Inference on heavy tails from dependent data

S.Y.Novak

Siberian Adv. Math., 2002, v. 12, No 2, 73–96.

Abstract

Statistics of extremes has been well developed for the case of independent and identically distributed (i.i.d.) observations. In a growing number of applications, however, the data appears dependent and heavy–tailed.

We deal with problems of tail index and extreme quantile estimation from a sample of dependent random variables. Consistency and asymptotic normality of the corresponding estimators are established under mild mixing conditions. The accuracy of estimation is shown to be of the same order as if the data were independent.

We suggest an approach to bias reduction. Besides limit theorems, we present a procedure of practical estimation.

Key words and phrases: tail index, extreme quantile, heavy tails, dependence, Value–at–Risk.

1 Introduction

Statistics of extremes aims to estimate the tail probability $\mathbb{P}(X > x)$ when x is "large". The other quantities of interest are the tail index and extreme quantiles. These problems have important applications in finance, insurance, network modelling, meteorology, etc. (see, e.g., [4, 9, 28] and references therein; the fact that financial data often exhibits heavy tails is discussed in textbooks, cf. [17], ch. 11).

In the case of a parametric family of distributions and i.i.d. data, the maximum likelihood approach yields natural estimators of the tail probabilities (cf. [9]). Unfortunately, one cannot be confident that the distribution belongs to a particular parametric family. Besides, the assumption of independence appears unrealistic in a growing number of applications.

The present paper deals with the problems of non–parametric tail index and extreme quantile estimation from a sample of dependent data. We show that, under mild mixing conditions, the accuracy of estimation is of the same order as if the data were independent.

Our main tool is the *ratio estimator* of the tail index. This estimator seems to have advantages, both practical and theoretical, over Hill's and some other tail index estimators (see conclusions below).

In section 2 we recall the main properties of the ratio estimator (in the i.i.d. case) before presenting the results. In Section 3 we present results on tail index and tail constant estimation from a sample of dependent heavy–tailed random variables (r.v.s).

Our approach differs from those in [6, 14, 29, 30, 36]. Roughly speaking, we assume that a mixing coefficient tends to zero not slower than $(\ln l)^{-c}$ for some constant $c > 1$. This condition is fulfilled in many particular parametric models, including the popular ARCH/GARCH model (for the definition of ARCH/GARCH, see [9]). We suggest also a procedure of bias reduction.

Section 4 is devoted to the problems of extreme quantile and *Expected Shortfall* estimation. These problems arise when one measures risk of heavy losses of a portfolio of risky assets. In particular, *Value–at–Risk* (VaR) is usually defined as the $q\%$–quantile for a small q. Expected Shortfall (ES) is the corresponding conditional expectation: if VaR equals y then (we prefer to deal with the right–hand tail)

$$E(y) = \mathbb{E}\{X - y | X > y\}\,.$$

There are two different approaches to quantile estimation. The classical one suggests using the empirical quantile $Q_n = F_n^{-1}$, where F_n is the empirical distribution function. Sharp results on asymptotics of Q_n in the case of i.i.d. data can be found in [7, 8, 33], see also references therein. Normal approximation for a weighted empirical quantile process in the case of dependent data is presented in [6].

The empirical quantile $Q_n(q)$ becomes unreliable if q is small, which is not a rare situation when one estimates risk of heavy losses. In such cases, the Extreme Value Theory (EVT) approach suggests using the features of the distribution (1) when constructing a quantile estimator. The EVT approach seems to be the only one that works for a "small" q (we allow $q = q(n) \to 0$ as the sample size $n \to \infty$).

The EVT approach has been discussed in the literature on extremes for a long time (see [34] and references therein). In the case of a parametric family of distributions and i.i.d. data, the EVT approach is presented in [9], Section 6.5, and [18].

We apply the non–parametric EVT approach to the problem of extreme quantile estimation from a sample of dependent data. We introduce a new quantile estimator and suggest simple sufficient conditions for the consistency and asymptotic normality of estimators of the tail index, tail constant and extreme quantiles. We present also a procedure of practical estimation.

Unless otherwise specified, the limits are as $n \to \infty$ and $\sum^n$ denotes summation from 1 to n. We write $a_n \sim b_n$ if $a_n/b_n \to 1$, $a_n \ll b_n$ if $a_n/b_n \to 0$, and $a_n \asymp b_n$ if $0 < \liminf a_n/b_n < \limsup a_n/b_n < \infty$; $\Phi_{a;b}$ is the distribution function of the normal $\mathcal{N}(a;b)$ law, $\Phi = \Phi_{0;1}$.

2 The ratio estimator

We say that the distribution has a *heavy tail* if

$$G(x) \equiv \mathbb{P}(X > x) = L(x)x^{-1/a} \qquad (a > 0) \tag{1}$$

for all large enough x, where the (unknown) function L is slowly varying at infinity:

$$\lim_{x\to\infty} L(xt)/L(x) = 1 \qquad (\forall t > 0)\,. \tag{2}$$

The number $1/a$ is called the *tail index*. In the case $L(x) = C + o(1)$, C is called the *tail constant.*

Distributions that obey (1) form a *non–parametric family* of probability laws. Our purpose is to estimate the index a (equivalently, the tail index $1/a$), extreme quantile (VaR), Expected Shortfall (ES) and the tail constant (when it exists).

In this section we assume that $X_1, \ldots, X_n$ are independent random variables distributed according to (1).

The *ratio estimator*

$$a_n \equiv a_n(x_n) = \sum^n \ln(X_i/x_n)\mathbb{1}\{X_i > x_n\} \Big/ \sum^n \mathbb{1}\{X_i > x_n\} \tag{3}$$

was introduced by Goldie and Smith [11]. In the case of a parametric family of distributions $\mathbb{P}(X > x) = x^{-1/a}$ $(x \geq 1, a > 0)$, $a_n(1)$ is the *maximum likelihood estimator* (MLE). The ratio estimator is also a *least squares estimator* in the following sense: the function $g(a) = \sum^n (\ln(X_i/x) - a)^2\, \mathbb{1}_i$ takes on its minimum at $a = a_n(x)$.

The threshold level x_n needs to be chosen properly. If x_n is too small then the bias of the ratio estimator is large; if x_n is too large then the bias is small but the variance is large (since only a small part of a sample contributes to the inference). The assumption

$$p_n \to 0\,,\ np_n \to \infty \tag{4}$$

where $p_n = \mathbb{P}(X > x_n)$, means that x_n is neither too small nor too large. It guarantees the consistency of the ratio estimator in the case of independent data (see [20]).

Let $X_{(n)} \leq ... \leq X_{(1)}$ be the sample order statistics. Denote $x_n^* = X_{(k_n+1)}$, where k_n is an integer number. The statistic $a_n(x_n^*)$ is Hill's estimator

$$a_n^H = k_n^{-1} \sum_{i=1}^{k_n} \ln(X_{(i)}/X_{(k_n+1)})\,.$$

A number of other estimators of the tail index can be found in [3, 4, 28]. A comparison of the asymptotic performance of some tail index estimators is given in [12, 19]. Concerning Hill's and the ratio estimator, the comparison of rates of tail index estimation in [19] is in favour of the ratio estimator.

Denote

$$a^* = \mathbb{E}\{\ln(X/x_n)|X > x_n\}\,, \ v = a^*/a - 1$$

(we suppress the dependence of a^* and v on x_n). The ratio estimator (3) is the sample analog of $a^*(x_n)$. According to the relation (7) below, $a^*(x) \to a$ as $x \to \infty$.

It is shown in [19, 20] that

$$\sqrt{np_n}(a_n/a - 1) \Rightarrow \mathcal{N}(0;1) \tag{5}$$

if and only if $np_n v^2 \to 0$; if $v\sqrt{np_n} \to b$ then $\sqrt{np_n}(a_n/a-1) \Rightarrow \mathcal{N}(b;1)$. In these limit theorems, $\sqrt{np_n}$ may be replaced by $N_n^{1/2}$, where

$$N_n \equiv N_n(x_n) = \sum\nolimits_{i=1}^n \mathbb{I}\{X_i > x_n\}$$

is the *number of exceedances* over the threshold x_n.

If $v\sqrt{np_n} \to 0$ then $\left[a_n\Big/\left(1+q_\varepsilon N_n^{-1/2}\right); a_n\Big/\left(1-q_\varepsilon N_n^{-1/2}\right)\right]$ with $\Phi(-q_\varepsilon) = \varepsilon/2$ is the *asymptotic confidence interval* (a.c.i.) of level $1-\varepsilon$ for the index a.

Asymptotic confidence intervals do not take into account the accuracy of normal approximation and hence may be far away from exact ones if the sample size is not large and the rate of convergence in the corresponding limit theorem is not fast. The *non–asymptotic confidence intervals* $I_\varepsilon = \left[a_n\Big/\left(1+y_\varepsilon N_n^{-1/2}\right); a_n\Big/\left(1-y_\varepsilon N_n^{-1/2}\right)\right]$ have been introduced in [25]. Here $\Phi(-y_\varepsilon) = \left(\varepsilon/2 - C_* N_n^{-1/2}\right)_+$ and $C_* < 0.8$ is the constant from the Berry–Esseen inequality.

The theoretically optimal threshold x_n^{opt} is the value x_n minimizing the main terms in the asymptotic expansion for the mean squared error $\mathbb{E}(a_n - a)^2 = bias^2 + variance$. The statistic (3) seems to be the only tail index estimator for which the asymptotics of the bias and the mean squared error (MSE) have been calculated (see [19, 21]):

$$\mathbb{E}(a_n/a - 1) = v + O\left((np_n)^{-2}\right)\,,\ \mathbb{E}(a_n/a-1)^2 \sim (np_n)^{-1} + v^2\,. \tag{6}$$

The condition $v\sqrt{np_n} \to b \neq 0$ balances the terms on the right–hand side of (6). Using the relation

$$\mathbb{E}\{\ln^k(X/x_n)|X > x_n\} = a^k k!(1+v_k) \sim a^k k! \qquad (k \in \mathbb{N})\,, \tag{7}$$

where

$$v_k \equiv v_k(x_n) = \int_0^\infty h_n(u)e^{-u}du^k/k!\,,\ \ h_n(u) = L^{-1}(x_n)L(x_n e^{au}) - 1 \tag{8}$$

(see [21]), an explicit expression for x_n^{opt} can be drawn under additional restrictions on the distribution (1).

Example 1. Consider the following non–parametric family of distributions:

$$\mathcal{P}_{a,b,c,d} = \left\{ \mathbb{P} : \mathbb{P}(X > x) = cx^{-1/a} \left(1 + dx^{-b/a} + o(x^{-b/a})\right)\right\} .$$

If $\mathbb{P} \in \mathcal{P}_{a,b,c,d}$ then, using (7), one gets $v(x) \sim -bd(1+b)^{-1}x^{-b/a}$. Hence the asymptotically optimal value of the threshold x_n is $x_n^{opt} = (2bcDn)^{\frac{a}{1+2b}}$, where $D = (bd/(1+b))^2$, and

$$\mathbb{E}\left(a_n(x_n^{opt})/a - 1\right)^2 \sim (1+2b)D^{\frac{1}{1+2b}}\,(2bcn)^{\frac{-2b}{1+2b}} \tag{9}$$

(cf. [23, 24]). The rate $n^{\frac{-2b}{1+2b}}$ is, in a sense, the best possible: a lower bound of that order can be deduced from Theorem 3.1 of Pfanzagl [27].

For instance, the standard Cauchy distribution belongs to the class $\mathcal{P}_{1,2,1/\pi,-1/3}$, $x_n^{opt} = (16n/81\pi)^{1/5}$ and $\mathbb{E}\,(a_n(x_n^{opt})/a-1)^2 \sim \frac{5}{4}\,(16/81)^{1/5}\,(\pi/n)^{4/5}$.

