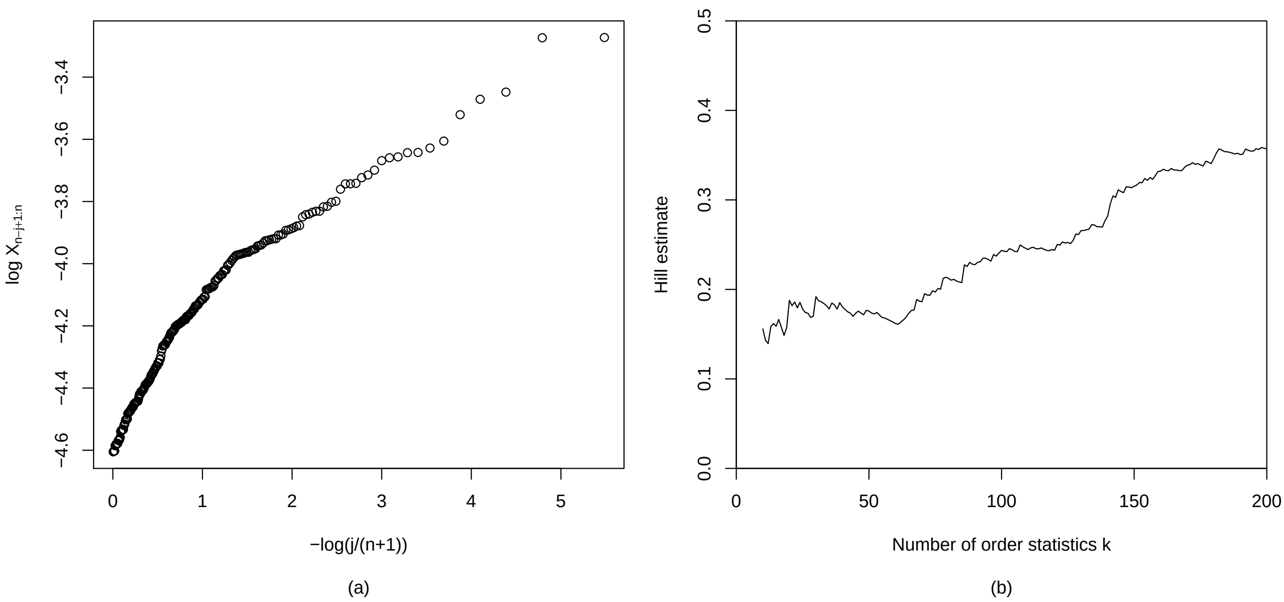

Figure 3: (a) Pareto quantile plot of shifted data. (b) Hill estimates as a function of upper order statistics k used in the estimation.

Given the very calm behavior of the log-returns during the training period one may have suspected that a break toward heavier tails is much more likely than one toward lighter tails. Additionally, it is vital for managing risk adequately to detect a break in the tail behavior quickly, because if it is registered to late the cost of hedging that risk may already have increased dramatically. For these two reasons we focus on the detectors W, which performed only slightly worse than the detectors V when there is break leading to heavier tails, yet detected those much faster.

Next, we verify that our two central assumptions, the stationary mixing assumption **(A2)** and the heavy tail assumption **(A5)**, are plausibly met by the time series in the training period. To check whether there is evidence for heavy tails we plot the Pareto quantile plot in Figure 3 (a), where the points $(-\log(j/(n+1)), \log X_{n-j+1:n})$, $j = 1, \ldots, n$, are plotted. See Beirlant, Vynckier and Teugels (1996) for more on Pareto quantile plots. In order for all $\log X_{n-j+1:n}$'s to be well-defined we shifted the observations to the positive half-line by adding the absolute value of the smallest return plus 0.01. An upward sloping linear trend, like the one that can be seen in Figure 3 (a) from $-\log(j/(n+1)) \approx 1.5$ onwards, for some $j = 1, \ldots, k+1$ in the plot indicates a good fit of the tail to (strict) Pareto behavior. An estimate of $1/\alpha$ can then be obtained as the slope of the line from the point $(-\log((k+1)/(n+1)), \log X_{n-k:n})$ onwards, where the slope seems to be roughly 0.2. This is confirmed in the (slightly upward trending) Hill plot in Figure 3 (b), which displays the Hill estimates of the shifted data as a function of the upper order statistics k used in the estimation. As for the mixing assumption **(A2)** the best ARCH(p)-model (by AIC) was an ARCH(1). Using an ARCH-LM test however, we could not reject the null of no conditional heteroscedasticity for this model (p-value of 0.86). Routine testing and plotting of the autocorrelation function of the raw and squared log-losses also revealed no dependence in the data, such that the data may reasonably be regarded as independent. Further, applying the retrospective tests of Hoga (2016+) and Hoga (2015) we found no evidence of extreme quantile or tail index breaks during that period which would violate the stationarity assumption. Hence, we proceed with our monitoring procedure.

Since their inception by Engle (1982) (G)ARCH-models have arguably become the most popular models for returns on risky assets. So the absence of ARCH-effects in the training period may be surprising. However, Stărică and Granger (2005) argue for models of returns that are locally i.i.d. In our case the period from 2005 to 2006 seems to be a period, where returns behave like an i.i.d. sequence.

The results are shown in the middle part of Figure 4. As in the simulations we choose $k/n = 0.2$ and $t_0 = 0.2$. All procedures terminate at the 5%-level if the value of 45.4 is exceeded by the detector. We see that the procedure testing for a change in the 10%-quantile of the log-returns terminates first in November of 2007, followed later by the detection of a break in the 1%-quantile in August 2008. A 0.1%-quantile break is detected last in early 2009. However, we find no evidence of a tail index break in the observation period. The lower part of Figure 4 sheds some light on this phenomenon. There, the rolling window extreme value index estimates based on samples of length $\lfloor nt_0 \rfloor = 100$, that the detectors are based on, are presented. The estimates hover around the value of 0.6 during the whole period, which is roughly the extreme value index estimate of 0.52 based on the training period. (Due to the location dependence of Hill estimates, the extreme value index in the training period was estimated to be roughly 0.2 for the shifted data in Figure 3 (b) and 0.52 now, based on non-shifted returns. The extreme value index itself is of course shift-invariant.) This contrasts with the behavior

