Pareto approximation of the tail by local exponential modeling

Ion Grama
Université de Bretagne Sud
rue Yves Mainguy, Tohannic
56000 Vannes, France
email: ion.grama@univ-ubs.fr

Vladimir Spokoiny
Weierstrass Institute
Mohrenstr. 39
D-10117 Berlin, Germany,
email: spokoiny@wias-berlin.de

This version of: February 20, 2003.

Abstract

We give a new adaptive method for selecting the number of upper order statistics used in the estimation of the tail of a distribution function. Our approach is based on approximation by an exponential model. The selection procedure consists in consecutive testing for the hypothesis of homogeneity of the estimated parameter against the change-point alternative. The selected number of upper order statistics corresponds to the first detected change-point. Our main results are non-asymptotic.

1 Introduction

This paper is concerned with the adaptive estimation of the tail of a distribution function (d.f.) F. A popular estimator for use in the extreme value theory was proposed by Hill (1975). Given a sample $X_1, ..., X_n$ from the d.f. F the Hill estimator is defined as

$$\widehat{\alpha}_{n,k} = \frac{1}{k} \sum_{i=1}^{k} \log \frac{X_{n,i}}{X_{n,k+1}},$$

where $X_{n,1} \geq ... \geq X_{n,n}$ are the order statistics pertaining to $X_1, ..., X_n$ and k is the number of upper order statistics used in the estimation. There is a vast literature on the asymptotic properties of the Hill estimator. Suppose that d.f. F is regularly varying with index of regular variation β [see for example Bingham, Goldie and Teugels (1987)]. Weak consistency for estimating β was established by Mason (1982), under the conditions that $k \to \infty$ and $k/n \to 0$ as $n \to \infty$. Asymptotic normality of the Hill estimator was proved by Hall (1982). A strong consistency result can be found in Deheuvels, Haeusler and Mason (1988). Further properties concerning the efficiency have been studied in Dress (2001). For extensions to dependent observations see, for instance, Resnik and Starica (1998) and the references therein. The asymptotic results mentioned above do not give any recipe about selecting the parameter k in practical applications, while the behavior of the error estimation depends essentially on it. Different approaches for data driven choices of k have been proposed in the literature, mainly based on the idea of balancing the bias and the asymptotic variance of the Hill estimator. We refer to Hall and Welsh (1985), Danielson, de Haan, Peng, Vries (2001), Beirlant, Teugels and Vinysaker (1996), Resnik and Starica (1997), Dress and Kaufman (1998), among many others. However the bias of the Hill estimator for estimating the parameter of regular variation as a rule diminishes very slowly, which makes any choice of the parameter k not very efficient from the practical point of view. A striking example is the so called Hill Horror plot (see Figure 1, left).

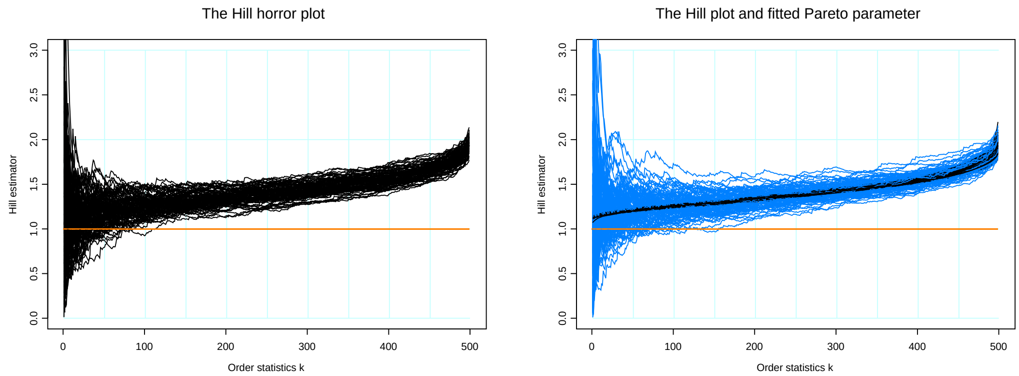

Figure 1: Left: 100 realizations of the Hill estimator for Pareto-log d.f. $F(x) = 1 - (x/e)^{-1/\beta}\log x,\ x \geq e$, where the parameter $\beta = 1$ is expected to be estimated. Right: 100 realizations of the Hill estimator for Pareto-log d.f. and the fitted Pareto parameter. Here the dark lines represent the fitted Pareto index computed from the approximation formulas (3.5), (3.1) and the light ones are the corresponding Hill plots.

For more insight on the problem the reader is referred to the book by Embrechts, Klüppelberg and Mikosch (1997), from which we cite on the page 351: "On various occasions we hinted at the fact that the determination of the number k of upper order statistics finally used remains a delicate point in the whole set-up. Various papers exist which offer a semi-automatic or automatic, so-called "optimal", choice of k. ... We personally prefer a rather pragmatic approach realizing that, whatever method one chooses, the "Hill horror plot" ... would fool most, if not all. It also serves to show how delicate a tail analysis in practice really is." An interesting exchange of opinions on this subject may be found in the survey paper by Resnik (1997) and in the supplied discussion.