Adaptive versions of x_n^{opt} may be constructed by replacing the numbers a, b, c, d with their consistent estimators $\hat{a}, \hat{b}, \hat{c}, \hat{d}$ such that $|\hat{a} - a| + |\hat{b} - b| = o_p(1/\ln n)$.

Note that the ratio estimator (as well as some other tail index estimators) is not shift invariant. It is easy to suggest a modification that has the shift invariance property. For instance, consider the estimator $a_n^*(x) = \sum^n \ln((X_i - m_n)/x)\mathbb{1}_i^* \Big/ \sum^n \mathbb{1}_i^*$, where $m_n = X_{[n/2],n}$ is the sample median and $\mathbb{1}_i^* = \mathbb{1}\{X_i - m_n > x\}$. Then $\mathbb{P}(X_i - m_n > x)$ obeys (1), $\sum^n \mathbb{1}_i^* \Big/ \sum^n \mathbb{1}_i \xrightarrow[p]{} 1$ and $\sum^n \ln((X_i - m_n)/x)\mathbb{1}_i^* \Big/ \sum^n \ln(X_i/x)\mathbb{1}_i \xrightarrow[p]{} 1$ as $x = x_n \to \infty$ by the LLN and the properties of slowly varying functions. Hence $a_n^*(x)/a_n(x) \xrightarrow[p]{} 1$. Existing applications, however, present data with natural origin points, and hence do not provide evidence in support to estimators with the shift invariance property.

3 Tail index estimation

Given a sample $X_1, \ldots, X_n$ from a (strictly) stationary sequence $X, X_1, X_2, \ldots$ of random variables (r.v.s) with marginal distribution (1), we want to estimate the tail index and the tail constant.

Recall the definition of the mixing coefficients $\rho(\cdot)$ and $\varphi(\cdot)$:

$$\begin{aligned} \rho(l) &= \sup_i \sup \left\{ \mathrm{corr}(\xi\eta) \ : \xi \in \mathcal{F}_{1,i},\ \eta \in \mathcal{F}_{i+l,\infty},\ \mathbb{E}(\xi^2 + \eta^2) < \infty \right\}, \\ \varphi(l) &= \sup_i \sup \left\{ |\mathbb{P}(B|A) - \mathbb{P}(B)| : A \in \mathcal{F}_{1,i},\ B \in \mathcal{F}_{i+l,\infty} \right\}, \end{aligned}$$

where $\mathcal{F}_{1,i} = \sigma\{X_1, ..., X_i\}$, $\mathcal{F}_{i,\infty} = \sigma\{X_{i+1}, ...\}$.

Conditions. Throughout the paper we assume (4) and the following mixing condition:

$$\sum_{i \geq 1} i^{-1}\rho(i) < \infty . \tag{10}$$

Besides, in all statements except Propositions 1 and 5 we assume $\varphi(l) \to 0$ as $l \to \infty$.

Mixing conditions of this type are typical in the literature on sums of dependent r.v.s (cf. [26, 37, 38]). Condition (10) is satisfied even if $\rho(l)$ decays like $(\ln l)^{-c}$ for some $c > 1$. Since $\rho(l) \le 2\varphi^{1/2}(l)$ (see [2]), condition (10) is valid if

$$\sum_{i\ge 1} i^{-1}\varphi^{1/2}(i) < \infty \,. \tag{11}$$

In many models (like popular ARCH/GARCH processes) $\varphi(\cdot)$ decays exponentially fast (see [5]).

Proposition 1 *The ratio estimator is consistent:*

$$a_n \xrightarrow[p]{} a \,.$$

If $\lim_{x\to\infty} L(x) = C$ *and* $(\ln x_n)^2(v^2 + 1/np_n) \to 0$ *then*

$$\hat{C}_n \equiv \hat{C}_n(x_n) = x_n^{1/a_n} n^{-1} \sum^{n} 1\!\!\mathrm{I}\{X_i > x_n\}$$

is a consistent estimator of the tail constant: $\hat{C}_n \xrightarrow[p]{} C$.

The estimator $\hat{C}_n$ has been introduced by Goldie and Smith [11]. In the case of i.i.d. data, sufficient conditions for consistency and asymptotic normality of $\hat{C}_n$ are given in [11, 19, 20].

Denote $1\!\!\mathrm{I}_i = 1\!\!\mathrm{I}\{X_i > x_n\}$, and let

$$Y_i = \ln(X_i/x_n)1\!\!\mathrm{I}_i \,, \;\; Y_i^* = Y_i - a^* 1\!\!\mathrm{I}_i \,.$$

Theorem 2 *Suppose that* $(a^* - a)\sqrt{np_n} \to \mu$ *and* $\mathbb{E}\left(\sum^n Y_i^*\right)^2 \sim \sigma^2 np_n$ *for some* $\sigma \in (0;\infty)$, $\mu \in \mathbb{R}$. *Then*

$$\frac{a_n - a}{\sigma}\sqrt{np_n} \Longrightarrow \mathcal{N}(\mu/\sigma; 1)\,. \tag{12}$$

Remark 1. In Theorems 2, 7 and Corollary 3 below, $\sqrt{np_n}$ may be replaced by $N_n^{1/2}$.

In the i.i.d. case we have $\sigma = a$, and (12) becomes

$$(a_n/a - 1)\sqrt{np_n} \Longrightarrow \mathcal{N}(\mu/a; 1)\,. \tag{13}$$

According to (12), $a_n = a + \xi_n/\sqrt{np_n}$, where the distribution of the r.v. ξ_n converges to a normal one. If (12) holds together with the convergence of the second moment and

$$\mathbb{P}(X > x) = cx^{-1/a}\left(1 + O\left(x^{-b/a}\right)\right) \qquad (\exists b > 0) \tag{14}$$

then (8) and (12) imply

$$\mathrm{MSE}\,(a_n) \equiv \mathbb{E}(a_n - a)^2 = O\left(n^{-2b/(1+2b)}\right)$$

if $x_n \asymp n^{a/(1+2b)}$. In other words, *the rate of approximation* $a_n \approx a$ *is the same as if the data were independent.*

In the case of dependent data, σ usually is not known, and one would want to replace σ by a consistent estimator $\hat{\sigma}$.

Denote $T_{k,j} = \sum_{l=(j-1)r+1}^{jr} Y_l^k \mathbb{1}_l$, and let

$$\hat{\sigma}_k^2 \equiv \hat{\sigma}_k^2(n) = N_n^{-1} \sum_{j=1}^{[n/r]} T_{k-1,j}^2 \,,\ \hat{\sigma}_{lm} \equiv \hat{\sigma}_{lm}(n) = N_n^{-1} \sum_{j=1}^{[n/r]} T_{l-1,j} T_{m-1,j}$$

($0 \le l < m, k \ge 1$, $1 \ll r = r(n) \ll n$). It follows from (7) and moment inequalities for sums of dependent r.v.s (see [26, 37]) that there exist constants c_k such that $\mathrm{Var}\left(\sum^n Y_i^{k-1}\mathbb{1}_i\right) \le c_k n p_n$ for every $k \in \mathbb{N}$. In the i.i.d. case,

$$\mathrm{Var}\left(\sum\nolimits^n Y_i^{k-1}\mathbb{1}_i\right) \sim \sigma_k^2 n p_n \,,\ \mathbb{E}\sum\nolimits^n \mathbb{1}_i \sum\nolimits^n \bar{Y}_j \sim \sigma_{12} n p_n \tag{15}$$

for some $\sigma_k \in (0;\infty)$, $\sigma_{12} \in \mathbb{R}$, where $\bar{Y}_i = Y_i - a^* p_n$. One can expect that (15) holds also in the case of weakly dependent observations. Sufficient conditions for (15) can be drawn using results in [16].

Corollary 3 *Suppose that* $\sigma^2 \equiv (a\sigma_1)^2 + \sigma_2^2 - 2a\sigma_{12} > 0$, $1 \ll r = r(n) \ll n$ *and* $(a^* - a)\sqrt{np_n} \to \mu$. *If (15) holds for* $k = 1, 2$ *then* $Var\,(\sum^n Y_i^*) \sim \sigma^2 n p_n$ *and*

$$\frac{a_n - a}{\hat{\sigma}} \sqrt{np_n} \Longrightarrow \mathcal{N}(\mu/\sigma; 1)\,, \tag{16}$$

where $\hat{\sigma}^2 \equiv \hat{\sigma}^2(n) = (a_n\hat{\sigma}_1)^2 + \hat{\sigma}_2^2 - 2a_n\hat{\sigma}_{12}$.

Bias reduction. From a practical point of view, it can sometimes be preferable to drop the accuracy of approximation in order to eliminate the asymptotic bias μ/σ . Note that the accuracy of normal approximation in (12) reduces when we use the estimator $\hat{\sigma}$ instead of the unknown σ since $\hat{\sigma} - \sigma = O_p\Big((np_n/r)^{-1/2}\Big)$. Therefore, though

$$\sup_y \left| \mathbb{P}\left(\frac{a_n - a}{\sigma}\sqrt{np_n} < y\right) - \Phi_{\frac{\mu}{\sigma};1}(y)\right| \le C_*(np_n)^{-1/2} \tag{17}$$

(at least in the i.i.d. case, cf. [21, 25]), the right–hand side of (17) may become $C_+(np_n/r)^{-1/2}$ if σ is replaced by $\hat{\sigma}$ in the left–hand side. Thus, we do not lose much if we switch to an estimator $a_{n,r}$ such that the rate of normal approximation $a_{n,r} \approx a$ is, in a sense, $O\Big((np_n/r)^{-1/2}\Big)$.

Denote

$$N_{n,r} = \sum\nolimits_{i=1}^{[n/r]} \mathbb{1}_{ir} \,,\ a_{n,r} \equiv a_{n,r}(x_n) = \sum\nolimits_{i=1}^{[n/r]} Y_{ir}/N_{n,r} \,.$$

Theorem 4 *Suppose that* $r = r(n) \in \{1, ..., n\}$ *is chosen so that*

$$r \to \infty\,,\;\; np_n/r \to \infty\,,\;\; v^2 np_n = o(r)\,. \tag{18}$$

Then

$$\left(\frac{a_{n,r}}{a} - 1\right)\sqrt{np_n/r} \Longrightarrow \mathcal{N}(0;1)\,,\;\; \left(\frac{a_{n,r}}{a} - 1\right)\sqrt{N_{n,r}} \Longrightarrow \mathcal{N}(0;1)\,. \tag{19}$$

If we assume conditions of Corollary 3 then (18) holds, e.g., when $r = g(np_n)$, where $g(x) \to \infty$, $g(x)/x \to 0$ as $x \to \infty$. According to (19),

$$I_{n,r} = \left[a_{n,r}\Big/\left(1 + q_\varepsilon N_{n,r}^{-1/2}\right); a_{n,r}\Big/\left(1 - q_\varepsilon N_{n,r}^{-1/2}\right)\right]$$

is the a.c.i. of level $1 - \varepsilon$, where $\Phi(-q_\varepsilon) = \varepsilon/2$.

The important question is how to choose the threshold x_n. A simple practical approach is to plot $a_n(\cdot)$ and then choose the estimate from an interval in which the function $a_n(\cdot)$ demonstrates stability. The background for this approach is provided by the consistency result. Indeed, if the sequence $\{x_n\}$ obeys (4) then so does $\{tx_n\}$ for every $t > 0$. Hence there must be an interval of threshold levels $[x_-; x_+]$ (formed by a significant number of sample points) such that $a_n(x) \approx a$ for all $x \in [x_-; x_+]$.

