Report 99-043

Inference on heavy tails from dependent data

S.Y.Novak

ISSN: 1389-2355

Inference on heavy tails from dependent data

S.Y.Novak
EURANDOM

Research Report No 99-043

Abstract

The statistics of extremes has been well developed for the case of 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, tail constant and quantile estimation from a sample of dependent random variables. Consistency and asymptotic normality of the corresponding estimators is established under mild mixing conditions.

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 (defined below) and quantiles. These problems have important applications in finance, insurance, network modelling, meteorology, etc.; they attracted significant interest from researchers (see, e.g., [4, 12, 28] and references therein).

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 (see [12] and references therein). 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 quantile estimation from a sample of dependent data. We show that, under mild mixing conditions, the estimators have the same accuracy as if the data were independent, while asymptotic variances may be larger.

Our main tool is the *ratio estimator* of the tail index.

We recall the main properties of the ratio estimator (in the i.i.d. case) before presenting the results.

---

[1] *AMS 1991 Subject Classification:* primary 62G05, secondary 60G70.

Key words and phrases: tail index estimation, quantile estimation, dependence.

Unless otherwise specified, the limits are as $n \to \infty$. 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 $0 < \liminf a_n/b_n < \limsup a_n/b_n < \infty$.

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:

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

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

Distributions that obey (1) form a non-parametric class of probability laws. Our purpose is to estimate the index a, the quantiles, and the tail constant (if it exists) from a sample $X_1, \ldots, X_n$ of random variables (r.v.s) distributed according to (1).

The ratio estimator

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

was introduced by Goldie and Smith [14]. 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 (see Figure 1 below); 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).

Let $X_{(n)} \leq \ldots \leq X_{(1)}$ be the sample order statistics. Denote $x_n^* = X_{(k_n+1)}$, where k_n is an integer number. Then 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 [4, 5, 28, 33]. A comparison of the asymptotic performance of some tail index estimators is given in [15].

Denote $p_n = \mathbb{P}(X > x_n)$,

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

The ratio estimator (3) is the sample analog of $a^*(x_n)$. According to the relation (7) below, $a^*(x) \to a$ as $x \to \infty$. The assumption

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

guarantees the consistency of the ratio estimator (see [21]). It is shown in [20, 21] 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.

The important question is how to choose the threshold x_n. The theoretically optimal threshold x_n^{opt} is the value x_n minimising the main terms in the asymptotic expansion for the mean squared error $\mathbb{E}(a_n - a)^2$. The ratio estimator (3) seems to be the only tail index statistic for which the asymptotics of the bias and the mean squared error have been calculated (see [20, 22]):

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

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

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

where

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

(see [25]), an explicit expression for x_n^{opt} can be drawn under additional restrictions on the distribution (1).

Consider, for instance, the following particular subclass of the non–parametric family (1):

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

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

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

(see [25]). 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 $\mathcal{P}_{1,2,1/\pi,-1/3}$, $x_n^{opt} = (16n/81\pi)^{1/5} \approx 2.29$ and $\mathbb{E}(a_n(x_n^{opt})/a - 1)^2 \sim \frac{5}{4}\left(\frac{16}{81}\right)^{1/5}\left(\frac{\pi}{n}\right)^{4/5}$. 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 $|\hat{a} - a| + |\hat{b} - b| = o_p(1/\ln n)$.

A simple practical approach is to plot $a_n(\cdot)$ and then choose x_n from an interval in which the function $a_n(\cdot)$ demonstrates stability. The background for this approach is provided by our 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 $[x_-; x_+]$ such that $a_n(x) \approx a$ for all $x \in [x_-; x_+]$.

Embrechts et al. ([12], p. 355) recommend choosing the minimal value y such that the empirical mean excess function

$$\mathcal{M}_n(y) = \sum^n (X_i - y)\mathbb{I}\{X_i > y\}\big/N_n(y)$$

is approximately linear (assuming $a < 1$). This approach makes the variance of the estimator "small", but the bias may become "large".

Denote

$$\mathcal{M}_{n,t}(y) = \sum^{n} X_i^t \mathbb{I}\{X_i > y\} \Big/ N_n(y) .$$

Let $[x_n^-; x_n^+]$ be the interval where the function $\mathcal{M}_{n,t}(x)$ is approximately linear.

We suggest choosing the average value $\hat{a}$ of $a_n(x)$ from this interval. To justify this approach, we recall that

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

where the random variable η_a has the exponential distribution with $\mathbb{E}\eta_a = a$. From (7) and the dominated convergence theorem,

$$\mathbb{E}\{(X/y)^t|X > y\} = \mathbb{E}\{e^{t\ln(X/y)}|X > y\} \to \mathbb{E}e^{t\eta_a} = 1/(1-at)$$

if $at < 1$. Hence $\mathbb{E}^{1/t}\{X^t|X > y\}$ and its empirical counterpart $\mathcal{M}_{n,t}^{1/t}(y)$ are asymptotically linear. Our simulation results (see Section 4 below) demonstrate that $\hat{a}$ seems to approximate the index a better than $a_n(x_n^-)$.

Since a is not known, the question is which t to choose. We suggest the following 3–step procedure: (1) construct a rough estimate $\tilde{a}_n$ of a; (2) with $t := 1/2\tilde{a}_n$, choose $[x_n^-; x_n^+]$ as described above; (3) get the final estimator.

If one believes that $v\sqrt{np_n} \to 0$, 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 of level $1-\varepsilon$ for the index a. The asymptotic confidence intervals do not take into account the accuracy of normal approximation, and hence may be too far 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

$$I_\varepsilon = \left[a_n \big/ \left(1 + y_\varepsilon N_n^{-1/2}\right); a_n \big/ \left(1 - y_\varepsilon N_n^{-1/2}\right)\right]$$

were introduced in [24]. Here $\Phi(-y_\varepsilon) = \left(\varepsilon/2 - C_* N_n^{-1/2} - |2v_1^* - v_2^*|/\sqrt{2\pi e}\right)_+$ and $C_* \leq 0.8$ is the constant from the Berry–Esseen inequality.

While efforts of researchers were concentrated mainly on the case of independent observations, increasing amount of data sets exhibiting heavy tails and dependence have been encountered in finance, teletraffic engineering, meteorology, hydrology, etc. (see [12, 28]). This stimulated recent studies of statistics of extremes in the case of dependent data.

Hsing [18] and Resnick & Starica [29] suggested sufficient conditions for consistency of Hill's estimator of the tail index in the case of m–dependent sequences and some stationary processes. Complicated sufficient conditions for the asymptotic normality of Hill's estimator in the case of dependent data are given by Starica [34] and Drees [8]. The latter paper presents also conditions of the asymptotic normality of empirical quantiles.