The aim of the present paper is to give a natural resolution to the "Hill horror plot" paradox and to rehabilitate the Hill estimator, for finite sample sizes, by looking at the problem from the point of view of selecting an appropriate tail. In Section 3 we shall see that, for finite sample sizes, the Hill estimator is close to another quantity which can be interpreted as the parameter of the approximating Pareto distribution and which we shall call the fitted Pareto index [see (2.4) for the definition of this quantity]. In Figure 1, right, we give a simulation for the Pareto-log d.f.; other examples are presented in the Appendix 8. The importance of this interpretation, perhaps, is justified by the fact that it allows new approaches for selecting the number k of retained upper order statistics. For estimating the fitted Pareto index we propose a method based on successive testing of the hypothesis that the first k normed log-spacings follow exponential distributions with homogeneous parameters. The idea goes back to Spokoiny (1998). However our procedure is different in several aspects. First, our test is based on the likelihood ratio test statistic for testing homogeneity of the estimated parameters against the change-point alternative. Second, in our procedure the number k is selected to be the detected change-point. We also refer the reader to Picard and Tribulieu (2002) where the change point Pareto model (see Pareto-CP d.f. in the Appendix) is used for estimation in the parametric context.

Our main results are non-asymptotic. We establish an "oracle" inequality for the adaptive estimator of the fitted index. The result claims that the risk of the adaptive estimator is only within come constant factor worse than the risk of the best possible estimator for the given model.

The paper is organized as follows. In Sections 2 and 3 we formulate the problem and give the approximation by the exponential model. The adaptive procedure is presented in Section 4. Section 5 illustrate the numerical performances of the method on some artificial data sets. The results and the proofs are given in Sections 6 and 7.

2 The model and the problem

Let $X_1, ..., X_n$ be i.i.d. observations with common d.f. $F(x)$ supported on (a, ∞), where $a > 0$ is a fixed real number. Assume that the function F is strictly increasing and has a continuous density f. Since $F(a) = 0$, the d.f. F can be represented as

$$F(x) = 1 - \exp\left(-\int_a^x \lambda(t)\, dt\right), \quad x \geq a, \tag{2.1}$$

where

$$\lambda(x) = \frac{f(x)}{1 - F(x)}, \quad x \geq a$$

is the hazard rate. Note that if $\lambda(x) = \frac{1}{\alpha x}$, then the d.f. F is Pareto with index $1/\alpha$, which is a typical fat tail distribution. To allow more general laws with heavy tails we shall assume that

$$\lambda(x) = \frac{1}{\alpha(x)\, x}, \tag{2.2}$$

where the function $\alpha(x)$, $x > a$, can be approximated by a constant for big values of x. For instance, this is the case when there exists an $\beta > 0$ such that

$$\lim_{x\to\infty} \alpha(x) = \beta. \tag{2.3}$$

Many regularly varying at infinity d.f.'s F satisfy the assumptions (2.1), (2.2) and (2.3), see representation theorems in Seneta (1976) or Bingham, Goldie and Teugels (1987). If this is the case, then the limit in (2.3) is nothing else but the index of regular variation.

Our problem can be formulated as follows. Let $X_{n,1} > ... > X_{n,n}$ be the order statistics pertaining to $X_1, ..., X_n$. The goal is to find a natural number k such that on the set $\{X_{n,1}, ..., X_{n,k}\}$ the function $\alpha(x)$, $x \geq a$, can be well approximated by the value $\alpha(X_{n,1})$ and to estimate this value. The intuitive meaning of this is to find a Pareto approximation for the tail of the d.f. F on the data set $\{X_{n,1}, ..., X_{n,k}\}$. Note that this problem is different from that of estimating the index of regular variation β defined by the limit (2.3). As it was stressed in the Introduction the main advantage of the present setting is, perhaps, the fact that it allows new algorithms for the choice of the nuisance parameter k. The approach adopted in this paper is based on the approximation by an exponential model which is presented in the next section.

Before to proceed with this, we shall point out the connection of the function $\alpha(\cdot)$ to the logarithmic mean excess of F :

$$\nu(t) = \int_t^\infty \log \frac{x}{t} \frac{F(dx)}{1 - F(t)}, \quad t \geq a. \tag{2.4}$$

Integration by parts gives, for any $t \geq a$,

$$\int_t^\infty \alpha(x) \frac{F(dx)}{1 - F(t)} = \nu(t). \tag{2.5}$$

By straightforward calculations it can be seen that the number $\nu(t)$ is the minimizer of the Kullback-Leibler distance between Pareto d.f. $P_\alpha(x) = 1 - x^{-1/\alpha}$, $x \geq 1$ and the excess d.f. $F(x|t) = 1 - (1 - F(xt)) / (1 - F(t))$, $x \geq 1$. Thus the number $\nu(t)$ can be interpreted as the parameter of the best Pareto fit to the tail of the d.f. F on the interval $[t, \infty)$. We shall call the function $\nu(t)$, $t \geq a$ the fitted Pareto index.

3 Approximation by exponential model

The function $\alpha(\cdot)$ will be estimated from the approximating exponential model. Our motivation is somewhat similar to that of Hill (1975) [see also Beirlant, Dierskx, Goegebeur et Matthys (2000) for another exponential approximation]. The construction of the approximating exponential model employs the following lemma, called Renyi representation of order statistics.

Lemma 3.1. *Let $X_1, ..., X_n$ be i.i.d. r.v.'s with common strictly increasing d.f. F and $X_{n,1} > ... > X_{n,n}$ be the order statistics pertaining to $X_1, ..., X_n$. Then the r.v.'s*

$$\xi_i = i \log \frac{1 - F(X_{n,i+1})}{1 - F(X_{n,i})}, \quad i = 1, ..., n-1.$$

are i.i.d. standard exponential.