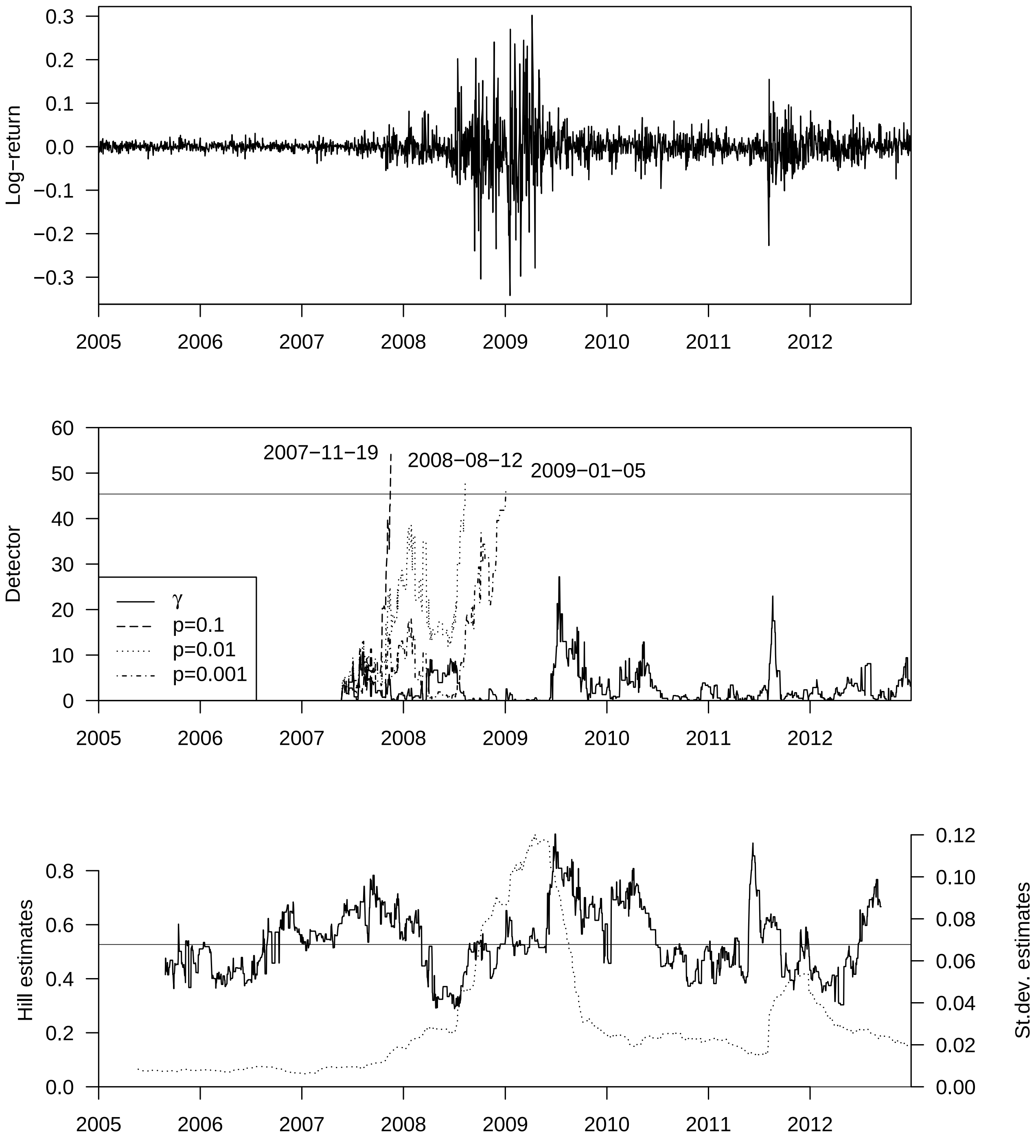

Figure 4: Top panel: Log-returns of Bank of America stock. Middle panel: Values of detectors $W^{\widehat{\gamma}}$ (solid), and $W^{\widehat{x}_p}$ for $p = 0.1, 0.001, 0.0001$ (dashed, dotted, dash-dotted) and value of 5%-threshold (horizontal solid line). Bottom panel: Rolling Hill estimate (jagged solid line), Hill estimate based on training period (straight solid line), and standard deviation estimates (dotted line).

of the standard deviation estimates based on the same rolling windows, where we see a marked spike peaking in early 2009. Hence, we find indications that the change in the extreme quantiles is not caused by a change in the tail index but rather by a change in the scale of the log-returns. Largely, the above results are consistent with the simulations under the alternative, where a variance change occurred. Procedures based on $W^{\widehat{x}_p}$ detect mere variance changes more easily for larger values of p, while that based on $W^{\widehat{\gamma}}$ did not pick up a tail index change.

Appendix

Proof of Theorem 1: In the following let K, K_1 be positive constants that may change from line to line. The proof of *(i)* mainly rests on a time shifted version of the (weighted) weak convergence established in Hoga (2016+, Theorem 5); see also the proof of Theorem 1 in Hoga (2016+). That is, for any $t_0 > 0$ we have that for some $\delta > 0$ and any $\nu \in \big[0, 1/2\big)$,

$$\frac{1}{y^\nu} e_n(t,y) := \frac{\sqrt{k}}{y^\nu}\left\{\frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor}\left[I_{\{U_i > 1-\frac{k}{n}y\}} - \frac{k}{n}y\right]\right\} \underset{(n\to\infty)}{\xrightarrow{\mathcal{D}}} \frac{1}{y^\nu} W(t,y) \quad \text{in } D\left([t_0, T]\times[0, y_0+\delta]\right) \quad \text{(A.1)}$$

for a sequence of uniformly distributed random variables $U_i \sim \mathcal{U}[0,1]$ satisfying **(A2)**-**(A4)**, where $W(t,y)$ is a continuous centered Gaussian process with covariance function

$$\operatorname{Cov}\left(W(t_1, y_1), W(t_2, y_2)\right) = \min(t_1, t_2)\, r(y_1, y_2)$$

and $r(\cdot,\cdot)$ defined in **(A3)**. For the continuous mapping theorem (CMT) to imply (A.8) below, we need to extend the convergence in (A.1) to $D\left([0,T]\times[0, y_0+\delta]\right)$. To do so, it suffices to show by Billingsley (1968, Thm. 4.2) that

$$\lim_{t_0\downarrow 0}\limsup_{n\to\infty} P\left\{\sup_{\substack{t\in[0,t_0]\\ y\in(0,\widetilde{y}_0]}}\left|\frac{e_n(t,y)}{y^\nu}\right| > \varepsilon\right\} = 0, \tag{A.2}$$

$$\lim_{t_0\downarrow 0} P\left\{\sup_{\substack{t\in[0,t_0]\\ y\in(0,\widetilde{y}_0]}}\left|\frac{W(t,y)}{y^\nu}\right| > \varepsilon\right\} = 0, \tag{A.3}$$

where $\widetilde{y}_0 := y_0 + \delta$. We first show (A.3) by using the inequality in Lin and Choi (1999, Lem. 2.1) similarly as in the proof of Hoga (2016+, Thm. 5). More specifically, we set

$$\begin{aligned}
\mathbb{D}_j &:= [0, t_0] \times [\widetilde{y}_0 e^{-(j+1)}, \widetilde{y}_0 e^{-j}],\\
\Gamma_j^2 &:= \sup_{t\in\mathbb{D}_j} \mathrm{E}[W(t,y)]^2 = \sup_{t\in\mathbb{D}_j} tr(y,y) \le K t_0 \widetilde{y}_0 e^{-(j+1)},\\
\lambda_j &:= t_0 \widetilde{y}_0 e^{-(j+1)}[e-1]
\end{aligned}$$

