Submitted to the Annals of Statistics
arXiv: math.PR/0000000

ADAPTIVE ESTIMATION OF THE EXCESS OF A DISTRIBUTION FUNCTION

By Ion Grama and Vladimir Spokoiny

University of South Brittany and Weierstrass-Institute

We use the fitted Pareto law to construct an accompanying approximation of the excess distribution function. A selection rule of the location of the excess distribution function is proposed based on a stagewise lack-of-fit testing procedure. Our main result is an oracle type inequality for the Kullback-Leibler loss of the obtained adaptive estimator.

1. Background and outline of main results. Let $X_1, ..., X_n,$ be i.i.d. observations with continuous d.f. F supported on the interval $[x_0, \infty)$, $x_0 \geq 0$. Assume that d.f. F is "heavy tailed", i.e. that F belongs to the domain of attraction of the Fréchet law $\Phi_{1/\gamma}(x) = \exp\left(-x^{-1/\gamma}\right)$, $x \geq 0$, with parameter $1/\gamma$. By Fisher-Trippet-Gnedenko theorem (see Bingham et all. [2]) this is equivalent to saying that for any $x \geq 1$,

$$F_t(x) \to P_\gamma(x) \text{ as } t \to \infty, \tag{1.1}$$

where $F_t(x)$ is the excess d.f. over the threshold $t > x_0$ defined by

$$F_t(x) = 1 - \frac{1 - F(xt)}{1 - F(t)}, \; x \geq 1$$

AMS 2000 subject classifications: Primary 62G32, 62G08; secondary 62G05

Keywords and phrases: nonparametric adaptive estimation, extreme values, Hill estimator, probabilities of rare events, high quantiles.

and $P_{\theta}(x) = 1 - x^{-1/\theta}$, $x \geq 1$ is the standard Pareto d.f. with parameter $\theta > 0$. (1.1) suggests using $P_{\gamma}(x)$ with estimated γ (called index of regular variation), as an approximation of $F_t(x)$ for a given x and large t. However, (1.1) can be misleading in cases when the convergence to the limit distribution is too slow. This is easily seen by inspecting the trajectories of the *Hill estimator* (1.2) computed from samples drawn from the loggamma distribution $F(x)$, see Figure 1. It is sometimes called the Hill horror plot, because of the important discrepancy between the Hill estimator and the estimated parameter γ, even for very large sample sizes (see Embrechts et all. [7] or Resnick [20]). The explanation lies in the fact that the Hill estimator merely fits a Pareto distribution to the data thereby providing an approximation of the excess d.f. F_t rather than for γ itself. Despite these evidences the prob-

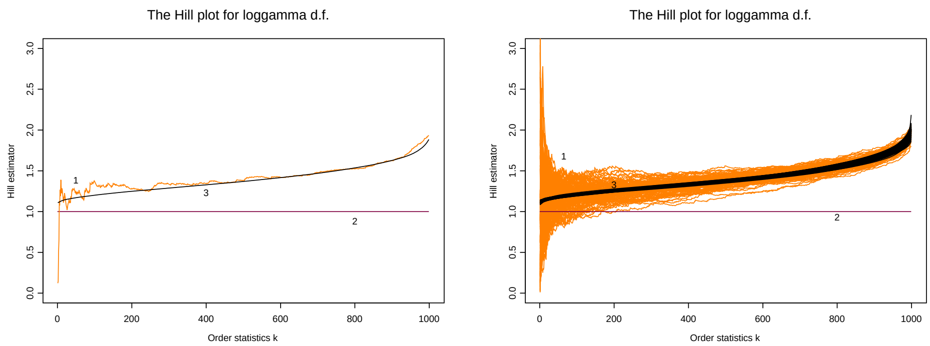

FIG 1. *1 – The Hill estimator $\widehat{h}_{n,k}, k = 1, ..., n$, for loggamma d.f. with rate parameter 1 and shape parameter 2. 2 – Index of regular variation $\gamma = 1$ which is expected to be estimated. 3 – The fitted Pareto parameter $\theta_t(F)$ computed from the approximation formulas (4.4), (4.8). Left: 1 realization; Right: 100 realizations.*

lem of estimating the excess d.f. F_t regardless of the limit P_{γ} is less studied in the literature.

The goal of the present paper is twofold. First of all, we shall consider the problem of recovering the excess d.f. F_t from the data $X_1, ..., X_n$ directly, and secondly, we shall propose an adaptive procedure of the choice of the location of the tail t. Motivated by (1.1), we assume that for large values of $t \geq x_0$ the excess d.f. F_t can be approximated by a Pareto law P_{θ_t} with some index $\theta_t > 0$ possibly depending on the location t and generally different from γ. The statistical problem is that of recovering F_t by constructing a family of estimators $\widehat{\theta}_{n,t}$, $t \geq x_0$ of the parameters θ_t, $t \geq x_0$ and proposing an adaptive rule for choosing the location threshold t.

Some consequences of main results of the paper are formulated below. Let $X_{n,1} > X_{n,2} > ... > X_{n,n}$ be the order statistics pertaining to $X_1, ..., X_n$ and $\widehat{h}_{n,k}$, $k = 2, ..., n$ be the family of Hill estimators, where

$$\widehat{h}_{n,k} = \frac{1}{k-1} \sum_{i=1}^{k-1} \log \frac{X_{n,i}}{X_{n,k}}, \tag{1.2}$$

see Hill [14]. Denote $\widehat{n}_t = \sum_{i=1}^{n} 1\left(X_{n,i} > t\right)$ and $\widehat{\theta}_{n,t} = \widehat{h}_{n,\widehat{n}_t+1}$, where $\widehat{\theta}_{n,t} = 0$ if $\widehat{n}_t = 0$.

The discrepancy between two equivalent probability laws P and Q is measured by the Kullback-Leibler divergence $\mathcal{K}\left(P, Q\right) = \int \log \frac{dP}{dQ} dP$ and by the χ^2-divergence $\chi^2\left(P, Q\right) = \int \frac{dP}{dQ} dP - 1$. For any $t \geq x_0$ the best approximation of the excess d.f. F_t is defined by looking for the "closest" element in the set of Pareto distributions. Let $\theta_t\left(F\right) = \arg\min_{\theta>0} \mathcal{K}\left(F_t, P_\theta\right)$ be the minimum Kullback-Leibler divergence Pareto parameter, called in the sequel for short *fitted Pareto index.* Assume that F admits an accompanying Pareto tail, which is expressed by the condition $\chi^2\left(F_t, P_{\theta_t(F)}\right) \to 0$ as $t \to \infty$. This condition is not very restrictive and in fact defines a class of d.f.'s which can be related to those defined in Hall and Welsh [12] and Drees [6].

According to our Theorem 4.4

$$\mathcal{K}\left(F_{\tau_n}, P_{\widehat{\theta}_{n,\tau_n}}\right) = O_{\mathbf{P}_F}\left(\frac{\log n}{n\left(1-F\left(\tau_n\right)\right)}\right) \text{ as } n\to\infty,$$

for any sequence $\{\tau_n\}$ obeying

$$\chi^2\left(F_{\tau_n}, P_{\theta_{\tau_n}(F)}\right) = O\left(\frac{\log n}{n\left(1-F\left(\tau_n\right)\right)}\right) \to 0 \text{ as } n\to\infty. \tag{1.3}$$

Here and thereafter $\mathbf{P}_F$ denotes the probability measure corresponding to the i.i.d. observations $X_1, ..., X_n$ with d.f. F. The sequence $\{\tau_n\}$ in the definition of the estimator $\widehat{\theta}_{n,\tau_n}$ being, generally, unknown, we give an automatic selection rule $\widehat{k}_n$ (Section 3) such that the adaptive estimator $\widehat{\theta}_n = \widehat{h}_{n,\widehat{k}_n}$ satisfies

$$\mathcal{K}\left(F_{\tau_n}, P_{\widehat{\theta}_n}\right) = O_{\mathbf{P}_F}\left(\frac{\log n}{n\left(1-F\left(\tau_n\right)\right)}\right) \text{ as } n\to\infty, \tag{1.4}$$

for any sequence of locations $\{\tau_n\}$ obeying (1.3), see Theorem 4.10. According to (1.4) the adaptive estimator $P_{\widehat{\theta}_n}$, constructed without any knowledge of the location τ_n, provides the same rate of convergence as the non-adaptive estimator $\widehat{\theta}_{n,\tau_n}$ which is constructed under the additional information that τ_n is known. From the results in Hall and Welsh [12] and Drees [6] it follows that the estimators $\widehat{\theta}_{n,\tau_n}$ and $\widehat{\theta}_n$ attain optimal or suboptimal rate in some classes of functions (see Section 5 for details).

Many results on the adaptive choice of the number of upper statistics to be involved in the estimation demand additional information on the structure of the unknown d.f. F. The adaptive procedure proposed in the paper doesn't require prior knowledge on the unknown d.f. F. In particular it applies for an arbitrary d.f. whose tail can be approximated by a Pareto law as described before. Such an approximation is plausible even for d.f.'s which are not regularly varying at infinity.

The brief outline of paper: In Section 2 we construct the local likelihood estimators. Section 3 presents the adaptive procedure for selecting the threshold t. Main results of the paper are presented in Section 4. Examples of computing the optimal rates of convergence are given in Section 5. In Sections 7 and 8 we prove exponential type bounds for the likelihood ratio used in the proofs of our main results and necessary auxiliary statements. We shall illustrate the performance of our results on some artificial data sets in Section 6.

2. Construction of the estimators. Let $\mathcal{F}$ be the set of all d.f. F having support on the interval $[x_0, \infty)$ with $x_0 \geq 0$, and admitting a strictly positive density f_F w.r.t. Lebesgue measure. For any $t \geq x_0$ define the *excess d.f.* over the threshold t as

$$F_t(x) = 1 - \frac{1 - F(tx)}{1 - F(t)}, \ x \geq 1. \tag{2.1}$$

It is easy to see that

$$F_t(x) = 1 - \exp\left(- \int_t^{tx} \frac{du}{u\alpha_F(u)} \right),$$

where

$$\alpha_F(u) = \frac{1}{u\lambda_F(u)}, \ u > x_0 \tag{2.2}$$

and $\lambda_F(u) = \frac{f_F(u)}{1-F(u)}, u > x_0$ is the hazard rate. F_t admits the density

$$f_{F_t}(x) = \frac{tf_F(tx)}{1 - F(t)} = \frac{1}{x\alpha_F(tx)} \exp\left(- \int_t^{tx} \frac{du}{u\alpha_F(u)} \right), \ x \geq 1. \tag{2.3}$$

Note that, according to the von Mises theorem, if there exists a constant $\alpha > 0$ such that $\alpha_F(x) \to \alpha$ as $x \to \infty$, then F is regularly varying with the index of regular variation α, see Beirlant et all. [1].

Recall that given $X_{n,k+1} = t$ the observations $X_{n,1}/t, ..., X_{n,k}/t$ are the order statistics of an i.i.d. sequence with common density f_{F_t} (see i.e. Reiss [19]). Motivated by this we define the *local* log-*likelihood function*

$$L_{n,t}(F) = \sum_{i:X_i>t} \log f_{F_t}(X_i/t). \tag{2.4}$$

Let $\mathcal{K}(\theta', \theta) = \mathcal{K}(P_{\theta'}, P_\theta)$ be the Kullback-Leibler divergence between P_θ and $P_{\theta'}$,

$$\mathcal{K}(\theta', \theta) = \int \log \frac{dP_{\theta'}}{dP_\theta} dP_{\theta'} = G\left(\frac{\theta'}{\theta} - 1\right), \ \theta', \theta > 0, \tag{2.5}$$

where $G(x) = x - \log(1+x)$. We extend this definition by setting $\mathcal{K}(\theta', \theta) = \infty$ if at least one of $\theta' = 0$ or $\theta = 0$ holds. Lemma 8.1 implies

$$\mathcal{K}(\theta_1, \theta_2) \asymp \left(\frac{\theta_1}{\theta_2} - 1\right)^2 \quad \text{as} \quad \frac{\theta_1}{\theta_2} - 1 \to 0$$

2.1. *Pareto type tails .* Let $\mathcal{F}_t$ be the set of functions $F \in \mathcal{F}$ satisfying $\alpha_F(x) = \theta$, for $x \in (t, \infty)$, where $\theta > 0$ and $t \geq x_0$. If $F \in \mathcal{F}_t$, then the d.f. F_t is exactly Pareto d.f. P_θ. Maximization of the local log-likelihood (2.4) over $\mathcal{F}_t$ gives the maximum local quasi-likelihood estimator

$$\widehat{\theta}_{n,t} = \frac{1}{\widehat{n}_t} \sum_{i:X_i \in (t,\infty)} \log \frac{X_i}{t}, \tag{2.6}$$

where $\widehat{n}_t = \sum_{i=1}^n 1(X_i > t)$ denotes the number of observations in the interval (t, ∞). Here and in the sequel the indeterminacy $0/0$ arising in the definition of the estimators is understood as 0, that is, for $t \geq X_{n,1}$ the estimator $\widehat{\theta}_{n,t}$ is defined to be 0. Although $\widehat{\theta}_{n,t}$ is not exactly the Hill estimator, it is closely related. In fact, if $t = X_{n,k}$, where $X_{n,1} \geq ... \geq X_{n,n}$ are the order statistics pertaining to $X_1, ..., X_n$, and $2 \leq k \leq n$, then $\widehat{\theta}_{n,t} = \widehat{\theta}_{n,X_{n,k}}$ coincides with the Hill estimator $\widehat{h}_{n,k}$, see Hill [14].