In this paper, 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 quantiles. In contrast to the results of the papers [34, 8], the accuracy of estimators appears to be of the same order as if the data were independent.

In Section 2 we develop procedures of tail index and tail constant estimation from a sample of dependent random variables. Our approach differs from those in [8, 18, 29, 30, 34]. 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 models, including the popular GARCH model.

We suggest a procedure of bias reduction in the CLT for the ratio estimator. Consistency of the tail constant estimator in the case of dependent data seems to be established for the first time.

Section 3 is devoted to the problem of quantile estimation. The problem arises, for example, when one measures the risk of heavy losses of a portfolio. In particular, *Value at Risk* is usually defined as the $m\%$–quantile for a small m.

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 [3, 9, 11, 32], see also references therein. Normal approximation for a weighted empirical quantile process in the case of dependent data is presented in [8].

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 is the only one that works for "very small" q (in particular, for $q < 1/n$, when $F_n \equiv 0$).

The EVT aproach to the problem of quantile estimation seems to have been introduced by Smith [33]. Although Smith dealt with the problem of tail estimation, the link to quantile estimation is evident. In the case of a parametric family of distributions and i.i.d. data, heuristic explanation of the EVT approach is given in [12], Section 6.5, and [19].

We apply the non–parametric EVT approach to the problem of quantile estimation from a sample of dependent data. A new quantile estimator is introduced. We show that, under mild assumptions, our estimator is consistent and asymptotically normally distributed.

In Section 4, we illustrate our results by examples of simulated data. Proofs are assembled in Section 5.

2 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, the tail constant (when it exists) and quantiles.

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

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

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

Proposition 1 *If*

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

then $a_n \xrightarrow[p]{} a$ *. If* $L(x) = C(1 + o(1))$ *as* $x \to \infty$ *and* $(\ln x_n)^2(v^2 + 1/np_n) \to 0$ *then*

$$\hat{C}_n = x_n^{1/a_n} n^{-1} \sum^n \mathbb{I}\{X_i > x_n\}$$

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

The estimator $\hat{C}_n$ was introduced by Goldie and Smith [14]. In the case of i.i.d. data, sufficient conditions for consistency and asymptotic normality of $\hat{C}_n$ are given in [14, 20, 21].

Denote $\mathbb{I}_i = \mathbb{I}\{X_i > x_n\}$, and let

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

Throughout the paper we assume that (10) holds and $\varphi(l) \to 0$ as $l \to \infty$. Since $\rho(l) \leq 2\varphi^{1/2}(l)$ (see [2]), this condition is valid if

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

Condition (10) is satisfied even if $\rho(l)$ decays like $(\ln l)^{-c}$ for some $c > 1$. In many models (like popular GARCH processes) $\varphi(\cdot)$ (and hence $\rho(\cdot)$) decay exponentially fast (see [7]).

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*

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

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.*

It follows from (7) and moment inequalities for sums of dependent r.v.s (see Utev [35]) that $\mathrm{Var}\left(\sum_{i=1}^n Y_i^{k-1}\mathbb{I}_i\right) \le c_k np_n$ for every $k \in \mathbb{N}$. In the i.i.d. case,

$$\mathrm{Var}\left(\sum_{i=1}^n Y_i^{k-1}\mathbb{I}_i\right) \sim \sigma_k^2 np_n\,,\ \mathbb{E}\left(\sum_{i=1}^n \mathbb{I}_i\right)\left(\sum_{i=1}^n \bar{Y}_i\right) \sim \sigma_{12}np_n \tag{15}$$

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.

Denote $T_{k,j} = \sum_{l=(j-1)r+1}^{jr} Y_l^k\mathbb{I}_l$, and let ($0 \le l < m,\ k \ge 1,\ 1 \le r \le n$)

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

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* $Var\left(\sum_{i=1}^n Y_i^*\right) \sim \sigma^2 np_n$ *and*

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

where $\hat\sigma^2 \equiv \hat\sigma^2(n) = (a_n\hat\sigma_1)^2 + \hat\sigma_2^2 - 2a_n\hat\sigma_{12}$.

Remark 1. In Theorem 2, Corollary 3 and Theorem 6 below, $\sqrt{np_n}$ may be replaced by $N_n^{1/2}$.

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 $\hat\sigma - \sigma = O_p\left((np_n/r)^{-1/2}\right)$. Therefore, though (at least in the i.i.d. case, cf. [22, 24])

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

where $\Phi_{a;b}$ is the distribution function of the normal $\mathcal{N}(a;b)$ law, 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 loose much if we switch to an estimator $a_{n,r}$ such that the rate of normal approximation for $a_{n,r} - a$ is, in a sense, $O\left((np_n/r)^{-1/2}\right)$.

Let

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

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

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

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., with $r = \sqrt{np_n}$.

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]). Using "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 Berstein's blocks method. Typically, conditions are formulated in terms of the mixing coefficient $\alpha(\cdot)$ (though in [34, 8], conditions are given in terms of the stronger coefficient $\beta(\cdot)$). In particular, Starica ([34], 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 $1 \ll l \le r \ll n$, $1 \ll k \ll n$.

Our condition (11) is preferable if the mixing coefficients $\beta(\cdot)$ and $\varphi(\cdot)$ 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 $k = o((\ln l)^3)$. Therefore, $rk^{-1/2+\varepsilon} \gg r(\ln l)^{-4.5+3\varepsilon} \ge l(\ln l)^{-4.5+3\varepsilon} \to \infty$. Hence (21) does not hold while (11) is evidently valid.

Sufficient conditions for the asymptotic normality of the ratio estimator (in terms of the mixing coefficient $\alpha(\cdot)$) can be deduced from the results of Rootzen et al. [30]. Sufficient conditions for the asymptotic normality of Hill's estimator (in terms of the mixing coefficient $\beta(\cdot)$) are obtained by Starica [34] for the stationary solution of a stochastic difference equation

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

where $(A_i, B_i), i \ge 1$ is an i.i.d. sequence of r.v.s. According to Goldie [13], (14) holds under some natural assumptions on the r.v.s A_i, B_i. Starica [34] shows 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 [16, 17]) 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 by Drees [8] imply the asymptotic normality of Hill's estimator under mixing conditions similar to those of Starica [34] 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.

3 Quantile estimation

This section is devoted to the problem of quantile estimation in the case of dependent data.

Let $y_q = \inf\{t : G(t) \leq q\}$. Given $q = q(n)$ "close" to zero, we want to construct an estimator $\hat{y}_q = \hat{y}_q(n)$ such that

$$\hat{y}_q/y_q \underset{p}{\longrightarrow} 1\,, \quad G(\hat{y}_q)/q \underset{p}{\longrightarrow} 1\,. \tag{24}$$

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 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 $G(y_q) \approx \frac{N_n}{n}\left(\frac{y_q}{x_n}\right)^{-1/a}$.

