Inference on the tail process with application to financial time series modelling

Richard A. Davis
Columbia University
Department of Statistics
1255 Amsterdam Ave. New York, NY 10027, USA
rdavis@stat.columbia.edu

Holger Drees
University of Hamburg
Department of Mathematics
Bundesstraße 55, 20146 Hamburg, Germany
holger.drees@math.uni-hamburg.de

Johan Segers   Michał Warchoł
Université catholique de Louvain
Institut de Statistique, Biostatistique et Sciences Actuarielles
Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium
johan.segers@uclouvain.be, michal.warchol@uclouvain.be

March 26, 2017

Abstract

To draw inference on serial extremal dependence within heavy-tailed Markov chains, Drees, Segers and Warchoł [Extremes (2015) 18, 369–402] proposed nonparametric estimators of the spectral tail process. The methodology can be extended to the more general setting of a stationary, regularly varying time series. The large-sample distribution of the estimators is derived via empirical process theory for cluster functionals. The finite-sample performance of these estimators is evaluated via Monte Carlo simulations. Moreover, two different bootstrap schemes are employed which yield confidence intervals for the pre-asymptotic spectral tail process: the stationary bootstrap and the multiplier block bootstrap. The estimators are applied to stock price data to study the persistence of positive and negative shocks.

Keywords: Financial time series; Heavy–tails; Multiplier block bootstrap; Regular variation; Shock persistence; Stationary time series; Tail process.

1 Introduction

The typical modelling paradigm for a time series often starts by choosing a flexible class of models that captures salient features present in the data. Of course, *features* depends on the type of characteristics one is looking for. For a financial time series consisting of say log-returns of some asset, the key features, often referred to as *stylized facts*, include heavy-tailed marginal distributions and serially uncorrelated but dependent data. These characteristics are readily detected using standard diagnostics such as qq-plots of the marginal distribution and plots of the sample autocorrelation function (ACF) of the data and the squares of the data. The GARCH process (and its variants) as well as the stochastic volatility (SV) process driven by heavy-tailed noise exhibit these attributes and often serve as a starting point for building a model. More recently, considerable attention has been directed towards studying the extremal behavior of both financial and environmental time series, especially as it relates to estimating risk factors. Extremes for such time series can occur in clusters and getting a handle on the nature of clusters both in terms of size and frequency of occurrence is important for evaluating various risk measures. Ultimately, one wants to choose models that adequately describe various extremal dependence features observed in the data. The theme of this paper is to provide additional tools that not only give measures of extremal dependence, but can be used as a basis for assessing the quality of a model's fit to extremal properties present in the data.

The *extremal index* $\theta \in (0, 1]$ (Leadbetter, 1983) is one such measure of extremal dependence for a stationary time series. It is a measure of extremal clustering ($1/\theta$ is the mean cluster size of extremes) with $\theta < 1$ indicating clustering and $\theta = 1$ signifying no clustering in the limit. Unfortunately, θ is a rather crude measure and does not provide fine detail about extremal dependence. The extremogram, developed in Davis and Mikosch (2009), is an attempt to provide a measure of serial dependence among the extremes in a stationary time series. It was conceived to be used in much the same way as an ACF in traditional time series modelling, but only applied to extreme values.

In this paper, we will use the spectral tail process, as formulated by Basrak and Segers (2009) for heavy-tailed time series, to assess and measure extremal dependence. The spectral tail process provides a more in-depth description of the structure of extremal dependence than the extremogram. The first objective of this paper will be to establish limit theory for nonparametric estimates of the distribution of the spectral tail process for a class of heavy-tailed stationary time series. This builds on earlier work of Drees et al. (2015) for heavy-tailed Markov chains. The nonparametric estimates provide quantitative information about extremal dependence within a time series and as such can be used in both exploratory and confirmatory phases of modelling. As an example, it provides estimates of the probability that an extreme observation will occur at time t, given one has occurred at time 0, and that its absolute value will be even larger. These estimates can also be used for model confirmation, in much the same way that the ACF is used for assessing quality of fit for second-order models of time series. For example, one can compute a *pre-asymptotic* version (to be defined later) of the distribution of the spectral tail process from a GARCH process, which in most cases can be easily calculated via simulation. Then the estimated distribution of the spectral tail process can be compared with the *pre-asymptotic* version corresponding to a model for compatibility. A good fit would indicate the plausibility of using a GARCH model for capturing serial extremal dependence. The second main objective is then to provide a useful way of measuring compatibility, which we propose using resampling methods.

Recently, there has been increasing interest in the econometric literature for estimating quantities related to extremal dependence. For stochastic processes in continuous time, Bollerslev et al. (2013) define a χ-coefficient, derived from the extremogram, for assessing tail dependencies applied to financial time series. In a follow-up paper that explores tail risk premia, Bollerslev et al. (2015) make a connection between their estimates of the time-varying tail shape parameters and the extremogram. Linton and Whang (2007) (see also Han et al. (2016)) introduced the quantilogram, a diagnostic tool for measuring directional predictability in a time series. In some respects, our development can be viewed as the *quantilogram* for extreme quantiles. The theory, however, is different in that our quantiles are going to infinity. Nevertheless, our work does focus on a type of *directional predictability*, but only concentrated in the extremes. Tjøstheim and Hufthammer (2013) consider local Gaussian correlation and relate it to tail index dependence and the extremogram in a time series context. Their methodology is applied to financial time series.

The key object of study in this paper is the tail process and in particular, its normalized version – the spectral tail process. A *strictly* stationary univariate time series $(X_t)_{t\in\mathbb{Z}}$ is said to have a *tail process* $(Y_t)_{t\in\mathbb{Z}}$ if, for all integers $s \le t$, we have

$$\mathcal{L}\left(u^{-1}X_s, \ldots, u^{-1}X_t \mid |X_0| > u\right) \xrightarrow{d} \mathcal{L}\left(Y_s, \ldots, Y_t\right), \qquad u \to \infty, \tag{1.1}$$

with the implicit understanding that the law of $|Y_0|$ is non-degenerate. The law of $|Y_0|$ is then necessarily Pareto(α) for some $\alpha > 0$ and the function $u \mapsto \mathrm{P}[|X_0| > u]$ is regularly varying at infinity with index $-\alpha$:

$$\lim_{u\to\infty} \frac{\mathrm{P}[|X_0| > uy]}{\mathrm{P}[|X_0| > u]} = \mathrm{P}[|Y_0| > y] = y^{-\alpha}, \qquad y \in [1, \infty). \tag{1.2}$$

The existence of a tail process is equivalent to multivariate regular variation of the finite-dimensional distributions of $(X_t)_{t\in\mathbb{Z}}$ (Basrak and Segers, 2009, Theorem 2.1). In many respects, this condition can be viewed as the heavy-tailed analogue of the condition that a process is Gaussian in the sense that all the finite-dimensional distributions are specified to be of a certain type.

The *spectral tail process* is defined by $\Theta_t = Y_t/|Y_0|$, for $t \in \mathbb{Z}$. By (1.1), it follows that for all integers $s \le t$, we have

$$\mathcal{L}\left(X_0/u, X_s/|X_0|, \ldots, X_t/|X_0| \mid |X_0| > u\right) \xrightarrow{d} \mathcal{L}\left(Y_0, \Theta_s, \ldots, \Theta_t\right), \qquad u \to \infty. \tag{1.3}$$

The difference between (1.1) and (1.3) is that in the latter equation, the variables X_t have been normalized by $|X_0|$ rather than by the threshold u. Such auto-normalization allows the tail process to be decomposed into two stochastically independent components, i.e.,

$$Y_t = |Y_0|\,\Theta_t, \qquad t \in \mathbb{Z}.$$

Independence of $|Y_0|$ and $(\Theta_t)_{t\in\mathbb{Z}}$ is stated in Basrak and Segers (2009, Theorem 3.1). The random variable $|Y_0|$ characterizes the magnitudes of extremes, whereas $(\Theta_t)_{t\in\mathbb{Z}}$ captures serial dependence. The spectral tail process at time $t = 0$ yields information on the relative weights of the upper and lower tails of $|X_0|$: since $\Theta_0 = Y_0/|Y_0| = \operatorname{sign}(Y_0)$, we have

$$p = \mathrm{P}[\Theta_0 = +1] = \lim_{u\to\infty} \frac{\mathrm{P}[X_0 > u]}{\mathrm{P}[|X_0| > u]}, \qquad\qquad 1 - p = \mathrm{P}[\Theta_0 = -1]. \tag{1.4}$$

The distributions of the *forward* tail process $(Y_t)_{t\ge 0}$ and the *backward* tail process $(Y_t)_{t\le 0}$ mutually determine each other (Basrak and Segers, 2009, Theorem 3.1). For all $i, s, t \in \mathbb{Z}$ with $s \le 0 \le t$ and for all measurable functions $f : \mathbb{R}^{t-s+1} \to \mathbb{R}$ satisfying $f(y_s, \ldots, y_t) = 0$ whenever $y_0 = 0$, we have, provided the expectations exist,

$$\mathrm{E}[f(\Theta_{s-i}, \ldots, \Theta_{t-i})] = \mathrm{E}\left[f\left(\frac{\Theta_s}{|\Theta_i|}, \ldots, \frac{\Theta_t}{|\Theta_i|}\right) |\Theta_i|^\alpha\, \mathbf{1}\{\Theta_i \neq 0\}\right] \tag{1.5}$$

The indicator variable $\mathbf{1}\{\Theta_i \neq 0\}$ can be omitted because of the presence of $|\Theta_i|^\alpha$, but sometimes, it is useful to mention it explicitly in order to avoid errors arising from division by zero. By exploiting the 'time-change formula' (1.5), we will be able to improve upon the efficiency of estimators of the spectral tail process.

Main interest in this paper is in the cumulative distribution function (cdf), $F^{(\Theta_t)}$, of Θ_t. If $F^{(\Theta_t)}$ is continuous at a point x, then

$$\lim_{u\to\infty} \mathrm{P}[X_t/|X_0| \le x \mid |X_0| > u] = \mathrm{P}[\Theta_t \le x] = F^{(\Theta_t)}(x). \tag{1.6}$$

We consider two estimates of $F^{(\Theta_t)}(x)$ based on forward and backward representations for the tail process. While these estimates are asymptotically normal, the expressions for the asymptotic variances are too complicated to be useful for constructing confidence regions. To overcome this limitation, inference procedures can be carried out using resampling methods. Two resampling methods for constructing confidence intervals, based on the stationary bootstrap as used in Davis et al. (2012), and the multiplier block bootstrap as described in Drees (2015), are applied to our estimates of $F^{(\Theta_t)}(y)$. In terms of coverage probabilities, the multiplier block bootstrap performed better than the stationary bootstrap procedure in all the cases we considered. However, both procedures require care when applied for very high thresholds.

We apply the methodology to study serial extremal dependence of daily log-returns on the S&P500 index and the P&G stock price. We distinguish between two sources of such dependence – positive and negative shocks – pointing out an asymmetric behavior. Specifically, we consider cases when extreme values (positive or negative) follow positive/negative shocks t time-lags later. In terms of the spectral tail process, this corresponds to the probabilities $\mathrm{P}[\pm\Theta_t > 1 \mid \Theta_0 = \pm 1]$. We illustrate how well the GARCH(1,1) model and an extension of it allowing for a leverage effect, the APARCH(1,1) model, can capture serial extremal dependence, as measured by these directional probabilities. These examples demonstrate how our methodology can provide useful information on the behavior of extremes that follow both positive and negative shocks, which, in turn, can be used in a model-building context.

The remainder of the paper is organized as follows: The two estimates of the tail process are described in Section 2, while the companion limit theory for these estimators is formulated in Section 3. The stationary bootstrap and multiplier bootstrap procedures are presented in Section 2. The validity of the proposed bootstrap methodology is established in Section 3 too. The finite-sample performance is investigated through Monte Carlo simulations in Section 4. The application of our methodology is provided in Section 5. The proofs of the main results are collected in Section 6.

2 Methodology

2.1 Estimators

The data consist of a stretch $X_{1-\tilde{t}}, \ldots, X_{n+\tilde{t}}$, where $\tilde{t}$ is fixed and corresponds to the maximal lag of interest, drawn from a regularly varying, stationary univariate time series with spectral tail process $(\Theta_t)_{t\in\mathbb{Z}}$ and index $\alpha > 0$.