and use $\phi(r) := K\sqrt{r}$. Then, Lemma 2.1 of Lin and Choi (1999) implies

$$\begin{aligned}
&P\left\{\sup_{\substack{t\in[0,t_0]\\ y\in(0,\widetilde{y}_0]}}\left|\frac{W(t,y)}{y^\nu}\right|>\varepsilon\right\}\\
&\le \sum_{j=0}^{\infty} P\left\{\sup_{\substack{t\in[0,t_0]\\ y\in\left(\widetilde{y}_0e^{-(j+1)},\widetilde{y}_0e^{-j}\right]}}\left|W(t,y)\right|>\varepsilon\left(\widetilde{y}_0e^{-(j+1)}\right)^\nu\right\}\\
&\le K\sum_{j=0}^{\infty}\frac{1}{t_0\widetilde{y}_0e^{-(j+1)}}\exp\left\{-\frac{1}{2}\left(\frac{\varepsilon\left(\widetilde{y}_0e^{-(j+1)}\right)^\nu}{\Gamma_j+(2\sqrt{2}+2)K_1\int_0^\infty\varphi(\sqrt{2}\lambda_j2^{-x^2})\mathrm{d}x}\right)^2\right\}\\
&\le Kt_0^{\frac{2\nu}{1-2\nu}}\sum_{j=0}^{\infty}\left(\left(t_0^{\frac{1}{1-2\nu}}\widetilde{y}_0e^{-(j+1)}\right)^{2\nu-1}\right)^{\frac{1}{1-2\nu}}\exp\left\{-K_1\left(t_0^{\frac{1}{1-2\nu}}\widetilde{y}_0e^{-(j+1)}\right)^{2\nu-1}\right\}\underset{(t_0\downarrow 0)}{\longrightarrow}0.
\end{aligned}$$

Before proving (A.2), put

$$S_s(y) := \frac{1}{y^\nu}\frac{1}{\sqrt{k}}\sum_{i=1}^{s}\left[I_{\left\{U_i>1-\frac{k}{n}y\right\}}-\frac{k}{n}y\right]\quad\text{and}\quad \|S_s\| := \sup_{y\in(0,\widetilde{y}_0]}\left|S_s(y)\right|.$$

Now use the Ottaviani-type inequality from Bücher (2015, Lem. 3) to obtain

$$\begin{aligned}
&P\left\{\sup_{\substack{t\in[0,t_0]\\ y\in(0,\widetilde{y}_0]}}\left|\frac{e_n(t,y)}{y^\nu}\right|>\varepsilon\right\}=P\left\{\max_{s=1,\dots,\lfloor nt_0\rfloor}\|S_s\|>\varepsilon\right\}\\
&\le\frac{P\left\{\left\|S_{\lfloor nt_0\rfloor}\right\|>\varepsilon\right\}+P\left\{\max_{\substack{r<s\in\{1,\dots,\lfloor nt_0\rfloor\}\\ s-r\le 2r_n}}\|S_s-S_r\|>\varepsilon\right\}+\frac{\lfloor nt_0\rfloor}{r_n}\beta(r_n)}{1-\max_{s=1,\dots,\lfloor nt_0\rfloor}P\left\{\left\|S_{\lfloor nt_0\rfloor}-S_s\right\|>\varepsilon\right\}}.
\end{aligned}\tag{A.4}$$

To start, consider the three terms in the numerator. First, $\frac{\lfloor nt_0\rfloor}{r_n}\beta(r_n)$ tends to zero as $n\to\infty$ by **(A2)**. Second, from (A.1) and the CMT,

$$P\left\{\left\|S_{\lfloor nt_0\rfloor}\right\|>\varepsilon\right\}\underset{(n\to\infty)}{\longrightarrow}P\left\{\sup_{y\in(0,\widetilde{y}_0]}\left|\frac{W(t_0,y)}{y^\nu}\right|>\varepsilon\right\},$$

where the limit tends to zero as $t_0\downarrow 0$ because of (A.3). Finally,

$$\begin{aligned}
&\max_{\substack{r<s\in\{1,\dots,\lfloor nt_0\rfloor\}\\ s-r\le 2r_n}}\|S_s-S_r\|=\max_{\substack{r<s\in\{1,\dots,\lfloor nt_0\rfloor\}\\ s-r\le 2r_n}}\sup_{y\in(0,\widetilde{y}_0]}\frac{1}{y^\nu}\left|\frac{1}{\sqrt{k}}\sum_{i=r+1}^{s}\left[I_{\left\{U_i>1-\frac{k}{n}y\right\}}-\frac{k}{n}y\right]\right|\\
&\le\max_{\substack{r<s\in\{1,\dots,\lfloor nt_0\rfloor\}\\ s-r\le 2r_n}}\sup_{y\in(0,\widetilde{y}_0]}\frac{1}{y^\nu}\left|\frac{1}{\sqrt{k}}\sum_{i=r+1}^{s}I_{\left\{U_i>1-\frac{k}{n}y\right\}}\right|+\sup_{y\in(0,\widetilde{y}_0]}\frac{1}{y^\nu}\left|\frac{2r_n}{\sqrt{k}}\frac{k}{n}y\right|=:A_n+B_n.
\end{aligned}$$

B_n tends to zero by **(A2)**. As for A_n, we have

$$P\left\{A_n > \varepsilon\right\} \leq P\left\{\max_{m\in\left\{0,\dots,\left\lfloor\frac{\lfloor nt_0\rfloor}{2r_n}\right\rfloor\right\}} \sup_{y\in(0,\widetilde{y}_0]} \frac{1}{y^\nu}\left|\frac{1}{\sqrt{k}}\sum_{i=\lfloor m2r_n\rfloor+1}^{\lfloor(m+2)2r_n\rfloor} I_{\left\{U_i>1-\frac{k}{n}y\right\}}\right| > \varepsilon\right\}$$

$$\leq \left\lfloor\frac{\lfloor nt_0\rfloor}{2r_n}\right\rfloor P\left\{\sup_{y\in(0,\widetilde{y}_0]} \frac{1}{y^\nu}\left|\frac{1}{\sqrt{k}}\sum_{i=1}^{4r_n} I_{\left\{U_i>1-\frac{k}{n}y\right\}}\right| > \varepsilon\right\} \leq t_0 K \underset{(t_0\downarrow 0)}{\longrightarrow} 0, \qquad \text{(A.5)}$$

where we have used strict stationarity in the second to last inequality and a result from the proof of Hoga (2016+, Thm. 5) in the last inequality. Thus, all terms in the numerator tend to zero.