Let $L_{n,t}(\theta',\theta) = L_{n,t}(P_{\theta'}) - L_{n,t}(P_\theta)$ be the log of the local likelihood ratio of $P_{\theta'}$ w.r.t. P_θ. By elementary calculation one can see that

$$L_{n,t}\left(\widehat{\theta}_{n,t},\theta\right) = \widehat{n}_t \mathcal{K}\left(\widehat{\theta}_{n,t},\theta\right). \tag{2.7}$$

2.2. *Pareto change point type tails.* Let $\mathcal{F}_{t,\tau}$ be the set of functions $F \in \mathcal{F}$ having the change point structure: $\alpha_F(x) = \theta_1$, for $x \in [t,\tau)$, $\alpha_F(x) = \theta_2$, for $x \in [\tau,\infty)$, where $\theta_1,\theta_2 > 0$ and $1 \le t \le \tau < \infty$. Of course $\mathcal{F}_t \subset \mathcal{F}_{t,\tau}$. If $F \in \mathcal{F}_{t,\tau}$, then the d.f. F_t coincides with the Pareto change point d.f.

$$P_{\theta_1,\theta_2,\tau/t}(x) = 1 - \exp\left(\int_1^x \frac{du}{\alpha'(u)\,u}\right),$$

where $\alpha'(x) = \theta_1$, for $x \in [1,\tau/t)$, $\alpha'(x) = \theta_2$, for $x \in [\tau/t,\infty)$. For given $t \le X_{n,1}$ and $\tau \ge 1$ maximization of the local likelihood (2.4) over $\mathcal{F}_{t,\tau}$ gives the maximum likelihood estimator $\left(\widehat{\theta}_{n,t,\tau},\widehat{\theta}_{n,\tau}\right)$, where

$$\widehat{\theta}_{n,t,\tau} = \frac{\widehat{n}_t\widehat{\theta}_{n,t} - \widehat{n}_\tau\widehat{\theta}_{n,\tau}}{\widehat{n}_{t,\tau}}$$

and $\widehat{n}_{t,\tau} = \widehat{n}_t - \widehat{n}_\tau = \sum_{i=1}^n 1(t < X_i \le \tau)$ is the number of observations in the interval $(t,\tau]$. Recall that $0/0 = 0$ so that for $t \ge X_{n,1}$ the estimator $\widehat{\theta}_{n,t,\tau}$ is defined to be 0.

Denote by $L_{n,t}(\theta_1,\theta_2,\tau,\theta) = L_{n,t}(P_{\theta_1,\theta_2,t,\tau}) - L_{n,t}(P_\theta)$ the local log-likelihood *ratio* corresponding to Pareto change point model $P_{\theta_1,\theta_2,\tau/t}(x)$ with respect to the Pareto model P_θ. By straightforward calculations it is verified that

$$L_{n,t}\left(\widehat{\theta}_{n,t,\tau},\widehat{\theta}_{n,\tau},\tau,\theta\right) = \widehat{n}_{t,\tau}\mathcal{K}\left(\widehat{\theta}_{n,t,\tau},\theta\right) + \widehat{n}_\tau\mathcal{K}\left(\widehat{\theta}_{n,\tau},\theta\right). \tag{2.8}$$

3. Adaptive selection of the location of the tail. Several procedures have been proposed in the literature for the choice of the number of upper statistics to be used in the estimation of the index of regular variation. We refer to Beirlant et all. [1], and to the references therein ([13], [11], [5], [3], [8], [15]). However one should note that most of these procedures require some prior knowledge on the d.f. F.

To illustrate the problem let us recall the main result in Hall and Welsh [12] (see also Drees [6]). Let F be a d.f. with density

$$f_F(x) = d\alpha x^{-(\alpha+1)}(1 + r(x)), \ |r(x)| \le Ax^{-\alpha\rho}, \ \ x \ge 0, \tag{3.1}$$

where $|\alpha - \alpha_0| \le \varepsilon, \ |d - d_0| \le \varepsilon$ and $\alpha_0, \, d_0, \, \varepsilon, \, \rho, \, A > 0$. It is proved that the optimal rate of convergence that can be achieved for estimating $\alpha = 1/\beta$ is $n^{-\rho/(2\rho+1)}$. This optimal rate is attained for the *Hill estimator* $\widehat{h}_{n,k_n}$ with the choice $k_n \sim n^{2\rho/(2\rho+1)}$ depending on ρ. An adaptive estimator can be constructed by estimating ρ and implementing this estimate into the optimal k_n. This approach requires to know in advance the class of distributions F, or generally this information is not available in practice. It is also too conservative in the sense that it is oriented to the worst case in the given class but it may happen that particular distributions have nicer properties.

In this paper we will give a selection procedure which is distribution free and attains exactly or nearly optimal rates for each particular law F in contrast to minimax estimation which is oriented to the worst case in a given class of functions. These kind of results are usually related to the so called *oracle inequalities* (Donoho and Jonstone [4]).

The selection rule of the location of the tail τ which we propose is based on the stagewise lack-of-fit testing for the Pareto distribution (see also Grama and Spokoiny [9]). It can be compared with the adaptive procedures for selecting the bandwidth in nonparametric pointwise function estimation, see

Lepski [16], Lepski and Spokoiny [17]. Drees and Kaufmann [5] give a variant of the latter adapted to the tail index estimation. A stagewise procedure for testing Pareto d.f. has been proposed and its performance analyzed in Hall and Welsh [13], where it was shown that the choice based on the detected of lack-of-fit point introduce a significant bias. Our procedure differs from these approaches in that the point of lack-of-fit it taken as a pilot for the choice of k.

3.1. *The lack-of-fit test.* Denote $[a]$ the integer part of a. Let K_n, $n = 1, 2, ...$ be a sequence of natural numbers satisfying $K_n \leq n$ and $\lim_{n\to\infty} K_n = \infty$. Consider the uniform grid $\{r_1, ..., r_{K_n}\}$, where $r_i = r_i(n) = [in/K_n]$, $i = 1, ..., K_n$. Let k_0 be a natural number which is much smaller than n.

We shall choose the location of the tail of F in the random set $\{X_{n,k} : k = 1, ..., n\}$ and therefore the problem reduces to the choice of the natural number k. We shall proceed by local change-point detection, which consists in consecutive testing for the hypothesis H^0_{n,r_m} that conditionally on $X_{n,r_m+1} = s$ the observations $X_{n,1}/s, ..., X_{n,r_m}/s$ are the order statistics of an i.i.d. sample with a Pareto d.f. P_θ against the alternative H^1_{n,r_m} that conditionally on $X_{n,r_m+1} = s$ the observations $X_{n,1}/s, ..., X_{n,r_m}/s$ are the order statistics of a i.i.d. sample with a Pareto change-point d.f. $P_{\theta_1,\theta_2,\tau/s}$, for all $m = r_{k_0}, ..., r_{K_n}$.

For testing H^0_{n,r_m} against H^1_{n,r_m} we shall make use of the likelihood ratio statistic $T_n(t,\tau)$ which is defined by

$$T_n(t,\tau) = \sup_{F\in\mathcal{F}_t^{cp}} L_{n,t}(F) - \sup_{F\in\mathcal{F}_t} L_{n,t}(F) = L_{n,t}\left(\widehat{\theta}_{n,t,\tau}, \widehat{\theta}_{n,\tau}, \tau, \widehat{\theta}_{n,t}\right), \tag{3.2}$$

for $x_0 \leq t \leq \tau$. Taking into account (2.8) one gets

$$T_n(t,\tau) = T_n^{(1)}(t,\tau) + T_n^{(2)}(t,\tau), \; t < \tau, \tag{3.3}$$

where

$$T_n^{(1)}(t,\tau) = \widehat{n}_{t,\tau}\mathcal{K}\left(\widehat{\theta}_{n,t,\tau}, \widehat{\theta}_{n,t}\right), \; T_n^{(2)}(t,\tau) = \widehat{n}_{\tau}\mathcal{K}\left(\widehat{\theta}_{n,\tau}, \widehat{\theta}_{n,t}\right).$$

For each m and $k \leq m$ consider the test statistics

$$T_{n,m} = \max_{\rho m \leq k \leq (1-\delta)m} T_{n,m,k}, \; T_{n,m,k} = T_{n,m,k}^{(1)} + T_{n,m,k}^{(2)}, \tag{3.4}$$

where

$$T_{n,m,k}^{(i)} = T_n^{(i)}(X_{n,m}, X_{n,k}), \; i = 1, 2$$

and ρ and δ are constants satisfying $0 < \rho, \delta \leq \frac{1}{3}$. We shall suppose that δ is so large that $(1-\delta) r_i \leq r_{i-1}$, for all $i = k_0, ..., K_n$. Actually this condition is satisfied for any given $\delta > 0$ when n becomes sufficiently large. We shall also assume that $\rho r_{k_0} \geq r_1$.

The hypothesis H^0_{n,r_m} will be rejected if $T_{n,r_m} > \mathfrak{z}_n$, for some critical value $\mathfrak{z}_n = \mu \log n$, where μ is a positive constant.

3.2. *The adaptive procedure.* At this stage the required parameters are the number of the points on the grid K_n, the starting point k_0, two numbers ρ and δ which determine the size of the testing window and the critical value $\mathfrak{z}_n$.

The procedure of the adaptive choice of the value $\widehat{k}_n$ reads as follows.

INITIALIZE Set $i = k_0$.

STEP 1 Compute the test statistic T_{n,r_i} by (3.4) and (3.3).

STEP 2 If $i \le K_n$ and $T_{n,r_i} \le \mathfrak{z}_n$, increase i by 1 and repeat the procedure from Step 1. If $i \le K_n$ and $T_{n,r_i} > \mathfrak{z}_n$, define

$$\widehat{k}_n = \arg \max_{\rho r_i \le k \le (1-\delta) r_i,} T^{(2)}_{n,r_i,k} \tag{3.5}$$

and exit the procedure. If $i > K_n$ we define $\widehat{k}_n = n$ and exit the procedure.

The described procedure is equivalent to defining the adaptive value

$$\widehat{k}_n = \arg \max_{\rho \widehat{m}_n \le k \le (1-\delta) \widehat{m}_n,} T^{(2)}_{n,\widehat{m}_n,k}, \tag{3.6}$$

where

$$\widehat{m}_n = \min \{ r_i : T_{n,r_i} > \mathfrak{z}_n,\ i = k_0, ..., K_n \},$$

with the convention $\min \varnothing = r_{K_n}$. The adaptive location of the tail τ is then defined by $\widehat{\tau}_n = X_{n,\widehat{k}_n}$ and the adaptive estimator is set to

$$\widehat{\theta}_n = \widehat{h}_{n,\widehat{k}_n} \equiv \widehat{\theta}_{n,\widehat{\tau}_n}.$$

Remark 3.1 In the case of Pareto observations the test statistics (3.3) and (3.4) do not depend on the parameter of the Pareto law. For practical implementations, this suggests to compute the critical values $\mathfrak{z}_n$ by Monte-Carlo simulations from the homogeneous model with i.i.d. standard Pareto observations $X_i,\ i = 1, ..., n$. Our simulations show that the performance of the proposed adaptive procedure depends to some extent on the parameter ρ. In turn, parameters δ, k_0 and K_n influence less the procedure. The choice of parameters of the procedure is discussed in Section 6.

4. Main results .

Recall that $\widehat{n}_t$ the number of observations in the interval (t, ∞). Let $n_t = n (1 - F (t))$ be the expected number of observations in the same interval.

Let $X_{n,n+1} = x_0$ and $r_{K_n+1} = n+1$. Introduce the notation $\widetilde{n}_t = \min\{r : X_{n,r} \leq t,\ r = r_1, ..., r_{K_n+1}\}$, for $t \geq x_0$. Consider the family of estimators $\widetilde{\theta}_{n,t} = \widehat{h}_{n,\widetilde{n}_t} \equiv \widehat{\theta}_{n,X_{n,\widetilde{n}_t}}$, $t \geq x_0$, which we shall call *weak estimators,* where $\widehat{h}_{n,k}$ and $\widehat{\theta}_{n,t}$ are defined by (1.2) and (2.6) respectively. With $K_n = n$ we have $\widetilde{n}_t = \widehat{n}_t$ and $\widetilde{\theta}_{n,t} = \widehat{\theta}_{n,t}$, therefore all the results presented below concerning the weak estimators $\widetilde{\theta}_{n,t}$ hold true also for $\widehat{\theta}_{n,t}$.

For any equivalent probability measures P and Q denote by $\mathcal{K}(P,Q) = \mathbf{E}_P \log \frac{dP}{dQ}$ the Kullback-Leibler divergence. Additionally consider the χ^2-divergence $\chi^2(P,D) = \int \frac{dP}{dQ} dP - 1$. A simple application of the Jensen's inequality shows that $0 \leq \mathcal{K}(P,Q) \leq \log\left(1 + \chi^2(P,Q)\right)$.

We shall measure the discrepancy between two possible values $\theta_1 > 0$ and $\theta_2 > 0$ of the Pareto index in terms of the Kullback-Leibler divergence $\mathcal{K}(\theta_1, \theta_2)$ between two Pareto measures, see (2.5).

Thereafter $\mathbf{P}_F$ and $\mathbf{E}_F$ denote the probability and the expectation pertaining to the i.i.d. observations $X_1, ..., X_n$ with common d.f. F.

4.1. *Rates of convergence of nonadaptive estimators.* We say that the d.f. F admits an accompanying Pareto tail with tail index function θ_t, $t \geq x_0$ if for any $t \geq x_0$ there exists an index $\theta_t > 0$ such that θ_t is a continuous function of t and

$$\lim_{t \to \infty} \chi^2\left(F_t, P_{\theta_t}\right) = 0. \tag{4.1}$$

This definition can be viewed as an extension of the regular variation condition (1.1). The class of d.f. satisfying (4.1) is very large. Many d.f.'s are such that F_t is close in a given sense to a Pareto d.f. P_{θ_t} as $t \to \infty$, for some parameter $\theta_t > 0$ depending on t. This is the case of d.f.'s satisfying the Hall condition (3.1), log-gamma and Pareto d.f.'s with logarithmic type perturbations. We refer to Section 5, where θ_t is explicitly computed for these examples. The class of distributions defined by (4.1) includes d.f.'s which are not regularly varying. Examples are the normal and exponential d.f.'s with some $\theta_t \to 0$ as $t \to \infty$.

It is easy to see that if the d.f. F admits an accompanying Pareto tail with tail index function θ_t, $t \geq x_0$, then there exists a sequence $\{\tau_n\}$ such that

$$\chi^2 \left(F_{\tau_n}, P_{\theta_{\tau_n}} \right) = O\left(\frac{\log n}{n\left(1 - F\left(\tau_n\right)\right)} \right) \to 0 \text{ as } n \to \infty. \tag{4.2}$$

For the sake of brevity, a sequence of locations $\{\tau_n\}$ satisfying (4.2) is said to be *admissible.*

Theorem 4.1 *Assume that the d.f. F admits an accompanying Pareto tail with tail index function θ_t, $t \geq x_0$. Then, for any admissible sequence of locations τ_n, $n = 1, 2, ...,$*

$$\mathcal{K}\left(\widetilde{\theta}_{n,\tau_n}, \theta_{\tau_n} \right) = O_{\mathbf{P}_F}\left(\frac{\log n}{n\left(1 - F\left(\tau_n\right)\right)} \right) \text{ as } n \to \infty.$$

Here and in the sequel the constant in $O_{\mathbf{P}_F}$ depends only on the constant in O in (4.2). This theorem is an immediate consequence of the more general Theorem 4.5 formulated below.