Proof. See for instance Reiss (1989) or Example 4.1.5 in Embrechts, Klüppelberg and Mikosch (1997)]. □

Let $Y_i = i \log \frac{X_{n,i}}{X_{n,i+1}}$, $i = 1, ..., n-1$. Then $Y_i = \alpha_i \xi_i, \quad i = 1, ..., n-1$, where

$$\alpha_i = -\log \frac{X_{n,i}}{X_{n,i+1}} / \log \frac{1 - F(X_{n,i})}{1 - F(X_{n,i+1})}. \tag{3.1}$$

It is easy to see that the function $\alpha(x)$ is defined through the d.f. F by the equations

$$\frac{1}{\alpha(x)} = x\lambda(x) = \frac{xf(x)}{1 - F(x)} = -\frac{\frac{d}{dx}\log(1 - F(x))}{\frac{d}{dx}\log x}, \quad x \geq a. \tag{3.2}$$

By identity (3.2) the value α_i can be regarded as an approximation of the value of the function $\alpha(\cdot)$ at the point $X_{n,i+1}$. More precisely, the mean value theorem implies

$$\alpha_i = \alpha\left(X_{n,i+1} + \theta_{n,i+1}\frac{X_{n,i} - X_{n,i+1}}{X_{n,i}}\right),$$

with some $\theta_{n,i+1} \in [0,1]$, for $i = 1, ..., n-1$. These simple considerations reduce the original model to the following inhomogeneous exponential model

$$Y_i = \alpha_i \xi_i, \quad i = 1, ..., n-1, \tag{3.3}$$

where $\alpha = (\alpha_1, ..., \alpha_{n-1})$ is a vector of unknown parameters. We assume *local homogeneity* of this model which stipulates that the components α_i's nearly equal α_1 within some interval $I = [1, k]$. In the sequel finding the Pareto approximation for the tail of the d.f. F will be viewed as the problem of choosing the interval $I = [1, k]$ and of estimating the component α_1 from the observations (3.3).

Under the assumption that

$$\alpha_1 = ... = \alpha_k, \tag{3.4}$$

the maximum likelihood estimator of α_1 is the sample mean

$$\widehat{\alpha}_k = \frac{1}{k}\sum_{i=1}^{k} Y_i,$$

which is the well-known Hill estimator. Our main concern is to choose appropriately the number k of upper order statistics used in the estimation.

If the condition (3.4) is not satisfied, then from the definition of the model (3.3) it follows that the Hill estimator $\widehat{\alpha}_k$ approximates without bias the quantity

$$\overline{\alpha}_k = \frac{1}{k} \sum_{i=1}^{k} \alpha_i, \tag{3.5}$$

which, in turn, is an approximation of the fitted Pareto index (2.4): $\overline{\alpha}_k \approx \nu(X_{n,k+1})$, for k big enough. The assumption of local homogeneity implies that the quantities $\overline{\alpha}_k$, α_k and $\alpha_1 = \overline{\alpha}_1$ are close to each other and thus under this assumption the Hill estimator also approximates the fitted Pareto parameter $\nu(t)$ at the point $t = X_{n,k+1}$. The simulations show a good concordance between the two latter quantities (see Figures 1, 4 and 5).

Although the above considerations shed some light on what does the Hill estimator estimate, the main problem, how to choose an appropriate value of k (even for the fitted Pareto index $\nu(X_{n,k+1})$ or equally for $\overline{\alpha}_k$) still remains open. Model selection based on the penalisation terms [see Barron, Birge and Massart (1999)] could be a reasonable alternative for defining the optimal and adaptive values of k. In this paper we take another adaptive approach. To avoid difficult interpretations with the choice of the optimal value k for the parameter $\overline{\alpha}_k$ we shall consider that the Hill estimator estimates the value α_1, which may be regarded as a constant approximation of the values α_i, $i = 1, ..., k$.

4 Adaptive selection of the parameter k

This section presents a method of selecting the parameter k in a data driven way. Throughout the paper we shall denote by $|I|$ the number of elements of the set I.

4.1 The adaptive procedure

Let $\mathcal{I}$ be a family of intervals of the form $I = [1, k]$, where $k \in \{1, ..., n-1\}$, such that $|I| \geq 2m_0$, for a prescribed natural number m_0, where m_0 is much smaller that $(n-1)/2$. A special case of the family $\mathcal{I}$ is given by the set of all the intervals $I = [1, k]$, satisfying this condition. Another example used later on in the simulations, is the set $\mathcal{I} = \mathcal{I}_q$ of intervals $I = [1, k]$, with k approximately lying in the geometric grid $\left\{ l : l \leq n,\ l = [m_0 + m_0 q^j],\ j = 1, 2, ... \right\}$, where $q > 1$. In the latter case the numbers m_0 and q will be parameters of the procedure.

The family $\mathcal{I}$ is naturally ordered by the length $|I|$ of $I \in \mathcal{I}$. The idea of our method is to test successfully the hypothesis of no change-point within the interval

I and to select k equal to the first detected change-point. The formal steps of the procedure for selecting the adaptive interval $\widehat{I}$ reads as follows:

INITIALIZATION Start with the smallest interval $I = I_0 \in \mathcal{I}$.

STEP 1 Take the next interval $I \in \mathcal{I}$.

STEP 2 From observations (3.3) test on homogeneity the vector α within the interval I against the change-point alternative, as described in Section 4.2.

STEP 3 If the change point was detected for the interval I, then define $\widehat{I}$ as the interval from one to the detected change-point and stop the procedure, otherwise repeat the procedure from the Step 1. If there was no change-point for all $I \in \mathcal{I}$, then define $\widehat{I} = [1, n-1]$.

The adaptive estimator is defined as $\widehat{\alpha} = \widehat{\alpha}_{\widehat{I}}$, where

$$\widehat{\alpha}_I = \frac{1}{|I|} \sum_{i \in I} Y_i, \tag{4.1}$$

for any interval I. The essential point in the above procedure is the Step 2 which stipulates testing the hypothesis of homogeneity for the interval I. It consists in applying the classical change-point test which is described in the next section.