In order to estimate $p = \mathrm{P}[\Theta_0 = 1]$, we simply take the empirical version of (1.4), yielding

$$\hat{p}_n = \frac{\sum_{i=1}^n \mathbf{1}\left(X_i > u_n\right)}{\sum_{i=1}^n \mathbf{1}\left(|X_i| > u_n\right)}.$$

For $\hat{p}_n$ to be consistent and asymptotically normal, the threshold sequence u_n should tend to infinity at a certain rate described in the next section.

To estimate the cdf, $F^{(\Theta_t)}$, of Θ_t, we propose the *forward estimator*

$$\hat{F}_n^{(\mathrm{f},\Theta_t)}(x) := \frac{\sum_{i=1}^n \mathbf{1}\left(X_{i+t}/|X_i| \le x,\ |X_i| > u_n\right)}{\sum_{i=1}^n \mathbf{1}\left(|X_i| > u_n\right)}. \tag{2.1}$$

This is just the empirical version of the left-hand side of (1.6). In equations (1.1) and (1.3), the conditioning event is $\{|X_0| > u\}$, making no distinction between positive extremes, $X_0 > u$, and negative extremes, $X_0 < -u$. However, these two cases can be distinguished by conditioning on the sign of Θ_0. In particular, we define

$$\hat{F}_n^{(\mathrm{f},\Theta_t|\Theta_0=\pm 1)}(x) := \frac{\sum_{i=1}^n \mathbf{1}\left(X_{i+t}/\left(\pm X_i\right) \le x,\ \pm X_i > u_n\right)}{\sum_{i=1}^n \mathbf{1}\left(\pm X_i > u_n\right)}. \tag{2.2}$$

The numerator in the estimator is a sum of indicator functions, most of which are zero. This often leads to a large variance. The time-change formula (1.5) yields a different representation of the law of Θ_t, motivating a different estimator than the one above. Depending on the value of x, the new estimator will involve more non-zero indicators, which receive weights instead. The simulation study reported in Section 4.1 will show that the resulting estimator may have a smaller variance than the one in (2.2), in particular if $|x|$ is large.

Lemma 2.1. *Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary univariate time series, regularly varying with index α and spectral tail process $(\Theta_t)_{t\in\mathbb{Z}}$. Then, for all integer $t \neq 0$,*

$$\mathrm{P}[\Theta_t \le x] = \begin{cases} 1 - \mathrm{E}\left[|\Theta_{-t}|^\alpha\, \mathbf{1}\left(\Theta_0/|\Theta_{-t}| > x\right)\right] & \text{if } x \ge 0, \\ \mathrm{E}\left[|\Theta_{-t}|^\alpha\, \mathbf{1}\left(\Theta_0/|\Theta_{-t}| \le x\right)\right] & \text{if } x < 0. \end{cases} \tag{2.3}$$

Moreover

$$\mathrm{P}[\Theta_t \le x \mid \Theta_0 = 1] = \begin{cases} 1 - \frac{1}{p}\,\mathrm{E}\left[\Theta_{-t}^\alpha\, \mathbf{1}\left(1/\Theta_{-t} > x,\ \Theta_0 = 1\right)\right] & \text{if } x \ge 0, \\ \frac{1}{p}\,\mathrm{E}\left[\Theta_{-t}^\alpha\, \mathbf{1}\left(-1/\Theta_{-t} \le x,\ \Theta_0 = -1\right)\right] & \text{if } x < 0, \end{cases} \tag{2.4}$$

and

$$\mathrm{P}[\Theta_t \le x \mid \Theta_0 = -1] = \begin{cases} 1 - \frac{1}{1-p}\, \mathrm{E}[(-\Theta_{-t})^\alpha\ \mathbf{1}\,(-1/\Theta_{-t} > x,\ \Theta_0 = 1)] & \textit{if } x \ge 0, \\ \frac{1}{1-p}\, \mathrm{E}[(-\Theta_{-t})^\alpha\ \mathbf{1}\,(1/\Theta_{-t} \le x,\ \Theta_0 = -1)] & \textit{if } x < 0. \end{cases} \tag{2.5}$$

If population quantities are replaced by their sample counterparts, Lemma 2.1 suggests the following *backward estimator* of the cdf of Θ_t:

$$\hat{F}_n^{(\mathrm{b},\Theta_t)}(x) := \begin{cases} 1 - \dfrac{\sum_{i=1}^n |\frac{X_{i-t}}{X_i}|^{\hat{\alpha}_n} \mathbf{1}\,(X_i/|X_{i-t}| > x,\ |X_i| > u_n)}{\sum_{i=1}^n \mathbf{1}\,(|X_i| > u_n)} & \text{if } x \ge 0, \\ \dfrac{\sum_{i=1}^n |\frac{X_{i-t}}{X_i}|^{\hat{\alpha}_n} \mathbf{1}\,(X_i/|X_{i-t}| \le x,\ |X_i| > u_n)}{\sum_{i=1}^n \mathbf{1}\,(|X_i| > u_n)} & \text{if } x < 0. \end{cases} \tag{2.6}$$

Here, $\hat{\alpha}_n$ is an estimator of the tail index, for which we will take the Hill-type estimator

$$\hat{\alpha}_n = \frac{\sum_{i=1}^n \mathbf{1}\,(|X_i| > u_n)}{\sum_{i=1}^n \log(|X_i|/u_n)\mathbf{1}\,(|X_i| > u_n)}. \tag{2.7}$$

Conditioning on an extreme value of a specific sign, we get

$$\hat{F}_n^{(\mathrm{b},\Theta_t|\Theta_0=\pm 1)}(x) := \begin{cases} 1 - \dfrac{\sum_{i=1}^n \left(\frac{\pm X_{i-t}}{X_i}\right)^{\hat{\alpha}_n} \mathbf{1}\,(\pm X_i/X_{i-t} > x,\ \ X_i > u_n)}{\sum_{i=1}^n \mathbf{1}\,(\pm X_i > u_n)} & \text{if } x \ge 0, \\ \dfrac{\sum_{i=1}^n \left(\frac{\mp X_{i-t}}{X_i}\right)^{\hat{\alpha}_n} \mathbf{1}\,(\pm X_i/X_{i-t} \le x,\ \ X_i < -u_n)}{\sum_{i=1}^n \mathbf{1}\,(\pm X_i > u_n)} & \text{if } x < 0. \end{cases}$$

The asymptotic and finite-sample distributions of these estimators will be investigated in the following sections.

2.2 Resampling

We explore two different bootstrap schemes that yield confidence intervals for $F^{(\Theta_t)}(x)$, or rather, for the pre-asymptotic version $\mathrm{P}[X_t/|X_0| \le x \mid |X_0| > u]$: the stationary bootstrap and the multiplier block bootstrap. We apply each of the two resampling schemes to both the forward and backward estimators at various levels x and at different lags t.

The stationary bootstrap goes back to Politis and Romano (1994) and is an adaptation of the block bootstrap by allowing for random block sizes. The resampling scheme was applied to the extremogram in Davis et al. (2012). It consists of generating pseudo-samples $X_1^*, \ldots, X_n^*$, drawn from the sample $X_1, \ldots, X_n$ by taking the first n values in the sequence

$$X_{K_1}, \ldots, X_{K_1+L_1-1}, X_{K_2}, \ldots, X_{K_2+L_2-1}, \ldots,$$

where $K_1, K_2 \ldots$ is an iid sequence of random variables uniformly distributed on $\{1, \ldots, n\}$ and $L_1, L_2, \ldots$ is an iid sequence of geometrically distributed random variables (independent of $(K_j)_{j\in\mathbb{N}}$) with distribution $\mathrm{P}[L_1 = l] = p\,(1-p)^{l-1}$, $l = 1, 2, \ldots$ for some $p = p_n \in (0,1)$ such that $p_n \to 0$ and $np_n \to \infty$. If the index t thus obtained exceeds the sample size n, we replace t by $(t - 1 \bmod n) + 1$, i.e., we continue from the beginning of the sample. The estimators are then applied to $X_{1-\tilde{t}}^*, \ldots, X_{n+\tilde{t}}^*$.

The multiplier block bootstrap method was applied to cluster functionals in Drees (2015). It consists of splitting the data set into $m_n = \lfloor n/r_n \rfloor$ blocks of length r_n and multiplying the cluster functionals of each block by a random factor. (Here $\lfloor x \rfloor$ denotes the integer part of x.)

Specifically, for iid random variables ξ_j, independent of $(X_t)_{t\in\mathbb{Z}}$, with $\mathrm{E}[\xi_j] = 0$ and $\mathrm{var}\,[\xi_j] = 1$, the bootstrapped forward estimator can be written as

$$\hat{F}_n^{*(\mathrm{f},\Theta_t)}(x) := \frac{\sum_{j=1}^{m_n}(1+\xi_j)\sum_{i\in I_j}\mathbf{1}\left(\frac{X_{i+t}}{|X_i|}\leq x,\ |X_i|>u_n\right)}{\sum_{j=1}^{m_n}(1+\xi_j)\sum_{i\in I_j}\mathbf{1}\left(|X_i|>u_n\right)},$$

where $I_j = \{(j-1)r_n+1,\dots,jr_n\}$ denotes the set of indices belonging to the jth block. Similarly, the bootstrapped backward estimator for $x>0$ with estimated index of regular variation is

$$\hat{F}_n^{*(\mathrm{b},\Theta_t)}(x) := 1 - \frac{\sum_{j=1}^{m_n}(1+\xi_j)\sum_{i\in I_j}\left|\frac{X_{i-t}}{X_i}\right|^{\hat{\alpha}_n^*}\mathbf{1}\left(\frac{X_i}{|X_{i-t}|}>x,\ |X_i|>u_n\right)}{\sum_{j=1}^{m_n}(1+\xi_j)\sum_{i\in I_j}\mathbf{1}\left(|X_i|>u_n\right)} \tag{2.8}$$

with

$$\hat{\alpha}_n^* := \frac{\sum_{j=1}^{m_n}(1+\xi_j)\sum_{i\in I_j}\mathbf{1}(|X_i|>u_n)}{\sum_{j=1}^{m_n}(1+\xi_j)\sum_{i\in I_j}\log(|X_i|/u_n)\mathbf{1}(|X_i|>u_n)}. \tag{2.9}$$

If the threshold u_n is high, it may be advisable to construct bootstrap confidence intervals based on lower thresholds and then scale accordingly; see the explanation after Theorem 3.3.

2.3 Testing for dependence of extreme observations

For iid random variables, the spectral tail process simplifies to $\Theta_t \equiv 0$ a.s. for all nonzero t. If this occurs for a stationary, regularly varying time series, then we say that the series exhibits *serial extremal independence*. The opposite case is referred to as *serial extremal dependence*, i.e., at least one of the variables Θ_t for $t\neq 0$ is not degenerate at 0. Since the convergence of the pre-asymptotic distribution can be arbitrarily slow, one cannot formally test for extremal dependence within the present framework.

However, if one wants to test whether the exceedances over a given high threshold u are independent then one may check whether the lower bound of a confidence interval for, say, $\mathrm{P}[|X_t|\geq|X_0|\mid|X_0|>u]$ constructed by one of the bootstrap methodologies is larger than this probability under the assumption of exact independence of $|X_0|\mathbf{1}(|X_0|>u)$ and $|X_t|\mathbf{1}(|X_t|>u)$, which is easily shown to equal $\mathrm{P}[|X_0|>u]/2$.

If one prefers to work with exceedances of the original time series (instead of its absolute values), then the probability under the assumption of independence depends on the relative weights of the upper and lower tails, and can thus not be calculated analytically. In that case, it seems natural to calculate this probability by Monte Carlo simulation. To this end, one generates (conditionally) iid samples according to the empirical distribution of the original time series by drawing with replacement from the observations, which corresponds to the classical bootstrap procedure for iid data. The considered probability under independence can be approximated by the pertaining relative frequency, which is then compared with the lower confidence bound for the probability estimated from the original time series. If the latter is larger this indicates that the exceedances in the time series exhibit a non-negligible serial dependence (see Figure 4 in Section 5).

Alternatively, one may compare the pre-asymptotic probability estimated from the observed time series using either the forward or the backward estimator with quantiles of the distribution of this estimator under independence. In a similar way as described above, the latter can be approximated by an empirical quantile obtained in Monte Carlo simulations with (conditionally) iid samples (cf. Figure 5).

3 Large-sample theory

Under certain conditions, the standardized estimation errors of the forward and the backward estimators converge jointly to a centered Gaussian process (Section 3.1). Convergence of the multiplier block bootstrap follows under the same conditions (Section 3.2).