Corollary 4.2 *Assume that F admits an accompanying Pareto tail with constant tail index function $\theta_t = \gamma$, $t \geq x_0$. Then by Theorem 4.1,*

$$\mathcal{K}\left(\widetilde{\theta}_{n,\tau_n}, \gamma \right) = O_{\mathbf{P}_F}\left(\frac{\log n}{n\left(1 - F\left(\tau_n\right)\right)} \right) \text{ as } n \to \infty, \tag{4.3}$$

with τ_n satisfying (4.2).

Let $\mathcal{P} = \{P_\theta : \theta > 0\}$ be the set of Pareto d.f.'s and $\mathcal{K}(Q,P) = \int \log \frac{dQ}{dP} dQ$ be the Kullback-Leibler entropy between Q and P, $P \ll Q$. For any $t \geq x_0$ we "project" F_t on the set $\mathcal{P}$ by choosing the closest element to F_t in $\mathcal{P}$, say $P_{\theta_t(F)}$, where

$$\theta_t(F) = \arg\min_{\theta>0} \mathcal{K}(F_t, P_\theta).$$

The parameter $\theta_t(F)$ will be called in the sequel *fitted Pareto index.* It can be easily computed and has the following explicit expression (see Figure 2 for a graphical representation):

$$\theta_t(F) = \int_1^\infty \log x F_t(dx) = \int_t^\infty \log \frac{x}{t} \frac{F(dx)}{1 - F(t)}, \ t \geq x_0. \tag{4.4}$$

The fitted Pareto index $\theta_t(F)$ is well defined for many d.f.'s F which are not necessarily regularly varying at ∞. However, if, in addition, F is regularly varying at ∞ with index of regular variation γ, then it is easy to verify that $\theta_t(F) \to \gamma$ as $t \to \infty$.

Corollary 4.3 *Assume that F admits an accompanying Pareto tail with tail index function $\theta_t = \theta_t(F)$, $t \geq x_0$. Then according to Theorem 4.1*

$$\mathcal{K}\left(\widetilde{\theta}_{n,\tau_n}, \theta_{\tau_n}(F)\right) = O_{\mathbf{P}_F}\left(\frac{\log n}{n(1 - F(\tau_n))}\right) \ \text{as } n \to \infty, \tag{4.5}$$

where τ_n satisfies (4.2).

Corollaries 4.2 and 4.3 can be compared with the consistency results for the Hill estimator established in Mason [18] (see also Hall [10]). Recall the main result of Mason [18]. If F is regularly varying with index of regular variation γ and k_n satisfies $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$, then

$$\widehat{h}_{n,k_n} \overset{\mathbf{P}_F}{\to} \gamma \text{ as } n \to \infty. \tag{4.6}$$

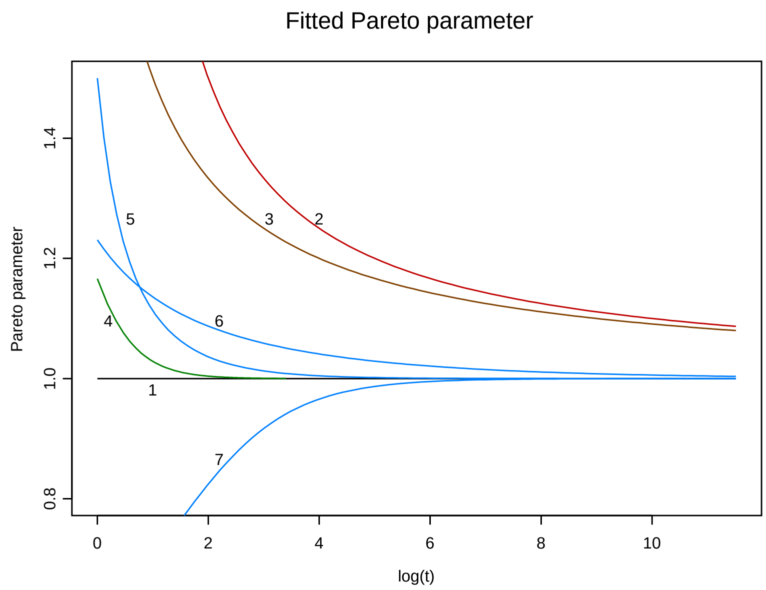

Fig 2. *Fitted Pareto index $\theta_t(F)$: 1 – Pareto d.f.; 2–log-perturbed Pareto d.f. (5.3); 3 – loggamma d.f.; 4 – Cauchy d.f.; 5,6,7 – Hall model (5.4).*

Our Corollary 4.3 improve upon this result by stating that if F admits an accompanying Pareto tail with tail index function $\theta_t = \theta_t(F)$, $t \geq x_0$, then for any τ_n satisfying (4.2),

$$\widehat{\theta}_{n,\tau_n} - \theta_{\tau_n}(F) = \widehat{h}_{n,\widehat{n}_{\tau_n}+1} - \theta_{\tau_n}(F) \overset{\mathbf{P}_F}{\to} 0 \text{ as } n \to \infty, \tag{4.7}$$

where $\widehat{n}_t = \sum_{i=1}^{n} 1(X_i > t)$. A comparison of the precision of the approximations (4.6) and (4.7) is given in Figure 1, where the realizations of the estimator $\widehat{h}_{n,k}$ are plotted as processes in k along with the fitted Pareto index $\theta_t(F)$, for $t = X_{n,1}, ..., X_{n,n}$. The underlying d.f. $F(x)$ is the loggamma one. From these graphs it is seen that for finite sample sizes the Hill estimator $\widehat{h}_{n,k}$ provides a satisfactory approximation of the quantity $\theta_{X_{n,k}}(F)$ while staying far away from the solid straight line corresponding to the parameter of regular variation $\gamma = 1$, except the cases when the fitted Pareto index itself is close to γ. These conclusions are confirmed also by simulation results reported in Figure 3.

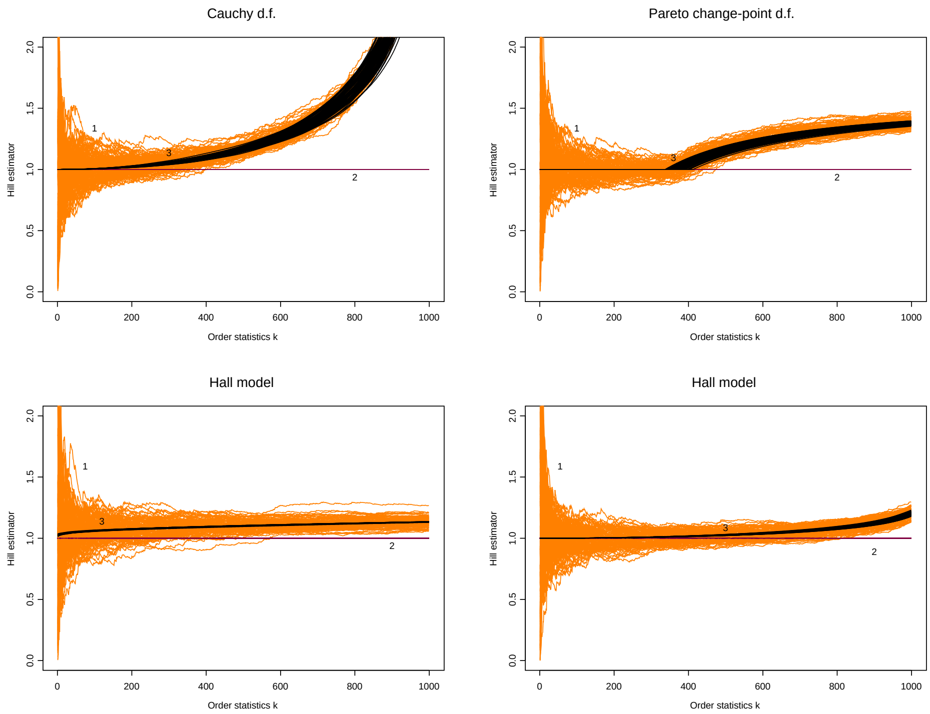

Fig 3. 1 – 100 realizations of the Hill estimator $\widehat{h}_{n,k}, k = 1, ..., n$, *for Cauchy d.f. (top left), Pareto change-point d.f. (top right), Hall model (5.4) (bottom left* $\alpha = 1, \beta = 3, c = 1.8$; *bottom right* $\alpha = 1, \beta = 1.2, c = 1.8$)*; 2 – Index of regular variation* $\gamma = 1$ *which is expected to be estimated. 3 – The fitted Pareto index* $\theta_t(F)$ *computed from (4.4), (4.8).*

Note that the fitted Pareto index $\theta_t\,(F)$ coincides with the mean value of the function α_F [see (2.2)] on the interval $[t, \infty)$ w.r.t. F_t :

$$\theta_t\,(F) = \int_1^\infty \alpha_F\,(tx)\,F_t\,(dx) = \int_t^\infty \alpha_F\,(x)\,\frac{F\,(dx)}{1 - F\,(t)}.$$

For numerical computations of the value $\theta_t\,(F)$ one can use the following approximation formula

$$\theta_{X_{n,k}}(F) \approx \frac{1}{k}\sum_{i=1}^{k} \alpha_F(X_{n,i}) . \tag{4.8}$$

Now we shall present an application of the bound (4.5) to the estimation of the excess d.f. F_{τ_n}.

Theorem 4.4 *Assume that the d.f. F admits an accompanying Pareto tail with tail index function $\theta_t = \theta_t(F)$, $t \geq x_0$. Then, for any admissible sequence of locations $\{\tau_n\}$,*

$$\mathcal{K}\left(F_{\tau_n}, P_{\widetilde{\theta}_{n,\tau_n}}\right) = O_{\mathbf{P}_F}\left(\frac{\log n}{n(1-F(\tau_n))}\right) \text{ as } n \to \infty.$$

Proof. For any $\theta > 0$ and any $s > x_0$,

$$\mathcal{K}(F_s, P_\theta) = \mathcal{K}\left(F_s, P_{\theta_s(F)}\right) + \mathcal{K}(\theta_s(F), \theta) . \tag{4.9}$$

The identity (4.9) follows immediately from the decomposition

$$\mathcal{K}(F_s, P_\theta) = \mathcal{K}\left(F_s, P_{\theta_s(F)}\right) + \int_1^\infty \log \frac{dP_{\theta_s(F)}}{dP_\theta} dF_s$$

and from

$$\int_1^\infty \log \frac{dP_{\theta_s(F)}}{dP_\theta} dF_s = \mathcal{K}(\theta_s(F), \theta) .$$

Using (4.9), one gets

$$\mathcal{K}\left(F_{\tau_n}, P_{\widetilde{\theta}_{n,\tau_n}}\right) = \mathcal{K}\left(F_{\tau_n}, P_{\theta_{\tau_n}(F)}\right) + \mathcal{K}\left(\theta_{\tau_n}(F), \widetilde{\theta}_{n,\tau_n}\right) .$$

Since by Lemma 8.1, $\mathcal{K}(\theta_1, \theta_2) \leq \frac{9}{4}\mathcal{K}(\theta_2, \theta_1)$, the assertion follows from the convergence result (4.5) and from the inequality

$$\mathcal{K}\left(F_{\tau_n}, P_{\theta_{\tau_n}(F)}\right) \leq \log\left(1 + \chi^2\left(F_{\tau_n}, P_{\theta_{\tau_n}(F)}\right)\right) = O\left(\frac{\log n}{n(1-F(\tau_n))}\right)$$

as $n \to \infty$. ■

The previous results are based on the following more general bound which is a simple application of an exponential bound for the maximum of the likelihood ratio.

Theorem 4.5 *Assume that* $\{\tau_n\}$ *is a sequence such that* $\tau_n \geq x_0$ *and* $\lim_{n\to\infty} n(1 - F(\tau_n)) = \infty$. *Then for any sequence* $\{\theta_n\}$ *of positive numbers the weak estimator* $\widetilde{\theta}_{n,\tau_n}$ *satisfies*

$$\mathcal{K}\left(\widetilde{\theta}_{n,\tau_n}, \theta_n\right) = O_{\mathbf{P}_F}\left(\frac{\log n}{n(1 - F(\tau_n))} + \chi^2(F_{\tau_n}, P_{\theta_n})\right) \text{ as } n \to \infty,$$

with an absolute constant in $O_{\mathbf{P}_F}$.

Proof. Letting $t = s = \tau_n$, $\theta = \theta_n$, $y = 4\log n + n_{\tau_n}\chi^2(F_{\tau_n}, P_{\theta_n})$ and using the fact that $\widehat{n}_{\tau_n}\mathcal{K}\left(\widehat{\theta}_{n,\tau_n}, \theta_n\right) \geq \widehat{n}_{\tau_n}\mathcal{K}\left(\widetilde{\theta}_{n,\tau_n}, \theta_n\right)$, by the first inequality of Proposition 7.3 one gets

$$\mathcal{K}\left(\widetilde{\theta}_{n,\tau_n}, \theta_n\right) = O_{\mathbf{P}_F}\left(\frac{\log n}{\widehat{n}_{\tau_n}} + \frac{n_{\tau_n}}{\widehat{n}_{\tau_n}}\chi^2(F_{\tau_n}, P_{\theta_n})\right) \text{ as } n \to \infty.$$

To obtain the requested we use the fact that by Lemma 8.3 it holds $\widehat{n}_{\tau_n} \overset{\mathbf{P}_F}{\asymp} n_{\tau_n}$ as $n \to \infty$, whenever $\lim_{n\to\infty} n_{\tau_n} = \infty$. ■

The rate of convergence $\frac{\log n}{n(1-F(\tau_n))}$ involved in the previous theorems depends on the unknown d.f. F and on the unknown location τ_n. The best possible rate of convergence for a given F is obtained by choosing τ_n from the balance equation (4.2). Explicit calculation of the resulting rates of convergence for some d.f.'s F are given in Section 5.