We now turn to the denominator in (A.4) and show that it tends to one. Using strict stationarity again, we get

$$\max_{s\in\left\{1,\dots,\lfloor nt_0\rfloor\right\}} P\left\{\left\|S_{\lfloor nt_0\rfloor} - S_s\right\| > \varepsilon\right\} = \max_{s\in\left\{1,\dots,\lfloor nt_0\rfloor\right\}} P\left\{\|S_s\| > \varepsilon\right\}$$

$$= \max_{m\in\left\{1,\dots,\lfloor nt_0\rfloor\right\}} P\left\{\sup_{y\in(0,\widetilde{y}_0]} \frac{1}{y^\nu}\left|\frac{1}{\sqrt{k}}\sum_{i=1}^{m}\left[I_{\left\{U_i>1-\frac{k}{n}y\right\}} - \frac{k}{n}y\right]\right| > \varepsilon\right\}.$$

Fix $\eta > 0$ small and set $i_n := \min\left\{i \in \mathbb{N}:\ \sqrt{k} \leq \eta\left(\widetilde{y}_0 e^{-(i+1)}\right)^{\nu-1}\right\}$. Decompose

$$\frac{1}{y^\nu}\frac{1}{\sqrt{k}}\sum_{i=1}^{m}\left[I_{\left\{U_i>1-\frac{k}{n}y\right\}} - \frac{k}{n}y\right] = \frac{1}{y^\nu}\sum_{w=0}^{\lfloor m/r_n\rfloor-1}\frac{1}{\sqrt{k}}\sum_{i=r_n+1}^{(w+1)r_n}\left[I_{\left\{U_i>1-\frac{k}{n}y\right\}} - \frac{k}{n}y\right]$$
$$+ \frac{1}{y^\nu}\frac{1}{\sqrt{k}}\sum_{i=\lfloor m/r_n\rfloor r_n+1}^{m}\left[I_{\left\{U_i>1-\frac{k}{n}y\right\}} - \frac{k}{n}y\right]. \qquad \text{(A.6)}$$

The second-term on the right-hand side of (A.6) contains at most r_n terms and hence its supremum can be bounded by

$$\sup_{y\in(0,\widetilde{y}_0]} \frac{1}{y^\nu}\frac{1}{\sqrt{k}}\sum_{i=\lfloor m/r_n\rfloor r_n+1}^{m} I_{\left\{U_i>1-\frac{k}{n}y\right\}} + \frac{r_n}{\sqrt{k}}\frac{k}{n}\widetilde{y}_0^{1-\nu} =: \widetilde{A}_n + \widetilde{B}_n.$$

Clearly, $\widetilde{B}_n = o(1)$ and that $\widetilde{A}_n = o_P(1)$ can be seen as for A_n in (A.5).

To show that the first term on the right-hand side of (A.6) is $o_P(1)$ it suffices to do so for $\widetilde{S}_{m_n}$, where $m_n = \left\lfloor\frac{m}{2r_n}\right\rfloor$, $\widetilde{S}_{m_n} := \frac{1}{y^\nu}\sum_{j=1}^{m_n} Y_{n,j}$ with $Y_{n,j}$ an i.i.d. stochastic process distributed as $\frac{1}{\sqrt{k}}\sum_{i=1}^{r_n}\left[I_{\left\{U_i>1-\frac{k}{n}y\right\}} - \frac{k}{n}y\right]$; see the proof of Drees (2000, Thm. 2.3).

Then, following the derivations in Drees (2000, Eq. (5.6)) leads to

$$P\left\{\sup_{y\in(0,\widetilde{y}_0 e^{-i_n}]}\left|\widetilde{S}_{m_n}(y)\right| > \varepsilon\right\} \leq K\eta \underset{(\eta\to 0)}{\longrightarrow} 0,$$

where K is independent of m. It remains to prove

$$P\left\{\sup_{y\in(\widetilde{y}_0 e^{-i_n},\widetilde{y}_0]}\left|\widetilde{S}_{m_n}(y)\right| > \varepsilon\right\} \underset{(n\to\infty)}{\longrightarrow} 0.$$

By condition **(A4)** for the U_i we have $\mathrm{E}\left[Y_{n,1}(y)-Y_{n,1}(y)\right]^j \le Cr_n\frac{k}{n}k^{-j/2}(y-x)$, whence Burkholder's inequality implies

$$\begin{aligned}
&\mathrm{E}\left[y^{\nu}\left(\widetilde{S}_{m_n}(y)-\widetilde{S}_{m_n}(x)\right)\right]^4 \\
&\le K\left\{m_n\,\mathrm{E}\left[Y_{n,1}(y)-Y_{n,1}(x)\right]^4 + m_n(m_n-1)\left\{\mathrm{E}\left[Y_{n,1}(y)-Y_{n,1}(x)\right]^2\right\}^2\right\} \\
&\le K\left\{\frac{m}{n}\frac{1}{k}(y-x)+\left(\frac{m}{n}\right)^2(y-x)^2\right\}.
\end{aligned}$$

Hence, we obtain as in the proof of Drees (2000, Thm. 2.3) that

$$\begin{aligned}
P\left\{\sup_{y\in(\widetilde{y}_0e^{-i_n},\widetilde{y}_0]}\left|\frac{\widetilde{S}_{m_n}(y)}{y^{\nu}}\right|>\varepsilon\right\} &\le \left(\frac{m}{n}\right)^2 K\sum_{i=0}^{i_n-1}(\widetilde{y}_0e^{-i})^2\left[\widetilde{y}_0e^{-(i+1)}\right]^{-4\nu} \\
&\le \left(\frac{m}{n}\right)^2 K\int_0^{\widetilde{y}_0/e}u^{1-\nu}\mathrm{d}u \le Kt_0^2 \underset{(t_0\downarrow 0)}{\longrightarrow} 0,
\end{aligned}$$

finishing the proof of (A.2). Combining (A.1), (A.2) and (A.3) gives

$$\frac{\sqrt{k}}{y^{\nu}}\left\{\frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor}\left[I_{\{U_i>1-\frac{k}{n}y\}}-\frac{k}{n}y\right]\right\}\underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}}\frac{1}{y^{\nu}}W(t,y)\quad\text{in } D\left([0,T]\times[0,y_0+\delta]\right). \tag{A.7}$$

Then for X_i satisfying **(A2)-(A5)**, by the proof of Theorem 3.1 in Drees (2000), the uniformly distributed $U_i := F(X_i)$ satisfy **(A2)-(A4)** and

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

Hence, from (A.7),

$$\frac{\sqrt{k}}{y^{\nu}}\left\{\frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor}I_{\left\{X_i>U\left(\frac{n}{ky}\right)\right\}}-ty\right\}\underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}}\frac{1}{y^{\nu}}W(t,y)\quad\text{in } D\left([0,T]\times[0,y_0+\delta]\right)$$

The CMT then implies

$$\begin{aligned}
&\frac{\sqrt{k}}{y^{\nu}}\begin{pmatrix}\frac{1}{k}\sum_{i=n+1}^{\lfloor n\max(1+t_0,t)\rfloor}I_{\left\{X_i>U\left(\frac{n}{ky}\right)\right\}}-(\max(1+t_0,t)-1)y\\ \frac{1}{k}\sum_{i=\lfloor n(t-t_0)\rfloor+1}^{\lfloor nt\rfloor}I_{\left\{X_i>U\left(\frac{n}{ky}\right)\right\}}-t_0y\\ \frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor}I_{\left\{X_i>U\left(\frac{n}{ky}\right)\right\}}-ty\end{pmatrix}\\
&\qquad\underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}}\frac{1}{y^{\nu}}\begin{pmatrix}W(\max(1+t_0,t),y)-W(1,y)\\ W(t,y)-W(t-t_0,y)\\ W(t,y)\end{pmatrix}\quad\text{in } D^3\left([t_0,T]\times[0,y_0+\delta]\right).
\end{aligned}\tag{A.8}$$

