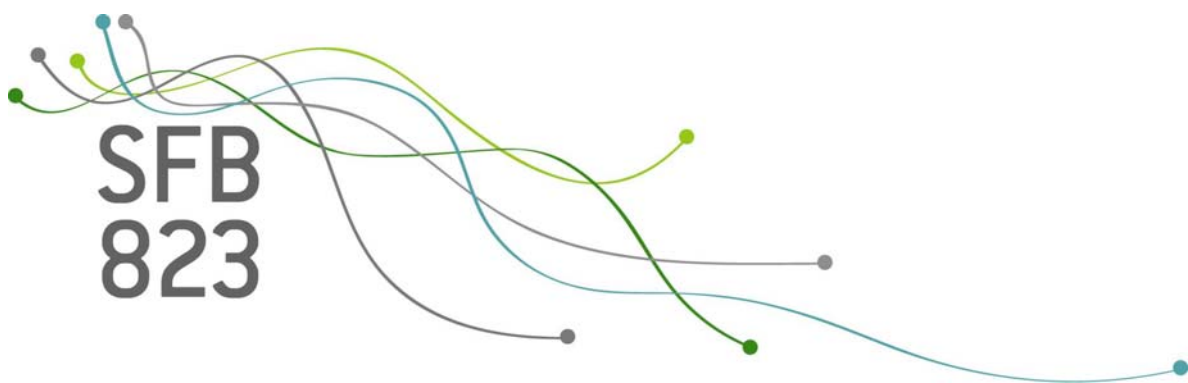# Sequential monitoring of the tail behavior of dependent data

Yannick Hoga, Dominik Wied

SFB
823

Sequential monitoring of the tail behavior of dependent data

Yannick Hoga,[*]   Dominik Wied[†]

October 15, 2015

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 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. Tests based on other estimators than the Hill (1975) estimator were first proposed by Einmahl, de Haan and Zhou (2015+) for independent and Hoga (2014) 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 (2014).

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. Simulations and an empirical application are presented in Sections 3 and 4. The proofs are collected in the 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 $\bar{F}_i(x) := 1 - F_i(x) = P\left(X_i > x\right)$, 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. Throughout $k = k(n) \in \mathbb{N}$ will denote a sequence satisfying $k \leq n - 1$,

$$k \underset{(n \to \infty)}{\longrightarrow} \infty \quad \text{and} \quad \frac{k}{n} \underset{(n \to \infty)}{\longrightarrow} 0, \tag{3}$$

controlling the number of largest 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 $U_i(1/p)$. For $t - s \geq 1/n$ and $y \in [0, 1]$ set

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

Under the assumption of strictly stationary X_i we write $\bar{F} = \bar{F}_i$ and $U = U_i$. Let $\widehat{\gamma}(s, t) := \widehat{\gamma}_n(s, t)$ $(0 \leq s < t < \infty, \ t - s \geq 1/n)$ 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 (see Hoga, 2015, for a motivation)

$$\widehat{x}_p(s, t) := X_k(s, t, 1)\left(\frac{np}{k}\right)^{-\widehat{\gamma}(s,t)} \tag{4}$$

is the so called Weissman (1978) estimator for the extreme quantile $U(1/p)$. To give an example of a tail index estimator we introduce the Hill (1975) estimator of γ

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

where $X_{n:n} \geq X_{n-1:n} \geq \ldots \geq X_{1:n}$ denote the order statistics of the sample $X_1, \ldots, X_n$.

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

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

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 (2007).

Write $D[a, b]$ for the space of cadlag functions on $[a, b]$ $(0 \leq a < b < \infty)$ endowed with the Skorohod topology and $D(I)$, $I \subset \mathbb{R}^l$ compact, for the multiparameter 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:

3

(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 (2014, 2015).

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

$$
\begin{aligned}
\mathcal{H}_{0,\gamma}: \quad & \gamma_1 = \ldots = \gamma_n = \gamma_{n+1} = \ldots \quad \text{vs.} \\
\mathcal{H}_{1,\gamma}^{\lessgtr}: \quad & \gamma_1 = \ldots = \gamma_n = \ldots = \gamma_{\lfloor nt^* \rfloor} \lessgtr \gamma_{\lfloor nt^* \rfloor + 1} = \gamma_{\lfloor nt^* \rfloor + 2} = \ldots,
\end{aligned}
\tag{5}
$$

and

$$
\begin{aligned}
\mathcal{H}_{0,U}: \quad & U_1\left(\frac{1}{p}\right) = \ldots = U_n\left(\frac{1}{p}\right) = U_{n+1}\left(\frac{1}{p}\right) = \ldots \quad \text{versus} \\
\mathcal{H}_{1,U}^{\lessgtr}: \quad & U_1\left(\frac{1}{p}\right) = \ldots = U_n\left(\frac{1}{p}\right) = \ldots = U_{\lfloor nt^* \rfloor}\left(\frac{1}{p}\right) \lessgtr U_{\lfloor nt^* \rfloor + 1}\left(\frac{1}{p}\right) = U_{\lfloor nt^* \rfloor + 2}\left(\frac{1}{p}\right) = \ldots
\end{aligned}
\tag{6}
$$

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

We use the following detectors for (5)

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

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

and for (6)

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

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

where $t_0 > 0$ defines the (minimal) fraction of n upon which the tail index estimators are based. We are inclined to reject $\mathcal{H}_0$ iff 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

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

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

4

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, Hörmann, Horváth, Hušková and Steinebach, 2012; Wied and Galeano, 2013).