4.2. *Stability property of the test statistic.* We say that the location t is *accepted* by the testing procedure if for any $r \in \mathcal{R}_n = \{r_{k_0}, ..., r_{K_n}\}$ satisfying

$X_{n,r} \geq t$ it holds $T_{n,r} \leq \mathfrak{z}_n$, where critical values $\mathfrak{z}_n$ are such that

$$\mathfrak{z}_n = \mu \log n \tag{4.10}$$

for some positive constant μ. Set

$$\Omega_{n,t} = \{t \text{ is accepted}\} = \cap_{r \in \mathcal{R}_n,\ X_{n,r} \geq t} \{T_{n,r} \leq \mathfrak{z}_n\}.$$

Theorem 4.6 *Assume that the d.f. F admits an accompanying Pareto tail with tail index function θ_t, $t \geq x_0$ and τ_n, $n = 1, 2, \ldots$ denotes an admissible sequence of locations. Then there exists a finite positive constant μ in (4.10) such that $\mathbf{P}_F$ (τ_n is accepted) $= \mathbf{P}_F(\Omega_{n,\tau_n}) \to 1$ as $n \to \infty$.*

Proof. First note that by Proposition 7.5

$$\mathbf{P}_F \left(\sup_{\tau_n \leq s \leq \tau} T_n(t, \tau) > z \right) \leq 2n^7 \exp(-y/2) + \frac{1}{n} \leq \frac{3}{n},$$

where $y = 16 \log n$ and $z = 2y + 2n_{\tau_n} \chi^2 \left(F_{\tau_n}, P_{\theta_{\tau_n}}\right)$. On the other hand, by (4.2) $z \leq \mathfrak{z}_n$, for some constant μ and n sufficiently large. Consequently $\Omega^c_{n,\tau_n} \subseteq \left\{ \sup_{\tau_n \leq t \leq \tau} T_n(t, \tau) > z \right\}$. This implies $\lim_{n\to\infty} \mathbf{P}_F \left(\Omega^c_{n,\tau_n} \right) = 0$. ∎

Remark 4.7 From the preceding proof it can be easily seen that the constant μ in the definition of the critical value $\mathfrak{z}_n$ depends only on the constant involved in the definition of O in (4.2), say λ. A simple tracking of constants shows that a crude upper bound for μ is $32 + 2\lambda e^{\lambda}$.

4.3. *Rates of convergence of the adaptive estimator.* In the sequel it is assumed that the critical values $\mathfrak{z}_n$ satisfy (4.10). First we compare the performance of the adaptive estimator $\widehat{\theta}_n$ with that of the weak estimator $\widetilde{\theta}_{n,\tau_n}$.

Theorem 4.8 *Assume that the d.f. F admits an accompanying Pareto tail with tail index function θ_t, $t \geq x_0$ and τ_n, $n = 1, 2, ...$ denotes an admissible sequence of locations. Then there exists a finite positive constant μ in (4.10) such that*

$$\mathcal{K}\left(\widehat{\theta}_n, \widetilde{\theta}_{n,\tau_n}\right) = O_{\mathbf{P}_F}\left(\frac{\log n}{n\left(1 - F\left(\tau_n\right)\right)}\right) \text{ as } n \to \infty.$$

Proof. Let $\Omega^*_{n,\tau_n} = \Omega_{n,\tau_n} \cap \left\{T_{n,r_{k_0}} \leq \mathfrak{z}_n\right\}$. Since by Theorem 4.6 $\mathbf{P}_F\left(\Omega_{n,\tau_n}\right) \to 1$ as $n \to \infty$ and by Lemma 8.4 $\mathbf{P}_F\left(X_{n,r_{k_0}} \geq \tau_n\right) \to 1$ as $n \to \infty$, it holds

$$\mathbf{P}_F\left(\Omega^*_{n,\tau_n}\right) \geq \mathbf{P}_F\left(\Omega_{n,\tau_n} \cap \left\{X_{n,r_{k_0}} \geq \tau_n\right\}\right) \to 1 \text{ as } n \to \infty. \tag{4.11}$$

The key point in the following is that by the definition of $\widehat{m}_n$ it holds $\widehat{m}_n > \widetilde{n}_{\tau_n}$ on the set Ω^*_{n,τ_n} (see Section 3.2). Denote by $\widetilde{m}_n$ the natural number in the set $\{r_1, ..., r_{K_n}\}$ preceding $\widehat{m}_n$. Thus $\widetilde{m}_n \geq \widetilde{n}_{\tau_n}$.

We split the further proof into two parts. First we shall compare $\widehat{h}_{n,\widetilde{m}_n}$ and $\widehat{h}_{n,\widetilde{n}_{\tau_n}}$. To this end define the sequence of natural numbers m_i, $i = 0, 1, ...$ such that $m_0 = \widetilde{m}_n$ and m_i is the smallest natural number on the grid $\mathcal{R}_n = \{r_{k_0}, ..., r_{K_n}\}$ exceeding $m_{i-1}/2$ for $i = 1, 2, ...i^*$, where i^* such that $\rho m_{i^*} \leq \widetilde{n}_{\tau_n} \leq (1 - \delta)\, m_{i^*}$. Let $m_{i^*+1} = \widetilde{n}_{\tau_n}$. Since, on the set Ω^*_{n,τ_n},

$$T_{n,k} \leq \mathfrak{z}_n \equiv \mu \log n, \text{ for } k \in \mathcal{R}_n, \ k \leq \widetilde{m}_n, \tag{4.12}$$

by (3.3), with $s = X_{n,m_{i-1}} \leq \tau = X_{n,m_i}$, one gets

$$m_i \mathcal{K}\left(\widehat{h}_{n,m_{i-1}}, \widehat{h}_{n,m_i}\right) \leq T_{n,m_{i-1},m_i} \leq \mu \log n, \ i = 1, ..., i^* + 1,$$

which in turn implies

$$\sum_{i=1}^{i^*} \sqrt{\mathcal{K}\left(\widehat{h}_{n,m_{i-1}}, \widehat{h}_{n,m_i}\right)} \leq \mu^{1/2} \log^{1/2} n \sum_{i=1}^{i^*} m_i^{-1/2}.$$

Taking into account that $m_i \geq m_{i-1}/2$, for $i = 1, ..., i^*$, we obtain

$$\sum_{i=1}^{i^*} m_i^{-1/2} \leq m_{i^*}^{-1/2} \sum_{i=1}^{i^*} 2^{-(i^*-i)/2} \leq 3.5 m_{i^*}^{-1/2}.$$

Since $\widetilde{n}_{\tau_n} \leq m_{i^*}$, by Lemma 8.2, on the set Ω^*_{n,τ_n} it holds

$$\sqrt{\mathcal{K}\left(\widehat{h}_{n,\widetilde{m}_n}, \widehat{h}_{n,\widetilde{n}_{\tau_n}}\right)} \leq \frac{3}{2} \sum_{i=1}^{i^*+1} \sqrt{\mathcal{K}\left(\widehat{h}_{n,m_{i-1}}, \widehat{h}_{n,m_i}\right)} \leq \frac{3 \cdot 4.5}{2} \mu^{1/2} \frac{\log^{1/2} n}{\widetilde{n}_{\tau_n}^{1/2}}. \tag{4.13}$$

Now we shall compare $\widehat{h}_{n,\widetilde{m}_n}$ and $\widehat{h}_{n,\widehat{k}_n}$. Recall that by the definition $\widehat{k}_n$ is a natural number satisfying $\rho \widehat{m}_n \leq \widehat{k}_n \leq (1-\delta)\, \widehat{m}_n \leq \widetilde{m}_n$ (see Section 3.2). Then, on the set Ω^*_{n,τ_n}, (4.12) implies $T_{n,\widetilde{m}_n,\widehat{k}_n} \leq \mathfrak{z}_n$. Since on the same set it holds $\widetilde{m}_n \geq \widetilde{n}_{\tau_n}$, we get

$$\sqrt{\mathcal{K}\left(\widehat{h}_{n,\widehat{k}_n}, \widehat{h}_{n,\widetilde{m}_n}\right)} \leq \frac{\mu^{1/2} \log^{1/2} n}{\widetilde{m}_n^{1/2}} \leq \mu^{1/2} \frac{\log^{1/2} n}{\widetilde{n}_{\tau_n}^{1/2}}. \tag{4.14}$$

Summing (4.13) and (4.14), by Lemma 8.2 it follows that on the set Ω^*_{n,τ_n},

$$\sqrt{\mathcal{K}\left(\widehat{h}_{n,\widehat{k}_n}, \widehat{h}_{n,\widetilde{n}_{\tau_n}}\right)} \leq (c\mu)^{1/2} \frac{\log^{1/2} n}{\widetilde{n}_{\tau_n}^{1/2}}, \tag{4.15}$$

where c is an absolute constant. Taking into account (4.11),

$$\mathbf{P}_F\left(\mathcal{K}\left(\widehat{\theta}_n, \widetilde{\theta}_{n,\tau_n}\right) \leq c\mu \frac{\log n}{\widetilde{n}_{\tau_n}}\right) \to 1 \text{ as } n \to \infty.$$

To get the requested assertion it suffices to replace the random rate of convergence $\widetilde{n}_{\tau_n}$ with the deterministic rate $n_{\tau_n} = n\,(1 - F\,(\tau_n))$ by Lemma 8.3.

■

Combining Theorem 4.8 with Theorem 4.1 one gets the following assertion which compare the adaptive estimator $\widehat{\theta}_n$ with the parameter θ_{τ_n}.

Theorem 4.9 *Assume that the d.f. F admits an accompanying Pareto tail with tail index function θ_t, $t \geq x_0$ and τ_n, $n = 1, 2, ...$ denotes an admissible* *sequence of locations. Then there exists a finite positive constant μ in (4.10) such that*

$$\mathcal{K}\left(\widehat{\theta}_n, \theta_{\tau_n}\right) = O_{\mathbf{P}_F}\left(\frac{\log n}{n\left(1 - F\left(\tau_n\right)\right)}\right) \text{ as } n \to \infty.$$

In particular if condition (4.1) is fulfilled with $\theta_t = \theta_t\left(F\right)$ one gets

$$\mathcal{K}\left(\widehat{\theta}_n, \theta_{\tau_n}\left(F\right)\right) = O_{\mathbf{P}_F}\left(\frac{\log n}{n\left(1 - F\left(\tau_n\right)\right)}\right) \text{ as } n \to \infty. \tag{4.16}$$

Another case of interest is when F is regularly varying with index of regular variation $\gamma > 0$. Assume that condition (4.1) is satisfied with $\theta_t = \gamma$, $t \geq x_0$. Then

$$\mathcal{K}\left(\widehat{\theta}_n, \gamma\right) = O_{\mathbf{P}_F}\left(\frac{\log n}{n\left(1 - F\left(\tau_n\right)\right)}\right) \text{ as } n \to \infty. \tag{4.17}$$

Now we are in position to formulate the result concerning the approximation of the excess d.f. .

Theorem 4.10 *Assume that the d.f. F admits an accompanying Pareto tail with tail index function $\theta_t = \theta_t\left(F\right)$, $t \geq x_0$ and τ_n, $n = 1, 2, ...$ denotes an admissible sequence of locations. Then there exists a finite positive constant μ in (4.10) such that*

$$\mathcal{K}\left(F_{\tau_n}, P_{\widehat{\theta}_n}\right) = O_{\mathbf{P}_F}\left(\frac{\log n}{n\left(1 - F\left(\tau_n\right)\right)}\right) \text{ as } n \to \infty.$$

Proof. The proof is similar to that of Theorem 4.4. The only changes are that $\widehat{\theta}_n$ replaces $\widehat{\theta}_{n,\tau_n}$ and that one uses (4.16) instead of (4.5). ■

5. Computation of the rates of convergence. In this section we shall compute explicitly optimal rates of convergence in two particular cases.

Introduce the distance $\rho_0(x,y) = \max\left\{\left|\log\frac{x}{y}\right|, \left|\frac{1}{x}-\frac{1}{y}\right|\right\},\ x,y>0$. From Proposition 8.6 it follows that the sequence $\{\tau_n\}$ is admissible if there exists a function $t \to \theta_t$ such that

$$\rho_{\tau_n}^2 \equiv \sup_{x\geq\tau_n} \rho_0\left(\alpha_F(x), \theta_{\tau_n}\right)^2 = O\left(\frac{\log n}{n\left(1-F(\tau_n)\right)}\right) \to 0 \text{ as } n\to\infty, \tag{5.1}$$

$$\sup_{m\geq n}\int_1^\infty (1+\log x)^2 x^{r_0} F_{\tau_m}(dx) = O(1) \text{ as } n\to\infty. \tag{5.2}$$

In turn this implies that the conclusions of Section 4 hold true. The optimal rate corresponds to minimal location τ_n satisfying (5.1).

5.1. *Perturbed Pareto model.* Assume that F has the form

$$F(x) = 1 - c_\beta x^{-1/\beta}\log x,\ x\geq x_0\geq e, \tag{5.3}$$

where $\beta\geq\beta_0>0$, x_0 and c_β are chosen such that $F(x)$ is strictly monotone and $F(x_0)=0$. By straightforward calculations $\theta_t(F) = \beta\left(1+\frac{\beta}{\log t}\right)$ and $\alpha_F(x) = \beta\left(1-\frac{\beta}{\log x}\right)^{-1}$. Since $\rho_{\tau_n}\leq\frac{\beta}{\log\tau_n}$ and $1-F(t) = c_\beta t^{-1/\beta}\log t$, for determining nearly optimal location τ_n we get the balance condition

$$\frac{\beta^2}{\log^2\tau_n} = O\left(\frac{\log n}{nc_\beta\tau_n^{-1/\beta}\log\tau_n}\right).$$

With the optimal choice $\tau_n \asymp \frac{n^\beta}{\log^\beta n}$, one gets $\frac{\log n}{n(1-F(\tau_n))} = O\left(\log^{-2} n\right)$ as $n\to\infty$. On the other hand condition (5.2) is satisfied since, by (2.3), $f_{F_{\tau_n}}(x) \leq \frac{1}{\beta}x^{-\frac{1-\varepsilon}{\beta}-1}$, for some $\varepsilon\in(0,1)$. Then according to the results in Section 4, $\mathcal{K}\left(\theta_{\tau_n}(F), \widehat{\theta}_{n,\tau_n}\right)$ and $\mathcal{K}\left(\theta_{\tau_n}(F), \widehat{\theta}_n\right)$ are $O_{\mathbf{P}_F}\left(\log^{-2} n\right)$ as $n\to\infty$. Taking $\theta_t=\beta$, in the same way one shows that $\mathcal{K}\left(\beta, \widehat{\theta}_{n,\tau_n}\right)$ and $\mathcal{K}\left(\beta,\widehat{\theta}_n\right)$ are $O_{\mathbf{P}_F}\left(\log^{-2} n\right)$ as $n\to\infty$. According to the results in Drees [6], Theorem 2.1, the best achievable rate for estimating β in L^2 norm is $\frac{1}{\log n}$, in a certain class of d.f.'s which includes the d.f. F satisfying (5.3) (we refer to Drees [6] for details). Our estimators $\widehat{\theta}_{n,\tau_n}$ and $\widehat{\theta}_n$ attain the same rate.