4.2 Test of homogeneity against the change-point alternative

The test of homogeneity against the change-point alternative is based on the likelihood ratio test statistic. For any interval $I \in \mathcal{I}$ denote by $\mathcal{J}_I$ the set of all subintervals $J \subset I$, $J \in \mathcal{I}$, such that $|I|/2 \le |J| \le |I| - m_0$. For every interval $J \in \mathcal{J}_I$ consider the problem of testing the hypothesis of homogeneity $\alpha_i = \theta$, $i \in I$ against the change-point alternative $\alpha_i = \theta_1$, $i \in J$ and $\alpha_i = \theta_2$, $i \in I \setminus J$ with $\theta_1 \neq \theta_2$. The likelihood ratio test statistic is defined by

$$\begin{aligned} T_{I,J} &= \sup_{\theta_1} L\left(Y_J, \theta_1\right) + \sup_{\theta_2} L\left(Y_{I \setminus J}, \theta_2\right) - \sup_{\theta} L\left(Y_I, \theta\right) \\ &= L\left(Y_J, \widehat{\alpha}_J\right) + L\left(Y_{I\setminus J}, \widehat{\alpha}_{I \setminus J}\right) - L\left(Y_I, \widehat{\alpha}_I\right), \end{aligned}$$

where $\widehat{\alpha}_I$ is the corresponding maximum likelihood estimator defined by (4.1) and

$$L\left(Y_I, \theta\right) = \sum_{i \in I} \log p\left(Y_i, \theta\right).$$

Since in the case under consideration $p\left(y, \theta\right) = \exp\left(-y/\theta\right)/\theta$, one gets

$$\begin{aligned} T_{I,J} &= -\sum_{i \in J} \left[\log \frac{\widehat{\alpha}_J}{\widehat{\alpha}_I} - Y_i \left(\frac{1}{\widehat{\alpha}_I} - \frac{1}{\widehat{\alpha}_J}\right)\right] + \sum_{i \in I \setminus J} \left[\log \frac{\widehat{\alpha}_{I\setminus J}}{\widehat{\alpha}_I} - Y_i \left(\frac{1}{\widehat{\alpha}_I} - \frac{1}{\widehat{\alpha}_{I \setminus J}}\right)\right] \\ &= |J|\, G\left(\frac{\widehat{\alpha}_J}{\widehat{\alpha}_I} - 1\right) + |I \setminus J|\, G\left(\frac{\widehat{\alpha}_{I\setminus J}}{\widehat{\alpha}_I} - 1\right), \end{aligned} \tag{4.2}$$

where $G(x) = x - \log(1+x)$, $x > -1$. The use of Taylor's expansion gives the approximating test statistic

$$\overline{T}_{I,J} = \frac{|J|}{2}\left(\frac{\widehat{\alpha}_J}{\widehat{\alpha}_I} - 1\right)^2 + \frac{|I \setminus J|}{2}\left(\frac{\widehat{\alpha}_{I\setminus J}}{\widehat{\alpha}_I} - 1\right)^2.$$

By simple algebra we can represent the latter statistic in the form

$$\overline{T}_{I,J} = \frac{|J| \cdot |I \setminus J|}{2\,|I|}\left(\frac{\widehat{\alpha}_J - \widehat{\alpha}_{I\setminus J}}{\widehat{\alpha}_I}\right)^2. \tag{4.3}$$

Now the test of homogeneity of α on the interval I can be based on the maximum of all such defined statistics $T_{I,J}$ or $\overline{T}_{I,J}$ over the set $\mathcal{J}_I$. The hypothesis of homogeneity on the interval I will be rejected if

$$T_I = \max_{J\in\mathcal{J}_I} T_{I,J} > t_\gamma, \quad \text{or} \quad \overline{T}_I = \max_{J\in\mathcal{J}_I} \overline{T}_{I,J} > \bar{t}_\gamma,$$

where the critical values t_γ and $\bar{t}_\gamma$ are defined to provide the prescribed rejection probability γ under the hypothesis of homogeneity within the interval I. These values can be computed by Monte-Carlo simulations from the homogeneous model with i.i.d. standard exponential observations Y_i, $i = 1, ..., n$. Here we utilize the fact that under the hypothesis of homogeneity the distributions of the test statistics T_I and $\overline{T}_I$ do not depend on α.

If the hypothesis of the homogeneity of α is rejected on the interval I then the detected change-point k^* corresponds to the length of the interval $J^* \in \mathcal{J}_I$ for which the statistic T_I attains its maximum, i.e.

$$k^* = |J^*|, \quad \text{where} \quad J^* = \arg\max_{J\in\mathcal{J}_I} T_{I,J}.$$

5 Simulation study

The aim of the present simulation study is to demonstrate the numerical performance of the proposed procedure. We focus on the quality of the selected interval I and of the corresponding adaptive estimator. The next figures present box-plots of the length of the selected interval $\widehat{I}$ and of the adaptive estimator $\widehat{\alpha}$ for different values of the parameter $\sqrt{t_\gamma}$ from 500 observations following Pareto and Pareto-log d.f.'s (see a list in the Appendix). The box-plots are obtained from 500 Monte-Carlo realizations. The set $\mathcal{I}$ is a geometric grid with parameters $m_0 = 25$, $q = 1.1$.

In Table 1 the mean absolute error (MAE) of the adaptive estimator $\hat{\alpha}$ w.r.t. the value $\alpha_1 = \alpha(X_{n,1})$ is computed for the d.f.'s introduced above.

The results clearly indicate that the increase of the parameter t_γ results in a smaller variability of the estimator but in a larger bias (in case when the model is not Pareto).