Now invoking a Skorohod construction (e.g., Wichura, 1970, Thm. 1) we get on a suitable probability space

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

Arguing component-wise as in the proof of Einmahl *et al.* (2016, Cor. 3) this implies for some $\widetilde{\delta}>0$ (using the second-order condition **(A5)**)

$$\sup_{\substack{t\in[t_0,T]\\ y\geq y_0^{-1/\gamma}-\widetilde{\delta}}} y^{\nu/\gamma} \left| \sqrt{k} \begin{pmatrix} \frac{1}{k}\sum_{i=n+1}^{\lfloor n\max(1+t_0,t)\rfloor} I_{\left\{X_i>yU\left(\frac{n}{k}\right)\right\}} - (\max(1+t_0,t)-1)y^{-1/\gamma} \\ \frac{1}{k}\sum_{i=\lfloor n(t-t_0)\rfloor+1}^{\lfloor nt\rfloor} I_{\left\{X_i>yU\left(\frac{n}{k}\right)\right\}} - t_0 y^{-1/\gamma} \\ \frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor} I_{\left\{X_i>yU\left(\frac{n}{k}\right)\right\}} - ty^{-1/\gamma} \end{pmatrix} - \begin{pmatrix} W(\max(1+t_0,t),y^{-1/\gamma})-W(1,y^{-1/\gamma}) \\ W(t,y^{-1/\gamma})-W(t-t_0,y^{-1/\gamma}) \\ W(t,y^{-1/\gamma}) \end{pmatrix} \right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0, \quad \text{(A.9)}$$

where Hoga (2016+, Prop. 2) justifies the final step of their proof in our case. Now retrace the proof of Hoga (2016+, Cor. 1) to see that

$$\sup_{\substack{t\in[t_0,T]\\ y\geq y_0^{-1/\gamma}}} y^{\nu/\gamma} \left| \sqrt{k} \begin{pmatrix} \frac{1}{k}\sum_{i=n+1}^{\lfloor n\max(1+t_0,t)\rfloor} I_{\left\{X_i>yX_k(1,\max(1+t_0,t),1)\right\}} - (\max(1+t_0,t)-1)y^{-1/\gamma} \\ \frac{1}{k}\sum_{i=\lfloor n(t-t_0)\rfloor+1}^{\lfloor nt\rfloor} I_{\left\{X_i>yX_k(t-t_0,t,1)\right\}} - t_0 y^{-1/\gamma} \\ \frac{1}{k}\sum_{i=1}^{\lfloor nt\rfloor} I_{\left\{X_i>yX_k(0,t,1)\right\}} - ty^{-1/\gamma} \end{pmatrix} - \begin{pmatrix} W(\max(1+t_0,t),y^{-1/\gamma})-W(1,y^{-1/\gamma})-y^{-1/\gamma}\left[W(\max(1+t_0,t),1)-W(1,1)\right] \\ W(t,y^{-1/\gamma})-W(t-t_0,y^{-1/\gamma})-y^{-1/\gamma}\left[W(t,1)-W(t-t_0,1)\right] \\ W(t,y^{-1/\gamma})-y^{-1/\gamma}W(t,1) \end{pmatrix} \right| \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} 0. \quad \text{(A.10)}$$

The key ingredient in this step justifying the replacement of $U(n/k)$ in (A.9) by the respective $X_k(s,t,1)$ in (A.10) is the generalized Vervaat lemma in Einmahl *et al.* (2010, Lem. 5), which gives a weak convergence result for $X_k(0,t,1)/U(n/k)$ from the third component of (A.9) (and similarly for the other components). The convergence in (A.10) is the key result with which one may deduce weak convergence results for various tail index estimators; see Examples 3 - 5 in Hoga (2016+). As already mentioned, we focus here on the Hill estimator $\widehat{\gamma} := \widehat{\gamma}_H$ defined in (6). Focus on the third component of (A.10) (the others can be dealt with similarly) and notice that for $j = 0, 1, \ldots, \lfloor kt \rfloor - 1$

$$\frac{1}{\lfloor kt \rfloor} \sum_{i=1}^{\lfloor nt \rfloor} I_{\{X_i > y X_k(0,t,1)\}} = \begin{cases} \frac{\lfloor kt \rfloor - j}{\lfloor kt \rfloor}, & y \in \left[\frac{X_{\lfloor nt \rfloor - \lfloor kt \rfloor + j : \lfloor nt \rfloor}}{X_{\lfloor nt \rfloor - \lfloor kt \rfloor : \lfloor nt \rfloor}}, \frac{X_{\lfloor nt \rfloor - \lfloor kt \rfloor + j + 1 : \lfloor nt \rfloor}}{X_{\lfloor nt \rfloor - \lfloor kt \rfloor : \lfloor nt \rfloor}} \right), \\ 0, & y \geq \frac{X_{\lfloor nt \rfloor : \lfloor nt \rfloor}}{X_{\lfloor nt \rfloor - \lfloor kt \rfloor : \lfloor nt \rfloor}}. \end{cases}$$

Then it is an easy exercise in integration to check that

$$\sqrt{k} \left(\widehat{\gamma}_H(0,t) - \gamma \right) = \sqrt{k} \int_1^\infty \left(\frac{1}{\lfloor kt \rfloor} \sum_{i=1}^{\lfloor nt \rfloor} I_{\{X_i > y X_k(0,t,1)\}} - y^{-1/\gamma} \right) \frac{\mathrm{d}y}{y}.$$

Similar representations hold for $\widehat{\gamma}(1, \max(1+t_0, t))$ and $\widehat{\gamma}(t - t_0, t)$, so that (A.10) implies

$$\sqrt{k} \begin{pmatrix} (\max(1+t_0,t) - 1) \left(\widehat{\gamma}(1, \max(1+t_0,t)) - \gamma \right) \\ t_0 \left(\widehat{\gamma}(t - t_0, t) - \gamma \right) \\ t \left(\widehat{\gamma}(0,t) - \gamma \right) \end{pmatrix} \xrightarrow[(n\to\infty)]{\mathcal{D}} \sigma_{\widehat{\gamma},\gamma} \begin{pmatrix} W(\max(1+t_0,t)) - W(1) \\ W(t) - W(t-t_0) \\ W(t) \end{pmatrix} \tag{A.11}$$

in $D^3[t_0, T]$, where $\sigma^2_{\widehat{\gamma},\gamma} = \gamma^2 \int_0^1 \int_0^1 \left\{ \frac{r(x,y)}{xy} - \frac{r(x,1)}{x} - \frac{r(1,y)}{y} + r(x,y) \right\} \mathrm{d}x\mathrm{d}y$ and $W(\cdot)$ a standard Brownian motion. Notice for this that by calculating covariances

$$\int_1^\infty \left[W(t, y^{-1/\gamma}) - y^{-1/\gamma} W(t,1) \right] \frac{\mathrm{d}y}{y} = \gamma \int_0^1 \left[W(t,u) - uW(t,1) \right] \frac{\mathrm{d}u}{u} \stackrel{\mathcal{D}}{=} \sigma_{\widehat{\gamma},\gamma} W(t).$$