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 (again, e.g., Aue *et al.*, 2012; Wied and Galeano, 2013). 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

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

where $\sigma_{\widehat{\gamma},\gamma}^2$ 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 (2014)). It turns out however, that in simulations values of even $n = 500$ for the training period are not sufficient to deliver accurate variance estimates, which distorts the sizes of our surveillance methods severely. This is why we opted for the self-normalized approach advocated in Shao and Zhang (2010), which delivers far superior size in simulations. Simulation evidence in Shao and Zhang (2010) suggests that for retrospective tests in finite samples 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

$$\lim_{n\to\infty} \frac{n}{r_n} \beta(l_n) + \frac{r_n}{\sqrt{k}} \log^2(k) = 0 \tag{7}$$

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 $x, y \in [0, y_0 + \varepsilon]$ $(y_0 \geq 1, \varepsilon > 0)$

$$\lim_{n\to\infty} \frac{n}{r_n k} \sum_{1 \leq i,j \leq r_n} Cov\left(I_{\left\{ X_i > U\left(\frac{n}{kx}\right) \right\}}, I_{\left\{ X_j > U\left(\frac{n}{ky}\right) \right\}} \right) = r(x, y). \tag{8}$$

(A4) For some constant $C > 0$

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

5

(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

$$\frac{U(1/p)}{U(n/k)} \left(\frac{np}{k}\right)^\gamma - 1 = o\left(\frac{1}{\sqrt{k}}\right). \tag{9}$$

The conditions **(A2)**-**(A5)** correspond to conditions **(C1)**-**(C4)** in Hoga (2014). For a discussion of these conditions, as well as some easier-to-verify sufficient conditions for **(A3)** and **(A4)** see Hoga (2014), where also some examples of time series satisfying these conditions are presented. Conditions **(A6)**-**(A7)** correspond to conditions **(C5)**-**(C6)** in Hoga (2015). **(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 $\lim_{n\to\infty} k^{-1/2} \log(np/k) = \lim_{n\to\infty} k^{-1/2} \log(np) = 0$ provides a lower bound, beyond which estimation is no longer feasible.

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. Here and henceforth we will assume that these stopping times are based on one of the tail index estimators from Examples 1 - 3 in Hoga (2014).

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

$$V_{t_0,T} := \frac{\sup_{t \in [1+t_0,T]} [W(t) - tW(1)]^2}{\int_{t_0}^1 [W(s) - sW(1)]^2 \, ds}, \tag{10}$$

$$W_{t_0,T} := \frac{\sup_{t \in [1+t_0,T]} [W(t) - W(t-t_0) - t_0 W(1)]^2}{\int_{t_0}^1 [W(s) - W(s-t_0) - t_0 W(1)]^2 \, ds},$$

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

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

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

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

(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*

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

respectively. Suppose further that

$$r_{\text{pre}}(tx, ty) = t r_{\text{pre}}(x, y) \quad \text{and} \quad r_{\text{post}}(tx, ty) = t r_{\text{post}}(x, y) \quad \forall t \in [0, 1], x, y \in [0, y_0]. \tag{11}$$

Then

(i)

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

$$\lim_{n\to\infty} P\left(\tau_n^{V^{\widehat{\gamma}}} < \infty\right) = P\left(V_{t_0,T} > c\right) \qquad\qquad \text{under } \mathcal{H}_{1,\gamma}^{>},$$

(ii)

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

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)

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

(iv)

$$\lim_{n\to\infty} P\left(\tau_n^{W^{\widehat{x}_p}} < \infty\right) = 1 \text{ if } t^* \in [1, T - t_0] \qquad\qquad \text{under } \mathcal{H}_{1,U}^{\lessgtr}.$$

Remark 2. (a) For a discussion of condition (11) and the choice of k we refer to Hoga (2014, Remark 8).

(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 $U_{\mathrm{pre}}/U_{\mathrm{post}} \in RV_{\gamma_{\mathrm{pre}}-\gamma_{\mathrm{post}}>0}$ and hence

$$\frac{U_{\mathrm{pre}}(1/p_n)}{U_{\mathrm{post}}(1/p_n)} \xrightarrow[(n\to\infty)]{} \infty,$$

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

(c) Under $\mathcal{H}_{1}^{>}$ the procedure based on $\tau_n^{V^{\widehat{\gamma}}}$ has power equal to size, which motivated the study of $\tau_n^{W^{\widehat{\gamma}}}$.

3 Simulations

In this section we investigate the finite-sample behavior of the monitoring procedures based on the stopping times $\tau_n^{W^{\widehat{\gamma}/\widehat{x}_p}}$ since, unlike with $\tau_n^{V^{\widehat{\gamma}/\widehat{x}_p}}$, they are generally consistent against both alternatives $\mathcal{H}_1^{\lessgtr}$. Throughout we simulate 5,000 times from time series with historical periods of length $n = 100, 500$ 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, which is in the class of kernel estimators in Example 1 in Hoga (2014). Simulation results were quite robust to the particular choice of k/n and the tail index estimator and are available from the authors upon request. The quantiles of the distribution of $V_{t_0,T}$ for $t_0 = 0.2$ and $T = 4$ given in (10) are tabulated in Table 1.