In order not to overload the presentation, we focus on nonnegative time series. We briefly indicate how the conditions and results must be modified in the real-valued case.

3.1 Asymptotic normality of the estimators

All estimators under consideration can then be expressed in terms of *generalized tail array sums*. These are statistics of the form $\sum_{i=1}^{n} \phi(X_{n,i})$, with

$$X_{n,i} := u_n^{-1} \left(X_{i-\tilde{t}}, \ldots, X_i, \ldots, X_{i+\tilde{t}}\right) \mathbf{1}(X_i > u_n). \tag{3.1}$$

Drees and Rootzén (2010) give conditions under which, after standardization, such statistics converge to a centered Gaussian process, uniformly over appropriate families of functions ϕ. From these results we will deduce a functional central limit theorem for the processes of forward and backward estimators defined in (2.1) and (2.6) with $\hat{\alpha}_n$ according to (2.7), respectively.

To ensure consistency, the threshold u_n must tend to infinity in such a way that

$$v_n := \mathrm{P}[X_0 > u_n]$$

tends to 0, but the expected number, nv_n, of exceedances tends to infinity. Moreover, we have to ensure that observations which are sufficiently separated in time are almost independent. The strength of dependence will be assessed by the β-mixing coefficients

$$\beta_{n,k} := \sup_{1\le l\le n-k-1} \mathrm{E}\Big[\sup_{B\in\mathcal{B}^n_{n,l+k+1}} \big|\mathrm{P}\big[B \mid \mathcal{B}^l_{n,1}\big] - \mathrm{P}[B]\big|\Big].$$

Here $\mathcal{B}^j_{n,i}$ is the σ-field generated by $(X_{n,l})_{i\le l\le j}$.

We assume that there exist sequences $l_n, r_n \to \infty$ and some $x_0 \ge 0$ such that the following conditions hold:

($\mathbf{A}(x_0)$) The cdf of Θ_t, $F^{(\Theta_t)}$, is continuous on $[x_0,\infty)$, for $t \in \{1,\ldots,\tilde{t}\}$.

(B) As $n\to\infty$, we have $l_n\to\infty$, $l_n = o(r_n)$, $r_n = o((nv_n)^{1/2})$, $r_nv_n \to 0$, and $\beta_{n,l_n} n/r_n \to 0$.

(C) For all $k \in \{0,\ldots,r_n\}$, there exists

$$s_n(k) \ge \mathrm{E}\Big[\log\Big(\frac{X_0}{u_n}\Big) \max\Big\{\log\Big(\frac{X_k}{u_n}\Big), \mathbf{1}(X_k>u_n)\Big\} \;\Big|\; X_0 > u_n\Big] \tag{3.2}$$

such that $s_\infty(k) = \lim_{n\to\infty} s_n(k)$ exists, $\lim_{n\to\infty}\sum_{k=1}^{r_n} s_n(k) = \sum_{k=1}^{\infty} s_\infty(k)$ holds and the last sum is finite.

Moreover, there exists $\delta > 0$ such that

$$\sum_{k=1}^{r_n} \Bigg(\mathrm{E}\Big[\Big(\log^+\Big(\frac{X_0}{u_n}\Big)\log^+\Big(\frac{X_k}{u_n}\Big)\Big)^{1+\delta} \;\Big|\; X_0>u_n\Big]\Bigg)^{1/(1+\delta)} = O(1), \qquad n\to\infty. \tag{3.3}$$

Without Condition ($\mathrm{A}(x_0)$) one cannot expect uniform convergence of the estimated cdf of Θ_t to the true cdf on $[x_0,\infty)$. Indeed, in this case even $\mathrm{P}[X_t/X_0 \le x \mid |X_0| > u]$ need not converge to $F^{(\Theta_t)}(x)$ for a point of discontinuity x. Condition (B) imposes restrictions on the rate at which v_n tends to 0 and thus on the rate at which u_n tends to ∞. Often, the β-mixing coefficients decay geometrically, i.e., $\beta_{n,k} = O(\eta^k)$ for some $\eta\in(0,1)$. Then one may choose $l_n = O(\log n)$, and Condition (B) is fulfilled for a suitably chosen r_n if $(\log n)^2/n = o(v_n)$ and $v_n = o(1/(\log n))$.

The technical Condition (C) rules out too large a cluster of extreme observations. Using integration by parts, the right-hand side of (3.2) can be bounded by

$$v_n^{-1}\int_1^\infty \Big(\mathrm{P}[X_0 > u_ns, X_k>u_n] + \int_1^\infty \mathrm{P}[X_0>u_ns, X_k>u_nt]\,t^{-1}\,dt\Big)s^{-1}ds.$$

Now one can use techniques employed in Drees (2000) and Drees (2003) to verify (3.2) for specific time series models like solutions to stochastic recurrence equations or suitable heavy-tailed linear time series. (Typically, the upper bounds $s_n(k)$ are of the form $\rho_k + \xi_n$ for a summable sequence ρ_k and $\xi_n = o(1/r_n)$.) The left-hand side of (3.3) can be rewritten in the form

$$\sum_{k=1}^{r_n} \left((1+\delta)^2 v_n^{-1} \int_1^\infty \int_1^\infty P[X_0 > u_n s, X_k > u_n t] (\log s \log t)^\delta (st)^{-1} \, ds \, dt \right)^{1/(1+\delta)},$$

which can then be bounded by similar techniques.

Under these conditions, one can prove the asymptotic normality of relevant generalized tail array sums (see Proposition 6.1 below) and thus the joint uniform asymptotic normality of the appropriately centered forward and the backward estimator of $F^{(\Theta_t)}$.

Theorem 3.1. *Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary, regularly varying process. If (A(x_0)), (B) and (C) are fulfilled for some $x_0 \geq 0$ and $y_0 \in [x_0,\infty) \cap (0,\infty)$, then*

$$(nv_n)^{1/2} \begin{pmatrix} \left(\hat{F}_n^{(\mathrm{f},\Theta_t)}(x_t) - P[X_t/X_0 \leq x_t \mid X_0 > u_n]\right)_{x_t\in[x_0,\infty)} \\ \left(\hat{F}_n^{(\mathrm{b},\Theta_t)}(y_t) - (1 - E[(X_{-t}/X_0)^\alpha \mathbf{1}(X_0/X_{-t} > y_t) \mid X_0 > u_n])\right)_{y_t\in[y_0,\infty)} \end{pmatrix}_{|t|\in\{1,\dots,\tilde{t}\}} \xrightarrow{d}$$
$$\begin{pmatrix} \left(Z(\phi_{2,x_t}^t) - \bar{F}^{(\Theta_t)}(x_t) Z(\phi_1)\right)_{x_t\in[x_0,\infty)} \\ \left(Z(\phi_{3,x_t}^t) - \bar{F}^{(\Theta_t)}(y_t) Z(\phi_1) + (\alpha^2 Z(\phi_0) - \alpha Z(\phi_1)) E[\log(\Theta_t) \mathbf{1}(\Theta_t > y_t)]\right)_{y_t\in[y_0,\infty)} \end{pmatrix}_{|t|\in\{1,\dots,\tilde{t}\}} \quad (3.4)$$

where Z is a centered Gaussian process, indexed by functions defined in (6.2), whose covariance function is given in (6.3), and $\bar{F}^{(\Theta_t)} := 1 - F^{(\Theta_t)}$ denotes the survival function of Θ_t. (Assertion (3.4) means that for suitable versions of the processes the convergence holds uniformly for all $x_t \geq x_0, y_t \geq y_0$ and $|t| \in \{1,\dots,\tilde{t}\}$ almost surely.)

Additional conditions are needed to ensure that the biases of the forward and the backward estimator of $F^{(\Theta_t)}$ are asymptotically negligible:

$$\sup_{x\in[x_0,\infty)} \left| P\left[\frac{X_t}{X_0} \leq x \,\middle|\, X_0 > u_n\right] - F^{(\Theta_t)}(x) \right| = o\left((nv_n)^{-1/2}\right), \quad (3.5)$$
$$\sup_{y\in[y_0,\infty)} \left| E\left[\left(\frac{X_{-t}}{X_0}\right)^\alpha \mathbf{1}(X_0/X_{-t} > y) \,\middle|\, X_0 > u_n\right] - \bar{F}^{(\Theta_t)}(y) \right| = o\left((nv_n)^{-1/2}\right), \quad (3.6)$$
$$E[\log(X_0/u_n) \mid X_0 > u_n] - 1/\alpha = o\left((nv_n)^{-1/2}\right), \quad (3.7)$$

for $t \in \{-\tilde{t},\dots,\tilde{t}\} \setminus \{0\}$ as $n \to \infty$. These conditions are fulfilled if nv_n tends to ∞ sufficiently slowly, because by definition of the spectral tail process, the regular variation of X_0 and by (2.3), the left-hand sides in (3.5)–(3.7) tend to 0 if $F^{(\Theta_t)}$ is continuous on $[x_0,\infty)$.

Corollary 3.2. *Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary, regularly varying process. If (A(x_0)), (B), (C), and (3.5)–(3.7) are fulfilled for some $x_0 \geq 0$ and $y_0 \in [x_0,\infty) \cap (0,\infty)$, then*

$$(nv_n)^{1/2} \begin{pmatrix} \left(\hat{F}_n^{(\mathrm{f},\Theta_t)}(x_t) - F^{(\Theta_t)}(x_t)\right)_{x_t\in[x_0,\infty)} \\ \left(\hat{F}_n^{(\mathrm{b},\Theta_t)}(y_t) - F^{(\Theta_t)}(y_t)\right)_{y_t\in[y_0,\infty)} \end{pmatrix}_{|t|\in\{1,\dots,\tilde{t}\}} \xrightarrow{d}$$
$$\begin{pmatrix} \left(Z(\phi_{2,x_t}^t) - \bar{F}^{(\Theta_t)}(x_t) Z(\phi_1)\right)_{x_t\in[x_0,\infty)} \\ \left(Z(\phi_{3,x_t}^t) - \bar{F}^{(\Theta_t)}(y_t) Z(\phi_1) + (\alpha^2 Z(\phi_0) - \alpha Z(\phi_1)) E[\log(\Theta_t) \mathbf{1}(\Theta_t > y_t)]\right)_{y_t\in[y_0,\infty)} \end{pmatrix}_{|t|\in\{1,\dots,\tilde{t}\}}$$

where Z is the centered Gaussian process defined in Theorem 3.1.

In general, it is difficult to compare the asymptotic variances of the backward and the forward estimator.

3.2 Consistency of the multiplier block bootstrap

Here we discuss the asymptotic behavior of the multiplier block bootstrap version of the forward and backward estimators. For the sake of brevity, we focus on estimators of $F^{(\Theta_t)}(x)$ for a fixed x.

Drees (2015) has shown convergence of bootstrap versions of empirical processes of tail array sums under the same conditions needed for convergence of the original empirical processes. Let P_ξ denote the probability w.r.t. $\xi = (\xi_j)_{j\in\mathbb{N}}$, i.e., the conditional probability given $(X_{n,i})_{1\le i\le n}$.

Theorem 3.3. *Let ξ_j, $j \in \mathbb{N}$, be iid random variables independent of $(X_t)_{t\in\mathbb{Z}}$ with $\mathrm{E}[\xi_j] = 0$ and $\mathrm{var}[\xi_j] = 1$. Then, under the conditions of Theorem 3.1, for all $x \ge x_0, y \ge y_0$,*

$$\begin{aligned}
\sup_{r,s\in\mathbb{R}^{2\tilde t}} \Bigg| \mathrm{P}_\xi \Big[&(nv_n)^{1/2} \Big(\hat F_n^{*(\mathrm{f},\Theta_t)}(x) - \hat F_n^{(\mathrm{f},\Theta_t)}(x) \Big) \le r_t, \\
&\qquad (nv_n)^{1/2} \Big(\hat F_n^{*(\mathrm{b},\Theta_t)}(y) - \hat F_n^{(\mathrm{b},\Theta_t)}(y) \Big) \le s_t,\ \forall |t| \in \{1,\dots,\tilde t\} \Big] \\
- \mathrm{P} \Big[&(nv_n)^{1/2} \Big(\hat F_n^{(\mathrm{f},\Theta_t)}(x) - F^{(\Theta_t)}(x) \Big) \le r_t, \\
&\qquad (nv_n)^{1/2} \Big(\hat F_n^{(\mathrm{b},\Theta_t)}(y) - F^{(\Theta_t)}(y) \Big) \le s_t,\ \forall |t| \in \{1,\dots,\tilde t\} \Big] \Bigg| \to 0
\end{aligned}$$

in probability.