From (A.11) and the CMT we obtain

$$V_n^{\widehat{\gamma}}(t) \xrightarrow[(n\to\infty)]{\mathcal{D}} \frac{\left[W(t) - tW(1) \right]^2}{\int_{t_0}^1 \left[W(s) - sW(1) \right]^2 \mathrm{d}s} \quad \text{in } D[1+t_0, T], \tag{A.12}$$

$$W_n^{\widehat{\gamma}}(t) \xrightarrow[(n\to\infty)]{\mathcal{D}} \frac{\left[W(t) - W(t-t_0) - t_0 W(1) \right]^2}{\int_{t_0}^1 \left[W(s) - W(s-t_0) - t_0 W(1) \right]^2 \mathrm{d}s} \quad \text{in } D[1+t_0, T].$$

The result is now proved via another application of the CMT.

For part *(ii)* we observe that it follows from (A.11) similarly as in the proof of Theorem 1 in Hoga (2015) that (for the general idea of how to derive convergence of $\widehat{x}_p$ from that of $\widehat{\gamma}$ see (A.17) below and the steps following it, in particular (A.18))

$$\frac{\sqrt{k}}{\log\left(k/(np)\right)} \begin{pmatrix} (\max(1+t_0,t) - 1) \log\left(\frac{\widehat{x}_p(1, \max(1+t_0,t))}{U(1/p)} \right) \\ t_0 \log\left(\frac{\widehat{x}_p(t-t_0,t)}{U(1/p)} \right) \\ t \log\left(\frac{\widehat{x}_p(0,t)}{U(1/p)} \right) \end{pmatrix}$$

$$\xrightarrow[(n\to\infty)]{\mathcal{D}} \sigma_{\widehat{\gamma},\gamma} \begin{pmatrix} W(\max(1+t_0,t)) - W(1) \\ W(t) - W(t-t_0) \\ W(1) \end{pmatrix} \quad \text{in } D^3[t_0, T], \tag{A.13}$$

whence with the CMT again

$$V_n^{\widehat{x}_p}(t) \xrightarrow[(n\to\infty)]{\mathcal{D}} \frac{\left[W(t) - tW(1) \right]^2}{\int_{t_0}^1 \left[W(s) - sW(1) \right]^2 \mathrm{d}s} \quad \text{in } D[1+t_0, T],$$

$$W_n^{\widehat{x}_p}(t) \xrightarrow[(n\to\infty)]{\mathcal{D}} \frac{\left[W(t) - W(t-t_0) - t_0 W(1) \right]^2}{\int_{t_0}^1 \left[W(s) - W(s-t_0) - t_0 W(1) \right]^2 \mathrm{d}s} \quad \text{in } D[1+t_0, T].$$

The conclusion follows as before. □

Proof of Theorem 2: We first prove the two statements in *(i)*. Under $\mathcal{H}_{1,\gamma}^{>}$ we get by an adaptation of the proof of Hoga (2016+, Thm. 3) similar to the one in the proof of Theorem 1

$$\sqrt{k}\left(\widehat{\gamma}(1,t)-\gamma_{\text{pre}}\right) \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} \frac{B_{\text{pre}}(t)-B_{\text{pre}}(1)}{t-1} \quad \text{in } D\left[1+t_0,T\right] \tag{A.14}$$

and by a further close inspection that even

$$\sqrt{k}\begin{pmatrix}\widehat{\gamma}(1,\max(1+t_0,t))-\gamma_{\text{pre}}\\ \widehat{\gamma}(0,t)-\gamma_{\text{pre}}\end{pmatrix} \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} \begin{pmatrix}(B_{\text{pre}}(\max(1+t_0,t))-B_{\text{pre}}(1))/(\max(1+t_0,t)-1)\\ B_{\text{pre}}(t)/t\end{pmatrix} \tag{A.15}$$

jointly in $D^2[t_0,T]$, where, setting $t_{\min}=\min(t,t^*)$,

$$B_{\text{pre}}(t)=\gamma_{\text{pre}}\int_0^1\left[W_{\text{pre}}(t_{\min},u\frac{t}{t_{\min}})-uW_{\text{pre}}(t_{\min},\frac{t}{t_{\min}})\right]\frac{\mathrm{d}u}{u}$$

and $W_{\text{pre}}(\cdot,\cdot)$ is as $W(\cdot,\cdot)$ in (A.1) with $r(\cdot,\cdot)$ replaced by $r_{\text{pre}}(\cdot,\cdot)$. Hence, (A.12) holds with $W(\cdot)$ replaced by $B_{\text{pre}}(\cdot)$, where $B_{\text{pre}}(\cdot)$ is also a continuous centered Gaussian process. The result follows.

For the other part of *(i)* it suffices to note that

$$\widehat{\gamma}(0,1) \underset{(n\to\infty)}{\overset{P}{\longrightarrow}} \gamma_{\text{pre}}<\gamma_{\text{post}} \underset{(n\to\infty)}{\overset{P}{\longleftarrow}} \widehat{\gamma}(1,T),$$

because by (A.14) $\widehat{\gamma}$ will converge in probability to the dominant tail index (i.e., $\max(\gamma_{\text{pre}},\gamma_{\text{post}})$) in a sample with one tail index break.

For *(ii)* simply note that

$$\widehat{\gamma}(0,1) \underset{(n\to\infty)}{\overset{P}{\longrightarrow}} \gamma_{\text{pre}}<\gamma_{\text{post}} \underset{(n\to\infty)}{\overset{P}{\longleftarrow}} \widehat{\gamma}(t^*,t^*+t_0).$$

If $\gamma_{\text{pre}}>\gamma_{\text{post}}$ in *(iii)* (the other case is similar) we can deduce from (A.13) that

$$\frac{\sqrt{k}}{\log\left(k/(np)\right)}\log\left(\frac{\widehat{x}_p(0,1)}{U_{\text{pre}}(1/p)}\right)=\mathcal{O}_P(1), \quad \text{and} \quad \frac{\sqrt{k}}{\log\left(k/(np)\right)}\log\left(\frac{\widehat{x}_p(1,T)}{U_{\text{pre}}(1/p)}\left(\frac{T}{t^*}\right)^{\gamma_{\text{pre}}}\right)=\mathcal{O}_P(1).$$