α	0.50	0.60	0.70	0.80	0.90	0.95	0.99
α-quantile	15.3	18.1	21.7	26.8	36.2	45.4	71.3

Table 1: α-quantiles of $V_{t_0,T}$ ($t_0 = 0.2$, $T = 4$)

We investigate size using data from a linear ARMA(1,1) and a non-linear ARCH(1) model. Specifically, we simulate $\{X_i\}_{i=1,\ldots,\lfloor nT \rfloor}$ from the two data generating processes (DGPs)

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

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

where t_{10} denotes Student's t-distribution with 10 degrees of freedom (i.e., tail index equal to 10). For the verification of the conditions **(A2)**-**(A7)** we refer to Hoga (2015) and the references cited therein. The $|X_i|$ generated from the ARMA(1,1)-model have tail index 10 because of Lemma 5.2 in Datta

DGP	Size	$\lfloor nT \rfloor = 500$				$\lfloor nT \rfloor = 2000$			
		$W^{\widehat{\gamma}}$	$W^{\widehat{x}_p}$			$W^{\widehat{\gamma}}$	$W^{\widehat{x}_p}$		
			$p=0.1$	0.01	0.001		$p=0.1$	0.01	0.001
(ARMA)	0.10	0.194	0.111	0.147	0.179	0.100	0.085	0.089	0.095
	0.05	0.129	0.068	0.097	0.120	0.054	0.042	0.049	0.055
(ARCH)	0.10	0.173	0.105	0.163	0.175	0.117	0.092	0.125	0.124
	0.05	0.110	0.063	0.107	0.116	0.067	0.051	0.077	0.073

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

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).

For both models tests are oversized for $\lfloor nT \rfloor = 500$ and rejections depend quite significantly on p. For $\lfloor nT \rfloor = 2000$ this dependence decreases and the size distortions vanish almost completely for the ARMA data and are markedly less severe for the ARCH data.

DGP	Break	Level	$\lfloor nT \rfloor = 500$				$\lfloor nT \rfloor = 2000$			
			$W^{\widehat{\gamma}}$	$W^{\widehat{x}_p}$			$W^{\widehat{\gamma}}$	$W^{\widehat{x}_p}$		
				$p=0.1$	0.01	0.001		$p=0.1$	0.01	0.001
(ARMA)	0.05	0.10	0.241	0.471	0.408	0.344	0.226	0.984	0.795	0.575
		0.05	0.170	0.354	0.309	0.252	0.145	0.958	0.686	0.450
	0.5	0.10	0.223	0.405	0.315	0.276	0.202	0.969	0.708	0.478
		0.05	0.150	0.290	0.232	0.201	0.132	0.932	0.590	0.359
(ARCH)	0.05	0.10	0.365	0.305	0.453	0.428	0.519	0.613	0.692	0.651
		0.05	0.265	0.228	0.361	0.337	0.408	0.496	0.597	0.546
	0.5	0.10	0.268	0.237	0.330	0.315	0.372	0.485	0.539	0.490
		0.05	0.185	0.164	0.252	0.228	0.277	0.374	0.438	0.389

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

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

$$X_{i,n} = \begin{cases} 0.3 \cdot X_{i-1} + Z_i + 0.7 \cdot Z_{i-1}, & i = 1, \ldots, \lfloor nt^* \rfloor, \\ 0.8 \cdot X_{i-1} + Z_i + 0.7 \cdot Z_{i-1}, & i = \lfloor nt^* \rfloor + 1, \ldots, \lfloor nT \rfloor, \end{cases} \qquad Z_i \overset{\text{i.i.d.}}{\sim} t_{10}, \qquad (12)$$

$$X_{i,n}^2 = \begin{cases} \left(0.01 + 0.3125 \cdot X_{i-1}^2\right) Z_i^2, & i = 1, \ldots, \lfloor nt^* \rfloor, \\ \left(0.01 + 0.5773 \cdot X_{i-1}^2\right) Z_i^2, & i = \lfloor nt^* \rfloor + 1, \ldots, \lfloor nT \rfloor, \end{cases} \qquad Z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \qquad (13)$$

where $n = 100, 500$ and $T = 4$ as before and $t^* = 1.15, 2.5$, corresponding to breaks after 5% and 50% of the observation period. 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.

Note that for the ARMA(1,1) model in (12) $\mathcal{H}_{0,\gamma}$ 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 $\tau_n^{W\widehat{\gamma}}$ than in Table 2. This is confirmed by the results in Table 3. Furthermore, the variance change is most frequently detected using $\tau_n^{W\widehat{x}_p}$ for $p = 0.1$. This may be explained by the higher variance of $\widehat{x}_p$ 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 occured.

If there is a break in the tail index and the variance as in the ARCH-case, one can see from Table 3 that the procedures based on the Weissman (1978) estimator clearly perform better than that based 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 (via $X_k(s,t,1)$, see (4)). Since in reality changes in the tail index will most likely result in changes of scale, one should use the tests based on $W^{\widehat{x}_p}$. Further, one should use $\tau_n^{W\widehat{x}_p}$ with $p = 0.01$, as this choice leads to the highest power.