Let $\hat{a}_n$ be a consistent estimator of the index a. Since $G(y_q) \sim q$ by formula (29) below, one can expect that the statistic

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

approximates y_q.

Proposition 5 *Suppose that*

$$1 \leq y_q/x_n \leq C_* \tag{26}$$

for some constant $C_* \in [1;\infty)$. *Then (24) holds.*

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

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

Theorem 6 *Suppose that (26) holds,* $\sigma^2 \equiv (a\sigma_1)^2 + \sigma_2^2 - 2a\sigma_{12} > 0$ *and*

$$\ln\left(G(x_n)/G(y_q)\right) \to d\,, \quad (a^* - a)\sqrt{np_n} \to \mu\,, \quad \left(y_q q^a L^{-a}(x_n) - 1\right)\sqrt{np_n} \to \nu \tag{27}$$

for some constants d, μ, ν. *Then*

$$\left(\hat{y}_q/y_q - 1\right)\sqrt{np_n} \Longrightarrow \mathcal{N}\left(d\mu - \nu, \sigma_{\mathbf{c}}^2\right), \tag{28}$$

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

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

Concerning the last relation in (27), note that

$$G(y_q) = G(G^{-1}(q)) \sim q \tag{29}$$

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 (27) may be rewritten as

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

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 $\hat{\sigma}_c^2 \equiv \hat{\sigma}_c^2(n) = \hat{\mathbf{c}}\hat{A}\hat{\mathbf{c}}^T$, where $\hat{\mathbf{c}} = (a_n, d_n)$ and $d_n = a_n^{-1}\ln(\hat{y}_q/x_n)$.

Corollary 7 *Assume the conditions of Theorem 6. If (15) holds then*

$$\left(\hat{y}_q/y_q - 1\right)\hat{\sigma}_c^{-1}N_n^{1/2} \Longrightarrow \mathcal{N}\left((d\mu - \nu)/\sigma_c, 1\right) . \tag{30}$$

We can eliminate the asymptotic bias $(d\mu - \nu)/\sigma_c$ but at a cost of a slower rate of normal approximation. Denote $\hat{y}_{q,r} = \hat{y}_q(a_{n,r})$.

Theorem 8 *Assume the conditions of Theorem 6. If (15) and (18) hold then*

$$\left(\hat{y}_{q,r}/y_q - 1\right)(a_n d_n)^{-1} N_{n,r}^{1/2} \Longrightarrow \mathcal{N}(0,1) . \tag{31}$$

4 Examples

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 $\mathbb{P}_0$.

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). If $x \in [0.5; 17]$ then $a_n(x)$ ranges from 0.9235 to 1.3481. 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, $a_n(x)$ ranges from 0.9235 to 1.1824 .

It is reasonable to pick up the estimate of the index a from the interval $[1; 14]$ — say, taking the average value $\hat{a}$ of $a_n(x)$ in the interval $x \in [1; 14]$: $\hat{a} = 0.9983$.

The plot of the tail constant estimator $\hat{C}_n(\cdot)$ is presented in the first picture of Figure 2. 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.

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

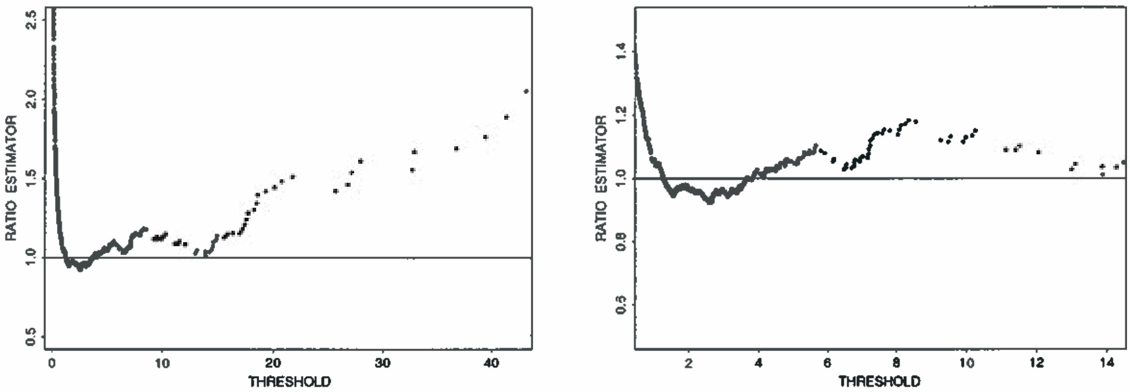

Figure 1: Tail index estimation from the distribution $\mathbb{P}_0$, $n = 1000$. The average value $\hat{a}$ of $a_n(x)$ in the interval $x \in [1; 14]$ is $\hat{a} = 0.9983$.

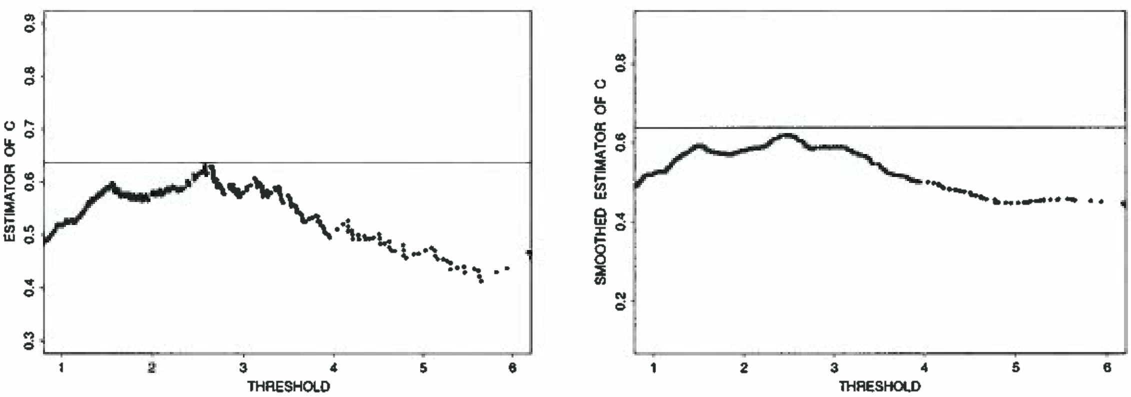

Figure 2: Tail constant estimation. Estimator $\hat{C}_n(\cdot)$ seems to be stable when $x \in [1.5; 3.5]$. The corresponding fragment of the curve is formed by 229 points, the average value of $\hat{C}_n(\cdot)$ in that interval is 0.5854 (the actual value of C is $2/\pi \approx 0.6366$).