In particular, if a and b are such that $\mathrm{P}_\xi \left[\hat F_n^{*(\mathrm{b},\Theta_t)}(y) \in [a,b] \right] = \beta$, then

$$\left[2\hat F_n^{(\mathrm{b},\Theta_t)}(y) - b,\ 2\hat F_n^{(\mathrm{b},\Theta_t)}(y) - a \right]$$

is a confidence interval for $F^{(\Theta_t)}(y)$ with approximative coverage level β. However, if the number of exceedances over a given threshold is too small, one may prefer to construct confidence intervals based on bootstrap estimators corresponding to lower thresholds. Let $\tilde u_n$ denote another threshold sequence, let $\tilde v_n = \mathrm{P}[X_0 > \tilde u_n]$ denote the corresponding exceedance probabilities, and let $\hat{\tilde F}_n^{(\mathrm{b},\Theta_t)}(y)$ and $\hat{\tilde F}_n^{*(\mathrm{b},\Theta_t)}(y)$ denote the backward estimator and the bootstrap version thereof, respectively, based on the exceedances over $\tilde u_n$. The conditional distribution of

$$(n\tilde v_n)^{1/2} \left(\hat{\tilde F}_n^{*(\mathrm{b},\Theta_t)}(y) - \hat{\tilde F}_n^{(\mathrm{b},\Theta_t)}(y) \right)$$

given the data is approximately the same as the unconditional distribution of

$$(nv_n)^{1/2} \left(\hat F_n^{(\mathrm{b},\Theta_t)}(y) - F^{(\Theta_t)}(y) \right)$$

Hence, if a and b are such that $\mathrm{P}_\xi \left[\hat{\tilde F}_n^{*(\mathrm{b},\Theta_t)}(y) \in [a,b] \right] = \beta$ and if $\hat{\tilde v}_n/\hat v_n$ is a suitable estimator of $\tilde v_n/v_n$, then

$$\left[\left(\frac{\hat{\tilde v}_n}{\hat v_n} \right)^{1/2} \left(\hat{\tilde F}_n^{(\mathrm{b},\Theta_t)}(y) - b \right) + \hat F_n^{(\mathrm{b},\Theta_t)}(y),\quad \left(\frac{\hat{\tilde v}_n}{\hat v_n} \right)^{1/2} \left(\hat{\tilde F}_n^{(\mathrm{b},\Theta_t)}(y) - a \right) + \hat F_n^{(\mathrm{b},\Theta_t)}(y) \right] \tag{3.8}$$

is a confidence interval for $F^{(\Theta_t)}(y)$ with approximative coverage probability β. In practice, one will often use large order statistics as thresholds, say the k_n-th and $\tilde k_n$-th largest observations, respectively. In that case, $\tilde v_n/v_n$ can be replaced by $\tilde k_n/k_n$. A similar approach, namely to use a variance estimator which is based on a lower threshold, has successfully been employed in Drees (2003, Section 5).

Of course, confidence intervals based on the bootstrap version of the forward estimator can be constructed analogously.

Remark 3.4. It is possible to generalize Theorem 3.3 to cover the joint limit distribution of the bootstrap estimators for all $x \geq x_0$ and $y \geq y_0$. Technically, this requires to endow the space of probability measures on spaces of bounded functions from $[x_0, \infty)$ (resp. $[y_0, \infty)$) to $\mathbb{R}^{2\tilde{t}}$ with a metric that induces weak convergence, e.g., the bounded Lipschitz metric. This is the approach in Drees (2015) to establish the consistency of a bootstrap method for estimating the extremogram, a close cousin of the tail process. Based on such a result, one may construct uniform confidence bands for the function $F^{(\Theta_t)}$ on $[x_0, \infty)$ or $[y_0, \infty)$, respectively, which in general will be considerably wider than the pointwise confidence intervals discussed above and will thus often be rather uninformative. For brevity, we omit the details.

Remark 3.5. For time series which may take on negative values too, the forward and backward estimators of $F^{(\Theta_t)}$ can be represented in terms of generalized tail array sums constructed from

$$X_{n,i}^{\tilde{t}} = u_n^{-1}\left(X_{i-\tilde{t}}, \ldots, X_i, \ldots, X_{i+\tilde{t}}\right) \mathbf{1}(|X_i| > u_n).$$

When $x < 0$, for example, the backward estimator $\tilde{F}_n^{(\mathrm{b},\Theta_t)}(x)$ is equal to the ratio of the generalized tail array sums pertaining to the functions

$$\begin{aligned}
&(y_{-\tilde{t}}, \ldots, y_0, \ldots, y_{\tilde{t}}) \mapsto |y_{-t}/y_0|^{\alpha}\, \mathbf{1}(y_0/|y_{-t}| \leq x, |y_0| > 1),\\
&(y_{-\tilde{t}}, \ldots, y_0, \ldots, y_{\tilde{t}}) \mapsto \mathbf{1}(|y_0| > 1).
\end{aligned}$$

Limit theorems can be obtained by the same methods as in the case of non-negative observations under obvious analogues to the conditions (A(x_0)), (B) and (C) with $v_n := \mathrm{P}[|X_0| > u_n]$.

4 Finite-sample performance

In Section 4.1, we show results from a numerical simulation study designed to test the performance of the forward (2.1) and the backward (2.6) estimators. We continue in Section 4.2 by evaluating the performance of two bootstrap schemes, the multiplier block bootstrap and the stationary bootstrap, described in Section 2.2.

The simulations are based on pseudo-random samples from two widely used models for financial time series. Both models are of the form $X_t = \sigma_t Z_t$ where σ_t and Z_t are independent. First, we consider the GARCH(1,1) model with $\sigma_t^2 = 0.1 + 0.14 X_{t-1}^2 + 0.84\sigma_{t-1}^2$, the innovations Z_t being independent t_4 random variables, standardized to have unit variance. The second model is the stochastic volatility (SV) process with $\log \sigma_t = 0.9 \log \sigma_{t-1} + \epsilon_t$, with independent standard normal innovations ϵ_t and independent innovations Z_t with common distribution $t_{2.6}$. The parameters have been chosen to ensure that both time series are regularly varying with index $\alpha = 2.6$ (Davis and Mikosch, 2001; Mikosch and Stărică, 2000).

4.1 Forward and backward estimators

We estimate $\mathrm{P}[\Theta_t \leq x]$ for both the GARCH(1,1) and the SV model, for various arguments x and lags t, via the forward and the backward estimator, with estimated tail index α. The threshold is set at the empirical 95% quantile of the absolute values of a time series of length $n = 2\,000$. We do 1 000 Monte Carlo repetitions and calculate bias, standard deviation, and root mean squared error (RMSE) with respect to the pre-asymptotic values $\mathrm{P}\left[X_t/|X_0| \leq x \mid |X_0| > F_{|X|}^{\leftarrow}(0.95)\right]$ in the forward representation. The true quantile $F_{|X|}^{\leftarrow}(0.95)$ of $|X_0|$ and the true pre-asymptotic values were calculated numerically via 10 000 Monte Carlo simulations based on time series of length 10 000.

It was already reported in the context of Markovian time series that for $t = 1$ and $|x|$ large, the backward estimators usually have a smaller variance than the forward estimators (Drees et al., 2015). Here, numerical simulations suggest that this is true for non-Markovian time series and for higher lags as well. The results are presented in Figure 1.

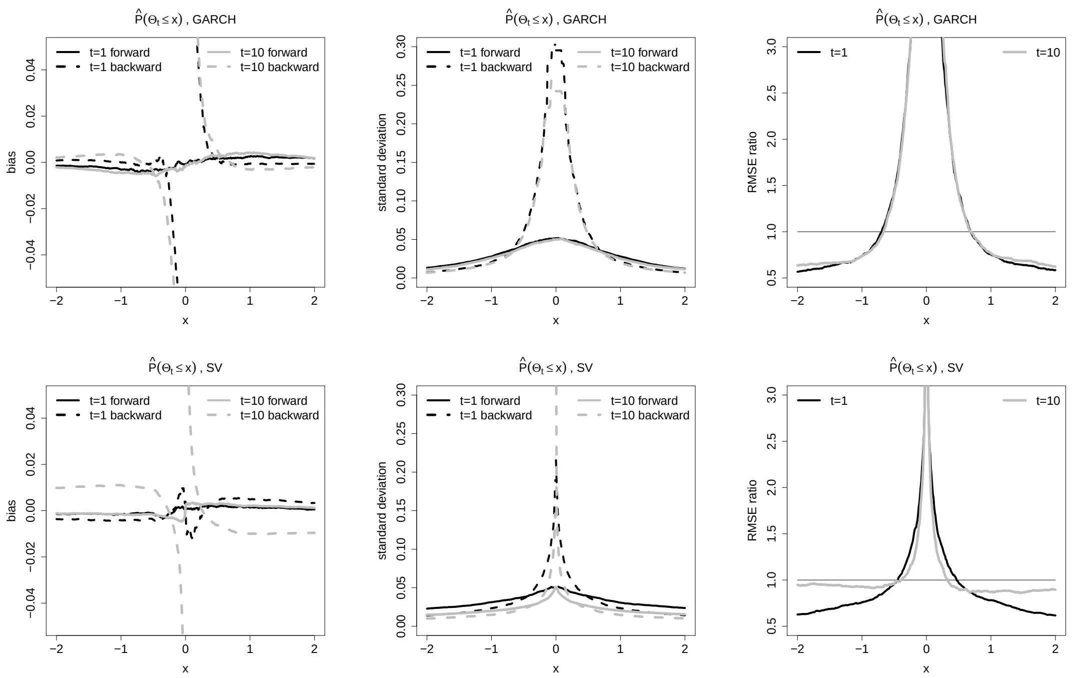

Figure 1: Performance of the forward and the backward estimators: bias (left), standard deviation (middle), ratio of root mean square errors (right) with respect to the pre-asymptotic values in the forward estimator, for GARCH(1,1) model (top) and SV model (bottom).

The right column, which shows the RMSE of the backward estimator divided by the RMSE of the forward estimators (both with respect to the pre-asymptotic values of the spectral tail process in the forward representation), shows that the backward estimator outperforms the forward estimator if x is sufficiently large in absolute value. This phenomenon was also observed at other lags (not shown). For some other models, however, such as certain stochastic recurrence equation or copula Markov models, the advantage of the backward estimator was observed only for smaller lags ($t = 1, \ldots, 4$).

4.2 Bootstrapped spectral tail process

We asses the performance of the two bootstrap schemes, the stationary bootstrap and the multiplier block bootstrap. To do so, we estimate the coverage probability of the bootstrapped confidence intervals with respect to the true pre-asymptotic spectral tail process in the forward representation. We focus on probabilities of the form $\mathrm{P}[|\Theta_t| > 1]$. This particular value can be of interest due to its interpretation as the probability of a shock being followed by an even larger aftershock, i.e., $|X_t|$ being larger than $|X_0|$ conditionally on $|X_0|$ exceeding some threshold already. The true pre-asymptotic values in the forward representation were calculated numerically via 10 000 Monte Carlo simulations with time series of length 10 000.