We suggest choosing the *average value* $\hat{a} = \text{mean}\{a_n(x) : x \in [x_-; x_+]\}$. Then $x_n \in [x_-; x_+]$ can be chosen as a point such that $a_n(x_n) = \hat{a}$. Despite fluctuations with the choice of x_- and x_+, the resulting estimate will be almost the same (this is the advantage of taking average). Examples in [23] show that this procedure works satisfactorily.

Remark 2. Weak dependence conditions are often expressed in terms of either α, β, φ or ρ mixing coefficients (the definitions of mixing coefficients can be found, e.g., in [2, 16]). Using Bernstein's "blocks" approach, one can check that $a_n \xrightarrow[p]{} a$ if (10) is replaced by the condition

$$\left(np_n^{1/2}\right)^{-1}\sum_{i=1}^{r}\alpha^{1/2}(i) + rp_n + [n/r](lp_n + \alpha(l)) \to 0 \tag{20}$$

for some sequences $l = l(n)$, $r = r(n)$ such that $1 \le l \le r \le n$.

Conditions of (20)–type appear when one uses Bernstein's method. Typically, conditions are formulated in terms of the mixing coefficient $\alpha(\cdot)$ (though in [36, 6], conditions are given in terms of the stronger coefficient $\beta(\cdot)$). In particular, Starica ([36], formulas (2.20) and (3.2)) assumes that

$$nr^{-1}\beta(l) + rk^{-1/2+\varepsilon} + krn^{-1} \to 0 \tag{21}$$

for some $\varepsilon \in (0; 1/2)$ and some sequences $l = l(n)$, $k = k(n)$, $r = r(n)$ such that $1 \ll l \le r \ll n$, $1 \ll k \ll n$.

Condition (10) (or (11)) is preferable if the mixing coefficients have the same rate of decay. To illustrate this point, compare, for instance, (11) with (21) in the situation where $\beta(l) \asymp \varphi(l) \asymp (\ln l)^{-3}$. Since $nr^{-1}\beta(l) = o(1)$ and $k = o(n/r)$, we have $k = o\left((\ln l)^3\right)$. Therefore, $rk^{-1/2+\varepsilon} \gg r(\ln l)^{-4.5+3\varepsilon} \geq l(\ln l)^{-4.5+3\varepsilon} \to \infty$. Hence (21) does not hold while (11) is evidently valid.

Sufficient conditions for the asymptotic normality of the ratio estimator can be deduced also from the results of [30]. Sufficient conditions for the asymptotic normality of Hill's estimator (in terms of the mixing coefficient $\beta(\cdot)$) have been obtained in [36] for the stationary solution of a stochastic difference equation

$$X_i = A_i X_{i-1} + B_i\,, \tag{22}$$

where $\{(A_i, B_i), i \geq 1\}$ is a sequence of i.i.d.r.v.s. According to Goldie [10], (14) holds under some natural assumptions on the r.v.s A_i, B_i. Starica [36] showed that

$$\sqrt{k}\,(a_n^{H}/a - 1) \Longrightarrow \mathcal{N}(0; 1 - 2\delta) \tag{23}$$

if the sample fraction k_n obeys $(\ln n)^{2+\varepsilon} \ll k_n \ll n^{\kappa}$ for some $\varepsilon > 0$, where $\delta = \sum_{j=1}^{\infty} \int_0^1 \mathbb{P}(A_1 \times ... \times A_j > v^a)dv$ and $\kappa = (2/3 + \varepsilon) \wedge a/(a+1) \wedge b/(b+1) < 2b/(2b+1)$. We know from the results for the i.i.d. case (see [13]) that the optimal rate of the sample fraction is $k_n \asymp n^{2b/(2b+1)}$: it yields $\mathrm{MSE}\,(a_n^{H}) = O\left(n^{-2b/(1+2b)}\right)$. The assumption $k_n \ll n^{\kappa}$ means that the rate of approximation $a_n^{H} \approx a$ in (23) is worse than in the i.i.d. case.

Theorem 3.1 and Corollary 3.3 in [6] imply the asymptotic normality of Hill's estimator under mixing conditions similar to those of [36] plus the assumption that the sample fraction k_n obeys the condition $\ln^2 n \ln^4(\ln n) \ll k_n \ll n^{2b/(2b+1)}$. Hence the rate of approximation $a_n^{H} \approx a$ is again sub–optimal.

4 Quantile estimation

In this section we deal with the problem of extreme quantile (VaR) and Expected Shortfall estimation from a sample of dependent heavy–tailed data.

Let $y_q = \inf\{t : G(t) \leq q\}$ be the upper quantile of level q. Given $q = q(n) \to 0$, we want to construct an estimator $\hat{y}_q = \hat{y}_q(n)$ such that

$$\hat{y}_q/y_q \xrightarrow[p]{} 1\,,\ G(\hat{y}_q)/q \xrightarrow[p]{} 1\,. \tag{24}$$

Since $y_q = y_q(n)$ may be so large that only few elements of the sample exceed it, the sample quantile can hardly be regarded as a reliable estimator. The idea of the EVT approach is to use a "pilot" level x_n and the properties of a heavy–tail distribution when constructing an estimator of y_q.

More precisely, (1) entails the weak convergence

$$\mathcal{L}((X/y)|X > y) \Longrightarrow F_a\,,$$

where $F_a(x) = 1 - x^{-1/a}$ $(x \geq 1)$. Hence

$$G(y_q) = \mathbb{P}(X > x_n)\mathbb{P}(X > y_q | X > x_n) \approx \frac{N_n}{n} \left(\frac{y_q}{x_n}\right)^{-1/a} .$$

Let $\hat{a}_n$ be a consistent estimator of the index a. Since $G(y_q) \sim q$ by formula (30) below, one can expect that y_q may be well approximated by the statistic

$$\hat{y}_q \equiv \hat{y}_q(x_n) = (N_n/qn)^{\hat{a}_n} x_n . \tag{25}$$

Proposition 5 (consistency of extreme quantile/VaR estimator) *Suppose that*

$$(\hat{a}_n - a) \ln(p_n/q) \underset{p}{\longrightarrow} 0 , \; L(x_n)/L(y_q) \to 1 . \tag{26}$$

Then (24) holds.

Quantile estimator (25) is similar in nature to that of [39]. Assumption (26) holds, for instance, if

$$1 \leq y_q/x_n \leq C_* \tag{27}$$

for some constant $C_* \geq 1$. The practical procedure is to choose the *average value* $\text{mean}\{\hat{y}_q(x) : x \in [x_-; x_+]\}$, where $[x_-; x_+]$ is the interval in which the function $\hat{y}_q(x)$ demonstrates stability.

The distribution (1) has finite first moment if $0 < a < 1$. Denote

$$E_n = \hat{y}_q a_n/(1 - a_n) .$$

Since $E(y) \sim ya/(1-a)$ as $y \to \infty$, Propositions 1 and 5 entail the following result.

Corollary 6 (consistency of ES estimator) *If* $0 < a < 1$ *and (26) holds then* E_n *is a consistent estimator of the Expected Shortfall:* $E_n/E(y_q) \underset{p}{\longrightarrow} 1$.

For the situation where q is bounded away from 0, Scaillet [31] suggested a kernel–type estimator of $E(q)$. In such a case, the empirical estimator is believed to provide better accuracy of estimation. The feature of our result is that the level $q = q(n)$ is allowed to approach 0.

From now on, $\hat{y}_q = \hat{y}_q(a_n(x_n), x_n)$, where a_n is the ratio estimator (3). Denote

$$A_0 = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}, \; B = \begin{pmatrix} 1 & 0 \\ -a & 1 \end{pmatrix}, \; B_* = \begin{pmatrix} 1 & 0 \\ -a^* & 1 \end{pmatrix}.$$

Theorem 7 *Suppose that* $\sigma^2 \equiv (a\sigma_1)^2 + \sigma_2^2 - 2a\sigma_{12} > 0$ *and*

$$\ln(p_n/G(y_q)) \to d , \; (a^* - a)\sqrt{np_n} \to \mu , \; \left(y_q q^a L^{-a}(x_n) - 1\right) \sqrt{np_n} \to \nu \tag{28}$$

for some constants d, μ, ν. *Then*

$$(\hat{y}_q/y_q - 1) \sqrt{np_n} \Longrightarrow \mathcal{N}\left(d\mu - \nu, \sigma_{\mathbf{c}}^2\right) , \tag{29}$$

where $\sigma_{\mathbf{c}}^2 = \mathbf{c}A\mathbf{c}^T$, $\mathbf{c} = (a, d)$ *and* $A = BA_0B^T$.

The first condition in (28) is of the same style as (27). The second one balances the asymptotic bias and variance of the ratio estimator (cf. (6)). Concerning the last relation in (28), note that

$$G(y_q) = G(G^{-1}(q)) \sim q \tag{30}$$

as $q \to 0$ (see Theorem 1.5.12 in [1]). If $G(y)$ is strictly monotone for all large enough y then $q = G(y_q)$, and the last relation in (28) may be rewritten as

$$\left(L^a(y_q)/L^a(x_n) - 1\right)\sqrt{np_n} \to \nu\,.$$

Notice that $\sigma^2_{\mathbf{c}} = (a(1-d)\sigma_1)^2 + (d\sigma_2)^2 + 2ad(1-d)\sigma_{12}$ and $A = \begin{pmatrix} \sigma_1^2 & \tilde\sigma \\ \tilde\sigma & \sigma^2 \end{pmatrix}$, where $\tilde\sigma = \sigma_{12} - a\sigma_1^2$. In the i.i.d. case, $\sigma^2_{\mathbf{c}} = a^2(1+d^2)$.

Now we want to replace $\sigma_{\mathbf{c}}$ in (29) by its consistent estimator. Define $\hat A_0$, $\hat B$ and $\hat A$ similarly to A_0, B, A with σ_1, σ_2, σ_{12}, a replaced by $\hat\sigma_1$, $\hat\sigma_2$, $\hat\sigma_{12}$ and a_n. Denote $\hat\sigma^2_{\mathbf{c}} \equiv \hat\sigma^2_{\mathbf{c}}(n) = \hat{\mathbf{c}}\hat A\hat{\mathbf{c}}^T$, where $\hat{\mathbf{c}} = (a_n, d_n)$ and $d_n = a_n^{-1}\ln(\hat y_q/x_n)$.

Corollary 8 *Assume the conditions of Theorem 7. If (15) holds then*

$$\left(\hat y_q/y_q - 1\right)\hat\sigma_{\mathbf{c}}^{-1}N_n^{1/2} \Longrightarrow \mathcal{N}\left((d\mu - \nu)/\sigma_{\mathbf{c}}, 1\right)\,. \tag{31}$$

Remark 3. If $L(x) = C + o(1)$ as $x \to \infty$, where $C > 0$, then one can suggest the quantile estimator $y_q^* = \left(\hat C_n/q\right)^{a_n}$. Using (39), one can check that $y_q^*/y_q = 1 + O_p\left((\ln x_n)/\sqrt{np_n}\right)$ while (29) yields $\hat y_q/y_q = 1 + O_p\left(1/\sqrt{np_n}\right)$. Hence $\hat y_q$ is preferable to y_q^*.