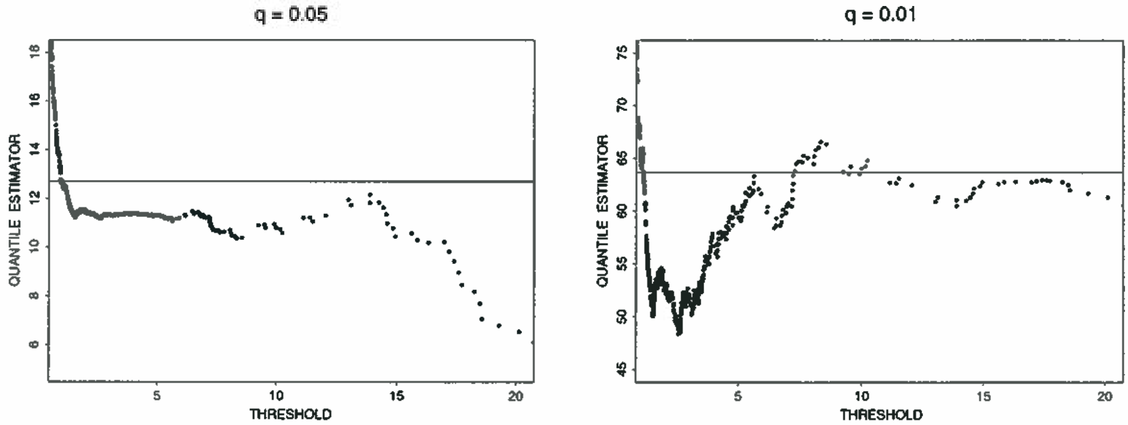

Figure 3: Quantile estimator (25). The first picture presents $\hat{y}_q$ for the case $q = 0.05$, the true value is $y_q = 12.706$. The plot demonstrates stability in the interval $x \in [1.5; 14]$ (formed by 345 points). The average value of $\hat{y}_q$ in that interval is 10.52. The empirical 0.95% quantile equals 9.91. The second picture displays $\hat{y}_q$ for the case $q = 0.01$, the true value is $y_q = 63.657$. The plot looks stabile 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 59.88. The empirical 0.99% quantile equals 41.34.

Let $\hat{a}$ be the tail index estimate obtained at the step of tail index estimation (in our example, $\hat{a} = 0.9983$). It can sometimes be worth using the estimator

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

instead of (25). The simulation results are presented in Figure 4 below.

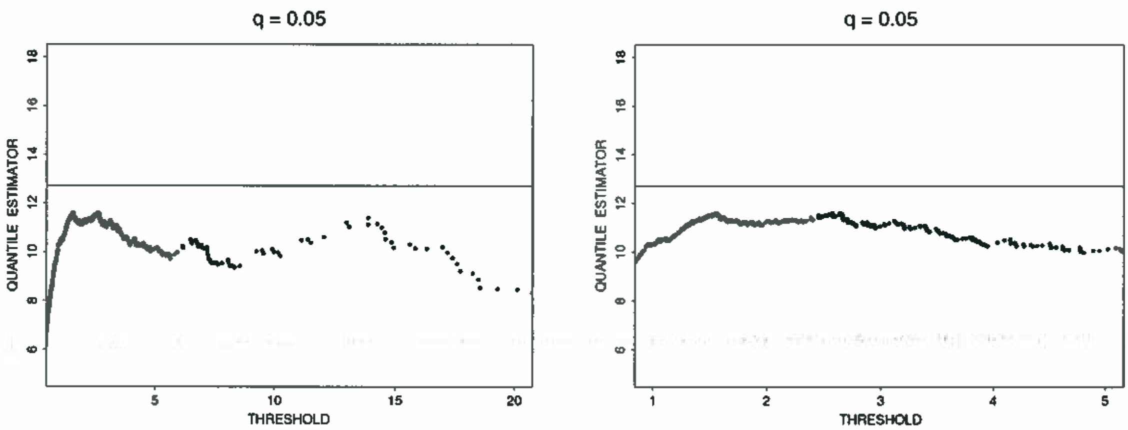

Figure 4: Quantile estimator $\tilde{y}_q$. The plots demonstrate stability in the interval $x \in [1.5; 4]$ (formed by 256 points). The average value of $\tilde{y}_q$ in that interval is 11.201, the true value is $y_q = 12.706$.

Example 2. Consider the following model:

$$X_1 = \xi_1\,, \;\; X_i = \alpha_i \xi_i + (1 - \alpha_i) X_{i-1} \qquad (i \geq 2)\,, \tag{32}$$

where $\xi_1, \xi_2, ..., \alpha_1, \alpha_2, ...$ are independent random variables, $\xi_i \stackrel{d}{=} X$ $(\forall i)$, the distribution of the random variable 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 [32]. It is a particular case of a stochastic difference equation (22). It is easy to see that (32) is a stationary Markov chain, the extremal index equals θ, and clusters have geometric distribution with the mean $1/\theta$. It is shown in [23] that $\varphi(k) \leq (1-\theta)^k$. Hence (11) and (10) hold.

We prove in Section 5 that

$$\mathrm{Var}\Big(\sum\nolimits^n Y_i^*\Big) = n\mathbb{E}(Y^*)^2\left[1 + 2(\theta^{-1}-1)(1-\kappa_n)\right] \sim np_n\left(2\theta^{-1}-1\right)a^2, \tag{33}$$
$$\mathrm{Var}\Big(\sum\nolimits^n Y_i^k \mathbb{I}_i\Big) \sim np_n\left(\frac{2}{\theta}-1\right)a^{2k}(2k)!\,,\ \mathbb{E}\Big(\sum\nolimits_{i=1}^n \mathbb{I}_i\Big)\Big(\sum\nolimits_{i=1}^n Y_i\Big) \sim np_n\left(\frac{2}{\theta}-1\right)a\,,$$

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

$$(a_n/a - 1)\sqrt{np_n} \Longrightarrow \mathcal{N}\left(m; 2\theta^{-1}-1\right) \tag{34}$$

if $v\sqrt{np_n} \to m$; $\sqrt{np_n}$ in (34) may be replaced by $N_n^{1/2}$.

This is a generalisation of the limit theorem (13): if $\alpha_i \equiv 1$ then (32) is a sequence of independent r.v.s, $\theta = 1$, and (34) 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 ratio estimator can be only larger.

We simulated the r.v.s $X_1, ..., X_{1000}$ according to the model (32) with the standard Cauchy marginal distribution and $\theta = 1/2$ (see Figure 5). The estimation results (based on absolute values $|X_i|$) are presented in Figure 5.