In Figure 2, we plot the results for the GARCH$(1, 1)$ model and for the SV model, for the forward and the backward estimators. The expected block size for the stationary bootstrap (represented by gray lines) was chosen as 100. For the multiplier block bootstrap (black lines), the block size was fixed at 100 and the multiplier variables ξ_j were drawn independently from the standard normal distribution. Estimates of the coverage probabilities are based on 1 000 simulations. In each such sample, we use 1 000 bootstrap samples for calculating the confidence intervals with nominal coverage probability 95%. We use two different thresholds, i.e., the 95% and 98% empirical quantiles of the absolute values of a time series of length $n = 2\,000$. For the higher

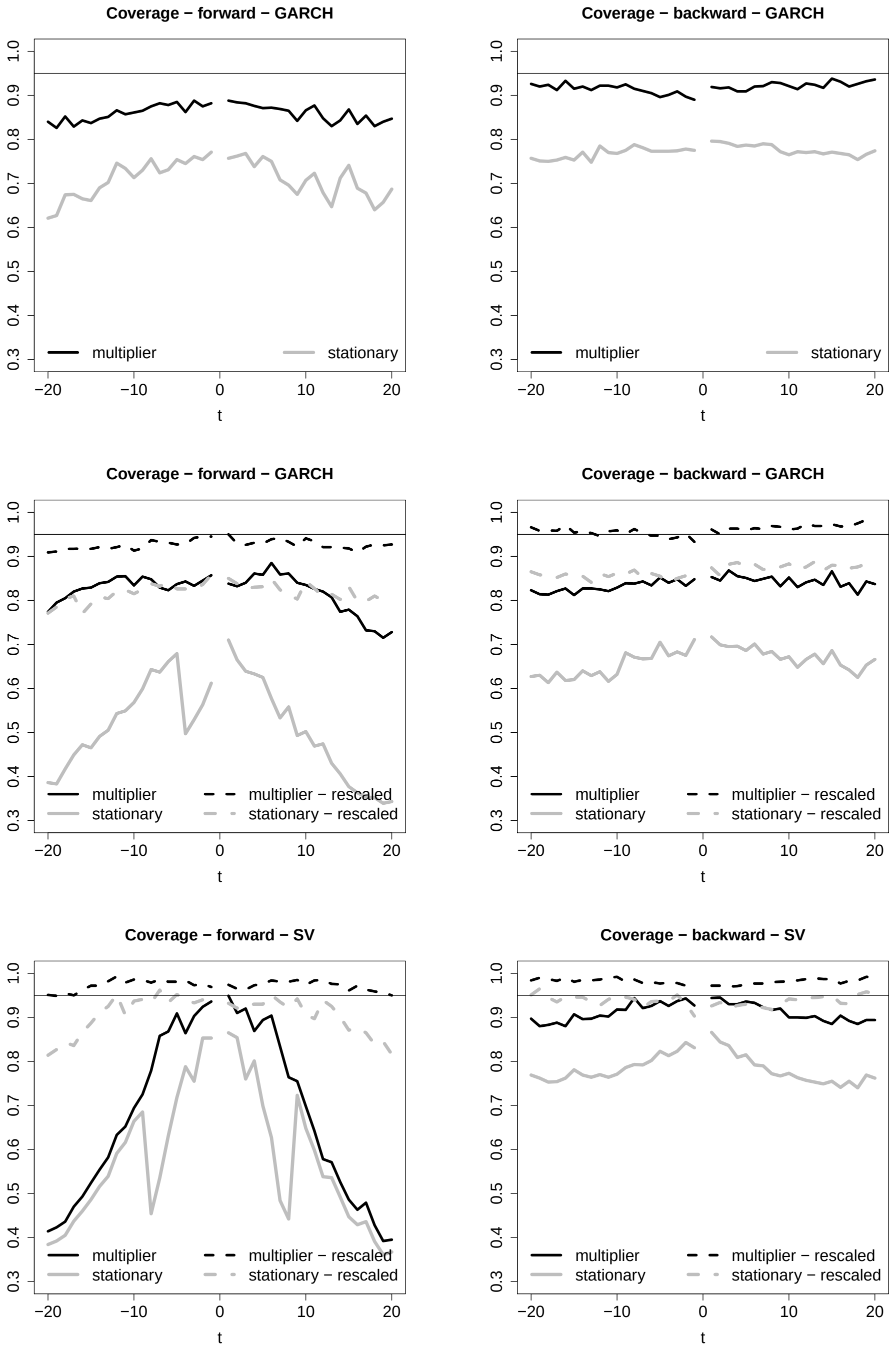

Figure 2: Coverage probabilities of confidence intervals for $P[|X_t/X_0| > 1 \mid |X_0| > u_n]$ (left: forward estimator; right: backward estimator) based on the stationary bootstrap (gray) and the multiplier block bootstrap (black). The top and the middle plots correspond to the GARCH(1, 1) model with thresholds set at the 95% and the 98% empirical quantiles, respectively. The bottom plots correspond to the SV simulation study with threshold set at the 98% empirical quantile. In the latter two cases, the dashed lines correspond to the coverage probabilities of the rescaled confidence intervals (3.8). The horizontal black line is the 0.95 reference line.

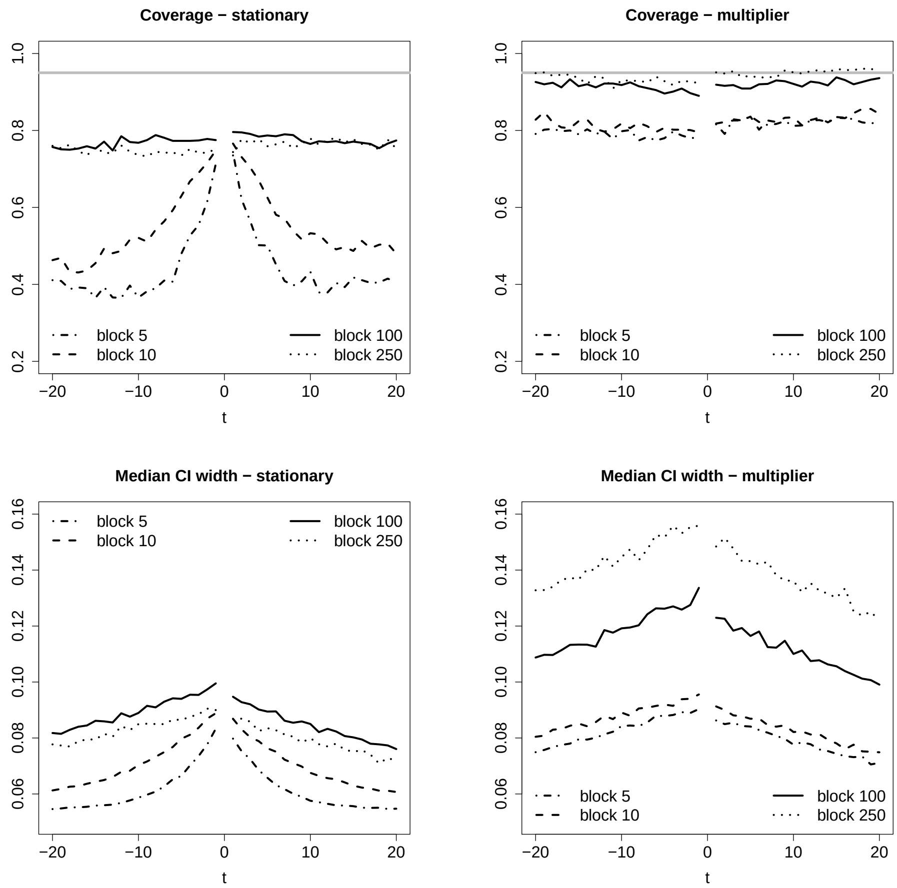

Figure 3: Coverage probabilities (top) and median widths (bottom) of confidence intervals for $\mathrm{P}[|X_t/X_0| > 1 \mid |X_0| > u_n]$ based on the stationary bootstrap (left column) and the multiplier block bootstrap (right column) for different block lengths. The dash–dotted, dashed, solid, and dotted lines represent (mean) block lengths 5, 10, 100, and 250 respectively. The plots correspond to the backward estimator and the GARCH$(1,1)$ model with thresholds set at the 95% empirical quantile. The horizontal black lines in the top plots are the 0.95 reference lines.

threshold, the confidence intervals were calculated either directly (indicated by the solid lines) or using a rescaled bootstrap estimator that was based on the exceedances over the 95% empirical quantile as in (3.8) (dashed lines).

In all cases, the multiplier block bootstrap produces a better coverage probability than the stationary bootstrap. Moreover, the backward estimator is more stable than the forward one, at least for $x = 1$, and this translates into higher stability of the bootstrapped confidence intervals. The effect is especially visible for higher thresholds, e.g., at the 98% quantile, leaving insufficiently many pairs of exceedances for accurate inference. Finally, rescaled confidence intervals (3.8) based on lower thresholds can have a much better coverage than confidence intervals based on higher thresholds.

In addition, in Figure 3 we show coverage probabilities and median confidence intervals widths for different block sizes. The multiplier block bootstrap is more robust to the choice of block length than the stationary bootstrap. In contrast to the stationary bootstrap, the multiplier block bootstrap produces confidence intervals whose coverage probabilities are fairly stable across different lags for a given block length.

It is important not to set the block length too low since it can lead to poor coverage probabilities, especially for higher lags. On the other hand, too large a block length can result in confidence

		ω	α_1	β_1	δ	γ_1
S&P500	GARCH	7×10^{-7}	0.062	0.932	-	-
		(2×10^{-7})	(0.006)	(0.007)		
	APARCH	5×10^{-5}	0.056	0.937	1.227	0.874
		(1×10^{-5})	(0.008)	(0.006)	(0.131)	(0.118)
P&G	GARCH	9×10^{-7}	0.04	0.957	-	-
		(2×10^{-7})	(0.004)	(0.004)		
	APARCH	17×10^{-5}	0.056	0.951	0.938	0.608
		(3×10^{-5})	(0.004)	(0.004)	(0.112)	(0.074)

Table 1: Parameters of the models fitted to daily log-returns of the S&P500 index (top) and the P&G stock price (bottom). Standard errors in parentheses.

intervals that are too wide.

5 Application

We first consider daily log-returns on the S&P500 stock market index between 1990-01-01 and 2010-01-01 taken from Yahoo Finance[1]. In Figure 4, we plot the sample spectral tail process probabilities $P[|\Theta_t| > 1 \mid \Theta_0 = \pm 1]$ and $P[\pm\Theta_t > 1 \mid \Theta_0 = \pm 1]$ based on the backward estimator with 98% empirical quantile taken as a threshold and the 80% pointwise confidence intervals from the multiplier bootstrap scheme rescaled via the 95% quantile as threshold as in (3.8). The estimated index of regular variation is $\hat{\alpha} = 3.17$.

The left-hand plots correspond to conditioning on a positive extreme at the current time instant, whereas the right-hand plots correspond to conditioning on a negative shock. The former plots indicate much weaker serial extremal dependence than the latter ones: negative shocks are more persistent than positive ones. This is indicated by the lower bounds of the confidence intervals being above the horizontal lines which correspond to probabilities under independence. In particular, the pattern of negative extremes followed by positive ones is clearly visible; see the right-hand plot on the second row. The above mentioned characteristics are shared by other stock's daily returns which were tested but not reported here.

Consider two widely used financial models of the type $X_t = \sigma_t Z_t$: first, the GARCH$(1,1)$ process, where

$$\sigma_t^2 = \omega + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

and second, the APARCH(1,1) process (Ding et al., 1993) with

$$\sigma_t^\delta = \omega + \alpha_1 \left(|X_{t-1}| - \gamma_1 X_{t-1}\right)^\delta + \beta_1 \sigma_{t-1}^\delta.$$

Both models allow for volatility clustering in the limit. Additionally, the APARCH model captures asymmetry in the volatility of returns. That is, volatility tends to increase more when returns are negative, as compared to positive returns of the same magnitude if $\gamma_1 > 0$. The asymmetric response of volatility to positive and negative shocks is well known in the finance literature as the *leverage effect* of the stock market returns (Black, 1976).

We fit those two models to daily log-returns of the S&P500 index. We use the garchFit function from the fGarch library available in R, the function being based on maximum likelihood estimation (Wuertz et al., 2013). The innovations, Z_t, are assumed to be standard normally distributed. The fitted parameters are given in the top part of Table 1.