If there are reasons to believe that the asymptotic bias $(d\mu - \nu)/\sigma_{\mathbf{c}}$ is negligible then (31) yields the asymptotic confidence interval $\left[\hat y_q\Big/\left(1 + \gamma_\varepsilon\hat\sigma_{\mathbf{c}}/N_n^{1/2}\right); \hat y_q\Big/\left(1 - \gamma_\varepsilon\hat\sigma_{\mathbf{c}}/N_n^{1/2}\right)\right]$ of level $1 - \varepsilon$, where $\Phi(-\gamma_\varepsilon) = \varepsilon/2$. In the i.i.d. case, this becomes

$$\left[\hat y_q\Big/\left(1 + \gamma_\varepsilon\sqrt{a_n^2 + \ln^2(\hat y_q/x_n)}\Big/N_n^{1/2}\right); \hat y_q\Big/\left(1 - \gamma_\varepsilon\sqrt{a_n^2 + \ln^2(\hat y_q/x_n)}\Big/N_n^{1/2}\right)\right]. \tag{32}$$

We can eliminate the asymptotic bias $(d\mu - \nu)/\sigma_{\mathbf{c}}$ but at a cost of a slower rate of normal approximation. Let $\hat y_{q,r}$ be defined by (25) with $\hat a_n = a_{n,r}$.

Theorem 9 *Assume the conditions of Theorem 7. If (15) and (18) hold then*

$$\left(\hat y_{q,r}/y_q - 1\right)(a_n d_n)^{-1}N_{n,r}^{1/2} \Longrightarrow \mathcal{N}(0,1)\,. \tag{33}$$

Example 2. Consider the following model:

$$X_1 = \xi_1\,,\ X_i = \alpha_i\xi_i + (1-\alpha_i)X_{i-1} \qquad (i \geq 2)\,, \tag{34}$$

where $\xi_1, \xi_2, ..., \alpha_1, \alpha_2, ...$ are independent random variables, $\xi_i \stackrel{d}{=} X$ $(\forall i)$, $\mathcal{L}(X)$ obeys (1) and $\mathbb{P}(\alpha_i = 1) = 1 - \mathbb{P}(\alpha_i = 0) = \theta \in (0;1)$ $(\forall i)$.

This model was introduced by Smith and Weissman [35]. It is a particular case of stochastic difference equation (22). It is easy to see that (34) is a stationary Markov chain, the extremal index equals θ and clusters have geometric distribution with the mean $1/\theta$. We have showed in [22] that $\varphi(k) \leq (1-\theta)^k$. Hence (11) and (10) hold.

We prove in the next section that

$$\mathrm{Var}\Big(\sum^n Y_i^*\Big) = n\mathbb{E}(Y^*)^2 \left[1 + 2(\theta^{-1} - 1)(1 - \kappa_n)\right] \sim np_n \left(2\theta^{-1} - 1\right) a^2 , \quad (35)$$

$$\mathrm{Var}\Big(\sum^n Y_i^k \mathbb{I}_i\Big) \sim np_n \left(\frac{2}{\theta} - 1\right) a^{2k}(2k)! , \ \mathbb{E}\left(\sum_{i=1}^n \mathbb{I}_i\right)\left(\sum_{i=1}^n \bar{Y}_i\right) \sim np_n \left(\frac{2}{\theta} - 1\right) a ,$$

where $\kappa_n = \frac{1-(1-\theta)^n}{n\theta}$. Hence the conditions of Theorem 2 and Corollary 3 are fulfilled, and the results of Sections 2 entail

$$(a_n/a - 1)\sqrt{np_n} \Longrightarrow \mathcal{N}\left(m; 2\theta^{-1} - 1\right) \quad (36)$$

if $v\sqrt{np_n} \to m$; $\sqrt{np_n}$ in (36) may be replaced by $N_n^{1/2}$.

This is a generalization of the limit theorem (13): if $\alpha_i \equiv 1$ then (34) is a sequence of independent r.v.s, $\theta = 1$, and (36) implies (13).

Notice that *the accuracy of the approximation* $a_n \approx a$ *is the same as if the data were independent*, but the asymptotic variance of the estimator can only be larger.

If there are reasons to believe that the data can be approximated by the model (34) then (36) provides an alternative way of constructing asymptotic confidence intervals. Namely, if θ_n is a consistent estimator of the extremal index θ (see [15, 40]) then (36) implies

$$(a_n/a - 1)\, N_n^{1/2}(2/\theta_n - 1)^{-1/2} \Longrightarrow \mathcal{N}(m_*; 1) ,$$

where $m_* = m(2/\theta - 1)^{-1/2}$. If the asymptotic bias m_* is believed to be negligible then

$$\left[a_n/\left(1 + \gamma_\varepsilon(2\theta_n^{-1} - 1)^{1/2}/N_n^{1/2}\right); a_n/\left(1 - \gamma_\varepsilon(2\theta_n^{-1} - 1)^{1/2}/N_n^{1/2}\right)\right]$$

is the a.c.i. of level $1 - \varepsilon$, where γ_ε is defined by the equation $\Phi(-\gamma_\varepsilon) = \varepsilon/2$.

In [23], we illustrate the results of Sections 2–4 by examples of simulated data.

Conclusion. Our simulation results in [23] show that the statistical procedure based on the ratio estimator perform satisfactorily. This can be a bit surprising in view of "Hill's horror plot" (see Figure 3 in [28]). A possible explanation is that the plot of Hill's estimator gives the same respect to 25% smallest and 25% largest elements of a sample — i.e., to its least and most informative parts.

In all our examples (see [23]), 25% smallest elements lie below the threshold $x = 0.5$ (and hence should not be given any attention). Approximately half of the sample elements lie below the threshold $x = 1$.

The feature of the ratio estimator plot is that it reduces the least informative part of a sample and highlights the most informative part.

The ratio estimator seems to be easier for theoretical investigation as well. For instance, the bias and the variance of a tail index estimator seems to be known only in the case of the ratio estimator. For a particular class of slowly varying functions L, the accuracy of estimation is sharper in the case of the ratio estimator (see [19]).

With no indicator in favor of other tail index estimators, the ratio estimator appears the basic tool of statistical analysis of heavy tails.

5 Proofs

Below, symbols c_i denote positive constants; a bar over a random variable means that it is centered by its expectation. We write $\xi_n \underset{p}{\sim} \eta_n$ or $\xi_n = \eta_n(1+o_p(1))$ if $\xi_n/\eta_n \underset{p}{\longrightarrow} 1$.

Proof of Proposition 1. The first part of the statement (consistency of the ratio estimator a_n) has been given in [24]. We present it for the sake of completeness.

One can check that (10) is equivalent to the condition $\sum_{i\geq 1}\rho(2^i) < \infty$ (in particular, this yields $\rho(l) \to 0$ as $l \to \infty$; (11) is equivalent to the condition $\sum_{i\geq 1}\varphi^{1/2}(2^i) < \infty$). We use Chebyshev's inequality, (7) and an estimate of the variance of a sum of dependent random variables (see Peligrad [26] or Utev [37]). Given $\varepsilon > 0$, denote $Z_i = Y_i^* - (\mathbb{1}_i - p_n)\varepsilon$. Then

$$\begin{aligned}\mathbb{P}(a_n - a^* > \varepsilon) &= \mathbb{P}\left(\sum\nolimits^n (Y_i - a^*)\mathbb{1}_i > \varepsilon \sum\nolimits^n \mathbb{1}_i\right) \\ &= \mathbb{P}\left(\sum\nolimits^n Z_i > \varepsilon n p_n\right) \leq (\varepsilon n p_n)^{-2}\,\mathrm{Var}\left(\sum\nolimits^n Z_i\right).\end{aligned}$$

By Theorem 1.1 in [37], there exists a constant c_ρ (depending only on $\rho(\cdot)$) such that

$$\mathrm{Var}\,N_n \leq c_\rho n p_n\,,\ \mathrm{Var}\left(\sum\nolimits^n Y_i^*\right) \leq c_\rho n p_n\,,\ \mathrm{Var}\left(\sum\nolimits^n Z_i\right) \leq c_\rho n \mathrm{Var}\,Z_1 \leq c n p_n \tag{37}$$

(we have used also (7)). Hence $\mathbb{P}(a_n - a^* > \varepsilon) \to 0$. Similarly one checks that $\mathbb{P}(a_n - a^* < -\varepsilon) \to 0$. Remind that $a^* \to a$ as $x_n \to \infty$. Hence $a_n \underset{p}{\longrightarrow} a$.

Now we show that $\hat{C}_n \underset{p}{\longrightarrow} C$. Chebyshev's inequality and (37) yield

$$N_n/np_n \underset{p}{\longrightarrow} 1\,. \tag{38}$$

Hence $\hat{C}_n = C x_n^{1/a_n - 1/a}(1+o_p(1))$. We have to prove that $(a_n - a)\ln x_n \underset{p}{\longrightarrow} 0$. Because of the assumption, $(a^* - a)\ln x_n \to 0$. It remains to check that $\left(\sum^n Y_i^*\right)(\ln x_n)/np_n \underset{p}{\longrightarrow} 0$. The latter follows from Chebyshev's inequality, the assumption and (37). □

Lemma 10 *If $\varphi(l) \to 0$ as $l \to \infty$ and (15) holds then*

$$(N_n/(np_n) - 1, a_n - a^*)\sqrt{np_n} \Longrightarrow \mathcal{N}(\mathbf{0}; A)\,. \tag{39}$$

Proof of Lemma 10. Note that

$$a_n - a^* = \frac{\sum^n Y_i^*}{\sum^n 1\!\!\mathrm{I}_i} = \frac{\sum^n Y_i^*}{np_n}\left(1 - \frac{\sum^n \bar{1\!\!\mathrm{I}}_i}{\sum^n 1\!\!\mathrm{I}_i}\right).$$

Taking into account (38), we shall check that

$$\left(\frac{N_n}{np_n} - 1, \frac{\sum^n Y_i^*}{np_n}\right)\sqrt{np_n} = \left(\frac{\sum^n \bar{1\!\!\mathrm{I}}_i}{\sqrt{np_n}}, \frac{\sum^n Y_i^*}{\sqrt{np_n}}\right) \Longrightarrow \mathcal{N}(\mathbf{0}; A).$$

Notice that $(\bar{1\!\!\mathrm{I}}_i, Y_i^*)^T = B_*\zeta_i$, where $\zeta_i = (\bar{1\!\!\mathrm{I}}_i, \bar{Y}_i)^T$. In order to check that $\sum_{i=1}^n \zeta_i/\sqrt{np_n} \Rightarrow \mathcal{N}(\mathbf{0}; A_0)$, we apply the following result of Utev [38].