For the loggamma d.f. we obtain the same rate of convergence since it has essentially the same behavior as the d.f. F defined by (5.3).

5.2. *Hall model.* Assume that F is the of the form

$$F(x) = 1 - c_\beta x^{-1/\beta} - c_\gamma x^{-1/\gamma}, \quad x \geq x_0, \tag{5.4}$$

where $\beta = \gamma + \alpha \geq \beta_0 > 0,\ \alpha, \beta, \gamma > 0$ and x_0, c_β and c_γ are such that $F(x)$ is increasing on $[x_0, \infty)$. Also it is not exactly the model proposed by Hall, we shall call it Hall model. By straightforward calculations $\theta_t(F) = \frac{\beta c_\beta t^{-1/\beta} + \gamma c_\gamma t^{-1/\gamma}}{c_\beta t^{-1/\beta} + c_\gamma t^{-1/\gamma}}$ and $\alpha_F(x) = \frac{c_\beta x^{-1/\beta} + c_\gamma x^{-1/\gamma}}{\beta^{-1} c_\beta x^{-1/\beta} + \gamma^{-1} c_\gamma x^{-1/\gamma}}$. It is easy to check that for all sufficiently large t there are constants k_0, k_1 depending only on β, γ such that $\rho_{\tau_n} = O(k_1 \tau_n^{-\frac{1}{\gamma} + \frac{1}{\beta}})$. Since $1 - F(t) = c_\beta t^{-1/\beta} + c_\gamma t^{-1/\gamma}$, for determining the nearly optimal location τ_n we get the balance condition

$$k_1^2 \tau_n^{-\frac{2}{\gamma} + \frac{2}{\beta}} \leq r_2 \frac{\log n}{n \left(c_\beta \tau_n^{-1/\beta} + c_\gamma \tau_n^{-1/\gamma} \right)}.$$

The optimal choice $\tau_n \asymp \left(\frac{n}{\log n} \right)^{\frac{\beta\gamma}{2\beta - \gamma}}$, implies $\frac{\log n}{n(1 - F(\tau_n))} = O\left(\left(\frac{\log n}{n} \right)^{\frac{2\alpha}{\beta + \alpha}} \right)$ as $n \to \infty$. As in the previous example one can show that (5.2) is satisfied. Then according to the results in Section 4, $\mathcal{K}\left(\theta_{\tau_n}(F), \widehat{\theta}_{n,\tau_n} \right)$ and $\mathcal{K}\left(\theta_{\tau_n}(F), \widehat{\theta}_n \right)$ are $O_{\mathbf{P}_F}\left(\left(\frac{\log n}{n} \right)^{\frac{2\alpha}{\beta + \alpha}} \right)$ as $n \to \infty$. Taking $\theta_t = \beta$ we have that $\mathcal{K}\left(\beta, \widehat{\theta}_{n,\tau_n} \right)$ and $\mathcal{K}\left(\beta, \widehat{\theta}_n \right)$ are $O_{\mathbf{P}_F}\left(\left(\frac{\log n}{n} \right)^{\frac{2\alpha}{\beta + \alpha}} \right)$ as $n \to \infty$. By the results in Hall and Welsh [12] (see also Drees [6], Theorem 2.1, for a more general result), the optimal rate of convergence that can be achieved for estimating $\alpha = 1/\beta$ in L^2 norm is $n^{\rho/(2\rho+1)}$, in the class of d.f. F having the density (3.1). The d.f. F defined by (5.4) satisfies this condition with $\gamma = \beta^{-1}(1+\rho)^{-1}$. Since $\frac{\alpha}{\beta+\alpha} = \frac{2\beta-2\gamma}{2\beta-\gamma} = \frac{\rho}{2\rho+1}$, the estimators $\widehat{\theta}_{n,\tau_n}$ and $\widehat{\theta}_n$ attain this rate for β up to an additional $\log^{\frac{\alpha}{\beta+\alpha}} n$ factor.

6. Numerical results.

6.1. *Choice of the parameters of the adaptive procedure.* An important parameter in the proposed adaptive procedure is the sequence of critical values $\mathfrak{z}_n$. According to Remark 3.1 the test statistic does not depend on the parameter of the Pareto law if the observations follow a Pareto model. Therefore we propose to compute the critical values $\mathfrak{z}_n$ by Monte-Carlo simulations from the homogeneous model with i.i.d. standard Pareto observations X_i, $i = 1, ..., n$.

We simulated 2000 realizations with three different sample sizes $n = 200, 500, 1000$ and with the grid length K_n set to 200. The size of testing windows in $T_{n,m}$ is determined by $\rho = 1/4$ and $\delta = 1/20$. The empirical d.f. of the statistic $T_n = \max_{m=r_1,\ldots,r_{K_k}} T_{n,m}$ has been computed and it was found that in all simulations the critical value $\mathfrak{z}_n = 10$ corresponds to a 99% confidence level. The same critical value $\mathfrak{z}_n = 10$, corresponding to a 99% confidence level, has been found from 2000 realizations with $n = 1000$, $\rho = 1/4$, $\delta = 1/20$ and with different grid lengths $K_n = 100, 200, 300$. Additional simulations show that finite sample properties of the test statistic T_n do depend very little on the parameters ρ and δ. The value $\mathfrak{z}_n = 10$ which approximately corresponds to a 99% confidence level in all cases and the grid length $K_n = 200$ have been retained.

Further simulations show that the finite sample performance of the adap-

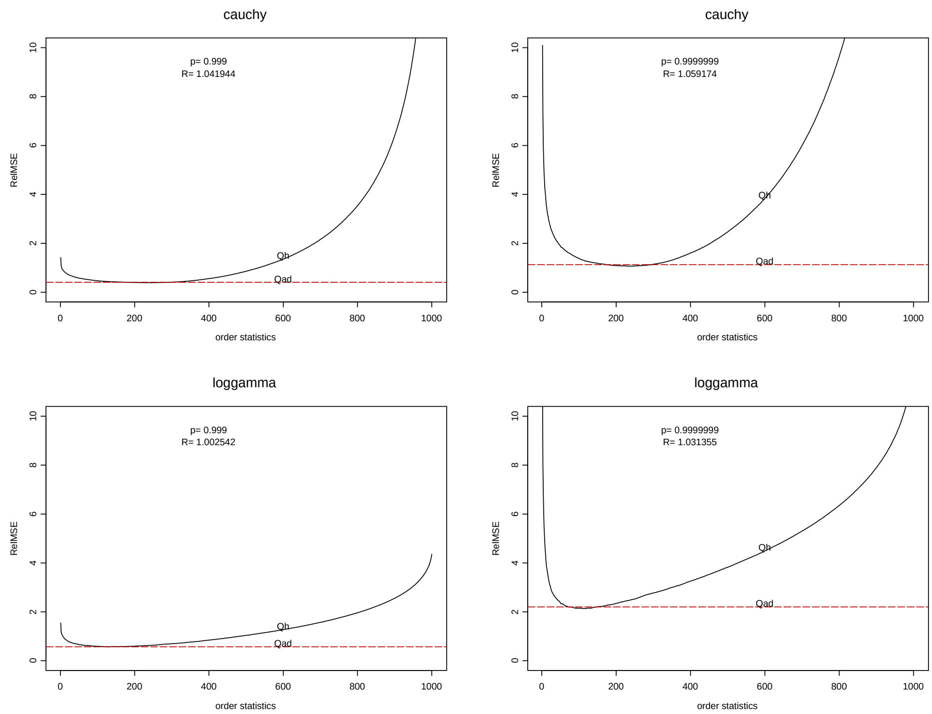

FIG 4. *$Qh = \sigma\left(\widehat{q}_{n,k,p}, q_p\right)$ – RelMSE of the Weissman estimator for $k = 1, ..., n$ and $Qad = \sigma\left(\widehat{q}_{n,p}, q_p\right)$ – RelMSE of the adaptive estimator; $R = r_{n,p}$. Top: Cauchy observations; Bottom: loggamma observations.*

tive estimator depends mainly on the parameter ρ which plays the same role as the bandwidth in the nonparametric kernel density estimation. The choice of ρ, in turn, depends on the class of functions in hands. In the simulations below we fix the following values $\delta = 1/20$, $k_0 = n/20$, $K_n = 200$, $\mathfrak{z}_n = 10$. As to the value of ρ it will be fixed to $1/4$. This choice is motivated by the desire to minimize the relative mean squared error for some given heavy tailed laws. In the simulations below we shall consider the following distributions.

- The positive part of Cauchy d.f. $F(x) = \frac{2}{\pi} \arctan x$, $x \geq 0$.

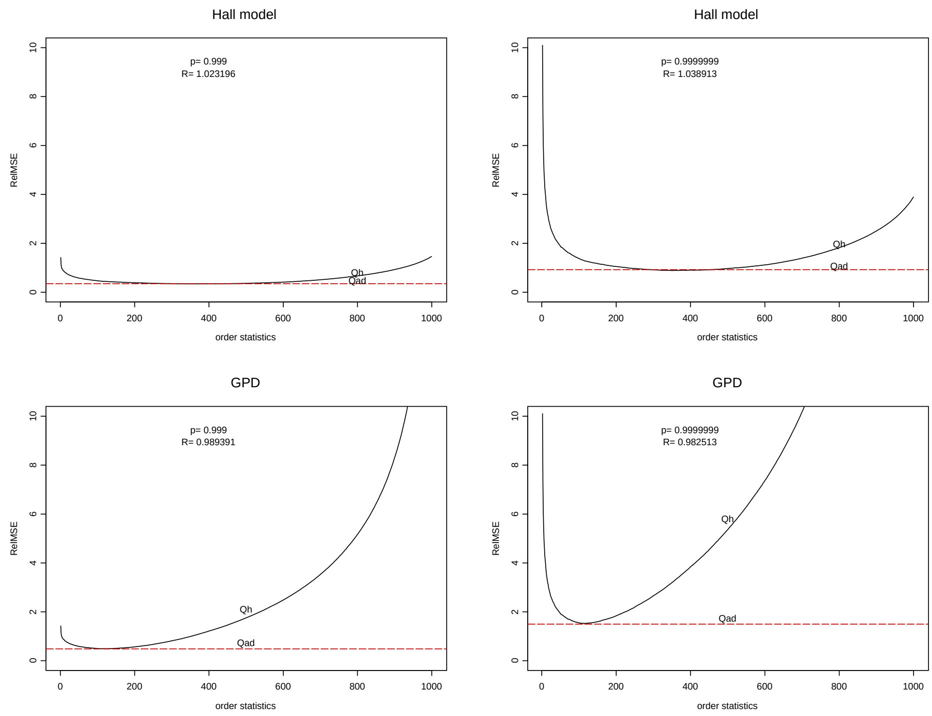

FIG 5. *$Qh = \sigma\left(\widehat{q}_{n,k,p}, q_p\right)$ – RelMSE of the Weissman estimator for $k = 1, ..., n$ and $Qad = \sigma\left(\widehat{q}_{n,p}, q_p\right)$ – RelMSE of the adaptive estimator; $R = r_{n,p}$. Top: Observations from Hall's model; Bottom: GPD observations.*

- Loggamma d.f. $F(x) = G_{1,2}(\log x)$, $x \geq 1$, where $G_{\lambda,\alpha}(x)$, $x \geq 0$ is gamma d.f. with parameters $\lambda, \alpha > 0$.
- log perturbed Pareto d.f. $F(x) = 1 - x^{-1}\log x$, $x \geq x_0 = e$.
- Hall's model $F(x) = 1 - 2x^{-1} + x^{-2.5}$, $x \geq x_0 > 0$, where x_0 is the solution of the equation $2x_0^{-1} - x_0^{-2.5} = 1$.
- GPD $F(x) = 1 - (1+x)^{-1}$, $x \geq 0$.

6.2. *Estimation of extreme quantiles.* We shall demonstrate the performance of the adaptive estimator $\widehat{\theta}_n = \widehat{h}_{n,\widehat{k}_n}$ by presenting the results of a simulation study for estimating extreme quantiles. We consider two oppo-

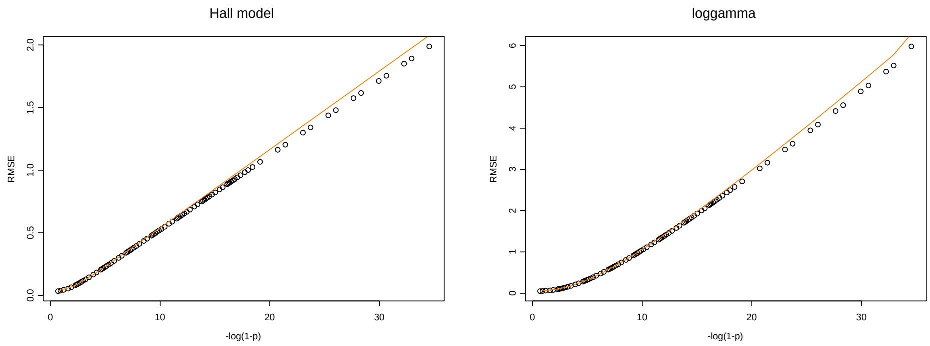

FIG 6. *Minimal error* $\min_k \sigma\left(\widehat{q}_{n,k,p}, q_p\right)$ *(points) and* $\sigma\left(\widehat{q}_{n,p}, q_p\right)$ *(solid line) as functions of* p.

site cases: observations from d.f.'s whose tails are close to Pareto model in the range of big order statistics (such as Cauchy d.f., GPD, some of the Hall models) and observations from d.f.'s whose tails are not well approximated by a Pareto model in the range of the large order statistics at least for samples of reasonable size (such as loggamma d.f. and log perturbed Pareto d.f.). For many d.f.'s our simulations show a behavior in-between the latter two types. We performed 2000 Monte-Carlo simulations of $n = 1000$ observations. The quantiles of F are estimated by solving for $x \geq \tau_n$ in the following approximation formula:

$$1 - P_{\widehat{\theta}_n}\left(\frac{x}{\tau_n}\right) = \left(\frac{x}{\tau_n}\right)^{-1/\widehat{\theta}_n} \approx 1 - F_{\tau_n}\left(\frac{x}{\tau_n}\right) = \frac{1 - F(x)}{1 - F(\tau_n)} = \frac{1-p}{1 - F(\tau_n)}.$$

If $x < \tau_n$ we determine x from the equality $p = F(x)$. The unknown location parameter τ_n has to be replaced with the adaptive value $\widehat{\tau}_n = X_{n,\widehat{k}_n}$ and F with the empirical d.f. $\widehat{F}_n$, which leads to the following adaptive estimate of the quantiles of F :

$$\widehat{q}_{n,p} = \widehat{q}_{n,\widehat{k}_n,p} = \begin{cases} X_{n,[n(1-p)]}, & \text{if } p < 1 - \frac{\widehat{k}_n}{n}, \\ X_{n,\widehat{k}_n} \left(\frac{\widehat{k}_n}{n(1-p)} \right)^{\widehat{\theta}_n}, & \text{otherwise}, \end{cases} \quad p \in (0,1). \tag{6.1}$$

Here and in the sequel $\widehat{q}_{n,k,p}$ denotes the quantile estimator

$$\widehat{q}_{n,k,p} = \begin{cases} X_{n,[n(1-p)]}, & \text{if } p < 1 - \frac{k}{n}, \\ X_{n,k} \left(\frac{k}{n(1-p)} \right)^{\widehat{h}_{n,k}}, & \text{otherwise}, \end{cases} \quad p \in (0,1),\ k = 2, ..., n, \tag{6.2}$$

which combines the sample quantile estimator for low quantiles and the estimator introduced by Weissman [21] for high quantiles.