To assess how promptly changes are detected, Figures 1 and 2 show histograms of the (finite) realizations of $\tau_n^{W\widehat{\gamma}}$ and $\tau_n^{W\widehat{x}_p}$ at the 10% level for the ARMA and the ARCH models given in (12) and (13) respectively with $\lfloor nT \rfloor = 2000$. There are 19 bars in all plots at points $1 + l \cdot 0.05 \cdot 3$ $(l = 1, \ldots, 20)$. The value of t^* at which the changes are located are given by $1.15 = 1 + 1 \cdot 0.05 \cdot 3$ $(l = 1)$ and $2.5 = 1 + 10 \cdot 0.05 \cdot 3$ $(l = 10)$.

The results for the ARMA model are displayed in Figure 1. The high false detection rate for the tail index-based method seems largely to be due to false detections just shortly after the break, as can be seen from the distinctive peak in panel (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. 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 variance before one can reject the null here.
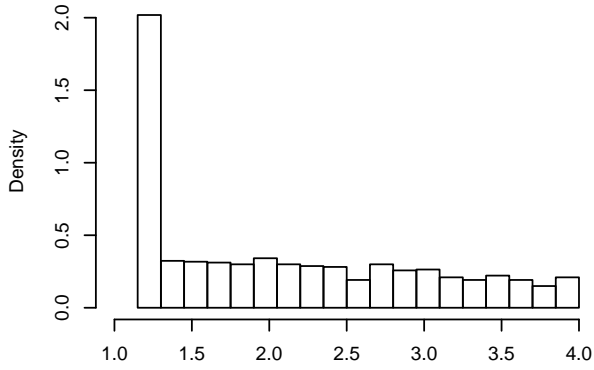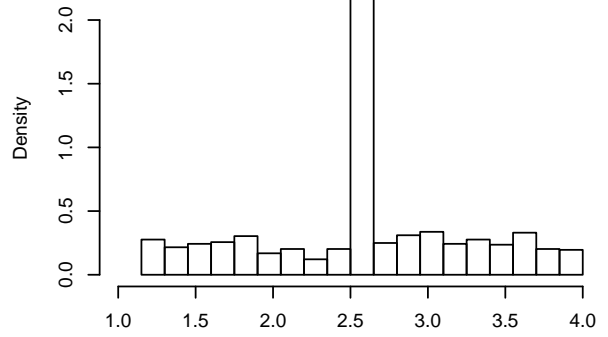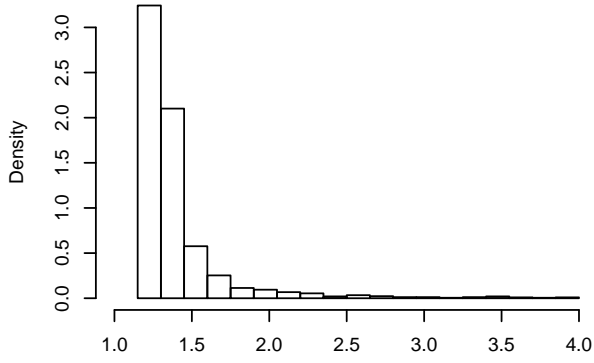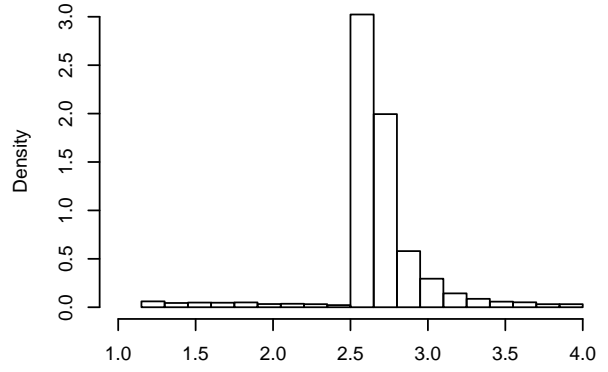
10
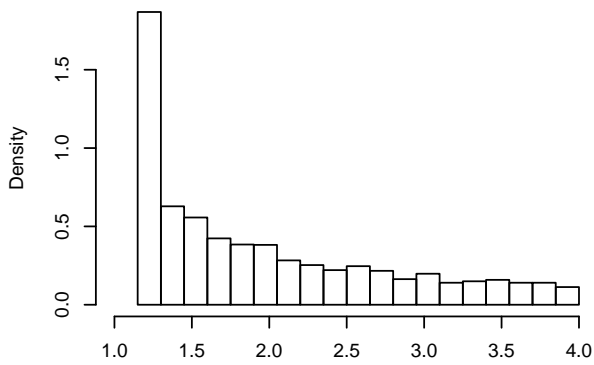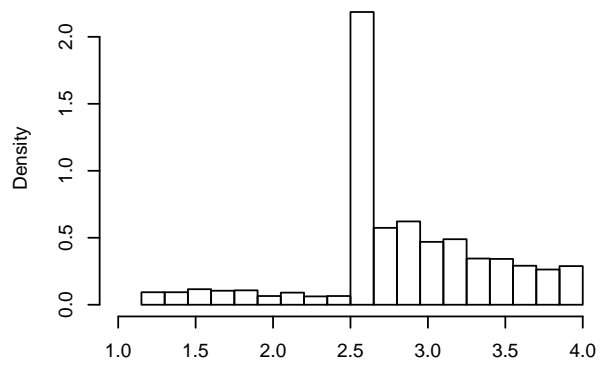
Figure 1: Histograms of detection times $\tau_n^{W^{\widehat{\gamma}}}$, $\tau_n^{W^{\widehat{x}_p}}$ for $p = 0.1, 0.001$ for (12) and $t^* = 1.15$ ((a1), (b1), (c1)), $t^* = 2.5$ ((a2), (b2), (c2)) for $\lfloor nT \rfloor = 2000$
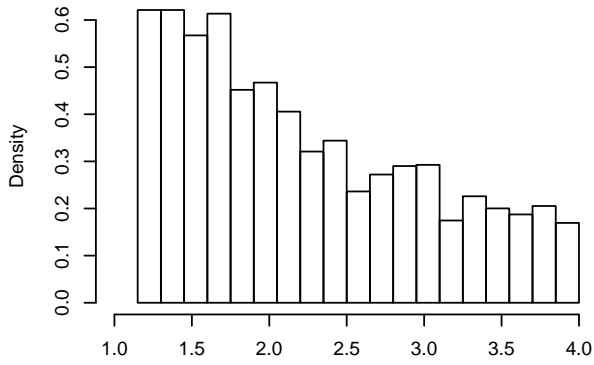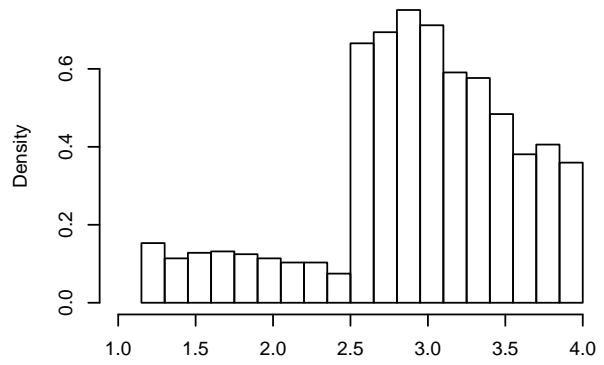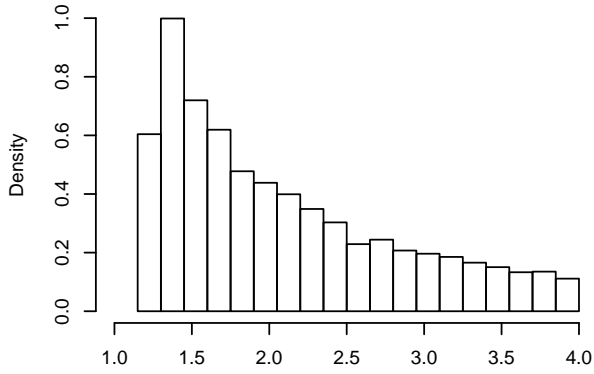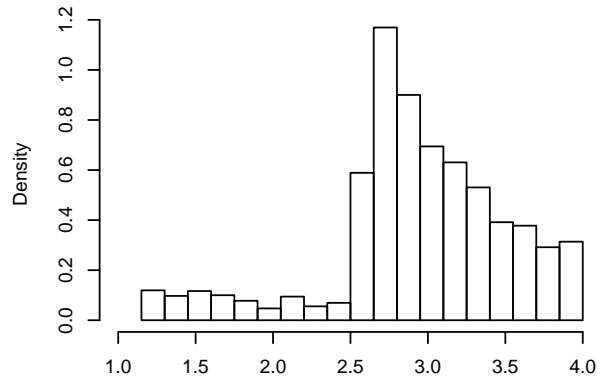
11
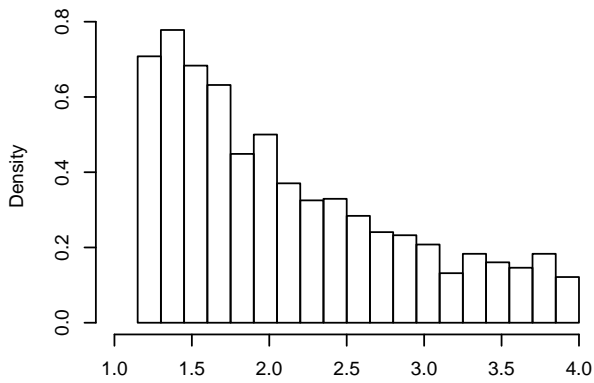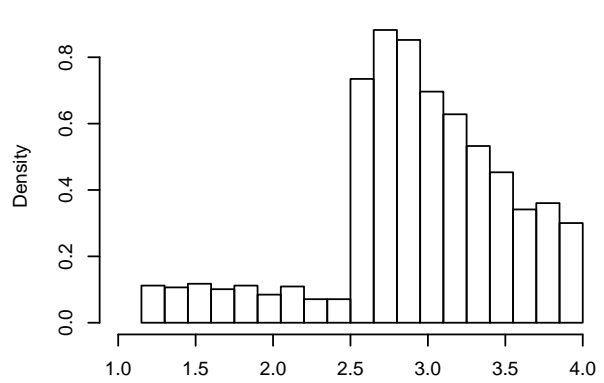
Figure 2: Histogram of detection times $\tau_n^{W^{\widehat{\gamma}}}$, $\tau_n^{W^{\widehat{x}_p}}$ for $p = 0.1, 0.001$ for (13) and $t^* = 1.15$ ((a1), (b1), (c1)), $t^* = 2.5$ ((a2), (b2), (c2)) for $\lfloor nT \rfloor = 2000$