From this the result follows if we can establish the right-hand side result, because

$$V_n^{\widehat{x}_p}(T)=\frac{\left[(T-1)\frac{\sqrt{k}}{\log(k/(np))}\left\{\log\left(\frac{\widehat{x}_p(1,T)}{U_{\text{pre}}(1/p)}\right)-\log\left(\frac{\widehat{x}_p(0,1)}{U_{\text{pre}}(1/p)}\right)\right\}\right]^2}{\int_{t_0}^1\left[s\frac{\sqrt{k}}{\log(k/(np))}\log\left(\frac{\widehat{x}_p(0,s)}{\widehat{x}_p(0,1)}\right)\right]^2\mathrm{d}s}. \tag{A.16}$$

Decompose

$$\begin{aligned}\left(\frac{t}{t_{\min}}\right)^{\gamma_{\text{pre}}}\widehat{x}_p(1,t)-U_{\text{pre}}\left(\frac{1}{p}\right)=&\left[\left(\frac{t}{t_{\min}}\right)^{\gamma_{\text{pre}}}X_k(1,t,1)-U_{\text{pre}}\left(\frac{n}{k}\right)\right]\left(\frac{np}{k}\right)^{-\widehat{\gamma}(1,t)}\\&+\left[\left(\frac{np}{k}\right)^{-\widehat{\gamma}(1,t)}-\left(\frac{np}{k}\right)^{-\gamma_{\text{pre}}}\right]U_{\text{pre}}\left(\frac{n}{k}\right)\\&+\left[U_{\text{pre}}\left(\frac{n}{k}\right)\left(\frac{np}{k}\right)^{-\gamma_{\text{pre}}}-U_{\text{pre}}\left(\frac{1}{p}\right)\right]\\=&I+II+III.\end{aligned} \tag{A.17}$$

Before considering these three terms separately, observe that by the mean value theorem, using $\frac{\partial}{\partial\tau}\left(x^{\tau}\right)=x^{\tau}\log\left(x\right)$, there exists a $\xi\in[-1,1]$ such that

$$\begin{aligned}\left(\frac{np}{k}\right)^{-\widehat{\gamma}(1,t)}-\left(\frac{np}{k}\right)^{-\gamma_{\text{pre}}}&=\left(-\widehat{\gamma}(1,t)+\gamma_{\text{pre}}\right)\left(\frac{np}{k}\right)^{-\gamma_{\text{pre}}+\xi\left(\gamma_{\text{pre}}-\widehat{\gamma}(1,t)\right)}\log\left(\frac{np}{k}\right)\\&=\left(\frac{np}{k}\right)^{-\gamma_{\text{pre}}}\left(\gamma_{\text{pre}}-\widehat{\gamma}(1,t)\right)\left(\frac{np}{k}\right)^{\xi\left(\gamma_{\text{pre}}-\widehat{\gamma}(1,t)\right)}\log\left(\frac{np}{k}\right).\end{aligned}$$

Then use

$$\left(\frac{np}{k}\right)^{\xi\left(\gamma_{\text{pre}}-\widehat{\gamma}(1,t)\right)}=\exp\left[\xi\left(\gamma_{\text{pre}}-\widehat{\gamma}(1,t)\right)\log\left(\frac{np}{k}\right)\right]=\exp\left[\xi\mathcal{O}_P\left(\frac{1}{\sqrt{k}}\right)\log\left(\frac{np}{k}\right)\right]\underset{(n\to\infty)}{\overset{P}{\longrightarrow}}1$$

uniformly in t (by **(A6)** and (A.14)) and **(A7)** to get

$$\begin{aligned}\frac{U_{\text{pre}}\left(\frac{n}{k}\right)}{U_{\text{pre}}\left(\frac{1}{p}\right)}\left[\left(\frac{np}{k}\right)^{-\widehat{\gamma}(1,t)}-\left(\frac{np}{k}\right)^{-\gamma_{\text{pre}}}\right]&=\frac{U_{\text{pre}}\left(\frac{n}{k}\right)}{U_{\text{pre}}\left(\frac{1}{p}\right)}\left[\left(\frac{np}{k}\right)^{-\gamma_{\text{pre}}}\left(\gamma_{\text{pre}}-\widehat{\gamma}(1,t)\right)\left(\frac{np}{k}\right)^{\xi\left(\gamma_{\text{pre}}-\widehat{\gamma}(1,t)\right)}\log\left(\frac{np}{k}\right)\right]\\&=\left(1+o_P\left(1\right)\right)\left(\gamma_{\text{pre}}-\widehat{\gamma}(1,t)\right)\log\left(\frac{np}{k}\right).\end{aligned}\tag{A.18}$$

Furthermore applying the functional delta method (e.g., van der Vaart and Wellner, 1996, Theorem 3.9.4) to (a slight adaptation of) Hoga (2016+, Eq. (70)) yields

$$\begin{aligned}&\sqrt{k}\left\{\frac{X_k(1,t,1)}{U_{\text{pre}}\left(\frac{n}{k}\right)}-\left(\frac{t}{t_{\min}}\right)^{-\gamma_{\text{pre}}}\right\}\underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}}\\&\qquad-\gamma_{\text{pre}}\left(\frac{t}{t_{\min}}\right)^{-(\gamma_{\text{pre}}+1)}\left[\frac{W_{\text{pre}}\left(t_{\min},\frac{t}{t_{\min}}\right)}{t_{\min}}-W_{\text{pre}}\left(1,\frac{t}{t_{\min}}\right)\right]\quad\text{in }D[1+t_0,T],\end{aligned}\tag{A.19}$$

where we used that

$$\phi:D[1+t_0,T]\to D[1+t_0,T],\quad\phi(f(\cdot))=f^{-\gamma_{\text{pre}}}(\cdot)$$

is Hadamard-differentiable tangentially to $C[1+t_0,T]$ in $t/t_{\min}$ with derivative

$$\phi'_{t/t_{\min}}(f(\cdot))=-\gamma_{\text{pre}}\left(\frac{t}{t_{\min}}\right)^{-(\gamma_{\text{pre}}+1)}f(\cdot).$$

Combining we get

$$
\begin{aligned}
& \frac{I}{U_{\mathrm{pre}}(1/p)\log\left(\frac{k}{np}\right)} \\
= \quad & \frac{1}{U_{\mathrm{pre}}\left(\frac{1}{p}\right)}\left(\frac{np}{k}\right)^{-\widehat{\gamma}(1,t)}\frac{1}{\log\left(\frac{k}{np}\right)}\left[X_k(1,t,1)-U_{\mathrm{pre}}\left(\frac{n}{k}\right)\right] \\
= \quad & \frac{U_{\mathrm{pre}}\left(\frac{n}{k}\right)}{U_{\mathrm{pre}}\left(\frac{1}{p}\right)}\left[\left(\frac{np}{k}\right)^{-\widehat{\gamma}(1,t)}-\left(\frac{np}{k}\right)^{-\gamma_{\mathrm{pre}}}+\left(\frac{np}{k}\right)^{-\gamma_{\mathrm{pre}}}\right]\frac{1}{\log\left(\frac{k}{np}\right)}\left[\frac{X_k(1,t,1)}{U_{\mathrm{pre}}\left(\frac{n}{k}\right)}-1\right] \\
\underset{(\mathrm{A}.18)}{\overset{\textbf{(A7)}}{=}} \quad & \left[(1+o_P(1))(\gamma_{\mathrm{pre}}-\widehat{\gamma}(1,t))\log\left(\frac{np}{k}\right)+1+o\left(\frac{1}{\sqrt{k}}\right)\right]\frac{1}{\log\left(\frac{k}{np}\right)}\left[\frac{X_k(1,t,1)}{U_{\mathrm{pre}}\left(\frac{n}{k}\right)}-1\right] \\
\overset{(\mathrm{A}.19)}{=} \quad & \left[(1+o_P(1))\sqrt{k}(\gamma_{\mathrm{pre}}-\widehat{\gamma}(1,t))k^{-1/2}\log\left(\frac{np}{k}\right)+1+o\left(\frac{1}{\sqrt{k}}\right)\right]\frac{1}{\log\left(\frac{k}{np}\right)}\mathcal{O}_P(1/\sqrt{k}) \\
\underset{(\mathrm{A}.14)}{\overset{\textbf{(A6)}}{=}} \quad & o_P\left(1/\sqrt{k}\right)
\end{aligned}
$$