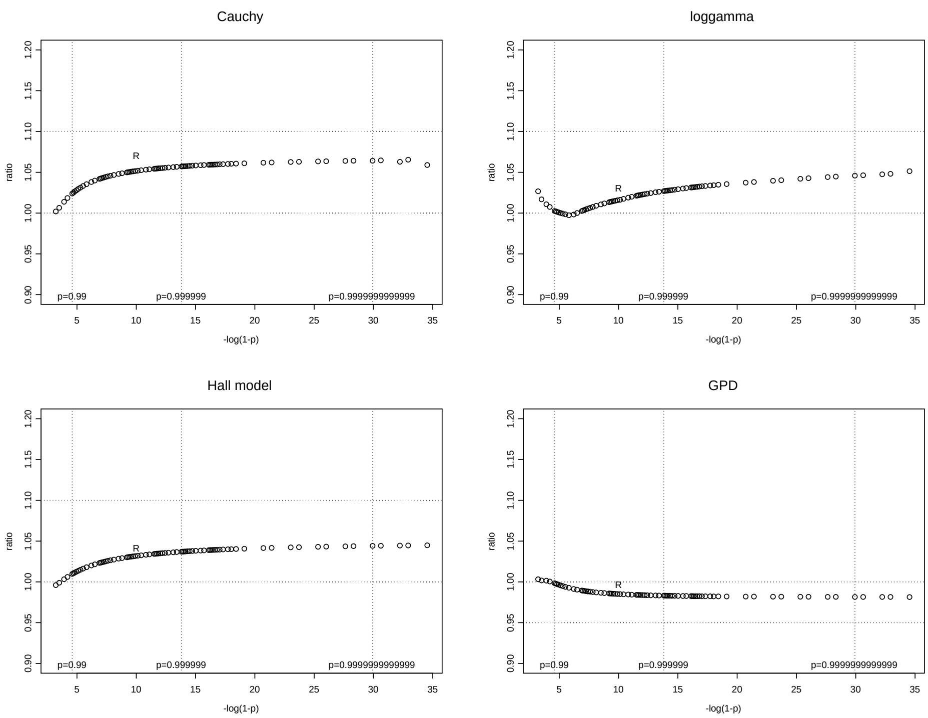

FIG 7. *The ratio $r_{n,p}$ as a function of p.*

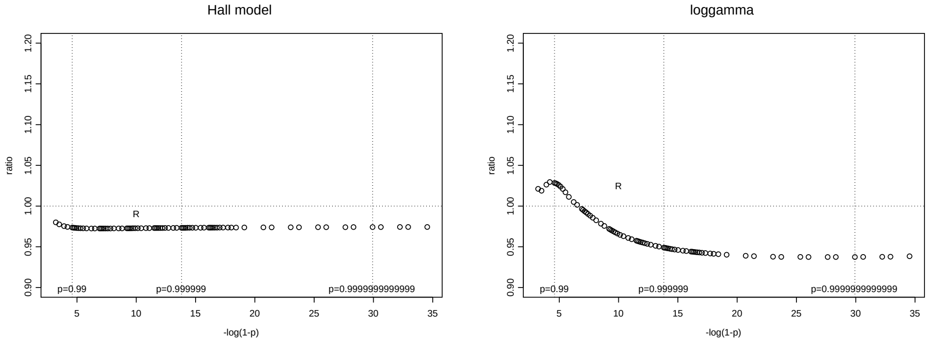

FIG 8. *The ratio $r_{n,p}$ as a function of p. Adaptive procedure is performed with $\rho = 1/2$ (left) and $\rho = 1/10$ (right).*

6.2.1. *The performance of the adaptive estimator.* For any estimator $\widehat{\alpha}$ of α let

$$\sigma^2(\widehat{\alpha}, \alpha) = \frac{1}{n}\sum_{i=1}^{n} \log^2 \frac{\widehat{\alpha}}{\alpha}$$

be the relative mean squared error (RelMSE) of $\widehat{\alpha}$. We compare $\sigma(\widehat{q}_{n,p}, q_p)$ with $\sigma(\widehat{q}_{n,k,p}, q_p)$. Figures 4 and 5 plot these quantities for $p = 1 - 1/n = 0.999$ and $p = 0.9999999$ as a function of k. It is useful to compare RelMSE $\sigma(\widehat{q}_{n,p}, q_p)$ with minimal RelMSE $\min_k \sigma(\widehat{q}_{n,k,p}, q_p)$ as a function of p (see Figure 6). The ratio $r_{n,p} = \sigma(\widehat{q}_{n,p}, q_p) / \min_k \sigma(\widehat{q}_{n,k,p}, q_p)$ regarded as a function of p is plotted in Figure 7 (see also Table1 for a more precise evaluation.)

These simulations show that the proposed adaptive procedure captures nearly the best choice in k which depends on the unknown d.f. F. The procedure gives reasonable results in both cases, for d.f. with Pareto like tails as well as with d.f. which exhibits large perturbations from these tails. Table 1 hints that the increase of RelMSE introduced by the adaptive procedure for estimating high quantiles q_p with $p \in [0.9, 0.9999999999]$ does not exceed 7%. For log perturbed Pareto d.f. the results are very similar to those of

TABLE 1
Values of $r_{n,p}$

p	0.9	0.99	0.999	0.9999	0.99999
Cauchy	1.017966	1.023952	1.041944	1.049905	1.054291
loggamma	1.042706	1.002527	1.002542	1.013393	1.021253
Hall model	0.996002	1.009698	1.023196	1.030144	1.034276
GPD	1.094321	0.998349	0.989391	0.985767	0.984071
p	0.999999	0.9999999	0.99999999	0.999999999	0.9999999999
Cauchy	1.057159	1.059174	1.060642	1.061758	1.062635
loggamma	1.026952	1.031355	1.03472	1.037275	1.039637
Hall model	1.036994	1.038913	1.040339	1.041438	1.042312
GPD	0.983118	0.982513	0.982184	0.981981	0.981829

loggamma d.f. and therefore will not be presented here.

We would like to point out that for GPD the ratio $r_{n,p}$ is even less than 1, which means that the adaptive quantile estimator $\widehat{q}_{n,p}$ improves the performance of individual quantile estimators $\widehat{q}_{n,k,p}$, $k = 2, ..., n$. This improvement can be observed for other d.f. with an appropriate choice of the parameter ρ. The corresponding plots of the ratio $r_{n,p}$ for loggamma d.f. with $\rho = 1/10$ and for Hall model with $\rho = 1/2$ are given in Figure 8.

The high variability for extreme quantiles $q_p,\ p > 1 - 0.1/n$ (see Figure 6) is mainly explained by the bias introduced by the Pareto model and less by the variability introduced by the adaptive procedure. The bias reducing techniques can be applied under some additional assumptions on the underlying d.f. F. Our adaptive values $\widehat{k}_n$ and $\widehat{\theta}_n$ can be applied with these type of bias reduced estimators to construct new adaptive quantile estimators, however this issue will not be discussed here. For further details on this subject we refer to Danielsson et all. [3], Gomes and Oliveira [8]; see also Chapter 4.7 in Beirlant et all. [1].

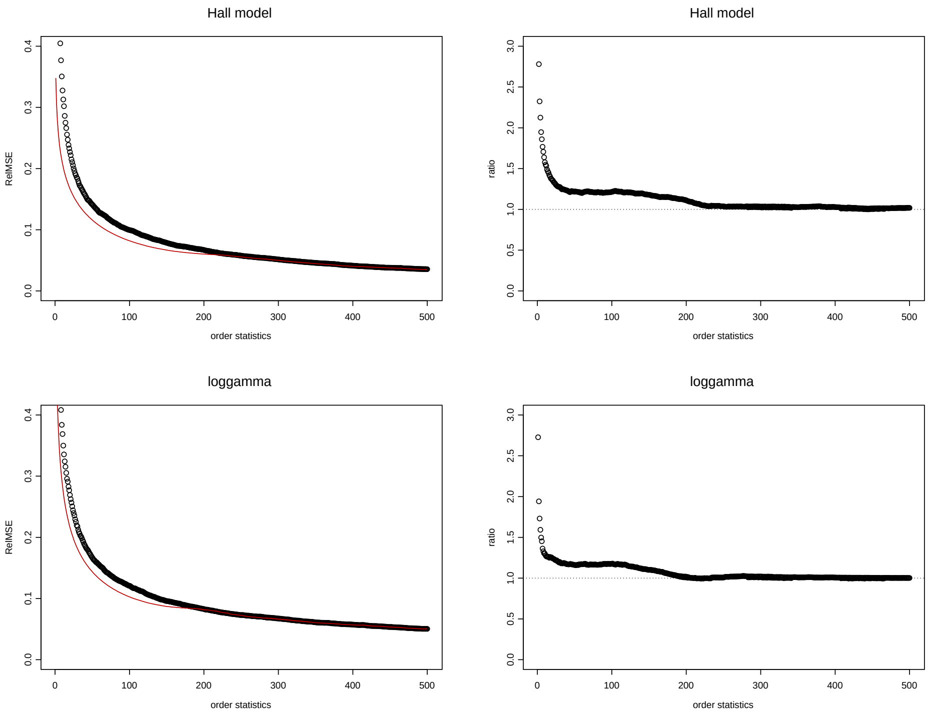

FIG 9. *Left:* $\sigma(X_{n,k}, q_{p_{n,k}})$ *(points) and* $\sigma(\widehat{q}_{n,p_{n,k}}, q_{p_{n,k}})$ *(solid line) as functions of* k*; Right:* $r^0_{n,k} = \sigma(X_{n,k}, q_{p_{n,k}})/\sigma(\widehat{q}_{n,p_{n,k}}, q_{p_{n,k}})$ *as function of* k*; Top: Hall model; Bottom: loggamma d.f. .*

6.2.2. *Comparison with sample quantiles.* For any $k = 1, ..., n$ the sample quantile $X_{n,k}$ is considered as an estimate of the true quantile $F^{-1}(p_{n,k})$, where $p_{n,k} = 1 - k/n$. We shall compare the RelMSE of adaptive quantiles $\widehat{q}_{n,p_{n,k}}$, with those of sample quantiles $X_{n,k}$, for $k = 1, ..., 500$ by computing the ratio $r^0_{n,k} = \sigma(X_{n,k}, q_{p_{n,k}})/\sigma(\widehat{q}_{n,p_{n,k}}, q_{p_{n,k}})$. The results of the simulations are reported in Table 2 and Figure 9. They show the there is a substantial gain in variance if we use (6.1) for estimating large quantiles.

Figures reported in Tables 1 and 2 can be used to compare the performance of the adaptive estimator $\widehat{\theta}_n$ with other adaptive estimators.

TABLE 2
Values of $r^0_{n,k}$

k	1	2	3	4	5	10	20
Cauchy	3.5360	2.4100	2.0294	1.8671	1.7200	1.4226	1.2621
loggamma	2.7270	1.9417	1.7306	1.5924	1.4971	1.3010	1.2453
Hall model	4.1240	2.7809	2.3237	2.1246	1.9466	1.5772	1.3605
GPD	1.3117	1.2108	2.9563	2.0609	1.7629	1.6422	1.5288
k	30	40	50	60	70	80	90
Cauchy	1.2081	1.1852	1.1849	1.1755	1.1928	1.1860	1.1745
loggamma	1.1982	1.1724	1.1611	1.1696	1.1642	1.1622	1.1748
Hall model	1.2758	1.2324	1.2183	1.2001	1.2141	1.2088	1.2040
GPD	1.1813	1.1683	1.1675	1.1509	1.1557	1.1327	1.1033

6.3. *Estimation of the index of regular variation.* According to (4.17) the adaptive estimator $\widehat{\theta}_n$ converges to the index of regular variation γ. The performance of an estimator $\widetilde{\theta}_n$ w.r.t. γ will be measured using the root mean squared error (RMSE) $\sigma(\widetilde{\theta}_n) = E^{1/2}(\widetilde{\theta}_n - \gamma)^2$. The corresponding simulations of the RMSE's $\sigma(\widehat{\theta}_n)$ and $\sigma(\widehat{h}_{n,k})$ (as a function of k) are presented in Figure 10. In case of Cauchy d.f. the minimal value of RMSE of the Hill estimator is $\min_k \sigma(\widehat{h}_{n,k}) = 0.07385$, while the RMSE of the adaptive estimator is $\sigma(\widehat{\theta}_n) = 0.07899$, which gives the ratio $r^\gamma_n = \sigma(\widehat{\theta}_n)/\min_k \sigma(\widehat{h}_{n,k}) = 1.06966$. For loggamma d.f. the minimal value of RMSE of the Hill estimator is $\min_k \sigma(\widehat{h}_{n,k}) = 0.23112$, while the RMSE of the adaptive estimator is $\sigma(\widehat{\theta}_n) = 0.24804$, which gives the ratio $r^\gamma_n = \sigma(\widehat{\theta}_n)/\min_k \sigma(\widehat{h}_{n,k}) = 1.07321$. Thus for Cauchy and loggamma the adaptive estimator increases the minimal variance in the family of Hill estimators by 7.4%.