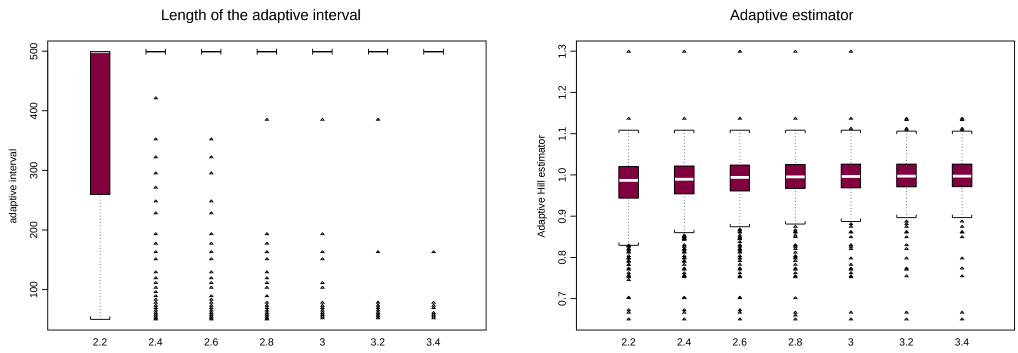

Figure 2: Box-plots of selected intervals and the adaptive estimators for Pareto d.f. from 500 realization.

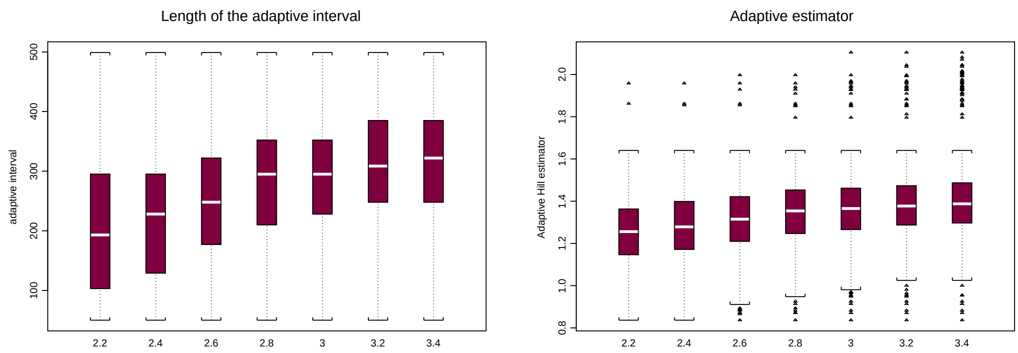

Figure 3: Box-plots of selected intervals and the adaptive estimators for Pareto-log d.f. from 500 realization.

A reasonable compromise is attained for $\sqrt{t_\gamma}$ about 2.6 leading to a relatively stable behavior of the procedure in the Pareto case and to a moderate bias in the non-Pareto case. The numerical simulation for the procedure with the parameter $\sqrt{t_\gamma} = 2.6$ for different values of the sample size n and different distributions (see a list in the Appendix 8) are summarized in Table 2. The other parameters are kept as in the previous case. In this table MAE is computed w.r.t. the value $\alpha_1 = \alpha(X_{n,1})$ for 500 simulations.

In the Appendix 8 we present the box-plots of the length (in %) of the selected interval $\widehat{I}$ and of the adaptive estimator $\widehat{\alpha}$ for different values of n from 500 simulations following different d.f.'s.

Table 1: MAE computed for 500 realizations

	t_γ=2.2	t_γ=2.4	t_γ=2.6	t_γ=2.8	t_γ=3.0	t_γ=3.2	t_γ=3.4	
Pareto	0.0642	0.0583	0.0546	0.0487	0.0459	0.0433	0.0395	
Cauchy-plus	0.1036	0.1076	0.1116	0.1166	0.1204	0.1232	0.1275	
Pareto-log	0.1838	0.2039	0.2231	0.2388	0.2581	0.2854	0.3106	
Pareto-CP	0.0746	0.0704	0.0697	0.0658	0.0642	0.0626	0.0615	

Table 2: MAE computed for 500 realizations

	n=200	n=300	n=400	n=500	n=800	n=1000	2000	n=3000
Pareto	0.0573	0.0507	0.0473	0.0521	0.0456	0.0495	0.0453	0.0415
Cauchy-plus	0.1483	0.1210	0.1133	0.1155	0.0846	0.0943	0.0720	0.0577
Pareto-log	0.2544	0.2309	0.2274	0.2178	0.1895	0.1828	0.1783	0.1713
GPD	0.2563	0.1829	0.1770	0.1564	0.1488	0.1301	0.1171	0.1095
Hall model	0.2498	0.2448	0.2377	0.2439	0.2344	0.2222	0.1961	0.1699
Pareto-CP	0.1001	0.0881	0.0737	0.0669	0.0566	0.0558	0.0432	0.0321
Standard Normal tail	0.2273	0.1718	0.1438	0.1242	0.0983	0.0941	0.0689	0.0654
Standard Exponential	0.2989	0.2370	0.1913	0.1707	0.1432	0.1373	0.1133	0.1007

6 Theoretical results

This section discusses some theoretical properties of the procedure presented in Section 4. Let $t_\gamma > 0$ and $\bar{t}_\gamma > 0$ be the critical values entering the definition of the change point tests from Section 4.2.

6.1 Properties of the selected interval

We start with results concerning the choice of the interval of homogeneity. We will ensure that the following two properties hold:

A. The intervals of homogeneity are accepted with high probabilities.

B. The intervals of non-homogeneity are rejected with high probabilities at least in some special cases, for instance, for the change-point model.