In Figure 4 we plot the pre-asymptotic spectral tail process probabilities based on the forward estimator for the fitted GARCH and APARCH models, together with the sample spectral tail

[1] http://finance.yahoo.com/

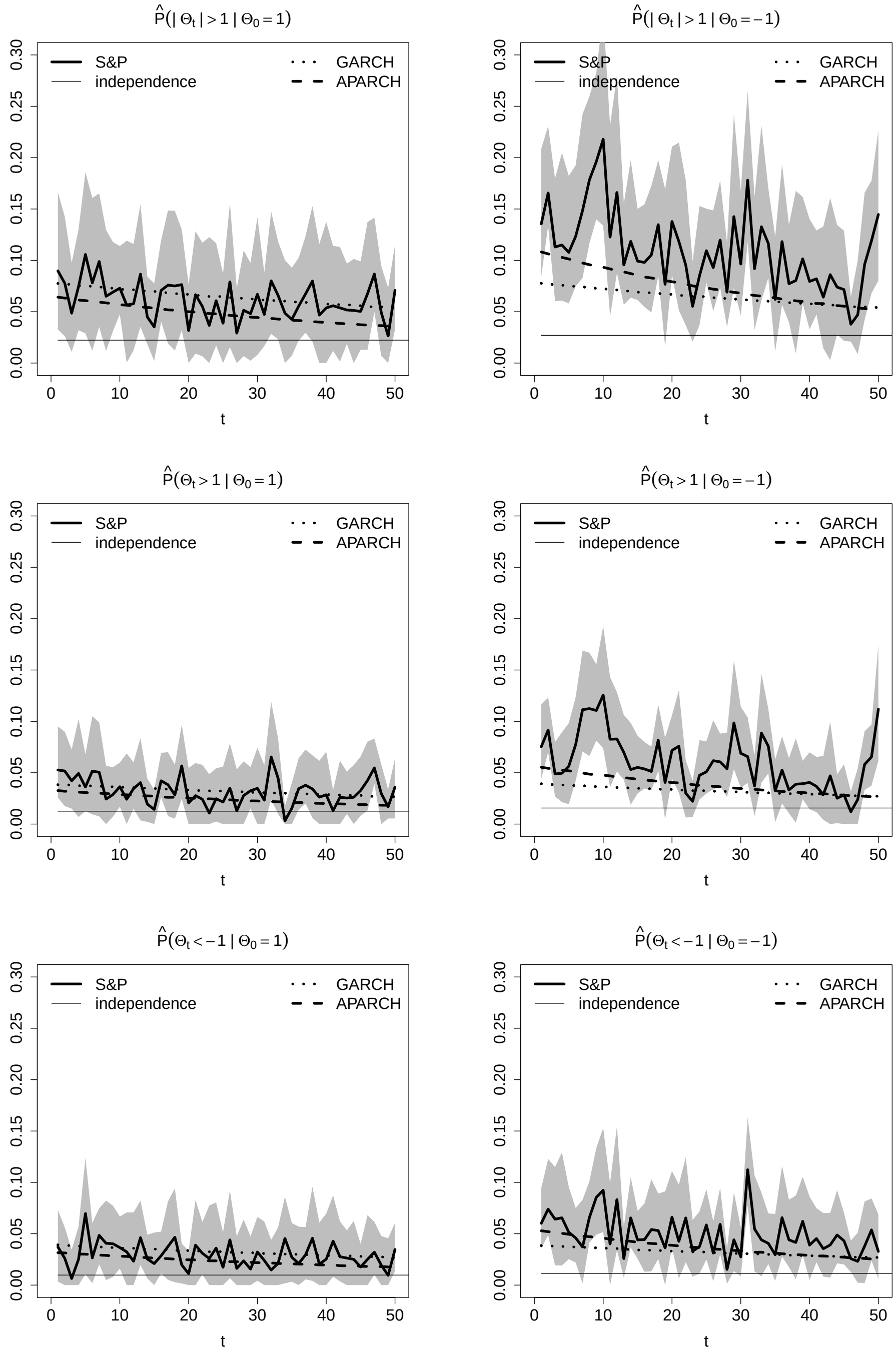

Figure 4: Sample spectral tail process probabilities (solid black bold line) for the S&P500 daily log-returns based on the backward estimator and the pre-asymptotic spectral tail process probabilities of the fitted GARCH(1,1) (dotted line) and APARCH(1,1) (dashed line) models. The gray area corresponds to the 80% pointwise confidence intervals for the pre-asymptotic spectral tail probabilities based on the multiplier bootstrap with 1 000 replications. Plots in the first column represent conditioning on a positive shock whereas in the second column one conditions on a negative shock. The horizontal line corresponds to the pre-asymptotic spectral tail probabilities under independence.

process probabilities for S&P500 daily log-returns estimated by the backward estimator. The pre-asymptotic values corresponding to the fitted models were calculated numerically via 10 000 Monte Carlo simulations with time series of length 10 000. Clearly, the APARCH model captures the asymmetry which the GARCH model cannot.

As a second example, we study daily log-returns on the P&G stock price between 1990-01-01 and 2010-01-01. The tail index is estimated at $\hat{\alpha} = 3.3$. We fit the GARCH(1,1) and APARCH(1,1) models to the time series and show the estimated parameters in the bottom part of Table 1.

In Figure 5 we plot the sample spectral tail process probabilities based on the daily log-returns themselves and on the residuals of the fitted GARCH(1,1) and APARCH(1,1) models obtained by the backward estimator. The top-right plot indicates that there is significant serial extremal dependence in the P&G daily log-returns triggered by the negative shocks. Due to high asymmetry in volatility, this feature is still present in the residuals of the fitted GARCH model whereas it is better removed by the APARCH filter.

6 Proofs

Proof of Lemma 2.1. To prove (2.3), apply the time-change formula (1.5) with $s = t = 0$, $i = -h$, and $f(y_0) = \mathbf{1}(y_0 \le x) - \mathbf{1}(0 \le x)$ to see that

$$\mathrm{P}[\Theta_h \le x] - \mathbf{1}(0 \le x) = \mathrm{E}[|\Theta_{-h}|^\alpha \mathbf{1}(\Theta_0/|\Theta_{-h}| \le x)] - \mathbf{1}(0 \le x)\,\mathrm{E}[|\Theta_{-h}|^\alpha].$$

For $x \ge 0$ in (2.4), apply the time-change formula (1.5) with $s = -h$, $t = 0$, $i = -h$ and $f(y_{-h}, \ldots, y_0) = \mathbf{1}(y_0 > x,\ y_{-h} = 1)$ to get

$$\mathrm{P}[\Theta_h > x,\ \Theta_0 = 1] = \mathrm{E}[|\Theta_{-h}|^\alpha\, \mathbf{1}(\Theta_0/|\Theta_{-h}| > x,\ \Theta_{-h} > 0)] = \mathrm{E}\big[\Theta_{-h}^\alpha\, \mathbf{1}(1/\Theta_{-h} > x,\ \Theta_0 = 1)\big],$$

whereas for $x < 0$, take $f(y_{-h}, \ldots, y_0) = \mathbf{1}(y_0 \le x,\ y_{-h} = 1)$ to obtain

$$\mathrm{P}[\Theta_h \le x,\ \Theta_0 = 1] = \mathrm{E}\big[\Theta_{-h}^\alpha\, \mathbf{1}(-1/\Theta_{-h} \le x,\ \Theta_{-h} > 0,\ \Theta_0 = -1)\big].$$

Similarly, in (2.5) choose $f(y_{-h}, \ldots, y_0) = \mathbf{1}(y_0 > x,\ y_{-h} = -1)$ and $f(y_{-h}, \ldots, y_0) = \mathbf{1}(y_0 \le x,\ y_{-h} = -1)$ for $x \ge 0$ and $x < 0$, respectively. $\square$

Next we turn to the asymptotic normality of the forward and backward estimators. Recall the definition of $X_{n,i}$ in (3.1). Consider the empirical process

$$\tilde{Z}_n(\psi) := (nv_n)^{-1/2} \sum_{i=1}^{n} \big(\psi(X_{n,i}) - \mathrm{E}[\psi(X_{n,i})]\big), \tag{6.1}$$

where ψ is one of the following functions:

$$\begin{aligned}
\phi_0\left(y_{-\tilde{t}}, \ldots, y_0, \ldots, y_{\tilde{t}}\right) &= \log^+(y_0), \\
\phi_1\left(y_{-\tilde{t}}, \ldots, y_0, \ldots, y_{\tilde{t}}\right) &= \mathbf{1}(y_0 > 1), \\
\phi_{2,x}^t\left(y_{-\tilde{t}}, \ldots, y_0, \ldots, y_{\tilde{t}}\right) &= \mathbf{1}(y_t/y_0 > x,\ y_0 > 1), \\
\phi_{3,x}^t\left(y_{-\tilde{t}}, \ldots, y_0, \ldots, y_{\tilde{t}}\right) &= (y_{-t}/y_0)^\alpha\, \mathbf{1}(y_0/y_{-t} > x,\ y_0 > 1)
\end{aligned} \tag{6.2}$$

for $|t| \in \{1, \ldots, \tilde{t}\}$ and $x \ge 0$. The asymptotic behavior of $\tilde{Z}_n$ can be derived from more general results by Drees and Rootzén (2010).

Proposition 6.1. *Let $(X_t)_{t\in\mathbb{Z}}$ be a non-negative, stationary, regularly varying time series with tail process $(Y_t)_{t\in\mathbb{Z}}$. Assume that conditions (A(x_0)), (B) and (C) are fulfilled for some $x_0 \ge 0$. Then, for all $y_0 \in [x_0, \infty) \cap (0, \infty)$, the sequence of processes*

$$\Big(\tilde{Z}_n(\phi_0),\ \tilde{Z}_n(\phi_1),\ \Big[(\tilde{Z}_n(\phi_{2,x}^t))_{x\in[x_0,\infty)},\ (\tilde{Z}_n(\phi_{3,y}^t))_{y\in[y_0,\infty)}\Big]_{|t|\in\{1,\ldots,\tilde{t}\}}\Big)$$

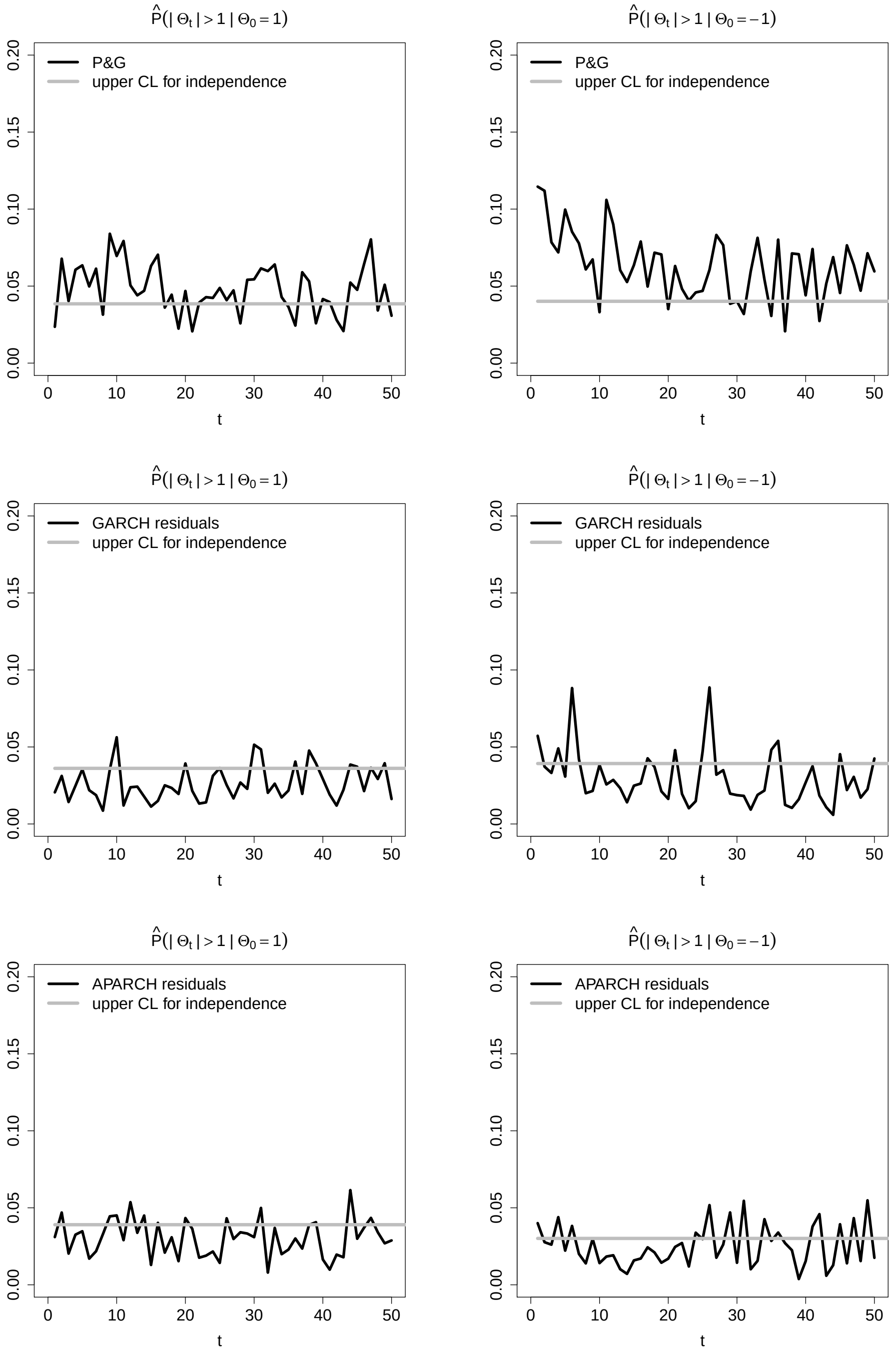

Figure 5: Sample spectral tail process (black line) for P&G daily log-returns (top), GARCH(1,1) residuals (middle), and APARCH(1,1) residuals (bottom) based on the backward estimator. Plots in the first column represent conditioning on a positive shock whereas in the second column one conditions on a negative shock. The horizontal gray lines correspond to the empirical 80% quantile of the backward estimator under independence obtained from 10 000 simulations.

converges weakly to a centered Gaussian process Z with covariance function given by

$$\operatorname{cov}(Z(\psi_1), Z(\psi_2)) = \sum_{j=-\infty}^{\infty} \mathrm{E}\big[\psi_1(Y_{-\tilde{t}}, \dots, Y_{\tilde{t}})\psi_2(Y_{j-\tilde{t}}, \dots, Y_{j+\tilde{t}})\big] =: c(\psi_1, \psi_2) \tag{6.3}$$

for all $\psi_1, \psi_2 \in \{\phi_0, \phi_1, \phi^t_{2,x}, \phi^t_{3,y} \mid x \geq x_0,\, y \geq y_0,\, |t| \in \{1, \dots, \tilde{t}\}\}$.