7. Proofs of the exponential bounds. For the local log-likelihood ratio $L_{n,t}(H,G) = L_{n,t}(H) - L_{n,t}(G)$ we have the representation

$$L_{n,t}(H,G) = \sum_{i:X_i>t} \log \frac{\alpha_G(X_i)}{\alpha_H(X_i)} + \int_{(t,X_i]} \left(\frac{1}{\alpha_G(u)} - \frac{1}{\alpha_H(u)} \right) \frac{du}{u}.$$

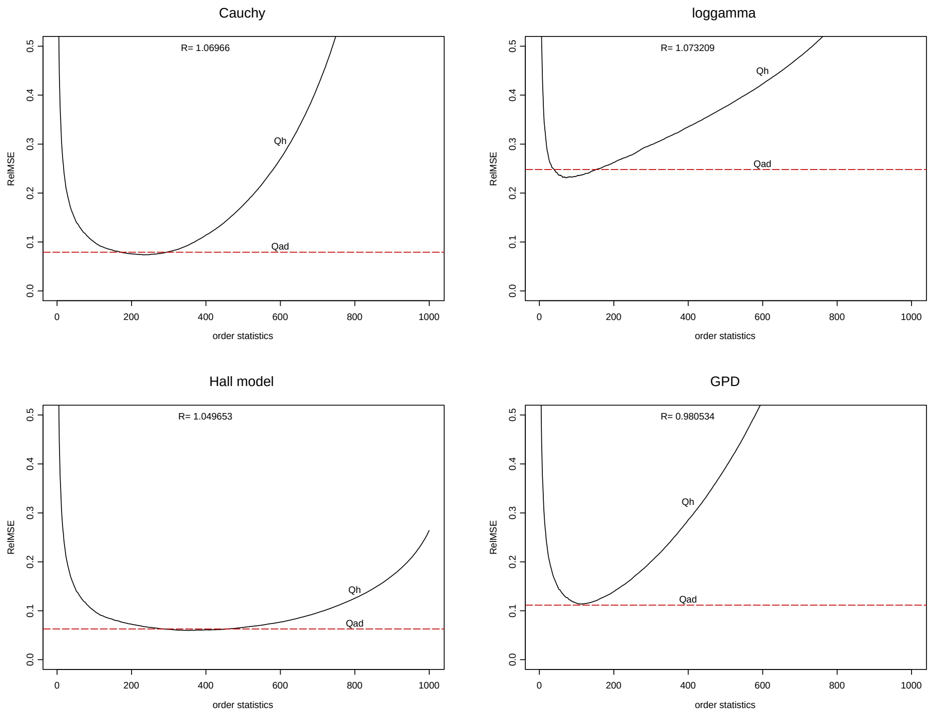

FIG 10. *Estimation of the index of regular variation γ: $Qh = \sigma(\widehat{h}_{n,k})$ – plot of the estimated RMSE of the Hill estimator $\widehat{h}_{n,k}$, for $k = 1, ..., n$; $Qad = \sigma(\widehat{\theta}_n)$ – RMSE of the adaptive estimator $\widehat{\theta}_n$ w.r.t. γ.*

Recall also the following notations: $n_t = n\,(1 - F\,(t))\,,\ \widehat{n}_t = \sum_{i=1}^n 1\,(X_i > t)$ and $\widehat{n}_{t,\tau} = \sum_{i=1}^n 1\,(t < X_i \le \tau)$ for $t \ge x_0$. Let $X_{n,1} \ge ... \ge X_{n,n}$ be the order statistics pertaining to $X_1, ..., X_n$ and $X_{n,0} = \infty$ and $X_{n,n+1} = x_0$. Set $r_{K_n+1} = n + 1$ and $\widetilde{n}_t = \min\{r : X_{n,r} \le t,\ r = r_0, ..., r_{K_n+1}\}$, for $t \ge x_0$.

We start with a bound for the log of the local likelihood ratio.

Proposition 7.1 *For any $F, G, H \in \mathcal{F}$ any $y > 0$ it holds,*

$$\mathbf{P}_F\,(L_{n,s}\,(H, G) > y) \le \exp\left(-\frac{y}{2} + \frac{n}{2}\,(1 - F\,(s))\,d_s\right), \tag{7.1}$$

where $d_s = \chi^2\,(F_s, G_s)$.

Proof. By exponential Tchebychev's inequality,

$$\mathbf{P}_F\left(L_{n,s}(H,G)>y\right)\leq \exp\left(-y/2+\log \mathbf{E}_F\left(\exp\left(L_{n,s}(H,G)/2\right)\right)\right).$$

Since the r.v.'s $X_1,...,X_n$ are i.i.d. , one gets

$$\log \mathbf{E}_F\left(\exp\left(\frac{1}{2}L_{n,s}(H,G)\right)\right)=n\log \mathbf{E}_F\left(\exp\left(\frac{1_{A_{i,s}}}{2}\log\frac{dH_s}{dG_s}(X_i/s)\right)\right),$$

where $A_{i,s}=\{X_i>s\}$. Since $\mathbf{E}_F\exp\left(1_{A_{i,s}}\log\frac{dH_s}{dF_s}\left(\frac{X_i}{s}\right)\right)=1$, by Hölder's inequality

$$\mathbf{E}_F\exp\left(\frac{1_{A_{i,s}}}{2}\log\frac{dH_s}{dG_s}\left(\frac{X_i}{s}\right)\right)\leq \mathbf{E}_F^{1/2}\exp\left(1_{A_{i,s}}\log\frac{dF_s}{dG_s}\left(\frac{X_i}{s}\right)\right).$$

Using

$$\mathbf{E}_F\exp\left(1_{A_{i,s}}\log\frac{dF_s}{dG_s}\left(\frac{X_i}{s}\right)\right)=F(s)+(1-F(s))(d_s+1),$$

one gets

$$\begin{aligned}\mathbf{P}_F\left(L_{n,s}(H,G)>y\right) &\leq \exp\left(-\frac{y}{2}+\frac{n}{2}\log\left\{1+(1-F(s))\,d_s\right\}\right)\\ &\leq \exp\left(-\frac{y}{2}+\frac{n}{2}(1-F(s))\,d_s\right).\end{aligned}$$

The inequality (7.1) follows. ∎

Corollary 7.2 *For any $\tau\geq t\geq s\geq x_0$ and $\theta',\theta>0$,*

$$\mathbf{P}_F\left(L_{n,t}(\theta',\theta)>y\right) \leq \exp\left(-y/2+\frac{n}{2}(1-F(s))\,d_s\right), \tag{7.2}$$

$$\mathbf{P}_F\left(L_{n,t,\tau}(\theta',\theta)>y\right) \leq \exp\left(-y/2+\frac{n}{2}(1-F(s))\,d_s\right), \tag{7.3}$$

where $d_s=\chi^2(F_s,P_\theta)$.

Proof. The first assertion follows from 7.1 when applied with H and G such that $\alpha_H(x)=\alpha_G(x)=\theta$ for $x\in[x_0,t)$ and $\alpha_H(x)=\theta'$, $\alpha_G(x)=\theta$

for $x \in [t, \infty)$. The second one is obtained with $\alpha_H(x) = \alpha_G(x) = \theta$ for $x \in [x_0, t) \cup [\tau, \infty)$ and $\alpha_H(x) = \theta'$, $\alpha_G(x) = \theta$ for $x \in [t, \tau)$. ■

Now we give an exponential bound for the maximum of the log-likelihood ratio.

Proposition 7.3 *For any $F \in \mathcal{F}$, $\theta > 0$, $y > 0$, $\tau \geq t \geq s \geq x_0$ it holds*

$$\begin{aligned}
\mathbf{P}_F\left(\widehat{n}_t \mathcal{K}\left(\widehat{\theta}_{n,t}, \theta\right) > y\right) &\leq 2n \exp\left(-\frac{y}{2} + \frac{n}{2}(1 - F(s))\, d_s\right), \\
\mathbf{P}_F\left(\widehat{n}_{t,\tau} \mathcal{K}\left(\widehat{\theta}_{n,t,\tau}, \theta\right) > y\right) &\leq 2n \exp\left(-\frac{y}{2} + \frac{n}{2}(1 - F(s))\, d_s\right),
\end{aligned}$$

where $d_s = \chi^2(F_s, P_\theta)$.

Proof. We shall prove only the second inequality the first one being proved in the same way.

First note that $\widehat{n}_{t,\tau} \mathcal{K}\left(\widehat{\theta}_{n,t}, \theta\right) = L_{n,t,\tau}\left(\widehat{\theta}_{n,t,\tau}, \theta\right)$. For the sake of brevity let $l_k(\alpha) = \left(y/k - \log \frac{\theta}{\alpha}\right) / \left(\frac{1}{\theta} - \frac{1}{\alpha}\right)$, $\alpha > \theta$. Since the function $l_k(\alpha)$ is continuous in α for $\alpha > \theta$ and $\lim_{\alpha \to \theta} l_k(\alpha) = \infty$, $\lim_{\alpha \to \infty} l_k(\alpha) = \infty$, there exists a finite point $\alpha_k^* > \theta$ which realize $\alpha_k^* = \arg\min_{\alpha > \theta} l_k(\alpha)$. Note that α_k^* is a function only on k, y, θ. With these notations, on the event $A_k = \{\widehat{\theta}_{n,t,\tau} > \theta,\ \widehat{n}_{t,\tau} = k\}$, we have that the inequality

$$L_{n,t,\tau}\left(\widehat{\theta}_{n,t,\tau}, \theta\right) = \widehat{n}_{t,\tau}\left(\log \frac{\theta}{\widehat{\theta}_{n,t,\tau}} - \left(1/\widehat{\theta}_{n,t,\tau} - 1/\theta\right)\widehat{\theta}_{n,t,\tau}\right) > y$$

is equivalent to $\widehat{\theta}_{n,t,\tau} \geq l_k\left(\widehat{\theta}_{n,t,\tau}\right)$ and the inequality $\widehat{\theta}_{n,t,\tau} \geq l_k(\alpha_k^*)$ is equivalent to $L_{n,t,\tau}(\alpha_k^*, \theta) > y$. Then

$$\begin{aligned}
\left\{L_{n,t,\tau}\left(\widehat{\theta}_{n,t,\tau}, \theta\right) > y\right\} \cap A_k &= \left\{\widehat{\theta}_{n,t,\tau} \geq l_k\left(\widehat{\theta}_{n,t,\tau}\right)\right\} \cap A_k \\
&\subseteq \left\{\widehat{\theta}_{n,t,\tau} \geq l_k(\alpha_k^*)\right\} \cap A_k \\
&\subseteq \{L_{n,t,\tau}(\alpha_k^*, \theta) > y\}.
\end{aligned}$$

In the same way

$$\left\{L_{n,t,\tau}\left(\widehat{\theta}_{n,t,\tau},\theta\right)>y\right\}\cap B_k\subseteq\left\{L_{n,t,\tau}\left(\alpha_k^{**},\theta\right)>y\right\},$$

where $B_k=\{\widehat{\theta}_{n,t,\tau}\leq\theta,\ \widehat{n}_{t,\tau}=k\}$ and $\alpha_k^{**}=\arg\max_{0<\alpha\leq\theta}l_k\left(\alpha\right)$ is a function only on k,y,θ. The latter implies

$$\begin{aligned}\mathbf{P}_F\left(L_{n,t,\tau}\left(\widehat{\theta}_{n,t,\tau},\theta\right)>y,\ \widehat{n}_{t,\tau}=k\right) &\leq \mathbf{P}_F\left(L_{n,t,\tau}\left(\alpha_k^{*},\theta\right)>y\right)\\ &\quad+\mathbf{P}_F\left(L_{n,t,\tau}\left(\alpha_k^{**},\theta\right)>y\right).\end{aligned}$$

Since by Corollary 7.2,

$$\mathbf{P}_F\left(L_{n,t,\tau}\left(\theta',\theta\right)>x\right)\leq\exp\left(-\frac{y}{2}+\frac{n}{2}\left(1-F\left(s\right)\right)d_se^{d_s}\right),$$

with $\theta'=\alpha_k^{*},\alpha_k^{**}$, one gets

$$\begin{aligned}\mathbf{P}_F\left(L_{n,t,\tau}\left(\widehat{\theta}_{n,t,\tau},\theta\right)>x\right) &= \sum_{k=1}^{n}\mathbf{P}_F\left(L_{n,t,\tau}\left(\widehat{\theta}_{n,t,\tau},\theta\right)>x,\ \widehat{n}_{t,\tau}=k\right)\\ &\leq 2n\exp\left(-\frac{y}{2}+\frac{n}{2}\left(1-F\left(s\right)\right)d_se^{d_s}\right),\end{aligned}$$

which completes the proof. ■

Proposition 7.4 *For any $F\in\mathcal{F}$ and $s\geq x_0$, $\theta>0$, $y>0$ it holds*

$$\begin{aligned}\mathbf{P}_F\left(\sup_{s\leq t}\widehat{n}_t\mathcal{K}\left(\widehat{\theta}_{n,t},\theta\right)>y\right) &\leq 2n^4\exp\left(-\frac{y}{2}+\frac{n}{2}\left(1-F\left(s\right)\right)d_s\right)+\frac{1}{n},\\ \mathbf{P}_F\left(\sup_{s\leq t\leq\tau}\widehat{n}_{t,\tau}\mathcal{K}\left(\widehat{\theta}_{n,t,\tau},\theta\right)>y\right) &\leq n^7\exp\left(-\frac{y}{2}+\frac{n}{2}\left(1-F\left(s\right)\right)d_s\right)+\frac{1}{n},\end{aligned}$$

where $d_s=\chi^2\left(F_s,P_\theta\right)$.

Proof. We shall give a proof only for the second inequality, the first one being proved in the same way.