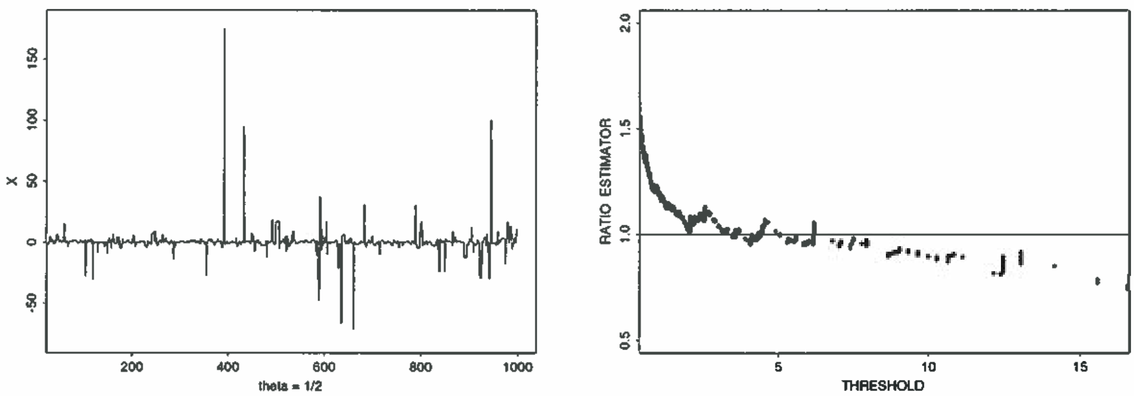

Figure 5: Process (32) with the standard Cauchy marginal distribution, $\theta = 1/2$, $n = 1000$. The ratio estimator demonstrates stability in the interval $x \in [1.5; 14]$ (formed by 322 points). The average value of $a_n(x)$ as $x \in [1.5; 14]$ is 1.0248.

The tail constant can be estimated as well.

Recall that $\hat{a}$ is the accepted tail index estimate (in our case, $\hat{a} = 1.0248$). It can sometimes be worth using the estimator $\tilde{C}_n(x) = x_n^{1/\hat{a}} N_n/n$ 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 $\hat{C}_n(x)$.

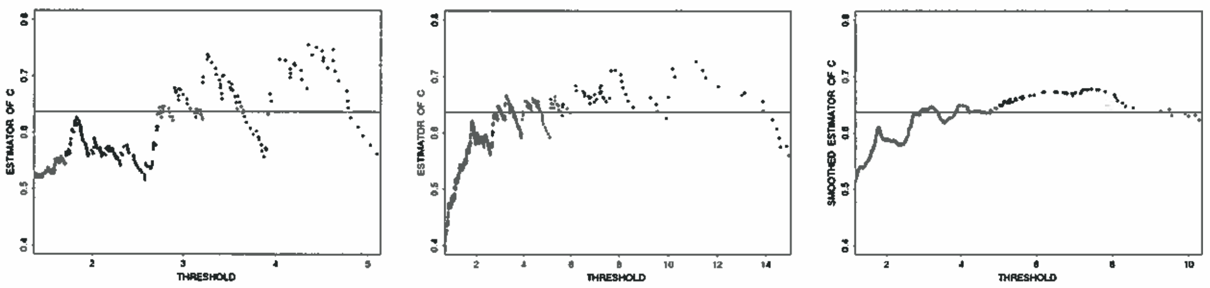

Figure 6: Tail constant estimators $\hat{C}_n(x)$, $\tilde{C}_n(x)$ and a smoothed version of $\tilde{C}_n(x)$. The average value of $\tilde{C}_n(x)$ as $x \in [2;12]$ is 0.6204, the interval is formed by 300 points. The true value is $C = 2/\pi \approx 0.6366$).

The results of quantile estimation are given in Figure 7. Both $\hat{y}_q$ and $\tilde{y}_q$ yield satisfactory estimates.

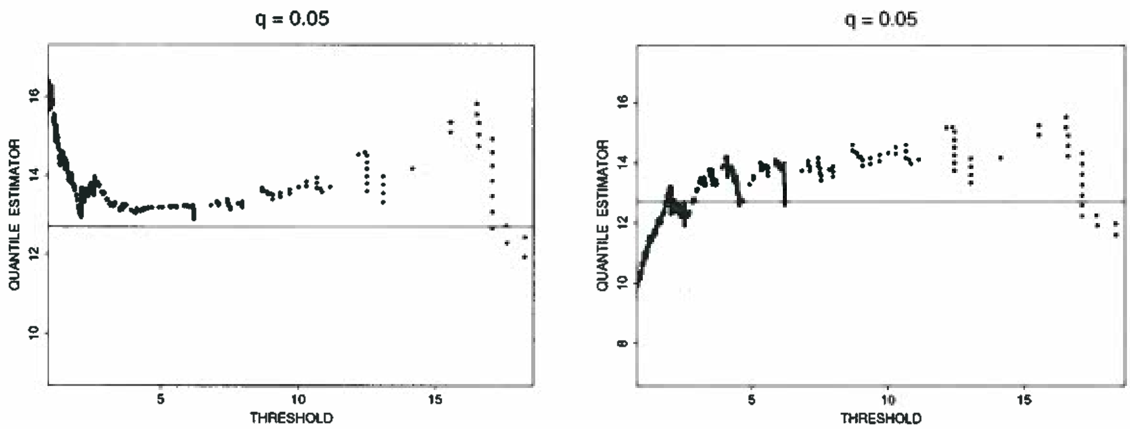

Figure 7: Quantile estimators $\hat{y}_q$ and $\tilde{y}_q$, $q = 0.05$. The plot of $\hat{y}_q$ looks stable in the interval $x \in [2;11]$ that is formed by 249 points. The average value of $\hat{y}_q$ in that interval is 13.34 (the true value is $y_q = 12.706$). The plot of $\tilde{y}_q$ is stable in the interval $x \in [2;18]$ that is formed by 279 points. The average value of $\tilde{y}_q$ in that interval is 13.28.

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

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

Explicit expressions for the constants a and C are given in [13] and [12], 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 $X_1 = Z_1$ (see Figure 8), and then estimated a from the absolute values of the last 1000 observations (which can be considered as a stationary sequence).

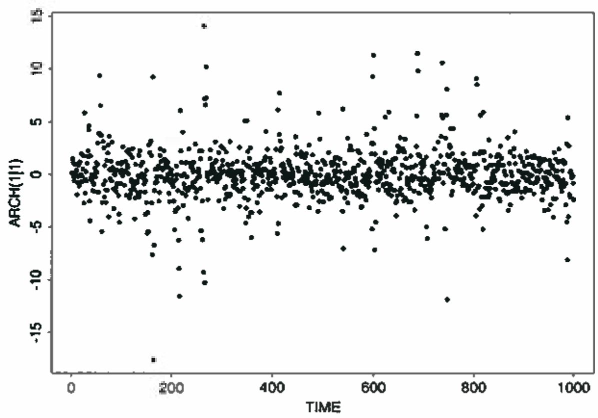

Figure 8: ARCH(1|1) process, $n = 1000$.