Consider first the property A. The assumption that the vector α is constant on some interval I can be quite restrictive for practical applications. Therefore the desirable property would be that the procedure accepts any interval $I \in \mathcal{I}$ for which α_i can be well approximated by a constant within the interval I. Let I be an interval and let α_I be the average of the α_i's over the interval I :

$$\alpha_I = \frac{1}{|I|} \sum_{i \in I} \alpha_i.$$

The non-homogeneity of the α_i's within the interval I can be naturally measured by the value

$$\Delta_I = \max_{i\in I}\left|\frac{\alpha_i}{\alpha_I} - 1\right|.$$

We say that I is a "good" interval if the value Δ_I is small. The next result claims that a "good" interval I will be accepted by the procedure with a high probability provided that the critical value t_γ was taken sufficiently large.

For every interval $I \in \mathcal{I}$, denote

$$S_I = \frac{1}{|I|}\sum_{i\in I}\alpha_i\left(\xi_i - 1\right) \quad \text{and} \quad V_I^2 = \sum_{i\in I}\alpha_i^2.$$

For given intervals $I \in \mathcal{I}$ and $J \in \mathcal{J}_I$, denote $J^c = I\diagdown J$ and, with a real $\lambda > 0$, define the events

$$\Omega_{I,J} = \left\{|S_I| \le \frac{\lambda V_I}{|I|},\ |S_J| \le \frac{\lambda V_J}{|J|},\ |S_{J^c}| \le \frac{\lambda V_{J^c}}{|J^c|}\right\}$$

and

$$\Omega_I = \bigcap_{J\in\mathcal{J}_I}\Omega_{I,J}\,.$$

The function $G\left(x\right)$ is defined for all $x > -1$. We extend it to the whole real line by defining $G\left(x\right) = +\infty$ for $x \le -1$.

Theorem 6.1. *A. Let $\gamma \in (0,1)$ and $I \in \mathcal{I}$. Let the numbers λ and m_0 be such that $\lambda \ge 2\sqrt{\log\frac{2|\mathcal{J}_I|+1}{\gamma}}$ and $\sqrt{m_0} > \frac{3}{2}\lambda\left(1+\Delta_I\right)$. Then $P\left(\Omega_I\right) \ge 1-\gamma$.*

B. Let $\gamma \in (0,1)$ and $I \in \mathcal{I}$. Let the numbers λ and m_0 be such that $\lambda \ge 2\sqrt{\log\frac{2|\mathcal{J}_I|+1}{\gamma}}$ and $\sqrt{m_0} > 3\lambda\left(1+\Delta_I\right)$. If Δ_I fulfills

$$G\left(-3\Delta_I - 3\lambda\left(1+\Delta_I\right)m_0^{-1/2}\right) \le \frac{4t_\gamma}{|I|}, \tag{6.1}$$

then on the set Ω_I it holds $T_I \le t_\gamma$.

C. Let $\gamma \in (0,1)$ and $I \in \mathcal{I}$. Let the numbers λ and m_0 be such that $\lambda \ge 2\sqrt{\log\frac{2|\mathcal{J}_I|+1}{\gamma}}$ and $\sqrt{m_0} > 3\lambda\left(1+\Delta_I\right)$. If Δ_I fulfills

$$\Delta_I \le \frac{\frac{2\sqrt{2}}{3}t_\gamma^{1/2}\left|I\right|^{-1/2} - \lambda m_0^{-1/2}}{1+\lambda m_0^{-1/2}},$$

then on the set Ω_I it holds $\overline{T}_I \le t_\gamma$.

Remark 6.2. *The condition on Δ_I from the part C of the theorem is similar to the condition (6.1) with the function $G(u)$ replaced by $u^2/2$. Moreover, the condition (6.1) follows from $\Delta_I \le (Ct_\gamma^{-1/2}\left|I\right|^{-1/2} - \lambda m_0^{-1/2})/(1+\lambda m_0^{-1/2})$ with some constant $C > 2\sqrt{2}/3$ provided that $3\Delta_I + 3\lambda\left(1+\Delta_I\right)m_0^{-1/2} < 1/2$, see Lemma 7.3.*

An immediate corollary of this result is an upper bound of the probability of rejecting a "good" interval I.

Corollary 6.3. *Under the conditions of the point B or C of Theorem 6.1 it holds respectively*

$$P\left(T_I > t_\gamma\right) < \gamma \quad or \quad P\left(\overline{T}_I > t_\gamma\right) < \gamma.$$

Now let us turn to the property B of the intervals of homogeneity. Consider the special case when the vector $\alpha = (\alpha_1, ..., \alpha_n)$ is piecewise constant. In this case an interval I is "good" if it does not contain a change point. The best choice of I can be defined as the interval $I^* = [1, k^*]$, where k^* is the first change point. Theorem 6.1 claims that the interval I^* will be accepted with high probability. The next result shows that all larger intervals will be rejected with high probability, thus implying that $\widehat{I}$ approximately equals I^*.