The weak convergence statements in Proposition 6.1 hold in the space of bounded functions on $\{\phi_0, \phi_1, \phi^t_{2,x}, \phi^t_{3,y} \mid x \geq x_0, y \geq y_0,\, |t| \in \{1, \dots, \tilde{t}\}\}$ equipped with the supremum norm; see van der Vaart and Wellner (1996, Section 1.5) for details.

Proof of Proposition 6.1. One can argue similarly as in the proof of Proposition B.1 of Drees et al. (2015), because the asymptotic equicontinuity of the process can be established for each t separately. Note that the discussion in Drees and Rootzén (2016) shows that part (ii) of condition (B) of Drees et al. (2015) is not needed.

By stationarity, the covariance of $Z(\psi_1)$ and $Z(\psi_2)$ is obtained as the limit of

$$\frac{1}{r_n v_n} \mathrm{E}\left[\sum_{i=1}^{r_n} \psi_1(X_{n,i}) \sum_{j=1}^{r_n} \psi_2(X_{n,j})\right] = \frac{1}{v_n} \sum_{k=-r_n+1}^{r_n-1} \left(1 - \frac{|k|}{r_n}\right) \mathrm{E}[\psi_1(X_{n,0})\psi_2(X_{n,k})].$$

This sum can be shown to converge to $c(\psi_1, \psi_2)$ using Pratt's lemma and Condition (C), as in Drees et al. (2015). □

Remark 6.2. The covariances can be expressed in terms of the spectral tail process. For example,

$$\begin{aligned} c(\phi^t_{3,x}, \phi_0) &= \sum_{j=-\infty}^{\infty} \mathrm{E}\big[\Theta^{\alpha}_{-t} \mathbf{1}(1/\Theta_{-t} > x) \log^+(Y_0\Theta_j)\big] \\ &= \sum_{j=-\infty}^{\infty} \mathrm{E}\big[\Theta^{\alpha}_{-t} \mathbf{1}(1/\Theta_{-t} > x)(\Theta^{\alpha}_j \wedge 1)(\log^+ \Theta_j + \alpha^{-1})\big]. \end{aligned}$$

Here we have used that Y_0 is independent of $(\Theta_s)_{s\in\mathbb{Z}}$ with distribution $\mathrm{P}[Y_0 > y] = y^{-\alpha}$ for $y \geq 1$.

Theorem 3.1 and Corollary 3.2 can now be proved in the same way as Theorem 4.5 in Drees et al. (2015). We omit the details, which can also be inferred from the more involved discussion of the bootstrap estimator below.

Drees (2015) has shown that under roughly the same conditions as used by Drees and Rootzén (2010), conditionally on the data, the following bootstrap version of the empirical process $\tilde{Z}_n$ has the same asymptotic behavior as $\tilde{Z}_n$:

$$Z_{n,\xi}(\psi) := (nv_n)^{-1/2} \sum_{j=1}^{m_n} \xi_j \sum_{i\in I_j} \big(\psi(X_{n,i}) - \mathrm{E}[\psi(X_{n,i})]\big), \tag{6.4}$$

with $I_j := \{(j-1)r_n + 1, \dots, jr_n\}$ and $m_n := \lfloor n/r_n \rfloor$. In what follows, the symbol E_ξ denotes the expectation w.r.t. $\xi = (\xi_j)_{j\in\mathbb{N}}$, i.e., the expectation conditionally on $(X_{n,i})_{1\leq i\leq n}$. Moreover, let BL_1 denote the set of all functions $g : \mathbb{R}^{4\tilde{t}+2} \to \mathbb{R}$ such that $\sup_{z\in\mathbb{R}^{4\tilde{t}+2}} |g(z)| \leq 1$ and $|g(z_1) - g(z_2)| \leq \|z_1 - z_2\|$ for all $z_1, z_2 \in \mathbb{R}^{4\tilde{t}+2}$.

Proposition 6.3. *Suppose that $(X_t)_{t\in\mathbb{Z}}$ is a non-negative, stationary, regularly varying time series and that the conditions (A(x_0)), (B) and (C) are fulfilled for some $x_0 \geq 0$. Then, for all $x \geq x_0$ and all $y_0 \in [x_0, \infty) \cap (0, \infty)$, one has*

$$\begin{aligned} &\Big(Z_{n,\xi}(\phi_0), Z_{n,\xi}(\phi_1), \big[Z_{n,\xi}(\phi^t_{2,x}), Z_{n,\xi}(\phi^t_{3,y})\big]_{|t|\in\{1,\dots,\tilde{t}\}}\Big) \\ &\xrightarrow{d} \Big(Z(\phi_0), Z(\phi_1), \big[Z(\phi^t_{2,x}), Z(\phi^t_{3,y})\big]_{|t|\in\{1,\dots,\tilde{t}\}}\Big) \end{aligned}$$

with Z as defined in Theorem 3.1. Moreover,

$$\sup_{g\in BL_1} \Big| \mathrm{E}_\xi g\Big(Z_{n,\xi}(\phi_0), Z_{n,\xi}(\phi_1), \big[Z_{n,\xi}(\phi^t_{2,x}), Z_{n,\xi}(\phi^t_{3,y})\big]_{|t|\in\{1,\dots,\tilde t\}}\Big) - \mathrm{E}g\Big(Z(\phi_0), Z(\phi_1), \big[Z(\phi^t_{2,x}), Z(\phi^t_{3,y})\big]_{|t|\in\{1,\dots,\tilde t\}}\Big)\Big| \to 0 \tag{6.5}$$

in probability.

Proposition 6.3 follows immediately from Drees (2015, Theorem 2.1), because in the proof of Proposition 6.1 (cf. the proof of Proposition B.1 of Drees et al. (2015)) it is shown that the assumptions of Drees (2015, Theorem 2.1) follow from the conditions of Proposition 6.3.

Now we are ready to prove the consistency of the multiplier block bootstrap procedure.

Proof of Theorem 3.3. We only prove consistency of the bootstrap version of the backward estimator, as the proof for the forward estimator is considerably simpler. For simplicity, we assume that $n = m_n r_n$. Let

$$\alpha_n := \frac{1}{\mathrm{E}\big[\log^+(X_0/u_n) \mid X_0 > u_n\big]} = \frac{v_n}{\mathrm{E}[\phi_0(X_{n,1})]}.$$

Recall $\tilde Z_n$ and $Z_{n,\xi}$ in (6.1) and (6.4) respectively, recall $I_j = \{(j-1)r_n+1,\dots,jr_n\}$, and recall $\hat\alpha_n$ and $\hat\alpha^*_n$ in (2.7) and (2.9), respectively. Then

$$\begin{aligned}(nv_n)^{1/2}(\hat\alpha^*_n - \hat\alpha_n) &= (nv_n)^{1/2}\frac{\sum_{j=1}^{m_n}\xi_j\sum_{i\in I_j}\mathbf{1}(X_i>u_n) - \hat\alpha_n\sum_{j=1}^{m_n}\xi_j\sum_{i\in I_j}\log^+(X_i/u_n)}{\sum_{j=1}^{m_n}(1+\xi_j)\sum_{i\in I_j}\log^+(X_i/u_n)}\\ &= \frac{Z_{n,\xi}(\phi_1) - \hat\alpha_n Z_{n,\xi}(\phi_0) + (r_nv_n)^{1/2}m_n^{-1/2}\sum_{j=1}^{m_n}\xi_j(1-\hat\alpha_n/\alpha_n)}{\alpha_n^{-1}(1+m_n^{-1}\sum_{j=1}^{m_n}\xi_j) + (nv_n)^{-1/2}\{\tilde Z_n(\phi_0) + Z_{n,\xi}(\phi_0)\}}.\end{aligned}$$

Since $m_n^{-1/2}\sum_{j=1}^{m_n}\xi_j$ and $\tilde Z_n$ are stochastically bounded and $\hat\alpha_n \to \alpha$ in probability, the assumptions $nv_n\to\infty$, $r_nv_n\to 0$, and $\alpha_n\to\alpha$, $m_n^{-1/2}$ and Proposition 6.3 ensure that

$$(nv_n)^{1/2}(\hat\alpha^*_n - \hat\alpha_n) = \alpha Z_{n,\xi}(\phi_1) - \alpha^2 Z_{n,\xi}(\phi_0) + o_P(1), \tag{6.6}$$

which converges weakly to $\alpha Z(\phi_1) - \alpha^2 Z(\phi_0)$. Moreover, conditionally on the data, it converges to the same limit weakly in probability in the sense of (6.5).

Next, recall $\hat F_n^{(\mathrm{b},\Theta_t)}(y)$ and $\hat F_n^{*(\mathrm{b},\Theta_t)}(y)$ in (2.6) and (2.8), respectively. For $y>0$, we have

$$\Big(1 - \hat F_n^{(\mathrm{b},\Theta_t)}(y)\Big)\sum_{i=1}^n \mathbf{1}(X_i>u_n) = \sum_{i=1}^n (X_{i-t}/X_i)^{\hat\alpha_n}\,\mathbf{1}(X_i/X_{i-t}>y,\, X_i>u_n).$$

It follows that

$$\begin{aligned}&\hat F_n^{(\mathrm{b},\Theta_t)}(y) - \hat F_n^{*(\mathrm{b},\Theta_t)}(y)\\ &= \Bigg[\sum_{i=1}^n\Bigg(\Big(\frac{X_{i-t}}{X_i}\Big)^{\hat\alpha^*_n} - \Big(\frac{X_{i-t}}{X_i}\Big)^{\hat\alpha_n}\Bigg)\mathbf{1}(X_i/X_{i-t}>y, X_i>u_n)\\ &\qquad + \sum_{j=1}^{m_n}\xi_j\sum_{i\in I_j}\Big(\frac{X_{i-t}}{X_i}\Big)^{\hat\alpha^*_n}\mathbf{1}(X_i/X_{i-t}>y, X_i>u_n)\\ &\qquad\qquad - \{1-\hat F_n^{(\mathrm{b},\Theta_t)}(y)\}\sum_{j=1}^{m_n}\xi_j\sum_{i\in I_j}\mathbf{1}(X_i>u_n)\Bigg] \Bigg/ \Bigg[\sum_{j=1}^{m_n}(1+\xi_j)\sum_{i\in I_j}\mathbf{1}(X_i>u_n)\Bigg].\end{aligned}$$

For any pair $(\underline{\alpha}, \overline{\alpha})$ such that $0 < \underline{\alpha} < \alpha < \overline{\alpha}$, there exists a constant $0 < C < \infty$ such that for all $\tilde{\alpha} \in [\underline{\alpha}, \overline{\alpha}]$ and, for suitable constants $\lambda = \lambda(\tilde{\alpha}) \in (0,1)$, we have, on the event $\{X_i/X_{i-t} > y_0\}$,

$$\begin{aligned}
&\left(\frac{X_{i-t}}{X_i}\right)^{\tilde{\alpha}} - \left(\frac{X_{i-t}}{X_i}\right)^{\alpha} - \left(\frac{X_{i-t}}{X_i}\right)^{\alpha} \log\left(\frac{X_{i-t}}{X_i}\right)(\tilde{\alpha}-\alpha) \\
&= \frac{1}{2}\left(\frac{X_{i-t}}{X_i}\right)^{\alpha+\lambda(\tilde{\alpha}-\alpha)} \log^2\left(\frac{X_{i-t}}{X_i}\right)(\tilde{\alpha}-\alpha)^2 \le C(\tilde{\alpha}-\alpha)^2.
\end{aligned}$$