The estimation results are presented in Figures 9 and 10.

Conclusion. Our simulation results show that the statistical procedures based on the ratio estimator perform quite satisfactory. This can be a bit surprising in view of Resnick [28]. The 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 25% least and 25% most informative parts. In all our examples, 25% smallest elements lie below the threshold 0.5 (and hence should not be given any attention), approximately half of the sample elements lie below the threshold 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.

5 Proofs

Below, symbols c_i denote positive constants; a bar over a random variable means that it is centred by its expectation.

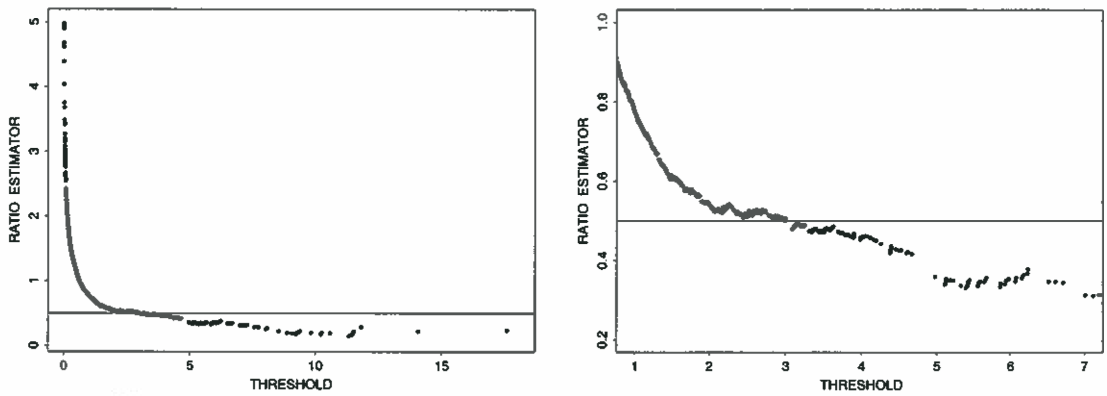

Figure 9: Tail index estimation from ARCH(1|1) process. The ratio estimator $a_n(x)$ behaves stable in the interval $x \in [2;4]$. This interval is formed by 179 points (out of 1000). The second picture focuses on that interval. The average value of the ratio estimator in that interval is $\hat{a} = 0.5096$. Another interval of stable behavior of $a_n(x)$ is [5;11]. We reject it since it is formed by 51 points only.

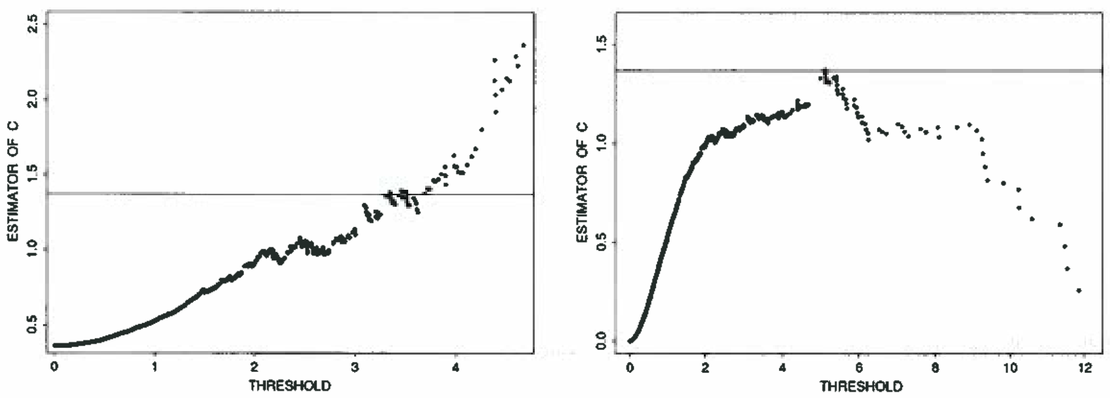

Figure 10: ARCH(1|1) process, tail constant estimators $\hat{C}_n(x)$ and $\tilde{C}_n(x)$. The interval [2; 3] (formed by 127 points) seems to be the only interval of stable behavior of $\hat{C}_n(x)$. The average value of this estimator as $x \in [2;3]$ is 0.9996 (the true value is $C = 1.3705$). The 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 1.0912.

Proof of Proposition 1. The first part of the statement (consistency of the ratio estimator a_n) is established in [25]. We present its proof 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 $l \to \infty$) and that (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 a variance of a sum of dependent random variables (see Peligrad [26] or Utev [35]). For any $\varepsilon > 0$, denote $Z_i = Y_i^* - (\mathbb{I}_i - p_n)\varepsilon$. Then

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

By Theorem 1.1 in [35], there exists a constant c_ρ (depending only on $\rho(\cdot)$) such that $\mathrm{Var}\,(\sum_{i=1}^n Z_i) \leq c_\rho n \mathrm{Var}\, Z_1 \leq c n p_n$ (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 $\hat{C}_n \underset{p}{\longrightarrow} C$. We apply Theorem 1.1 by Utev [35] to check that $\mathrm{Var}\,(\sum_{i=1}^n \mathbb{I}_i) \leq c_\rho n p_n$. An application of Chebyshev's inequality yields

$$N_n/np_n \underset{p}{\longrightarrow} 1\,. \tag{35}$$

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 $(\sum^n Y_i^*)(\ln x_n)/np_n \underset{p}{\longrightarrow} 0$. The latter follows from Chebyshev's inequality, the assumption and the estimate $\mathrm{Var}\,(\sum^n Y_i^*) \leq c_{\rho,a} n p_n$. □

Denote $G_n = N_n/n$.

Lemma 9 *If $\varphi_n(l) \to 0$ as $l \to \infty$ and (15) holds then*

$$\left(\frac{G_n}{G(x_n)} - 1, a_n - a^*\right)\sqrt{np_n} \Longrightarrow \mathcal{N}(\mathbf{0}; A)\,. \tag{36}$$

Proof of Lemma 9. Note that

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

Taking into account (35), we shell check that

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

Notice that $(\bar{\mathbb{I}}_i, Y_i^*)^T = B_*\zeta_i$, where $\zeta_i = (\bar{\mathbb{I}}_i, \bar{Y}_i)^T$. 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 [36].