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 2012 observations, $X_1, \ldots, X_{2012}$, the first quarter of which (roughly the years 2005 and 2006) make up the training period.
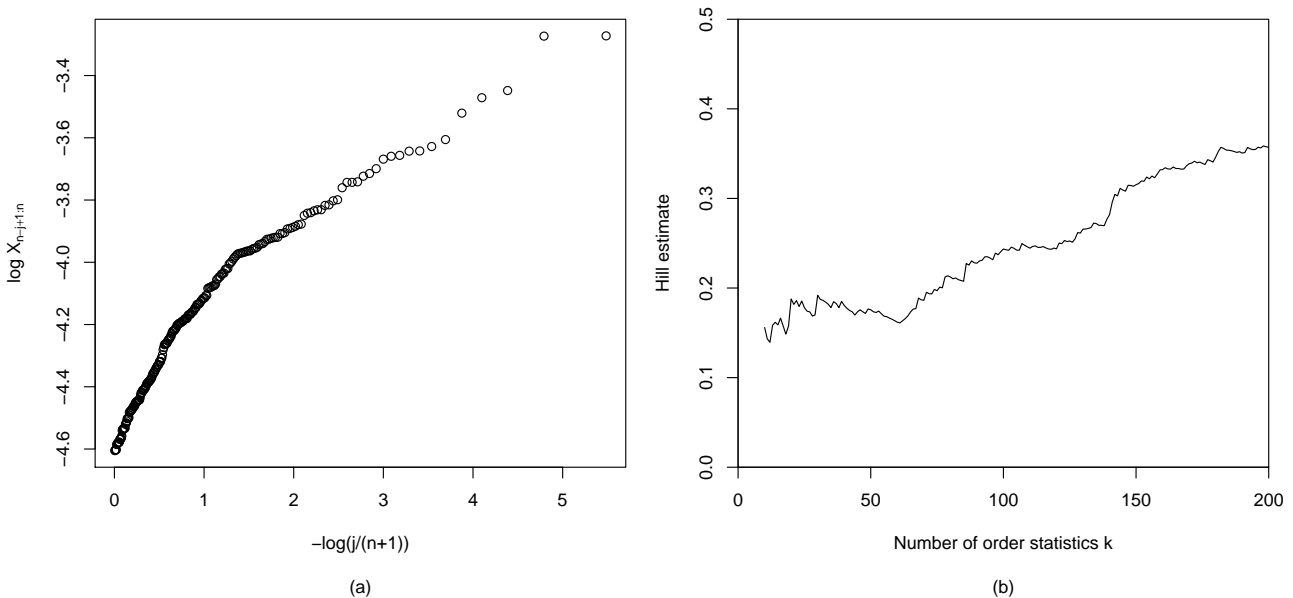

Figure 3: (a) Pareto quantile plot of shifted data. (b) Hill estimates as a function of upper order statistics k used in the estimation.

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 α can be obtained by the slope of the line from the point $(-\log((k+1)/(n+1)), \log X_{n-k:n})$ onwards, 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)** we remark that GARCH-models, which are widely used to model financial return data like the one considered here, are geometrically strong mixing, i.e., $\beta(k) = \eta^k$, $\eta \in (0, 1)$. Further, applying the retrospective tests of Hoga (2014) and Hoga (2015) we find 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.

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 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 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 occured. 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: The proof of *(i)* mainly rests on a time shifted version of the convergence established in Hoga (2014, Theorem 5) (cf. also the proof of Theorem 1 in Hoga (2014)). That is, for some $\delta > 0$ and $\nu \in (0, 1/2)$,