uniformly in t. Further, utilizing **(A6)** and **(A7)** for the third term,

$$
\frac{III}{U_{\mathrm{pre}}(1/p)\log\left(\frac{k}{np}\right)}=\frac{1}{\log\left(\frac{k}{np}\right)}\left[\frac{U_{\mathrm{pre}}\left(\frac{n}{k}\right)}{U_{\mathrm{pre}}\left(\frac{1}{p}\right)\left(\frac{np}{k}\right)^{-\gamma_{\mathrm{pre}}}}-1\right]=o\left(1/\sqrt{k}\right).
$$

Using (A.18) and (A.14) we get for the last term

$$
(t-1)\sqrt{k}\frac{II}{U_{\mathrm{pre}}\left(\frac{1}{p}\right)\log\left(\frac{k}{np}\right)}\xrightarrow[(n\to\infty)]{\mathcal{D}}B_{\mathrm{pre}}(t)-B_{\mathrm{pre}}(1)\quad\text{in } D\left[1+t_0,T\right],
$$

whence

$$
\frac{(t-1)\sqrt{k}}{\log\left(\frac{k}{np}\right)}\left\{\frac{\left(\frac{t}{t_{\min}}\right)^{\gamma_{\mathrm{pre}}}\widehat{x}_p(1,t)}{U_{\mathrm{pre}}\left(\frac{1}{p}\right)}-1\right\}\xrightarrow[(n\to\infty)]{\mathcal{D}}B_{\mathrm{pre}}(t)-B_{\mathrm{pre}}(1)\quad\text{in } D\left[1+t_0,T\right].
$$

Part *(iv)* is trivial, since

$$
\frac{\sqrt{k}}{\log\left(k/(np)\right)}\log\left(\frac{\widehat{x}_p(0,1)}{U_{\mathrm{pre}}(1/p)}\right)=\mathcal{O}_P(1),\quad\frac{\sqrt{k}}{\log\left(k/(np)\right)}\log\left(\frac{\widehat{x}_p(t^*,t^*+t_0)}{U_{\mathrm{post}}(1/p)}\right)=\mathcal{O}_P(1).
$$

The result follows using a similar expansion as in (A.16). □

References

Asai M, McAleer M, Yu J. 2006. Multivariate stochastic volatility: A review. *Econometric Reviews* **25**: 145–175.

Aschersleben P, Wagner M, Wied D. 2015. Monitoring euro area real exchange rates. In Steland A, Rafajłowicz E, Szajowski K (eds.) *Springer Proceedings in Mathematics and Statistics 122 (12th Workshop on Stochastic Models, Statistics and Their Applications)*, New York: Springer, pages 363–370.

Aue A, Hörmann S, Horváth L, Hušková M, Steinebach J. 2012. Sequential testing for the stability of high-frequency portfolio betas. *Econometric Theory* **28**: 804–837.

Beirlant J, Goegebeur Y, Segers J, Teugels J. 2004. *Statistics of Extremes*. Chichester: Wiley.

Beirlant J, Vynckier P, Teugels J. 1996. Tail index estimation, Pareto quantile plots, and regression diagnostics. *Journal of the American Statistical Association* **91**: 1659–1667.

Billingsley P. 1968. *Convergence of Probability Measures*. New York: Wiley.

Bradley R. 1986. Basic properties of strong mixing conditions. In Eberlein E, Taqqu M (eds.) *Dependence in Probability and Statistics*, Boston: Birkhäuser, pages 165–192.

Bradley R. 2007. *Introduction to strong mixing conditions Vol. I*. Heber City: Kendrick Press.

Breiman L. 1965. On some limit theorems similar to the arc-sine law. *Theory of Probability and its Applications* **10**: 351–360.

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

Chu JS, Stinchcombe M, White H. 1996. Monitoring structural change. *Econometrica* **64**: 1045–1065.

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

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

Datta S, McCormick W. 1998. Inference for the tail parameter of a linear process with heavy tail innovations. *Annals of the Institute of Statistical Mathematics* **50**: 337–359.

Davis R, Mikosch T. 1998. The sample autocorrelatios of heavy-tailed processes with applications to ARCH. *The Annals of Statistics* **26**: 2049–2080.

de Haan L, Ferreira A. 2006. *Extreme Value Theory*. New York: Springer.

Dierckx G, Teugels J. 2010. Change point analysis of extreme values. *Environmetrics* **21**: 661–686.

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.

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

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

Engle R. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**: 987–1007.

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

Hill J. 2011. Extremal memory of stochastic volatility with an application to tail shape inference. *Journal of Statistical Planning and Inference* **141**: 663–676.

Hoga Y. 2015. Testing for changes in (extreme) VaR. submitted to *Econometrics Journal*.

Hoga Y. 2016+. Change point tests for the tail index of β-mixing random variables. forthcoming in *Econometric Theory*.

Kelly B, Jiang H. 2014. Tail risk and asset prices. *Review of Financial Studies* **27**: 2841–2871.

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

Kulik R, Soulier P. 2011. The tail empirical process for long memory stochastic volatility sequences. *Stochastic Processes and their Applications* **121**: 109–134.

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

Quintos C, Fan Z, Phillips P. 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*. New York: Springer.

Shao X, Zhang X. 2010. Testing for change points in time series. *Journal of the American Statistical Association* **105**: 1228–1240.

Stărică C, Granger C. 2005. Nonstationarities in stock returns. *The Review of Economics and Statistics* **87**: 503–522.

van der Vaart A, Wellner J. 1996. *Weak Convergence and Empirical Processes*. New York: Springer.

Weissman I. 1978. Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association* **73**: 812–815.

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

Wied D, Galeano P. 2013. Monitoring correlation change in a sequence of random variables. *Journal of Statistical Planning and Inference* **143**: 186–196.

Zeileis A, Leisch F, Kleiber C, Hornik K. 2005. Monitoring structural change in dynamic econometric models. *Journal of Applied Econometrics* **20**: 99–121.