Let $N = n^3$ and $J = \{s_0, ..., s_N\}$ be the set of numbers satisfying $s_{i-1} < s_i$, $F([s_{i-1}, s_i)) = 1/N$ and $\cup_{i=1}^N [s_{i-1}, s_i) = [x_0, \infty)$. If we denote by $\mathfrak{A}_n$ the event that $X_{n,1}, ..., X_{n,n}$ will fall into disjoint intervals, then, for $K_n > 2$,

$$\begin{aligned} \mathbf{P}_F(\mathfrak{A}_n) &= \prod_{i=1}^n \left(1 - \frac{i-1}{N}\right) \geq 1 - \sum_{i=1}^n \log\left(1 - \frac{i-1}{N}\right) \\ &\geq 1 - \frac{3}{2}\sum_{i=2}^n \frac{i-1}{N} = 1 - \frac{3n(n-1)}{4n^3} \geq 1 - \frac{1}{n}. \end{aligned}$$

On the event $\mathfrak{A}_n$ it holds

$$\sup_{s \leq t \leq \tau} \widehat{n}_{t,\tau} \mathcal{K}\left(\widehat{\theta}_{n,t,\tau}, \theta\right) = \max_{s \leq t \leq \tau,\ t,\tau \in J} \widehat{n}_{t,\tau} \mathcal{K}\left(\widehat{\theta}_{n,t,\tau}, \theta\right).$$

Then

$$\begin{aligned} &\mathbf{P}_F\left(\sup_{s \leq t \leq \tau} \widehat{n}_{t,\tau} \mathcal{K}\left(\widehat{\theta}_{n,t,\tau}, \theta\right) > y\right) \\ &\leq \sum_{s \leq t \leq \tau,\ t,\tau \in J} \mathbf{P}_F\left(\widehat{n}_{t,\tau} \mathcal{K}\left(\widehat{\theta}_{n,t,\tau}, \theta\right) > y\right) + 1 - \mathbf{P}_F(\mathfrak{A}_n). \end{aligned} \tag{7.4}$$

According to Proposition 7.3

$$\mathbf{P}_F\left(L_{n,t,\tau}\left(\widehat{\theta}_{n,t,\tau}, \theta\right) \geq y\right) \leq 2n \exp(-y_s),$$

where $y_s = \frac{y}{2} - \frac{n}{2}(1 - F(s))\, d_s e^{d_s}$. Since $\sum_{s \leq t \leq \tau,\ t,\tau \in J} \leq n^6/2$, from (7.4) one gets

$$\mathbf{P}_F\left(\sup_{s \leq t \leq \tau} \widehat{n}_{t,\tau} \mathcal{K}\left(\widehat{\theta}_{n,t,\tau}, \theta\right) > y\right) \leq n^7 \exp(-y_s) + \frac{1}{n}.$$

The requested follows. ■

We end this section with an exponential bound for the statistic $T_n(t, \tau)$.

Proposition 7.5 *For any $F \in \mathcal{F}$ and $s \geq x_0$, $\theta > 0$, $y > 0$ it holds*

$$\mathbf{P}_F\left(\sup_{s \leq t \leq \tau} T_n(t, \tau) > 2y\right) \leq 2n^7 \exp\left(-\frac{y}{2} + \frac{n}{2}(1 - F(s))\, d_s\right) + \frac{2}{n},$$

where $d_s = \chi^2(F_s, P_\theta)$.

Proof. Let $\theta > 0$. Using (3.2) and the inequality $\sup_{F\in\mathcal{F}_t} L_{n,t}(F) \geq L_{n,t}(\theta)$, one gets $T_n(t,\tau) \leq L_{n,s}\left(\widehat{\theta}_{n,t,\tau}, \widehat{\theta}_{n,\tau}, \tau\right) - L_{n,t}(\theta) = L_{n,t}\left(\widehat{\theta}_{n,t,\tau}, \widehat{\theta}_{n,\tau}, \tau, \theta\right)$. The representation (2.8) implies $T_n(t,\tau) \leq \widehat{n}_{t,\tau}\mathcal{K}\left(\widehat{\theta}_{n,t,\tau}, \theta\right) + \widehat{n}_{\tau}\mathcal{K}\left(\widehat{\theta}_{n,\tau}, \theta\right)$. The requested inequality follows from Proposition 7.4. ∎

8. Auxiliary statements.

Lemma 8.1 *For any $\theta_1, \theta_2 > 0$ such that $\mathcal{K}(\theta_1, \theta_2) \leq \frac{1}{2}$ it holds*

$$\frac{1}{3}\log^2\frac{\theta_1}{\theta_2} \leq \mathcal{K}(\theta_1,\theta_2) \tag{8.1}$$

and for any $\theta_1, \theta_2 > 0$ such that $\log^2\frac{\theta_1}{\theta_2} \leq \frac{2}{3}$, it holds

$$\mathcal{K}(\theta_1,\theta_2) \leq \frac{3}{4}\log^2\frac{\theta_1}{\theta_2}. \tag{8.2}$$

In particular, for any $\theta_1, \theta_2 > 0$ such that $\mathcal{K}(\theta_1, \theta_2) \leq \frac{1}{2}$ it holds

$$\mathcal{K}(\theta_1,\theta_2) \leq \frac{9}{4}\mathcal{K}(\theta_2,\theta_1).$$

Proof. Note that $\frac{1}{3}\log^2(x+1) \leq G(x)$ for any x satisfying $G(x) \leq 1/2$ and $G(x) \leq \frac{3}{4}\log^2(x+1)$ for any x satisfying $\log^2(x+1) \leq \frac{2}{3}$, which in turn implies the requested. ∎

Lemma 8.2 *For any sequence of positive numbers $\theta_1, ..., \theta_M$ such that*

$$\sum_{i=1}^{M-1}\sqrt{\mathcal{K}(\theta_i,\theta_{i+1})} \leq \frac{1}{3}$$

it holds

$$\sqrt{\mathcal{K}(\theta_1,\theta_n)} \leq \frac{3}{2}\sum_{i=1}^{M-1}\sqrt{\mathcal{K}(\theta_i,\theta_{i+1})}. \tag{8.3}$$

Proof. To prove (8.3) note that by (8.1),

$$\left|\log\frac{\theta_1}{\theta_n}\right| \leq \sum_{i=1}^{M-1}\left|\log\frac{\theta_i}{\theta_{i+1}}\right| \leq \sqrt{3}\sum_{i=1}^{M-1}\sqrt{\mathcal{K}(\theta_i,\theta_{i+1})} \leq \frac{1}{\sqrt{3}}.$$

Then using (8.2),

$$\sqrt{\mathcal{K}\left(\theta_1, \theta_n\right)} \leq \frac{\sqrt{3}}{2}\left|\log \frac{\theta_1}{\theta_n}\right|,$$

which in conjunction with (8.1) proves (8.3). ■

Recall that $\widetilde{n}_t$ is the largest $r \in \mathcal{R}_n$ satisfying $X_{n,r} > t$. The following lemma gives a sufficient condition that $\widetilde{n}_{\tau_n} \overset{\mathbf{P}_F}{\asymp} n_{\tau_n}$.

Lemma 8.3 *If the sequence $\tau_n \geq x_0$, $n = 1, 2, \ldots$ is such that $n_{\tau_n} \to \infty$ as $n \to \infty$, then $\widehat{n}_{\tau_n} \overset{\mathbf{P}_F}{\asymp} n_{\tau_n}$ and $\widetilde{n}_{\tau_n} \overset{\mathbf{P}_F}{\asymp} n_{\tau_n}$ as $n \to \infty$.*

Proof. By Tchebychev's exponential inequality, for any $u > 0$ and $\varepsilon \in (0, 1)$,

$$\begin{aligned}\mathbf{P}_F\left(\widehat{n}_{\tau_n}/n_{\tau_n} < 1-\varepsilon\right) &\leq \exp\left(u\left(1-\varepsilon\right)n_{\tau_n} + n_{\tau_n}\left(e^{-u}-1\right)\right) \\ &\leq \exp\left(-u\varepsilon n_{\tau_n} + u^2 n_{\tau_n}\right).\end{aligned}$$

In the same way $\mathbf{P}_F\left(\widehat{n}_{\tau_n}/n_{\tau_n} > 1+\varepsilon\right) \leq \exp\left(-u\varepsilon n_{\tau_n} + u^2 n_{\tau_n}\right)$. Choosing $u = \varepsilon/2$ one gets

$$\mathbf{P}_F\left(\left|\frac{\widehat{n}_{\tau_n}}{n_{\tau_n}} - 1\right| > \varepsilon\right) \leq 2\exp\left(-\frac{\varepsilon^2}{4} n_{\tau_n}\right).$$

Since $n_{\tau_n} \to \infty$ one gets the first assertion. The second is obtained taking into account that $\frac{\widehat{n}_{\tau_n} - \widetilde{n}_{\tau_n}}{n_{\tau_n}} = O\left(\frac{n}{K_n n_{\tau_n}}\right) = O\left(K_n^{-1}\right)$ as $n \to \infty$. ■

Lemma 8.4 *For any sequence τ_n, $n = 1, 2, \ldots$ satisfying $n_{\tau_n} \to \infty$ as $n \to \infty$, it holds $\lim_{n\to\infty} \mathbf{P}_F\left(X_{n,k} > \tau_n\right) = 1$, for any given natural number k.*

Proof. By Lemma 8.3 $\mathbf{P}_F\left(\widehat{n}_{\tau_n} > k\right) \to 1$ as $n \to \infty$. Since $\mathbf{P}_F\left(X_{n,k} > \tau_n\right) = \mathbf{P}_F\left(\widehat{n}_{\tau_n} > k\right)$ the requested follows. ■

Lemma 8.5 *Assume that $P \sim Q$. Then*

$$\chi^2\left(P, Q\right) \leq \mathbf{E}_P\left(\log^2 \frac{dQ}{dP} \exp\left(\left|\log \frac{dQ}{dP}\right|\right)\right).$$

Proof. It is easy to see that $\chi^2(P,Q) = \int g\left(\frac{dQ}{dP}\right) dP$, where $g(x) = \frac{(x-1)^2}{x}$. Since $(x-1)^2 \le e^{2\log x}\log^2 x$, for $x > 1$ and $(x-1)^2 \le \log^2 x$, for $x \in (0,1)$, we get $g(x) \le \log^2 x \exp(|\log x|)$. The requested follows. ■

Recall that $\rho_0(x,y) = \max\left\{\left|\log \frac{x}{y}\right|, \left|\frac{1}{x} - \frac{1}{y}\right|\right\}$, $x, y > 0$.

Proposition 8.6 *Assume that d.f.'s F and G are such that it holds $d_t = \sup_{x\ge t}\rho_0(\alpha_F(x), \alpha_G(x)) \le \varepsilon_0$ and $\int_1^\infty (1+\log x)^2 x^{\varepsilon_0} F_t(dx) \le \varepsilon_1$. Then $\chi^2(F,G) \le C(\varepsilon_0,\varepsilon_1) d_t^2$, where $C(\varepsilon_0,\varepsilon_1) = \varepsilon_1 e^{\varepsilon_0}$.*

Proof. Since

$$\log \frac{dF_t(x)}{dG_t(x)} = \log \frac{\alpha_G(xt)}{\alpha_F(xt)} + \int_t^{xt}\left(\frac{1}{\alpha_G(u)} - \frac{1}{\alpha_F(u)}\right)\frac{du}{u},\ x \ge 1$$

it holds $\left|\log \frac{dF_t(x)}{dG_t(x)}\right| \le d_t(1+\log x)$. Using Lemma 8.5, with $P = G_t$ and $Q = G_t$ one gets $\chi^2(F,G) \le d_t^2 e^{d_t}\int_1^\infty (1+\log x)^2 x^{d_t} F_t(dx)$. The requested follows. ■

REFERENCES

[1] Beirlant, J., Goegebeur, Yu., Segers, J. and Teugels, J. (2004). *Statistics of extremes. Theory and applications.* Wiley.

[2] Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular variation.* Cambridge University Press, Cambridge.

[3] Danielsson, J. and de Haan, L., Peng, L., de Vries (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis,* **76**, 226-248.

[4] Donoho, D. and Jonstone, J. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrica.,* **81**, 425-455.

[5] Drees, H. and Kaufmann, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation. *Stoch. Processes Appl.,* **75**, 149-172.

[6] Drees, H. (1998). Optimal rates of convergence for estimates of the extreme value index. *Ann. Statist.,* **26**, 434-448.

[7] Embrechts, P., Klüppelberg, K., and Mikosch, T. (1997). Modelling extremal events. Springer.

[8] Gomes, M. I. and Oliveira, O. (2001). The bootstrap methodology in statistics of extremes – choice of the optimal sample fraction. *Extremes* **4**, 331-358.

[9] Grama, I. and Spokoiny, V. (2003). Tail index estimation by local exponential modeling. WIAS Preprint No. 819.

[10] Hall, P. (1982). On some simple estimates of an exponent of regular variation. *J. Roy. Statist. Soc.* Ser. B **44**, 37-42.

[11] Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis,* **32**, 177-203.

[12] Hall, P. and Welsh, A. H. (1984). Best attainable rates of convergence for estimates of parameters of regular variation. *Ann. Statist.,* **12**, 1079-1084.

[13] Hall, P. and Welsh, A. H. (1985). Adaptive estimates of regular variation. *Ann. Statist.,* **13**, 331-341.

[14] Hill, B. M., (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.,* **3**, 1163-1174.

[15] Huisman, R., Koedijk, K. G., Kool, C. J. M., and Palm, F. (2001). Tail-index estimates in small samples. Journal of Business & Economic Statistics. **19**, 208-216.

[16] Lepski, O. V. (1990) One problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35**, N 3, 459-470.

[17] Lepski, O. V. and Spokoiny, V. G. (1995) Local adaptation to inhomogeneous smoothness: resolution level. *Math. Methods Statist.* 4, 239-258.

[18] Mason, D. (1982). Laws of large numbers for sums of extreme values. *Ann. Probab.,* **10**, 754-764.

[19] Reiss, R.-D. (1989). *Approximate Distributions of Order Statistics: With Applications to Nonparametric Statistics.* Springer Series in Statistics. New York: Springer.

[20] Resnick, S. I. (1997). Heavy tail modelling and teletraffic data. *Ann. Statist.,* **25**, 1805-1869.

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

Université de Bretagne Sud
CERYC, 56000 Vannes, France
E-mail: ion.grama@univ-ubs.fr

Weierstrass-Institute
Mohrenstr. 39, 10117 Berlin, Germany
E-mail: spokoiny@wias-berlin.de