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

$$\lim_{n\to\infty} j_n z_n^{-2} \sum_{i=1}^{k_n} \mathbb{E}\xi_{i,n}^2 \mathbb{I}\{|\xi_{i,n}| > \varepsilon z_n/j_n\} = 0 \qquad (\forall \varepsilon > 0) \tag{37}$$

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

$$\sum^n c\zeta_i/\sqrt{np_n} \Longrightarrow \mathcal{N}(\mathbf{0}; cA_0c^T)\,. \tag{38}$$

Put $\xi_i = c_1\mathbb{I}_i + c_2(Y_i - a^*p_n)$ and $j_n = 1$. By the assumption, $\mathrm{Var}\,(\sum^n \xi_i) \sim \sigma^2 np_n$. To check (37), it suffices to show that

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

for any $\varepsilon > 0$. According to [1, 31],

$$L(xy)/L(x) \sim \exp\left(\int_x^{xy} \frac{w(z)}{z}dz\right) \qquad (x \to \infty)$$

uniformly in $y \ge 1$, where $w(z) \to 0$ as $z \to \infty$. 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 (7), we derive

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

Hence (37) holds, and Theorem A entails (38) and (36). □

Proof of Theorem 2. Arguments of the proof of Lemma 9 yield also that

$$\sum\nolimits_{j=1}^n Y_i^*\left(\mathrm{Var}\left(\sum\nolimits_{j=1}^n Y_i^*\right)\right)^{-1/2} \Longrightarrow \mathcal{N}(0;1)\,.$$

Taking into account (35) and the assumptions of the theorem, we get (12). □

Lemma 10 *If (15) holds then*

$$\hat\sigma_k \xrightarrow[p]{} \sigma_k\,,\ \hat\sigma_{12} \xrightarrow[p]{} \sigma_{12} \qquad (k \in \mathbb{N})\,. \tag{39}$$

Proof of Lemma 10. First of all, notice that

$$\mathbb{E}\left(\sum^r Y_i^{k-1}\mathbb{I}_i\right)^2 \sim \sigma_k^2 r p_n \qquad (k \in \mathbb{N})\,. \tag{40}$$

Indeed, denote $R_n = \mathrm{Var}\left(\sum^{[n/r]} T_{k,j}\right) - [n/r]\mathrm{Var}\, T_{k,1}$. By Utev's Theorem 1.1 [35], $R_n = o([n/r]\mathrm{Var}\, T_{k,1})$. Therefore, $\mathrm{Var}\left(\sum^{[n/r]} T_{k,j}\right) \sim [n/r]\mathrm{Var}\, T_{k,1} \le c_1 n p_n$, and

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

By the assumption, $r/n \to 0$. Thus, $\mathrm{Var}\, T_{k-1,1} \equiv \mathrm{Var}\left(\sum^r Y_i^{k-1}\mathbb{I}_i\right) \sim \sigma_k^2 r p_n$, and (40) follows.

We use Chebyshev's inequality to prove (39). Note that $\sigma_k^2 - [n/r]\mathbb{E}T_{k-1,1}^2/np_n = o(1)$. Using Theorem 1.1 and Corollary 2.3 by Utev [35], we get $\mathrm{Var}\, T_{k,1}^2 \le \mathbb{E}T_{k,1}^4 \le c_2 r p_n$ and

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

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

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

$(\forall \varepsilon > 0)$. 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 $\hat\sigma_{12} \underset{p}{\longrightarrow} \sigma_{12}$. Remaind that $Y_i^* = Y_i - a^*\mathbb{I}_i$. According to (15),

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

Similarly to (40), 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 Theorem 1.1 and Corollary 2.3 by Utev [35], we get $\mathrm{Var}\left(\sum_{j=1}^{[n/r]} T_{0,j}T_{1,j}\right) \sim \frac{n}{r}\mathrm{Var}\,(T_{0,1}T_{1,1}) \le c n p_n$. Note that

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

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

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

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 (35) and Lemma 9 imply

$$\zeta_n = \left(a_n - a^*\right)\sqrt{\sum_{i=1}^n \mathrm{I\!I}_i} \Longrightarrow \mathcal{N}(0;\sigma^2).$$

Since $(a^* - a)\sqrt{np_n} \to \mu$ as $n \to \infty$,

$$\left(a_n - a\right)\sqrt{\sum_{i=1}^n \mathrm{I\!I}_i} = \zeta_n + (a^* - a)\sqrt{np_n}\sqrt{\sum_{i=1}^n \mathrm{I\!I}_i/np_n} \Longrightarrow \mathcal{N}(\mu;\sigma^2).$$

Now (16) follows from this relation, (35) and Lemma 10. □

Proof of Theorem 4. Using Theorem 1.1 by Utev [35], we conclude that

$$\begin{aligned} \mathrm{Var}\left(\sum_{j=1}^{[n/r]} \mathrm{I\!I}_{jr}\right) &\sim [n/r]\mathrm{Var}\,\mathrm{I\!I}_1 \sim np_n/r\,, \\ \mathrm{Var}\left(\sum_{j=1}^{[n/r]} Y_{jr}^*\right) &\sim [n/r]\mathrm{Var}\,Y_1^* \sim a^2 np_n/r\,. \end{aligned}$$

Hence $N_{n,r}/\left(np_n/r\right) \underset{p}{\longrightarrow} 1$. Using the same arguments as in the proof of Lemma 9, we show that $\mathrm{I\!P}\left(Y > \varepsilon\sqrt{np_n/r}\middle|X > x_n\right) \to 0$. Theorem A with $\xi_i = Y_{ir}^*$ and $j_n = 1$ yields

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

Note that

$$(a^*/a - 1)\sqrt{N_{n,r}} = v\sqrt{np_n}\sqrt{N_{n,r}/np_n} \underset{p}{\longrightarrow} 0$$

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

Proof of Proposition 5. We write $\xi_n \underset{p}{\sim} \eta_n$ if $\xi_n = \eta_n(1 + o_p(1))$. First, we want to show that

$$G(\hat{y}_q)/G(y_q) \underset{p}{\longrightarrow} 1\,. \tag{41}$$

Evidently, $G(\hat{y}_q)/G(y_q)$ equals

$$\frac{G(\hat{y}_q)}{G(x_n)}\frac{G(x_n)}{G(y_q)} = \frac{G(x_n)}{G(y_q)}\frac{G((G_n/q)^{\hat{a}_n}x_n)}{G(x_n)} = \frac{G(x_n)}{G(y_q)}\left(\frac{G_n}{q}\right)^{-\hat{a}_n/a}\frac{L((G_n/q)^{\hat{a}_n}x_n)}{L(x_n)}.$$