Theorem 6.4. *Let $\gamma \in (0,1)$ and $2\sqrt{\log\frac{3}{\gamma}} \leq \lambda \leq \sqrt{m}$. Assume that $\alpha_i = \alpha$, for $i \in I^*$, and $\alpha_i = \beta$, for $i \in I \setminus I^*$, where $I = [1, k^* + m]$ and $\alpha \neq \beta$. If m satisfies $m \leq k^*$ and*

$$\sqrt{m} \geq \max\left\{d^{-1}\left(3\sqrt{t_\gamma} + \lambda\right), 4t_\gamma\right\}, \tag{6.2}$$

where $d = |\alpha - \beta| / (2\alpha + |\alpha - \beta|)$, then

$$P\left(T_I \leq t_\gamma\right) \leq \gamma \quad and \quad P\left(\overline{T}_I \leq t_\gamma/2\right) \leq \gamma.$$

6.2 Properties of the adaptive estimator $\widehat{\alpha}$.

Let $\widehat{I}$ be the interval computed by the adaptive procedure described in Section 4.1 with the test statistic $T_{I,J}$. The next assertions describe the accuracy of the adaptive estimator $\widehat{\alpha} = \widehat{\alpha}_{\widehat{I}}$ under the condition that $\widehat{I} \supset I^*$, where $I^* \in \mathcal{I}$ is a "good" interval.

Theorem 6.5. *Let $\gamma \in (0,1)$ and $I \in \mathcal{I}$. Let the numbers λ and m_0 be such that $\lambda \geq 2\sqrt{\log\frac{2|\mathcal{J}_I|+1}{\gamma}}$ and $\sqrt{m_0} > \max\left\{\sqrt{4t_\gamma}, \frac{3}{2}\lambda\left(1 + \Delta_I\right)\right\}$. Let the interval $I^* \in \mathcal{I}$ be such that $I^* \in \mathcal{J}_I$. If $T_I \leq t_\gamma$, then on the set Ω_I, it holds*

$$\left|\frac{\widehat{\alpha}_I - \widehat{\alpha}_{I^*}}{\widehat{\alpha}_{I^*}}\right| \leq \frac{\rho}{1-\rho},$$

where $\rho = 2\sqrt{t_\gamma\left|I^*\right|^{-1}}$.

From Theorem 6.5 it follows that if $\widehat{\alpha}_{I^*}$ provides a "good" estimate of α_{I^*}, then the adaptive estimator also provides a "good" estimate of α_{I^*}. A precise statement is given in the next corollary.

Corollary 6.6. *Let $\gamma \in (0,1)$ and $I \in \mathcal{I}$. Let the numbers λ and m_0 be such that $\lambda \geq 2\sqrt{\log \frac{2|\mathcal{J}_I|+1}{\gamma}}$ and $\sqrt{m_0} > \max\left\{\sqrt{4t_\gamma}, \frac{3}{2}\lambda\left(1+\Delta_I\right)\right\}$. Let the intervals $I^* \in \mathcal{I}$ and I be such that $I^* \in \mathcal{J}_{\widehat{I}(\omega)}$ and $\widehat{I}\left(\omega\right) \in \mathcal{J}_I$, for any $\omega \in \Omega_I$. Then on the set Ω_I the adaptive estimator $\widehat{\alpha}$ fulfills*

$$\frac{\left|\widehat{\alpha}-\alpha_{I^*}\right|}{\alpha_{I^*}} \leq \frac{1}{1-\rho}\frac{\lambda\left(1+\Delta_{I^*}\right)}{\sqrt{|I^*|}}+\frac{\rho}{1-\rho},$$

where $\rho = 2\sqrt{t_\gamma\left|I^*\right|^{-1}}$.

Similar properties can be established for the statistic $\overline{T}_{I,J}$.

7 Proofs of the main results

7.1 Auxiliary statements.

Lemma 7.1. *Let $\xi_1,...,\xi_m$ be i.i.d. standard exponential r.v.'s and the numbers $\beta_1,...,\beta_m$ satisfy the condition*

$$\left|\frac{\beta_i}{\overline{\beta}}-1\right| \leq \Delta, \quad i=1,...,m,$$

where $\overline{\beta} = (\beta_1+...+\beta_m)/m$ and $\Delta \in [0,1]$. Then, for every $\lambda \leq \frac{2}{3}\sqrt{m}/(1+\Delta)$,

$$P\left(\left|\sum_{i=1}^m \beta_i(\xi_i-1)\right| > \lambda V_m\right) \leq 2e^{-\lambda^2/4},$$

where $V_m^2 = \beta_1^2+...+\beta_m^2$.

Proof. By Chebyshev inequality, for any $u>0$,

$$P\left(\left|\sum_{i=1}^m \beta_i\xi_i\right| > \lambda V_m\right) \leq \frac{E\exp\left(u\sum_{i=1}^m \beta_i\left(\xi_i-1\right)\right)}{\exp\left(u\lambda V_m\right)}.$$

Since $\xi_1,...,\xi_n$ are independent, for any $u < \min\left\{\beta_i^{-1}\right\}$,

$$E\exp\left(u\sum_{i=1}^m \beta_i\left(\xi_i-1\right)\right) = \prod_{i=1}^m E\exp\left(u\beta_i\left(\xi_i-1\right)\right) = \prod_{i=1}^m \frac{\exp\left(-u\beta_i\right)}{1-u\beta_i}.$$

Therefore

$$P\left(\left|\sum_{i=1}^m \beta_i\xi_i\right| > \lambda V_m\right) \leq \exp\left(-u\lambda V_m - u\sum_{i=1}^m \beta_i - \sum_{i=1}^m \log\left(1-u\beta_i\right)\right).$$

This inequality with $u = \frac{\lambda}{2V_m}$ and the elementary inequality $-\log\left(1-x\right) \leq x+x^2$, for $x \leq 1/3$ yield

$$P\left(\left|\sum_{i=1}^m \beta_i\xi_i\right| > \lambda V_m\right) \leq \exp\left(-u\lambda V_m - u^2V_m^2\right) = \exp\left(-\frac{\lambda^2}{4}\right).$$

It remains to check that $\lambda \leq \frac{2\sqrt{m}}{3(1+\Delta)}$ implies that $u = \frac{\lambda}{2V_m} < \min\left\{\beta_i^{-1}\right\}$. Indeed $V_m^2 = \sum_{i=1}^m \beta_i^2 \geq m\overline{\beta}^2$ and therefore,

$$\beta_i u = \frac{\lambda \beta_i}{2V_m} \leq \frac{\lambda \beta_i}{2\overline{\beta}\sqrt{m}} \leq \frac{\lambda\left(1+\Delta\right)}{2\sqrt{m}} \leq \frac{1}{3},$$

which proves the lemma. □

In the proofs we shall use the following bounds. Recall that $G\left(x\right) = +\infty$, for $x \leq -1$.