Theorem A. *Let* $\{\xi_{i,n} : 1 \le i \le k_n\}_{n\ge1}$ *be a triangular array of r.v.s, and let* $\varphi_n(\cdot)$ *be the corresponding mixing coefficient. Denote* $S_n = \sum_{i=1}^{k_n} \xi_{i,n}$ *, and let* $z_n^2 = Var\, S_n$*. If* $\sup_n \varphi_n(lj_n) \to 0$ *as* $l \to \infty$ *for some sequence* $\{j_n\}$ *of integer numbers and*

$$\lim_{n\to\infty} j_n z_n^{-2} \sum_{i=1}^{k_n} \mathbb{E}\xi_{i,n}^2 1\!\!\mathrm{I}\{|\xi_{i,n}| > \varepsilon z_n/j_n\} = 0 \qquad (\forall \varepsilon > 0) \tag{40}$$

then $S_n/z_n \Longrightarrow \mathcal{N}(0; 1)$ *.*

Let $c = (c_1, c_2) \in \mathbb{R}^2$. We want to show that

$$\sum^n c\zeta_i/\sqrt{np_n} \Longrightarrow \mathcal{N}(\mathbf{0}; cA_0c^T). \tag{41}$$

Put $\xi_i = c_1\bar{1\!\!\mathrm{I}}_i + c_2(Y_i - a^*p_n)$ and $j_n = 1$. By the assumption, $\mathrm{Var}\,(\sum^n \xi_i) \sim \sigma^2 np_n$. To check (40), it suffices to show that

$$\mathbb{P}(Y > \varepsilon\sqrt{np_n}\,|X > x_n) \to 0\,,\ \mathbb{E}\left\{Y^2 1\!\!\mathrm{I}\{Y > \varepsilon\sqrt{np_n}\,\}\Big|X > x_n\right\} \to 0$$

for any $\varepsilon > 0$. According to [1, 32],

$$L(y)/L(x) \sim \exp\left(\int_x^y w(u)u^{-1}\,du\right) \tag{42}$$

as $x, y \to \infty$, where $w(u) \to 0$ as $u \to \infty$. Therefore,

$$\mathbb{P}(Y > \varepsilon\sqrt{np_n}|X > x_n) = \mathbb{P}(X > x_n e^{\varepsilon\sqrt{np_n}})/\mathbb{P}(X > x_n)$$

$$= L(x_n e^{\varepsilon\sqrt{np_n}})L^{-1}(x_n)e^{-\varepsilon\sqrt{np_n}/a} = e^{-(\varepsilon/a + o(1))\sqrt{np_n}} \to 0\,.$$

Using this relation and (7), we derive

$$\mathbb{E}^2\left\{Y^2 1\!\!\mathrm{I}\{Y > \varepsilon\sqrt{np_n}\,\}\Big|X > x_n\right\} \le \mathbb{E}\{Y^4|X > x_n\}\mathbb{P}\left(Y > \varepsilon\sqrt{np_n}\,|X > x_n\right) \to 0\,.$$

Hence (40) holds, and Theorem A entails (41) and (39). □

Proof of Theorem 2. Arguments of the proof of Lemma 10 yield also that

$$\sum^n Y_i^* \Big/ \sqrt{\operatorname{Var} \sum^n Y_i^*} \Longrightarrow \mathcal{N}(0;1)\,.$$

Taking into account (38) and the assumptions of the theorem, we get (12). □

Lemma 11 *If (15) holds then*

$$\hat\sigma_k \underset{p}{\longrightarrow} \sigma_k\,,\ \hat\sigma_{12} \underset{p}{\longrightarrow} \sigma_{12} \qquad (k \in \mathbb{N})\,. \tag{43}$$

Proof of Lemma 11. First, we notice that

$$\mathbb{E}\left(\sum^r Y_i^{k-1} \mathbb{I}_i\right)^2 \sim \sigma_k^2 r p_n \qquad (k \in \mathbb{N})\,. \tag{44}$$

Indeed, denote $R_n = \operatorname{Var}\left(\sum^{[n/r]} T_{k,j}\right) - [n/r]\operatorname{Var} T_{k,1}$. By Utev's Theorem 1.1 [37], $R_n = o([n/r]\operatorname{Var} T_{k,1})$. Therefore, $\operatorname{Var}\left(\sum^{[n/r]} T_{k,j}\right) \sim [n/r]\operatorname{Var} T_{k,1} \le c_1 n p_n$, and

$$\operatorname{Var}\left(\sum^n Y_i^k \mathbb{I}_i\right) = \operatorname{Var}\left(\sum^{[n/r]} T_{k,j}\right) + o(np_n) = [n/r]\operatorname{Var} T_{k,1} + o(np_n)\,.$$

By the assumption, $r/n \to 0$. Thus, $\operatorname{Var} T_{k-1,1} \equiv \operatorname{Var}\left(\sum^r Y_i^{k-1} \mathbb{I}_i\right) \sim \sigma_k^2 r p_n$, and (44) follows.

We use Chebyshev's inequality to prove (43). Note that $\sigma_k^2 - [n/r]\mathbb{E}T_{k-1,1}^2/np_n = o(1)$. Using Utev's [37] Theorem 1.1 and Corollary 2.3, we get $\operatorname{Var} T_{k,1}^2 \le \mathbb{E}T_{k,1}^4 \le c_2 r p_n$ and

$$\operatorname{Var}\left(\sum_{j=1}^{[n/r]} T_{k,j}^2\right) \le c_3 [n/r]\operatorname{Var} T_{k,1}^2 \le c_4 n p_n\,.$$

Hence the probability $\mathbb{P}(\hat\sigma_k^2 - \sigma_k^2 > 2\varepsilon)$ is not greater than

$$\mathbb{P}\left(\sum_{j=1}^{[n/r]} (T_{k-1,j}^2 - \mathbb{E}T_{k-1,j}^2) - (\sigma^2 + 2\varepsilon)\sum_{i=1}^n \bar{\mathbb{I}}_i > \varepsilon n p_n\right) \le \frac{c_\varepsilon}{np_n} \to 0$$

$(\forall \varepsilon > 0)$. Similarly we check that $\mathbb{P}(\hat\sigma_k^2 - \sigma_k^2 < -\varepsilon) \to 0$. Thus, $\hat\sigma_k \underset{p}{\longrightarrow} \sigma_k$.

It remains to show that $\hat\sigma_{12} \underset{p}{\longrightarrow} \sigma_{12}$. Recall that $Y_i^* = Y_i - a^* \mathbb{I}_i$. According to (15),

$$\sigma_{12} n p_n \sim \mathbb{E}\left(\sum^n \mathbb{I}_i\right)\left(\sum^n \bar{Y}_i\right) = \mathbb{E}\left(\sum_{j=1}^{[n/r]} T_{0,j}\right)\left(\sum_{j=1}^{[n/r]} T_{1,j}\right) + o(np_n)\,.$$

Similarly to (44), one can check that $\mathbb{E}\left(\sum^r \mathbb{I}_i\right)\left(\sum^r Y_{i+r}\right) \sim \sigma_{12} r p_n$. Using again Theorem 1.1 and Corollary 2.3 from [37], we get $\operatorname{Var}\left(\sum_{j=1}^{[n/r]} T_{0,j} T_{1,j}\right) \sim \frac{n}{r}\operatorname{Var}(T_{0,1}T_{1,1}) \le cnp_n$. Note that

$$\{\hat\sigma_{12} - \sigma_{12} > 2\varepsilon\} = \left\{\sum_{j=1}^{[n/r]} T_{0,j} T_{1,j} > (\sigma_{12} + 2\varepsilon)\sum^n \mathbb{I}_i\right\}$$
$$\subset \left\{\sum_{j=1}^{[n/r]} (T_{0,j}T_{1,j} - \mathbb{E}T_{0,j}T_{1,j}) - (\sigma_{12} + 2\varepsilon)\sum^n \bar{\mathbb{I}}_i > \varepsilon n p_n\right\}\,.$$

By Chebyshev's inequality, $\mathbb{P}\left(\hat\sigma_{12} - \sigma_{12} > 2\varepsilon\right)$ is not greater than

$$\frac{2\mathrm{Var}\left(\sum_{j=1}^{[n/r]} T_{0,j}T_{1,j}\right) + 2(\sigma_{12}+2\varepsilon)^2 \mathrm{Var}\left(\sum^n 1\!\!\mathrm{I}_i\right)}{(\varepsilon n p_n)^2} \le \frac{c_\varepsilon}{np_n} \to 0$$

$(\forall \varepsilon > 0)$. Similarly one checks that $\mathbb{P}\left(\hat\sigma_{12} - \sigma_{12} < -2\varepsilon\right) \to 0$. The proof is complete. □

Proof of Corollary 3. LLN (38) and Lemma 10 imply

$$\zeta_n = (a_n - a^*)\, N_n^{1/2} \Longrightarrow \mathcal{N}(0;\sigma^2)\,.$$

Since $(a^* - a)\sqrt{np_n} \to \mu$, we have

$$\Big(a_n - a\Big) N_n^{1/2} = \zeta_n + (a^* - a)\sqrt{np_n}\sqrt{N_n/np_n} \Longrightarrow \mathcal{N}(\mu;\sigma^2)\,.$$

Taking into account (38) and Lemma 11, we get (16). □

Proof of Theorem 4. Using Theorem 1.1 in [37] and (7), we conclude that

$$\begin{aligned} \mathrm{Var}\left(\sum\nolimits_{j=1}^{[n/r]} 1\!\!\mathrm{I}_{jr}\right) &\sim [n/r]\mathrm{Var}\, 1\!\!\mathrm{I}_1 \sim np_n/r\,, \\ \mathrm{Var}\left(\sum\nolimits_{j=1}^{[n/r]} Y^*_{jr}\right) &\sim [n/r]\mathrm{Var}\, Y^*_1 \sim a^2 np_n/r\,. \end{aligned}$$

Hence $N_{n,r}/\left(np_n/r\right) \xrightarrow[p]{} 1$. The same arguments as in the proof of Lemma 10 entail that $\mathbb{P}\left(Y > \varepsilon\sqrt{np_n/r}\,\middle|\, X > x_n\right) \to 0$. Theorem A with $\xi_i = Y^*_{ir}$ and $j_n = 1$ yields

$$\frac{\sum_{j=1}^{[n/r]}\left(Y_{jr} - a^* 1\!\!\mathrm{I}_{jr}\right)}{a\sqrt{np_n/r}} \Longrightarrow \mathcal{N}(0;1)\,.$$

Note that

$$(a^*/a - 1)\sqrt{N_{n,r}} \underset{p}{\sim} (a^*/a - 1)\sqrt{np_n/r} \xrightarrow[p]{} 0$$

by (18) and the LLN for $N_{n,r} = \sum_{j=1}^{[n/r]} 1\!\!\mathrm{I}_{jr}$. The result follows. □

Proof of Proposition 5. Denote $G_n = N_n/n$. According to (1) and [1, 32], $G^{-1}(z) = z^{-a}\ell(z)$, where ℓ is a slowly varying function. This and (42) entail

$$G(\hat y_q)/q \xrightarrow[p]{} 1 \iff \hat y_q/y_q \xrightarrow[p]{} 1\,. \tag{45}$$

Notice that $G(x_n) = p_n$ and

$$\frac{\hat y_q}{x_n} = \left(\frac{G_n}{q}\right)^{\hat a_n} = \left(\frac{G_n}{p_n}\right)^{\hat a_n}\left(\frac{p_n}{q}\right)^{\hat a_n} = \frac{p_n^a}{q^a}\left(\frac{G_n}{p_n}\right)^{\hat a_n}\left(\frac{p_n}{q}\right)^{\hat a_n - a}. \tag{46}$$

Taking into account (38) and the identity $p_n^a = x_n^{-1} L^a(x_n)$, we deduce

$$\hat{y}_q q^a / L^a(x_n) \underset{p}{\sim} (p_n/q)^{\hat{a}_n - a} .$$

This and (26) yield

$$\hat{y}_q q^a / L^a(x_n) \xrightarrow[p]{} 1 .$$

Since $y_q \sim q^{-a} L^a(y_q)$ according to (30), we have

$$\hat{y}_q / y_q \underset{p}{\sim} L^a(x_n)/L^a(y_q) \to 1$$

if (26) holds. The proof is complete. □

Proof of Theorem 7. From (46),

$$\begin{aligned} \hat{y}_q q^a / L^a(x_n) &= (G_n/p_n)^{a_n} (p_n/q)^{a_n - a} \\ &= 1 + (G_n/p_n - 1)\, a_n + (a_n - a) \ln (p_n/q) + \delta_n , \end{aligned}$$

where $\delta_n = o_p\left(|1 - G_n/p_n| + |a_n - a|\right)$. By Lemma 10,

$$(\hat{y}_q q^a / L^a(x_n) - 1) \sqrt{n p_n} \Longrightarrow \mathcal{N}(d\mu, \mathbf{c} A \mathbf{c}^T) .$$

Hence $(\hat{y}_q / y_q - 1) \sqrt{n p_n} \Longrightarrow \mathcal{N}(d\mu - \nu, \mathbf{c} A \mathbf{c}^T)$. □

Corollary 8 follows from Theorem 7 and Lemma 11. We should mention only that

$$\ln \frac{G(x_n)}{G(y_q)} = \frac{1}{a} \ln \frac{y_q}{x_n} + \ln \frac{L(x_n)}{L(y_q)} = \frac{1}{a} \ln \frac{y_q}{x_n} + o(1) .$$