$$\frac{\sqrt{k}}{y^\nu} \left\{ \frac{1}{k} \sum_{i=1}^{\lfloor nt \rfloor} \left[I_{\{\widetilde{U}_i > 1 - \frac{k}{n}y\}} - \frac{k}{n}y \right] \right\} \xrightarrow[(n\to\infty)]{\mathcal{D}} \frac{1}{y^\nu} W(t, y) \quad \text{in } D\left([t_0, T] \times [0, y_0 + \delta]\right)$$
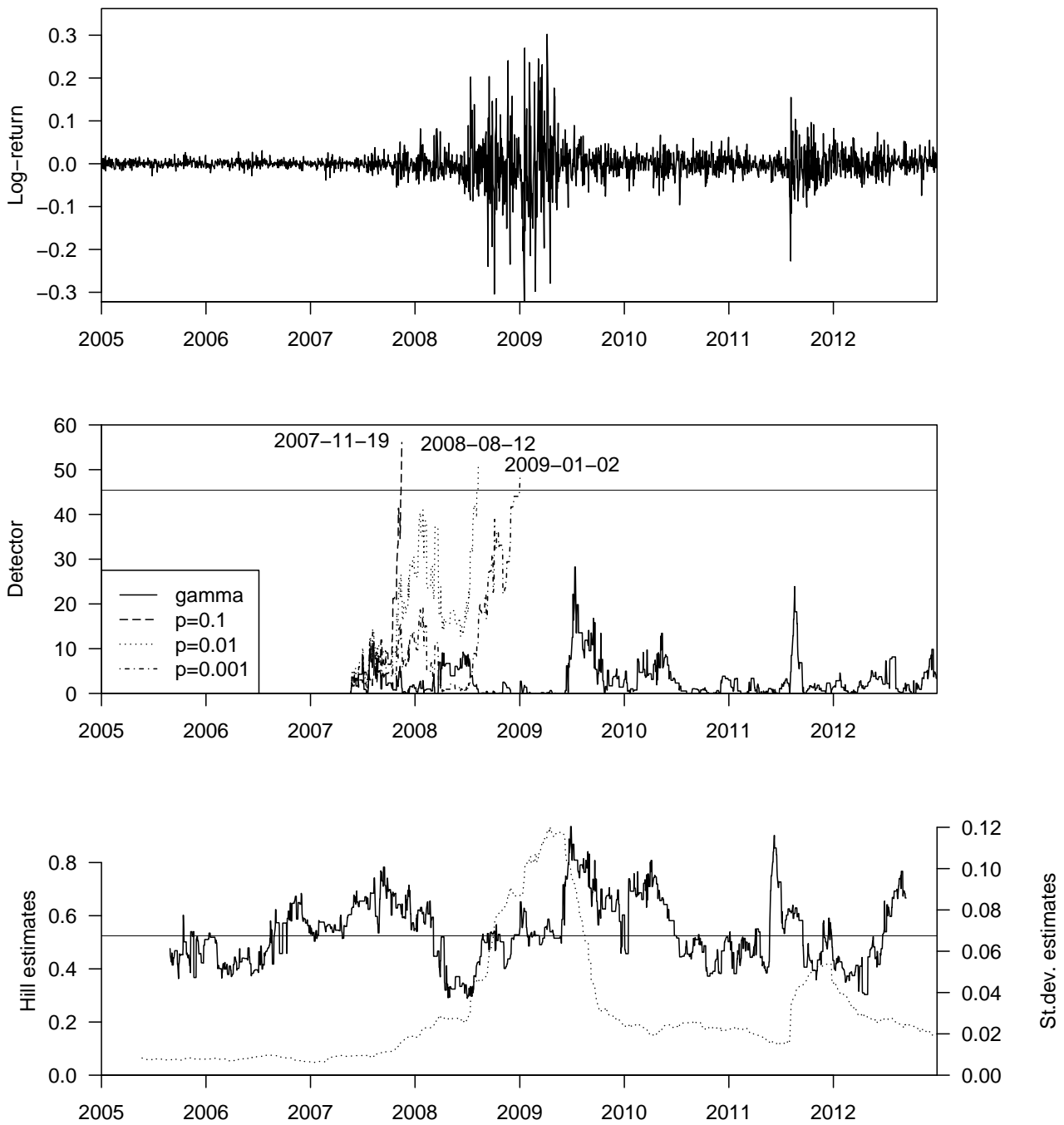
Figure 4: Top panel: Log-returns of Bank of America stock. Middle panel: Values of detectors $W^{\widehat{\gamma}}$ (gamma in solid), and $W^{\widehat{x}_p}$ for $p = 0.1, 0.001, 0.0001$ (dashed, dotted, dash-dotted) and value of 5%-threshold (horizontal solid line). Bottom panel: Rolling Hill estimate (jagged solid line), Hill estimate based on training period (straight solid line), and standard deviation estimates (dotted line).

for a sequence of random variables $\widetilde{U}_i \sim \mathcal{U}[0,1]$, s.t.

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

$W(t,y)$ is a centered Gaussian process with covariance function

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

Hence,

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

The continuous mapping theorem (CMT) then implies

$$\frac{\sqrt{k}}{y^\nu}\begin{pmatrix} \frac{1}{k}\sum_{i=n+1}^{\lfloor n\max(1+t_0,t)\rfloor} I_{\left\{X_i > U\left(\frac{n}{ky}\right)\right\}} - (\max(1+t_0,t)-1)y \\ \frac{1}{k}\sum_{i=\lfloor n(t-t_0)\rfloor}^{\lfloor nt \rfloor} I_{\left\{X_i > U\left(\frac{n}{ky}\right)\right\}} - t_0 y \\ \frac{1}{k}\sum_{i=1}^{\lfloor nt \rfloor} I_{\left\{X_i > U\left(\frac{n}{ky}\right)\right\}} - ty \end{pmatrix}$$

$$\xrightarrow[(n\to\infty)]{\mathcal{D}} \frac{1}{y^\nu}\begin{pmatrix} W(\max(1+t_0,t),y) - W(1,y) \\ W(t,y) - W(t-t_0,y) \\ W(t,y) \end{pmatrix}$$

in $D^3\left([t_0, T] \times [0, y_0 + \delta]\right)$. Following the steps of the proof of Theorem 1 in Hoga (2014) we get (on a suitable Skorohod probability space)

$$\sup_{\substack{t\in[t_0,T] \\ y \geq y_0^{-1/\gamma}-\delta}} \frac{1}{y^{\nu/\gamma}}\left| \sqrt{k}\begin{pmatrix} \frac{1}{k}\sum_{i=n+1}^{\lfloor n\max(1+t_0,t)\rfloor} I_{\left\{X_i > U\left(\frac{n}{ky}\right)\right\}} - (\max(1+t_0,t)-1)y^{-1/\gamma} \\ \frac{1}{k}\sum_{i=\lfloor n(t-t_0)\rfloor}^{\lfloor nt \rfloor} I_{\left\{X_i > U\left(\frac{n}{ky}\right)\right\}} - t_0 y^{-1/\gamma} \\ \frac{1}{k}\sum_{i=1}^{\lfloor nt \rfloor} I_{\left\{X_i > U\left(\frac{n}{ky}\right)\right\}} - ty^{-1/\gamma} \end{pmatrix}\right.$$

$$\left. - \begin{pmatrix} W(\max(1+t_0,t),y) - W(1,y) \\ W(t,y) - W(t-t_0,y) \\ W(t,y) \end{pmatrix}\right| \xrightarrow[(n\to\infty)]{\text{a.s.}} 0.$$