By (35), $(G_n/G(x_n))^{\hat{a}_n/a} \underset{p}{\longrightarrow} 1$. Properties of slowly varying functions and the assumption imply $|\ln(G(x_n)/G(y_q))| \le C_1 < \infty$. Hence $(1 - \hat{a}_n/a)\ln(G(x_n)/G(y_q)) \underset{p}{\longrightarrow} 0$. Taking into account (29), we deduce

$$\frac{G(\hat{y}_q)}{G(y_q)} \underset{p}{\sim} \left(\frac{G(x_n)}{G(y_q)}\right)^{1-\hat{a}_n/a}\frac{L((G_n/q)^{\hat{a}_n}x_n)}{L(x_n)} \underset{p}{\sim} \frac{L((G_n/q)^{\hat{a}_n}x_n)}{L(x_n)}.$$

Note that

$$\hat{y}_q \underset{p}{\sim} (G(x_n)/G(y_q))^{\hat{a}_n} x_n \geq x_n \underset{p}{\longrightarrow} \infty . \tag{42}$$

By the canonical representation of a slowly varying function (see [1, 31]),

$$\frac{L((G_n/q)^{\hat{a}_n} x_n)}{L(x_n)} \underset{p}{\sim} \exp\left(\int_{x_n}^{(G_n/q)^{\hat{a}_n} x_n} \frac{w(u)}{u} du\right),$$

where $w(u) \to 0$ as $u \to \infty$. Therefore, $L((G_n/q)^{\hat{a}_n} x_n)/L(x_n) = \exp\left(o(\hat{a}_n \ln(G_n/q))\right)$. Because of (26),

$$1 \leq \frac{G(x_n)}{G(y_q)} \leq \left(C_*^{1/a} + o(1)\right) \exp\left(\int_{x_n}^{y_q} \frac{|w(u)|}{u} du\right) = C_*^{1/a} + o(1).$$

Hence $L((G_n/q)^{\hat{a}_n} x_n)/L(x_n) \underset{p}{\longrightarrow} 1$, and (41) holds.

Taking into account (1), we rewrite (41) as $\hat{y}_q^{-1/a} L(\hat{y}_q) \underset{p}{\sim} y_q^{-1/a} L(y_q)$. This implies

$$\hat{y}_q / y_q \underset{p}{\sim} L^a(\hat{y}_q) L^{-a}(y_q). \tag{43}$$

Because of (2), (26) and (42),

$$\hat{y}_q \leq x_n \left(G(x_n)/G(C_* x_n)\right)^{\hat{a}_n} (1 + o_p(1)) \underset{p}{\sim} C_* x_n .$$

Hence $x_n(1 + o_p(1)) \leq \hat{y}_q \leq C_* x_n (1 + o_p(1))$. This and (26) entail

$$C_*^{-1} y_q (1 + o_p(1)) \leq \hat{y}_q \leq C_* y_q (1 + o_p(1)).$$

The first statement in (24) follows from (2) and (43). □

Actually, we have showed that $G(\hat{y}_q)/q \underset{p}{\longrightarrow} 1 \iff \hat{y}_q/y_q \underset{p}{\longrightarrow} 1$ in the assumptions of Theorem 5.

Proof of Theorem 6. Note that

$$\frac{\hat{y}_q}{x_n} = \left(\frac{G_n}{q}\right)^{a_n} = \left(\frac{G_n}{G(x_n)}\right)^{a_n} \left(\frac{G(x_n)}{q}\right)^{a_n} = \frac{G^a(x_n)}{q^a} \left(\frac{G_n}{G(x_n)}\right)^{a_n} \left(\frac{G(x_n)}{q}\right)^{a_n - a} .$$

Therefore,

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

where $\delta_n = o_p\left(|1 - G_n(x_n)/G(x_n)| + |a_n - a|\right)$. By Lemma 9,

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

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

Corollary 7 follows from Theorem 6 and Lemma 10. 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 8. Arguments similar to those in the proof of Theorem 6 yield

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

According to Lemma 9, $G_n(x_n)/G(x_n) - 1 = O_p\left(1/\sqrt{np_n}\right)$. Therefore,

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

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

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

by Theorem 4. The result follows. □

Proof of relation (33). By the well-known formula,

$$\mathbb{E}\left(\sum^n Y_i^*\right)^2 = n\left[\mathbb{E}(Y^*)^2 + 2\sum_{i=1}^n (1 - i/n)\,\mathbb{E}Y_1^*Y_{i+1}^*\right]. \tag{44}$$

Notice that

$$\mathbb{E}Y_1^*Y_{i+1}^* = (1-\theta)^i\mathbb{E}(Y^*)^2 \qquad (i \geq 1). \tag{45}$$

Indeed, if $\alpha_2 = \ldots = \alpha_{i+1} = 0$ 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 (33) follows from (44), (45) and (7).

By the same argument, $\mathbb{E}\bar{Y}_1^k\mathbb{I}_1\bar{Y}_{i+1}^k\mathbb{I}_{i+1} = (1-\theta)^i\mathbb{E}(\bar{Y}_1^k)^2\mathbb{I}_1$, and $\mathbb{E}\mathbb{I}_1\bar{Y}_{i+1} = \mathbb{E}\bar{Y}_1\mathbb{I}_{i+1} = (1-\theta)^i\mathbb{E}\mathbb{I}_1\bar{Y}_1 = (1-\theta)^i a^* p_n(1-p_n)$. Hence ($k \geq 0$)

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

Similarly one can check that

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

The proof is complete. □

Acknowledgements. The author is grateful to L. de Haan, T. Mikosch and C. de Vries for many helpful discussions.

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] Csörgö M. and Horvath L. (1993) *Weighted approximations in probability and statistics.* — New York: Wiley.

[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] Danielsson J., Jansen D.W. and de Vries C.G. (1996) The method of moments ratio estimator for the tail shape parameter. — Commun. Statist. Theory Meth., v. 25, No 4, 711–720.

[6] Danielsson J., de Haan L., Peng L. and de Vries C.G. (1998) Using a bootstrap method to choose the sample fraction in tail index estimation. — Preprint, Erasmus University Rotterdam.

[7] 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.

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

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

[10] Drees H. and Kaufman E. (1998) Selecting the optimal sample fraction in univariate extreme value estimation. — Stochastic Processes Appl., v. 75, 149–172.

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

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

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

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

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

[16] Hall P. (1982) On some simple estimates of an exponent of regular variation. — J. R. Statist. Soc. B, v. 44, No 1, 37–42.

[17] 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.

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

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

[20] 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.

[21] 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.

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

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

[24] Novak S.Y. (1998) Berry–Esseen inequalities for a ratio of sums of random variables. — University of Sussex: Research Report No 98/07 SMS/CSSM.

[25] Novak S.Y. (2000) Confidence intervals for a tail index estimator. — In: Proceedings Conf. "Measuring Risk in Complex Stochastic Systems", J. Franke, W. Haerdle and G. Stahl (eds), pp. 229–236. Berlin: Springer Verlag.

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

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

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

[34] 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

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

[36] 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.

EURANDOM
PO Box 513
Eindhoven 5600 MB
Netherlands