Hence $a_n^{-1} \ln(\hat{y}_q / x_n) \xrightarrow[p]{} d$. □

Proof of Theorem 9. Arguments similar to those in the proof of Theorem 7 yield

$$\frac{\hat{y}_{q,r} q^a}{L^a(x_n)} - 1 = \left(\frac{G_n}{p_n} - 1\right) a_{n,r} + (a_{n,r} - a) \ln \frac{p_n}{q} + o_p\left(\left|1 - \frac{G_n}{p_n}\right| + |a_{n,r} - a|\right) .$$

According to Lemma 10, $G_n/p_n - 1 = O_p\left(1/\sqrt{n p_n}\right)$. Therefore,

$$\left(\frac{\hat{y}_{q,r} q^a}{L^a(x_n)} - 1\right) \sqrt{n p_n / r} = (a_{n,r} - a) d \sqrt{n p_n / r}\; (1 + o_p(1)) + o_p(1) .$$

Because of the assumptions, $(y_q q^a L^{-a}(x_n) - 1)\sqrt{n p_n / r} \sim \nu / \sqrt{r} \to 0$. Hence

$$(\hat{y}_q / y_q - 1) \sqrt{n p_n / r} \Rightarrow \mathcal{N}(0, a^2 d^2)$$

by Theorem 4. The result follows. □

Proof of relation (35). By the well–known formula (cf. [16]),

$$\mathbb{E}\left(\sum^n Y_i^*\right)^2 = n\left[\mathbb{E}(Y^*)^2 + 2\sum^n (1-i/n)\,\mathbb{E}Y_1^*Y_{i+1}^*\right]. \tag{47}$$

Notice that

$$\mathbb{E}Y_1^*Y_{i+1}^* = (1-\theta)^i\mathbb{E}(Y^*)^2 \qquad (i \geq 1). \tag{48}$$

Indeed, if $\alpha_2 = ... = \alpha_{i+1} = 0$ then $Y_{i+1}^* = Y_1^*$, and $\mathbb{E}Y_1^*Y_{i+1}^* = \mathbb{E}(Y^*)^2$. Otherwise the random variables Y_1^* and Y_{i+1}^* are independent, and hence $\mathbb{E}Y_1^*Y_{i+1}^* = 0$. Relation (35) follows from (47), (48) and (7).

By the same argument, $\mathbb{E}\bar{Y}_1^k\mathbb{1}_1\bar{Y}_{i+1}^k\mathbb{1}_{i+1} = (1-\theta)^i\mathbb{E}(\bar{Y}_1^k)^2\mathbb{1}_1$, and $\mathbb{E}\mathbb{1}_1\bar{Y}_{i+1} = \mathbb{E}\bar{Y}_1\mathbb{1}_{i+1} = (1-\theta)^i\mathbb{E}\mathbb{1}_1\bar{Y}_1 = (1-\theta)^i a^* p_n(1-p_n)$. Hence

$$\begin{aligned}\mathrm{Var}\Big(\sum^n Y_i^k\mathbb{1}_i\Big) &= n\mathrm{Var}\Big(Y^k\Big)\left[1 + 2(\theta^{-1}-1)\left(1 - \frac{1-(1-\theta)^n}{n\theta}\right)\right] \\ &\sim n(2\theta^{-1}-1)\mathbb{E}Y^{2k} \sim np_n(2\theta^{-1}-1)a^{2k}(2k)!\end{aligned}$$

($k \geq 0$). Similarly one can check that

$$\mathbb{E}\sum^n \mathbb{1}_i\sum^n \bar{Y}_j = n\left[\mathbb{E}\mathbb{1}_1\bar{Y}_1 + \sum^n\left(1-\frac{i}{n}\right)\left(\mathbb{E}\mathbb{1}_1\bar{Y}_{i+1} + \mathbb{E}\mathbb{1}_{i+1}\bar{Y}_1\right)\right] \sim np_n\left(\frac{2}{\theta}-1\right)a\,.$$

The proof is complete. □

Remark 4. Smith ([34], pp. 1181-1182) claims the mean squared error of his estimator of the tail index can be smaller than that of the ratio estimator in the case of $\mathcal{P}_{a,a,c,d}$ family of distributions. This seems to be a mistake caused by the introduction of the artificial parameter σ which is linked to the tail index (denoted $-1/k$ in [34]) by the equation $\sigma = -ku$. The calculation of the MLE estimators $\hat{\sigma}$ and $\hat{k}$ in [34] is carried out as if σ and k were independent parameters. This leads to the contradiction: since $\sigma = -ku$, one has $\sigma\frac{\partial}{\partial\sigma} = k\frac{\partial}{\partial k}$ while $\sigma\frac{\partial}{\partial\sigma}\ln g(y;\sigma,k) \neq k\frac{\partial}{\partial k}\ln g(y;\sigma,k)$ on page 1204 of [34]. As a consequence,

$$\sigma\mathbb{E}\frac{\partial}{\partial\sigma}\ln g(Y;\sigma,k) \neq k\mathbb{E}\frac{\partial}{\partial k}\ln g(Y;\sigma,k)$$

on page 1180, and the arguments behind the bias calculation fail. Note that the derivation of the MLE using the density $g(y;-ku,k)$ instead of $g(y;\sigma,k)$ would yield the ratio estimator.

Acknowledgements. The author is grateful to L. de Haan, T. Mikosch and C. de Vries for helpful discussions, and to the referee for useful comments. The main part of the paper was written at Eurandom. Research was also supported in part by a scholarship from K.U.Leuven.

References

[1] Bingham N.H., Goldie C.M. and Teugels J.L. (1987) *Regular Variation.* — Cambridge: Cambridge University Press.

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

[3] Beirlant J., Vynckier P. and Teugels J.L. (1996) Excess functions and estimation of the extreme-value index. — Bernoulli, v. 2, No 4, 293–318.

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

[5] Davis R., Mikosch T. and Basrak B. (1999) Sample ACF of multivariate stochastic recurrence equations with applications to GARCH. — Preprint. University of Groningen, Department of Mathematics.

[6] Drees H. (1999) Weighted approximations of tail processes under mixing conditions. — University of Cologne: Preprint.

[7] Drees H. and de Haan L. (1999) Conditions for quantile process approximations. — Commun. Statist. Stochastic Models, v. 15, 485–502.

[8] Einmahl J. (1992) Limit theorems for tail processes with application to intermediate quantile estimation. — J. Statist. Plan. Infer., v. 32, 137–145.

[9] Embrechts P., Klüppelberg C. and Mikosch T. (1997) *Modelling Extremal Events for Insurance and Finance.* — Berlin: Springer Verlag.

[10] Goldie C.M. (1991) Implicit renewal theory and tails of solutions of random equations. — Ann. Appl. Probab., v. 1, 126–166.

[11] Goldie C.M. and Smith R.L. (1987) Slow variation with remainder: theory and applications. — Quart. J. Math. Oxford, v. 38, 45–71.

[12] de Haan L. and Peng L. (1998) Comparison of tail index etimators. — Statistica Neerlandica, v. 52, No 1, 60–70.

[13] Hall P. and Welsh A.H. (1984) Best attainable rates of convergence for estimates of parameters of regular variation. — Ann. Statist., v. 12, No 3, 1079–1084.

[14] Hsing T. (1991) On tail index estimation for dependent data. — Ann. Statist., v. 19, No 3, 1547–1569.

[15] Hsing T. (1991) Estimating the parameters of rare events. — Stochastic Processes Appl., v. 37, No 1, 117–139.

[16] Ibragimov I.A. and Linnik Y.V. (1971) *Independent and stationary sequences of random variables.* — Groningen: Wolters-Noordhoff Publishing, 443 pp.

[17] Luenberger D.G. (1998) *Investment Science.* — Oxford: Oxford University Press.

[18] McNeil A.J. (1997) Estimating the tails of loss severity distributions using extreme value theory. — Astin Bulletin, v. 27, No 1, 117-137.

[19] Novak S.Y. and Utev S.A. (1990) Asymptotics of the distribution of the ratio of sums of random variables. — Siberian Math. J., v. 31, No 5, 781–788.

[20] Novak S.Y. (1992) Inference about the Pareto–type distribution. — In: *Trans. 11-th Prague Conf. Inform Theory Statist. Decis. Func. Random Processes.* Prague: Academia, **B**, pp. 251–258.

[21] Novak S.Y. (1996) On the distribution of the ratio of sums of random variables. — Theory Probab. Appl., v. 41, No 3, 479–503.

[22] Novak S.Y. (1998) On the limiting distribution of extremes. — Siberian Adv. Math., v. 8, No 2, 70–95.

[23] Novak S.Y. (1999) Inference on heavy tails from dependent data. — Technical University of Eindhoven: Eurandom research report No 99–043.

[24] Novak S.Y. (2000) Confidence intervals for a tail index estimator. — In: "Measuring Risk in Complex Stochastic Systems", Lecture Notes Statist., v. 147, 215–222.

[25] Novak S.Y. (2000) On self–normalised sums. — Math. Methods Statist., v. 9, No 4, 415–436.

[26] Peligrad M. (1982) Invariance principle for mixing sequences. — Ann. Probab., v. 10, No 4, 968–981.

[27] Pfanzagl J. (2000) On local uniformity for estimators and confidence limits. — J. Statist. Plann. Inference, v. 84, 27–53.

[28] Resnick S.I. (1997) Heavy tail modeling and teletraffic data. — Ann. Statist., v. 25, No 5, 1805–1869.

[29] Resnick S. and Starica C. (1998) Tail index estimation for dependent data. — Ann. Appl. Probab., v. 8, No 4, 1156–1183.

[30] Rootzen H., Leadbetter M.R. and de Haan L. (1998) On the distribution of tail array sums for strongly mixing stationary sequences. — Ann. Appl. Probab., v. 8, No 3, 868–885.

[31] Scaillet O. (2000) Nonparametric estimation and sensitivity analysis of expected shortfall. — Universite Catholique de Louvain: Preprint.

[32] Seneta E. (1976) *Regularly Varying Functions.* — Lecture Notes Math., v. 508. Berlin: Springer–Verlag.

[33] Shorack G.R. and Wellner J.A. (1986) *Empirical processes with applications to statistics.* — New York: Wiley.

[34] Smith R.L. (1987) Estimating tails of probability distributions. — Ann. Statist., v. 15, No 3, 1174–1207.

[35] Smith R.L. and Weissman I. (1994) Estimating the extremal index. — J. R. Statist. Soc. B, v. 56, No 3, 515–528.

[36] Starica C. (1999) On the tail impirical process of solutions of stochastic difference equations. — Chalmers University: Preprint. http://www.math.chalmers.se/ starica/resume/publi1.html

[37] Utev S.A. (1989) Sums of φ-mixing random variables. — Trudy Inst. Mat. (Novosibirsk), v. 13, 78–100. (in Russian)

[38] Utev S.A. (1990) On the central limit theorem for φ-mixing triangle arrays of random variables. — Theory Probab. Appl., v. 35, no. 1, 131-139.

[39] Weissman I. (1978) Estimation of parameters and large quantiles based on the k largest observations. — J. Amer. Statist. Assoc., v. 73, 812–815.

[40] Weissman I. and Novak S.Y. (1998) On blocks and runs estimators of extremal index. — J. Statist. Planning Inference, v. 66, No 2, 281–288.

Appendix

In this section we illustrate the results of Sections 2–4 by examples of simulated data (cf. [23]). In Examples 1 and 2 below, the marginal distribution $\mathbb{P}_0$ is that of $|X|$, where X has the standard Cauchy distribution.