Hence

$$\begin{aligned}
&\hat{F}_n^{(\mathrm{b},\Theta_t)}(y) - \hat{F}_n^{*(\mathrm{b},\Theta_t)}(y) \\
&= \Bigg[\sum_{i=1}^{n}\left(\frac{X_{i-t}}{X_i}\right)^{\alpha} \log\left(\frac{X_{i-t}}{X_i}\right)(\hat{\alpha}_n^* - \hat{\alpha}_n)\mathbf{1}(X_i/X_{i-t} > y,\, X_i > u_n) \\
&\quad + \sum_{j=1}^{m_n}\xi_j \sum_{i\in I_j}\left\{\left(\frac{X_{i-t}}{X_i}\right)^{\alpha} + \left(\frac{X_{i-t}}{X_i}\right)^{\alpha}\log\left(\frac{X_{i-t}}{X_i}\right)(\hat{\alpha}_n^*-\alpha)\right\}\mathbf{1}(X_i/X_{i-t} > y,\, X_i > u_n) \\
&\quad - (1-\hat{F}_n^{(\mathrm{b},\Theta_t)}(y))\sum_{j=1}^{m_n}\xi_j\sum_{i\in I_j}\mathbf{1}(X_i > u_n) + R_n(y)\Bigg] \Bigg/ \Bigg[\sum_{j=1}^{m_n}(1+\xi_j)\sum_{i\in I_j}\mathbf{1}(X_i > u_n)\Bigg] \qquad (6.7)
\end{aligned}$$

with

$$\begin{aligned}
|R_n(y)| &\le C(\hat{\alpha}_n^* - \hat{\alpha}_n)^2 \sum_{i=1}^{n}\mathbf{1}(X_i/X_{i-t} > y,\, X_i > u_n) \\
&\quad + C(\hat{\alpha}_n^*-\alpha)^2\sum_{j=1}^{m_n}|\xi_j|\sum_{i\in I_j}\mathbf{1}(X_i/X_{i-t} > y,\, X_i > u_n) \\
&= O_P\big((nv_n)^{-1}nv_n + (nv_n)^{-1}m_n r_n v_n\big) = O_P(1), \qquad n\to\infty.
\end{aligned}$$

Consider the function

$$\phi_{4,x}^t(y_{-\tilde{t}},\ldots,y_0,\ldots,y_{\tilde{t}}) = (y_{-t}/y_0)^{\alpha}\log(y_{-t}/y_0)\mathbf{1}(y_0/y_{-t} > x,\, y_{-t} > 0,\, y_0 > 1).$$

One may show as in the proof of Proposition 6.1 that $\tilde{Z}_n(\phi_{4,y}^t)$ and $Z_{n,\xi}(\phi_{4,y}^t)$ both converge weakly to $Z(\phi_{4,y}^t)$. In particular, as $n\to\infty$,

$$\begin{aligned}
&(nv_n)^{-1}\sum_{i=1}^{n}\left(\frac{X_{i-t}}{X_i}\right)^{\alpha}\log\left(\frac{X_{i-t}}{X_i}\right)\mathbf{1}(X_i/X_{i-t} > y, X_i > u_n) \\
&= \mathrm{E}\left[\left(\frac{X_{-t}}{X_0}\right)^{\alpha}\log\left(\frac{X_{-t}}{X_0}\right)\mathbf{1}(X_0/X_{-t} > y) \mid X_0 > u_n\right] + O_P\big((nv_n)^{-1/2}\big) \\
&\to \mathrm{E}\big[\Theta_{-t}^{\alpha}\log(\Theta_{-t})\mathbf{1}(1/\Theta_{-t} > y)\big] \\
&= -\mathrm{E}[\log(\Theta_t)\mathbf{1}(\Theta_t > y)],
\end{aligned}$$

where the last step follows from the time-change formula (1.5) applied with $f(y_0) = -\log(y_0)\mathbf{1}(y_0 > y)$ and $(-t,0,-t)$ instead of (s,t,i). Therefore

$$\begin{aligned}
&\sum_{i=1}^{n}\left(\frac{X_{i-t}}{X_i}\right)^{\alpha}\log\left(\frac{X_{i-t}}{X_i}\right)(\hat{\alpha}_n^*-\hat{\alpha}_n)\mathbf{1}(X_i/X_{i-t} > y, X_i > u_n) \\
&= -(nv_n)^{1/2}\big(\mathrm{E}[\log(\Theta_t)\mathbf{1}(\Theta_t > y)] + o_P(1)\big)(nv_n)^{1/2}(\hat{\alpha}_n^*-\hat{\alpha}_n). \qquad (6.8)
\end{aligned}$$

Likewise, one can conclude that

$$(nv_n)^{-1/2}\sum_{j=1}^{m_n}\xi_j\sum_{i\in I_j}\left(\frac{X_{i-t}}{X_i}\right)^{\alpha}\log\left(\frac{X_{i-t}}{X_i}\right)\mathbf{1}(X_i/X_{i-t} > y, X_i > u_n)$$

$$
\begin{aligned}
&= Z_{n,\xi}(\phi^t_{4,y}) + (rv_n)^{1/2} m_n^{-1/2} \sum_{j=1}^{m_n} \xi_j \,\mathrm{E}\Big[\Big(\frac{X_{-t}}{X_0}\Big)^{\alpha} \log\Big(\frac{X_{-t}}{X_0}\Big) \mathbf{1}(X_0/X_{-t} > y) \,\Big|\, X_0 > u_n \Big] \\
&= O_P(1).
\end{aligned}
$$

As a consequence,

$$
\begin{aligned}
&\sum_{j=1}^{m_n} \xi_j \sum_{i \in I_j} \Big\{ \Big(\frac{X_{i-t}}{X_i}\Big)^{\alpha} + \Big(\frac{X_{i-t}}{X_i}\Big)^{\alpha} \log\Big(\frac{X_{i-t}}{X_i}\Big) \big(\hat{\alpha}^*_n - \alpha\big) \Big\} \mathbf{1}(X_i/X_{i-t} > y, X_i > u_n) \\
&= (nv_n)^{1/2} Z_{n,\xi}(\phi^t_{3,y}) + \sum_{j=1}^{m_n} \xi_j r_n v_n \big(\mathrm{E}\big[\Theta^{\alpha}_{-t} \mathbf{1}(1/\Theta_{-t} > y)\big] + o(1) \big) + O_P(1) \\
&= (nv_n)^{1/2} \Big(Z_{n,\xi}(\phi^t_{3,y}) + O_P\big((r_n v_n)^{1/2}\big) + O_P\big((nv_n)^{-1/2}\big) \Big). \qquad (6.9)
\end{aligned}
$$

Moreover, we find, as $r_n v_n \to 0$ and $\sum_{j=1}^{m_n} \xi_j = O_P(m_n^{1/2})$, that

$$
\begin{aligned}
\sum_{j=1}^{m_n} \xi_j \sum_{i \in I_j} \mathbf{1}(X_i > u_n) &= (nv_n)^{1/2} Z_{n,\xi}(\phi_1) + r_n v_n \sum_{j=1}^{m_n} \xi_j \\
&= (nv_n)^{1/2} \big(Z_{n,\xi}(\phi_1) + o_P(1)\big), \qquad n \to \infty. \qquad (6.10)
\end{aligned}
$$

The denominator of (6.7) equals $nv_n + O_P((nv_n)^{1/2})$. Combining (6.7)–(6.10) and (6.6) yields

$$
\begin{aligned}
&(nv_n)^{1/2} \big(\hat{F}_n^{(\mathrm{b},\Theta_t)}(y) - \hat{F}_n^{*(\mathrm{b},\Theta_t)}(y) \big) \\
&= -\mathrm{E}[\log(\Theta_t) \mathbf{1}(\Theta_t > y)] (nv_n)^{1/2} (\hat{\alpha}^*_n - \hat{\alpha}_n) + Z_{n,\xi}(\phi^t_{3,y}) - (1 - \hat{F}_n^{(\mathrm{b},\Theta_t)}(y)) Z_{n,\xi}(\phi_1) + o_P(1) \\
&= Z_{n,\xi}(\phi^t_{3,y}) - \bar{F}^{(\Theta_t)}(y) Z_{n,\xi}(\phi_1) - \mathrm{E}[\log(\Theta_t) \mathbf{1}(\Theta_t > y)] \big(\alpha Z_{n,\xi}(\phi_1) - \alpha^2 Z_{n,\xi}(\phi_0)\big) + o_P(1).
\end{aligned}
$$

Now the assertion is a direct consequence of Proposition 6.3 and Theorem 3.1. □

Acknowledgements

The authors wish to thank the editors and the referees for their careful reading and for various constructive comments and useful suggestions. J. Segers gratefully acknowledges funding by contract "Projet d'Actions de Recherche Concertées" No. 12/17-045 of the "Communauté française de Belgique" and by IAP research network Grant P7/06 of the Belgian government (Belgian Science Policy). The research of M. Warchoł was funded by a PhD grant of the "Fonds de la Recherche Scientifique" (F.R.S.-FNRS). H. Drees was partially supported by DFG research grant JA 2160/1. R. Davis was supported in part by ARO MURI grant W911NF-12-1-0385.

References

Basrak, B. and J. Segers (2009). Regularly varying multivariate time series. *Stochastic Processes and Their Applications 119*(4), 1055–1080.

Black, F. (1976). Studies of stock price volatility changes. In *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economics Section*, pp. 177–181.

Bollerslev, T., V. Todorov, and S. Z. Li (2013). Jump tails, extreme dependencies, and the distribution of stock returns. *Journal of Econometrics 172*(2), 307–324.

Bollerslev, T., V. Todorov, and L. Xu (2015). Tail risk premia and return predictability. *Journal of Financial Economics 118*(1), 113–134.

Davis, R. A. and T. Mikosch (2001). Point process convergence of stochastic volatility processes with application to sample autocorrelation. *Journal of Applied Probability 38A*, 93–104.

Davis, R. A. and T. Mikosch (2009). The extremogram: a correlogram for extreme events. *Bernoulli 15*(4), 977–1009.

Davis, R. A., T. Mikosch, and I. Cribben (2012). Towards estimating extremal serial dependence via the bootstrapped extremogram. *Journal of Econometrics 170*(1), 142–152.

Ding, Z., C. W. Granger, and R. F. Engle (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance 1*(1), 83–106.

Drees, H. (2000). Weighted approximations of tail processes for β-mixing random variables. *The Annals of Applied Probability 10*(4), 1274–1301.

Drees, H. (2003). Extreme quantile estimation for dependent data with applications to finance. *Bernoulli 9*(4), 617–657.

Drees, H. (2015). Bootstrapping empirical processes of cluster functionals with application to extremograms. *arXiv preprint arXiv:1511.00420*.

Drees, H. and H. Rootzén (2010). Limit theorems for empirical processes of cluster functionals. *The Annals of Statistics 38*(4), 2145–2186.

Drees, H. and H. Rootzén (2016). Correction note to "limit theorems for empirical processes of cluster functionals". *The Annals of Statistics* (to appear).

Drees, H., J. Segers, and M. Warchoł (2015). Statistics for tail processes of Markov chains. *Extremes 18*(3), 369–402.

Han, H., O. Linton, T. Oka, and Y.-J. Whang (2016). The cross-quantilogram: Measuring quantile dependence and testing directional predictability between time series. *Journal of Econometrics 193*(1), 251–270.

Leadbetter, M. (1983). Extremes and local dependence in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 65*(2), 291–306.

Linton, O. and Y.-J. Whang (2007). The quantilogram: With an application to evaluating directional predictability. *Journal of Econometrics 141*(1), 250–282.

Mikosch, T. and C. Stărică (2000). Limit theory for the sample autocorrelations and extremes of a GARCH (1,1) process. *The Annals of Statistics 28*(5), 1427–1451.

Politis, D. N. and J. P. Romano (1994). The stationary bootstrap. *Journal of the American Statistical association 89*(428), 1303–1313.

Tjøstheim, D. and K. O. Hufthammer (2013). Local gaussian correlation: a new measure of dependence. *Journal of Econometrics 172*(1), 33–48.

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

Wuertz, D., Y. C. with contribution from Michal Miklovic, C. Boudt, P. Chausse, and others (2013). *fGarch: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling.* R package version 3010.82.