And again retracing the proof of Hoga (2014, Corollary 1) we can see that

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

$$\left. - \begin{pmatrix} W(\max(1+t_0,t),y) - W(1,y) - y^{-1/\gamma}\left[W(\max(1+t_0,t),1) - W(1,1)\right] \\ W(t,y) - W(t-t_0,y) - y^{-1/\gamma}\left[W(t,1) - W(t-t_0,1)\right] \\ W(t,y) - y^{-1/\gamma}W(t,1) \end{pmatrix}\right| \xrightarrow[(n\to\infty)]{\text{a.s.}} 0.$$

From this the following joint convergence can be established similarly as in Examples 1 - 3 of Hoga (2014):

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

$$\tag{A.1}$$

in $D^3[t_0, T]$. From this and the CMT we obtain

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

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

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

For part *(ii)* we observe that it follows from (A.1) similarly as in the proof of Theorem 1 in Hoga (2015) that

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

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

whence with the CMT again

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

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

The conclusion follows as before. $\qquad \square$

Proof of Theorem 2: Under $\mathcal{H}_{1,\gamma}^>$ we get by an adaptation of the proof of Hoga (2014, Theorem 3)

$$\sqrt{k} \left(\widehat{\gamma}(1, t) - \gamma_{\mathrm{pre}} \right) \xrightarrow[(n \to \infty)]{\mathcal{D}} \sigma_{\widehat{\gamma}, \gamma} \frac{W(t) - W(1)}{t - 1} \quad \text{in } D[1 + t_0, T]$$

(cf. also the proof of Theorem 1) and by a further close inspection that even

$$\sqrt{k} \begin{pmatrix} \widehat{\gamma}(1, \max(1 + t_0, t)) - \gamma_{\mathrm{pre}} \\ \widehat{\gamma}(0, t) - \gamma_{\mathrm{pre}} \end{pmatrix} \xrightarrow[(n \to \infty)]{\mathcal{D}} \sigma_{\widehat{\gamma}, \gamma} \begin{pmatrix} (W(\max(1 + t_0, t)) - W(1))/(\max(1 + t_0, t) - 1) \\ W(t)/t \end{pmatrix}$$
(A.2)

jointly in $D[t_0, T] \times \mathbb{R}$. Hence, the result for $\tau_n^{V^{\widehat{\gamma}}}$ in *(i)* follows as in the proof of Theorem 1. For the other part of *(i)* it suffices to note that

$$\widehat{\gamma}(0, 1) \xrightarrow[(n \to \infty)]{P} \gamma_{\mathrm{pre}} < \gamma_{\mathrm{post}} \xleftarrow[(n \to \infty)]{P} \widehat{\gamma}(1, T),$$

because by Hoga (2014, Theorem 3) $\widehat{\gamma}$ will converge in probability to the dominant tail index (i.e., the larger value of γ) in a sample with one tail index break.

For *(ii)* simply note that

$$\widehat{\gamma}(0, 1) \xrightarrow[(n \to \infty)]{P} \gamma_{\mathrm{pre}} < \gamma_{\mathrm{post}} \xleftarrow[(n \to \infty)]{P} \widehat{\gamma}(t^*, t^* + t_0).$$

17

If $\gamma_{\text{pre}} \neq \gamma_{\text{post}}$ in case *(iii)* we can deduce by arguments similar to those in the proof of Theorem 2 in Hoga (2015) that

$$\frac{\widehat{x}_p(0,1)}{U_{\text{pre}}(1/p)} \xrightarrow[(n\to\infty)]{P} 1, \quad \frac{\widehat{x}_p(1,T)}{U_{\text{pre}}(1/p)} \xrightarrow[(n\to\infty)]{P} (t^* - 1)^{\gamma_{\text{pre}}}.$$

The result follows.

Part *(iv)* is trivial, since

$$\frac{\widehat{x}_p(0,1)}{U_{\text{pre}}(1/p)} \xrightarrow[(n\to\infty)]{P} 1, \quad \frac{\widehat{x}_p(t^*,t^*+t_0)}{U_{\text{post}}(1/p)} \xrightarrow[(n\to\infty)]{P} 1,$$

where $U_{\text{pre}}(1/p) \neq U_{\text{post}}(1/p)$. $\qquad\square$

References

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

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

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

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.

Einmahl J, de Haan L, Zhou C. 2015+. Statistics of heteroscedastic extremes. *Journal of the Royal Statistical Society: Series B* : 1–21.

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

Hoga Y. 2014. Change point tests for the tail index of β-mixing random variables. second resubmission to *Econometric Theory* pending.

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

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

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.

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

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