Example 1. We simulated 1000 i.i.d.r.v.s according to the distribution $\mathbb{P}_0$.

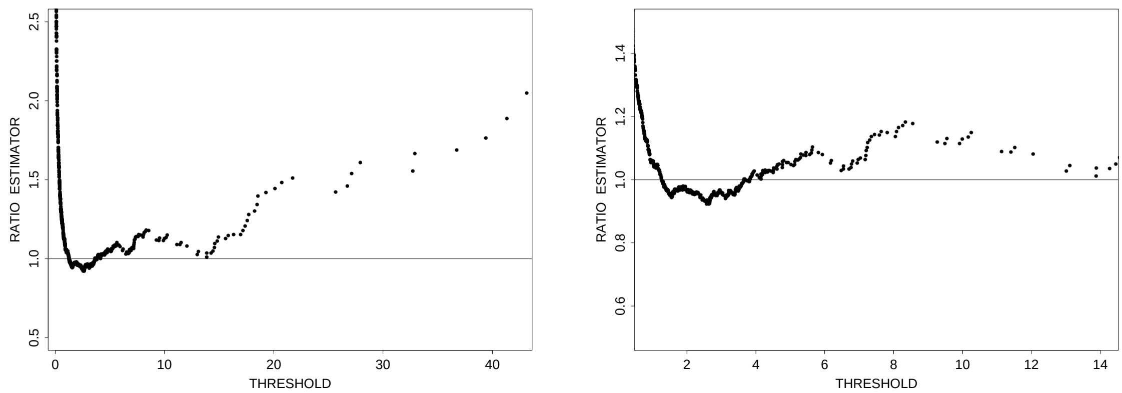

Figure 1: Tail index estimation from the distribution $\mathbb{P}_0$. The average value of $a_n(x)$, $x \in [1;14]$, is $\hat{a} = 0.998$. The asymptotic confidence interval of level 0.95 is $[0.91;1.10]$, and the non–asymptotic confidence interval is $[0.88;1.15]$.

The first picture of Figure 1 shows that the ratio estimator $a_n(x)$ behaves rather stable in the interval $x \in [0.5;17]$. The curve over the interval $[0.5;17]$ is formed by 701 points (out of 1000). The second picture is even more convincing: it demonstrates the behavior of the ratio estimator when the threshold x ranges in $[1;14]$. The corresponding fragment of the curve is formed by 479 points.

It is reasonable to pick up the estimate of the index a from the interval $[1;14]$. Following our procedure, we take the average value $\hat{a}$ of $a_n(x)$ in the interval $x \in [1;14]$: $\hat{a} = 0.998$.

Let x_n be the threshold corresponding to $\hat{a}$ (i.e., $a_n(x_n) = \hat{a}$). The corresponding asymptotic confidence interval (a.c.i.) of level 0.95 is $[0.91;1.10]$, and the non–asymptotic confidence interval is $[0.88;1.15]$.

A number of authors suggested estimating a by the value $\tilde{a}(x_*)$, where $\tilde{a}(\cdot)$ is a chosen tail index estimator, x is a "tuning" parameter and x_* is the left end–point of the interval $[x_*;x^*]$ where the curve $\tilde{a}(x)$ is "approximately flat" (or the mean excess function is "approximately linear"). The practical advantage of our procedure is that average is less sensitive to the choice of end–points of the interval of "regular behavior".

The plot of the tail constant estimator $\hat{C}_n(\cdot)$ is presented in the first picture of Figure 2. Estimator $\hat{C}_n(x)$ seems to be stable when $x \in [1.5;3.5]$. The corresponding fragment of the curve is formed by 229 points, the average value of $\hat{C}_n(x)$ in that interval is 0.585 ($C = 2/\pi \approx 0.637$). The plot of $\hat{C}_n(\cdot)$ looks undersmoothed. The plot of a smoothed version $C_n^*(\cdot)$ of the estimator $\hat{C}_n(\cdot)$ is shown in the second picture.

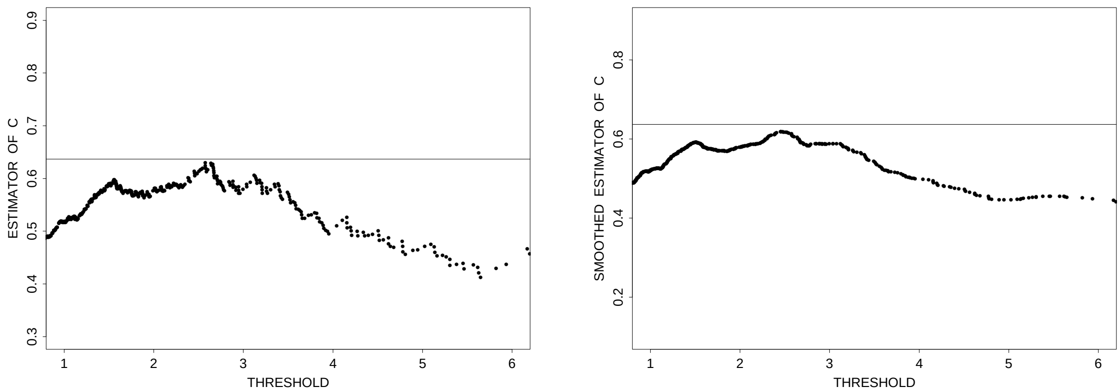

Figure 2: Tail constant estimator $\hat{C}_n = 0.585$ (the actual value of C is $2/\pi \approx 0.637$).

The plots of the quantile estimator (25) are given in Figure 3.

The first picture presents $\hat{y}_q$ for the case $q = 0.05$, the true value is $y_q = 12.7$. The plot demonstrates stability in the interval $x \in [1.5; 14]$ (formed by 345 points). The average value of $\hat{y}_q$ in that interval is 10.5 (the empirical 0.95% quantile equals 9.9).

The second picture displays $\hat{y}_q$ for the case $q = 0.01$; the true value is $y_q = 63.66$. The plot looks stable in the interval $[5.5; 18]$. The corresponding fragment of the curve is formed by 67 points, the average value of $\hat{y}_q$ in that interval is 59.9 (the empirical 0.99% quantile equals 41.3).

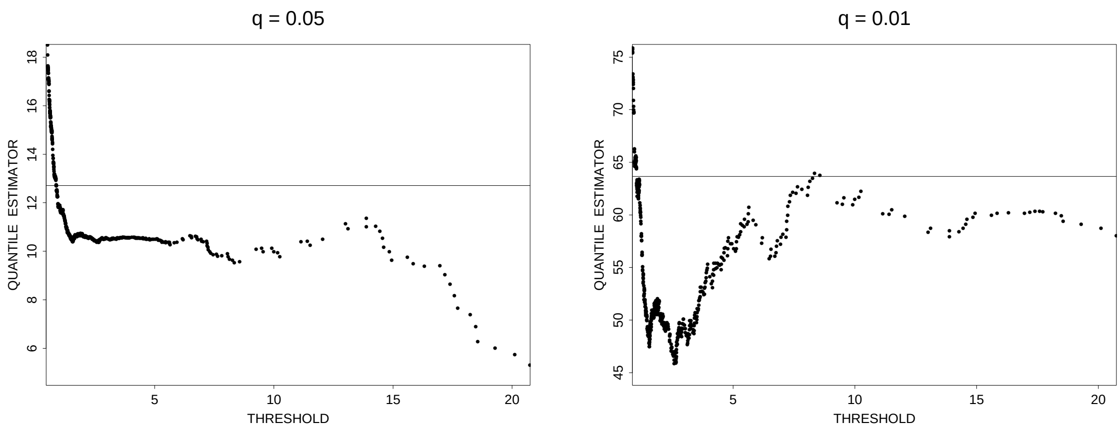

Figure 3: Quantile estimator (25). The first picture: $\hat{y}_q = 10.5$. The true value is $y_q = 12.7$, the empirical 0.95% quantile equals 9.91. The second picture: $\hat{y}_q = 59.9$. The true value is $y_q = 63.66$, the empirical 0.99% quantile equals 41.3.

Let $\hat{a}$ be the tail index estimate obtained at the step of tail index estimation (in our example, $\hat{a} = 0.998$). It can sometimes be worth using the estimator

$$\tilde{y}_q \equiv \tilde{y}_q(x_n) = (N_n/qn)^{\hat{a}}\, x_n$$

instead of (25). The simulation results are presented in Figure 4. The plots demonstrate stability in the interval $x \in [1.5; 4]$ (formed by 256 points). The average value of $\tilde{y}_q$ in that interval is 11.20 (the true value is $y_q = 12.7$), and the corresponding 0.95%–a.c.i. is $[9.10; 14.56]$ (the 0.95%–a.c.i. (32) is $[8.63; 13.46]$).

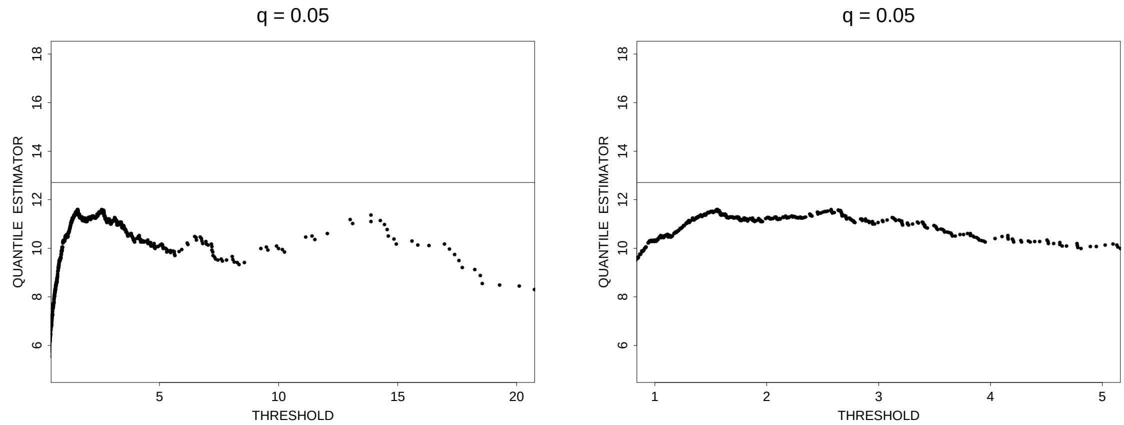

Figure 4: Quantile estimation: $\tilde{y}_q = 11.2$, the true value is $y_q = 12.7$. The asymptotic confidence interval of level 0.95 is $[9.10; 14.56]$.

Example 2. We have simulated 1000 r.v.s according to the model (34) with the marginal distribution $\mathbb{P}_0$ and $\theta = 1/2$. The plot of $a_n(x)$ is presented in Figure 5. The ratio estimator demonstrates stability in the interval $x \in [1.5; 14]$, formed by 322 points. The average value of $a_n(x)$ in that interval is 1.025.

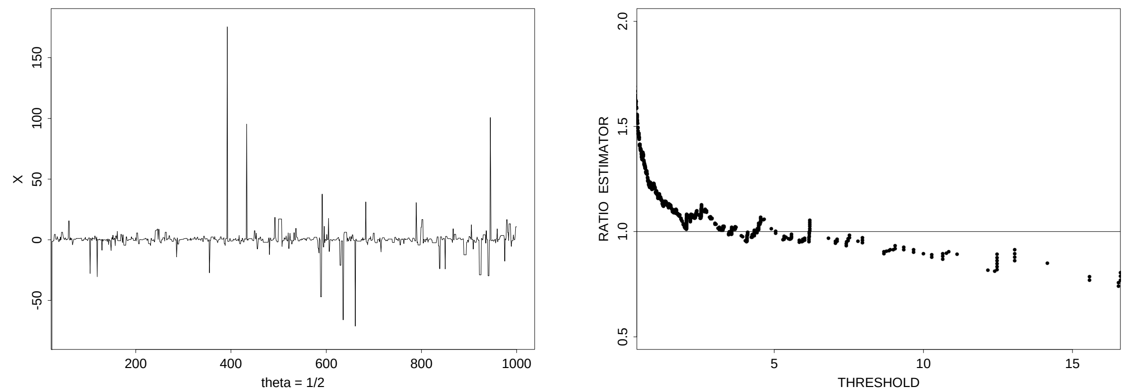

Figure 5: Process (34) with the standard Cauchy marginal distribution, $\theta = 1/2$, $n = 1000$. The ratio estimator $\hat{a} = 1.025$.

The tail constant $C = 2/\pi \approx 0.637$ can be estimated as well. Recall that $\hat{a}$ is the accepted tail index estimate (in our case, $\hat{a} = 1.025$). It can sometimes be worth using the estimator $\tilde{C}_n(x_n) = x_n^{1/\hat{a}} N_n/n$ instead of $\hat{C}_n(x)$.

The estimation results are presented in Figure 6. The plot of the estimator $\tilde{C}_n(x)$ is less volatile than that of $\hat{C}_n(x)$. The average value of $\tilde{C}_n(x)$ as $x \in [2; 12]$ is 0.62, the interval is formed by 300 points.

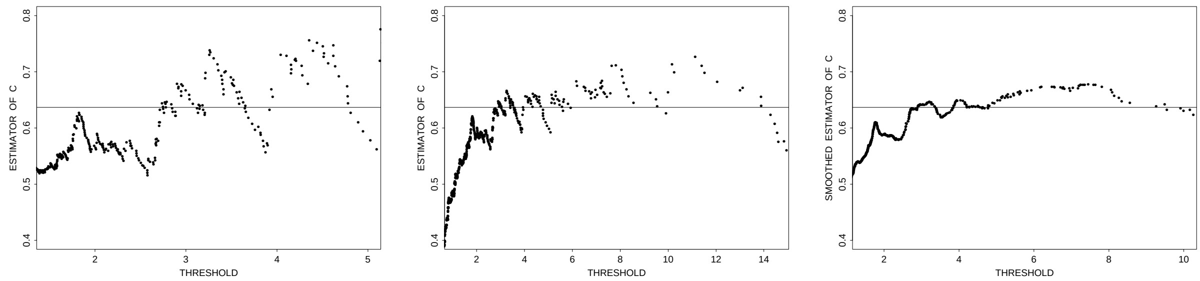

Figure 6: Tail constant estimators $\hat{C}_n(x)$, $\tilde{C}_n(x)$ and a smoothed version of $\tilde{C}_n(x)$. The average value $mean\{\tilde{C}_n(x) : x \in [2; 12]\} = 0.62$ (the true value is 0.637).

The results of quantile estimation are given in Figure 7. Both $\hat{y}_q$ and $\tilde{y}_q$ yield satisfactory estimates. The plot of $\hat{y}_q(x)$ is stable in the interval $x \in [2; 11]$ (formed by 249 points). The average value of $\hat{y}_q$ in that interval is 13.34 (the true value is $y_q = 12.7$). The plot of $\tilde{y}_q(x)$ is stable as $x \in [2; 18]$ (the interval is formed by 279 points), $mean\{\tilde{y}_q(x) : x \in [2; 18]\} = 13.28$.

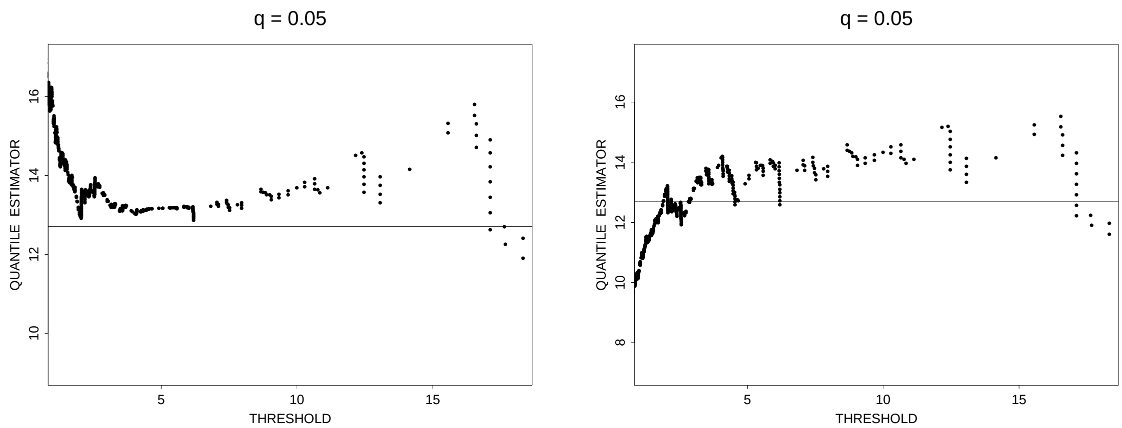

Figure 7: Quantile estimation: $\hat{y}_q = 13.34$, $\tilde{y}_q = 13.28$, the true value is $y_q = 12.706$.

The plot of the estimator $a_{n,r}$ is given in Figure 8 below. We simulated 1000 points according the model (34) with $\theta = 3/4$. The estimator looks stable in the interval $[2; 80]$ (formed by 316 points). The average value of $\hat{a}_{n,r} \equiv mean\{a_{n,r}(x) : x \in [2; 80]\} = 0.93$. The practical suggestion for the choice of r is to take the minimal integer exceeding the average cluster size ζ_n which can be estimated as $\zeta_n = 1/\theta_n$, where θ_n is a consistent estimator of the extremal index θ (cf. [15, 40]). This yields the 0.95%–a.c.i. $I_{n,r} = [0.81; 1.08]$.

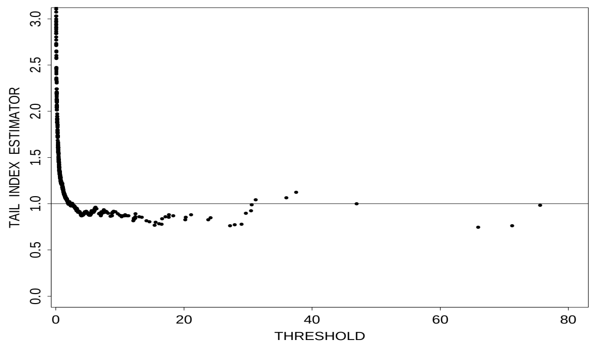

Figure 8: Process (34) with $\theta = 3/4$ and $n = 1000$. Tail index estimator $\hat{a}_{n,r} = 0.93$. The 0.95%–a.c.i. $I_{n,r} = [0.81; 1.08]$.

Example 3. The ARCH$(b|c)$ process is defined as a solution of the stochastic difference equation

$$X_n = Z_n \sqrt{b + cX_{n-1}^2} \qquad (n \geq 2)\,,$$

where $\{Z_i\}$ is a sequence of normal $\mathcal{N}(0;1)$ r.v.s, $b > 0$, $c \geq 0$. With a special choice of the initial r.v. X_1, the process is stationary, and

$$\mathbb{P}(|X| > x) \sim Cx^{-1/a} \qquad (x \to \infty)\,. \tag{49}$$

Explicit expressions for the constants a and C are given in [10] and [9], section 8.4. In particular, $a = 0.5$ and $C = 1.37$ if $b = c = 1$.

We simulated 10000 r.v.s from the ARCH(1|1) process with $X_1 = Z_1$ and then estimated a from the absolute values of the last 1000 observations (which can be considered as a stationary sequence). The estimation results are presented in Figures 9 and 10.

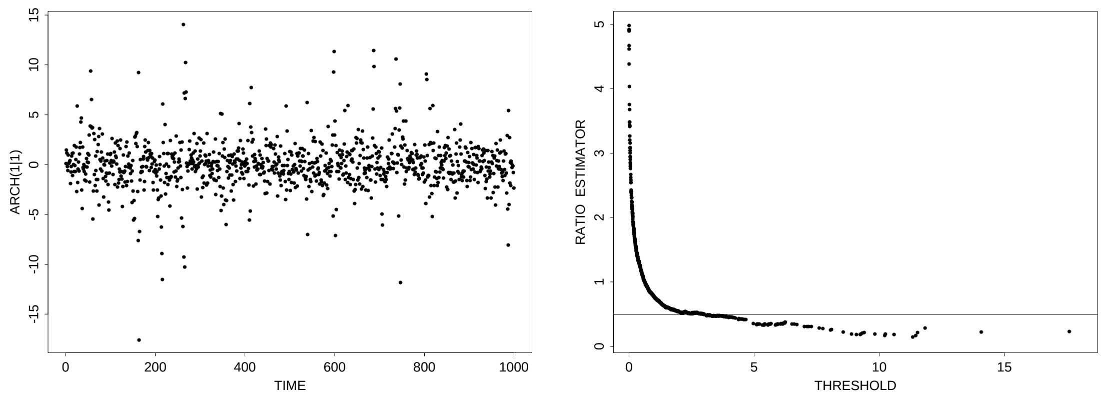

Figure 9: ARCH(1|1) process, $n = 1000$. The ratio estimator $\hat{a} = 0.51$ (the true value is $a = 0.5$).

The ratio estimator $a_n(x)$ behaves stable in the interval $x \in [2;4]$. This interval is formed by 179 points (out of 1000). The average value of the ratio estimator in that interval is $\hat{a} = 0.51$. Another interval of stable behavior of $a_n(x)$ is [5;11]. We reject it since it is formed by 51 points only.

Figure 10 presents the plots of tail constant estimators $\hat{C}_n(x)$ and $\tilde{C}_n(x)$. The interval $[2;3]$ (formed by 127 points) seems to be the only interval of stable behavior of $\hat{C}_n(x)$. The average value of this estimator as $x \in [2;3]$ is 0.9996 (the true value is $C = 1.37$). The estimator $\tilde{C}_n(x)$ is more stable in the interval $[2;9]$ (formed by 243 points). The average value of $\tilde{C}_n(x)$ as $x \in [2;9]$ is 1.09.

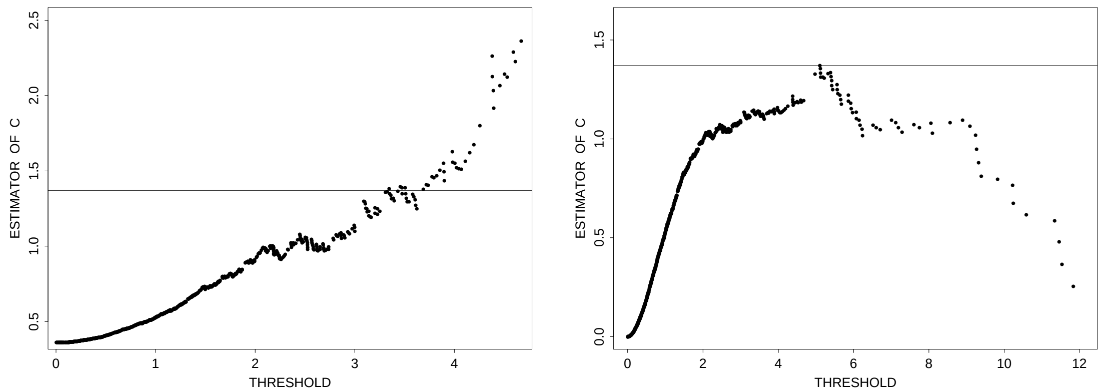

Figure 10: ARCH(1|1) process, tail constant estimators $\hat{C}_n = 0.9996$ and $\tilde{C}_n(x) = 1.09$ (the true value is $C = 1.37$).