Lemma 7.2. *For any $\delta \in [0,1]$ and any real x, the function $G\left(\cdot\right)$ fulfills*

$$\delta\left(1-\delta\right)G\left(|x|\right) \leq \delta G\left((1-\delta)x\right) + \left(1-\delta\right)G\left(-\delta x\right) \leq \delta\left(1-\delta\right)G\left(-|x|\right). \tag{7.1}$$

Proof. The proof of these bounds is based on the simple fact that the function

$$H\left(x\right) = 2G\left(x\right)/x^2, \quad x > -1, \tag{7.2}$$

is monotonously decreasing. □

Lemma 7.3. *Let $G_+^{-1}\left(x\right),\ x \geq 0$ be the inverse of the function $G\left(\cdot\right)$ on the interval $[0,\infty)$. Then*

$$G_+^{-1}\left(x\right) \leq 2\sqrt{x}, \quad 0 \leq x \leq 1/2.$$

Let $G_-^{-1}\left(x\right),\ x \geq 0$ be the inverse of the function $G\left(\cdot\right)$ on the interval $(-1,0]$. Then

$$-G_-^{-1}\left(x\right) \geq \sqrt{x}, \quad -1/2 \leq x \leq 0.$$

Proof. For any $a > 0$ and $x \in [0, G(a)]$ it holds $G_+^{-1}\left(x\right) \leq \sqrt{\frac{2x}{H(a)}}$, where $H\left(\cdot\right)$ is defined by (7.2). Taking $a = 1.4$ one gets the first inequality. If $a \in (-1,0]$ and $x \in [-G\left(a\right), 0]$ it holds $-G_-^{-1}\left(x\right) \geq \sqrt{\frac{2x}{H(a)}}$. The second inequality is obtained by putting $a = -0.7$. □

We shall also make use of the following bounds of the statistic $T_{I,J}$.

Lemma 7.4. *Let $\varepsilon = |J| \,/\, |I|$ and $R_{I,J} = \frac{\widehat{\alpha}_J - \widehat{\alpha}_{J^c}}{\widehat{\alpha}_I}$. Then the statistic $T_{I,J}$ satisfies*

$$\varepsilon\left(1-\varepsilon\right)|I|\,G\left(|R_{I,J}|\right) \leq T_{I,J} \leq \varepsilon\left(1-\varepsilon\right)|I|\,G\left(-|R_{I,J}|\right). \tag{7.3}$$

Proof. The trivial equality $|I|\,\widehat{\alpha}_I = |J|\,\widehat{\alpha}_J + |J^c|\,\widehat{\alpha}_{J^c}$ implies

$$\frac{\widehat{\alpha}_J}{\widehat{\alpha}_I} - 1 = \left(1-\varepsilon\right)R_{I,J} \quad \text{and} \quad \frac{\widehat{\alpha}_{J^c}}{\widehat{\alpha}_I} - 1 = -\varepsilon R_{I,J}. \tag{7.4}$$

Then the statistic $T_{I,J}$ can be written as

$$T_{I,J} = |I|\left[\varepsilon G\left((1-\varepsilon)R_{I,J}\right) + \left(1-\varepsilon\right)G\left(-\varepsilon R_{I,J}\right)\right]. \tag{7.5}$$

Using (7.1) one gets the required bounds. □

7.2 Proof of Theorem 6.1

Let $I \in \mathcal{I}$. For any $J \in \mathcal{J}_I$ denote $J^c = I \setminus J$. In the following J' denotes one of the intervals J, J^c or I. The definition of the sets $\mathcal{I}$ and $\mathcal{J}_I$ implies that $|J'| \geq m_0$.

Note that the estimator $\widehat{\alpha}_{J'}$ can be written as $\widehat{\alpha}_{J'} = \alpha_{J'} + S_{J'}$. Then, using Lemma 7.1, for any $\lambda \leq \frac{2}{3}\sqrt{m_0}/(1+\Delta_I)$, one gets

$$P(\Omega_I) \geq 1 - \sum_{J \in \mathcal{J}_I} P\left(\Omega_{I,J}^c\right) \geq 1 - (2|\mathcal{J}_I| + 1) \exp\left(-\lambda^2/4\right).$$

With $\lambda \geq 2\sqrt{\log \frac{2|\mathcal{J}_I|+1}{\gamma}}$, it holds

$$P(\Omega_I) \geq 1 - \gamma,$$

thus proving the part A of the theorem.

For the part B we have to show that on the random set Ω_I the statistics $T_{I,J}$ and $\overline{T}_{I,J}$ obey $|T_{I,J}| \leq t_\gamma$ and $\left|\overline{T}_{I,J}\right| \leq \bar{t}_\gamma$, for any $J \in \mathcal{J}_I$.

For the proof we need some inequalities. Note that each α_i satisfies $\alpha_i \leq \alpha_I (1 + \Delta_I)$, for $i \in I$, and by summing α_i^2 over $i \in J'$, it follows

$$V_{J'}^2 \leq (1+\Delta_I)^2 \alpha_I^2 \left|J'\right|. \tag{7.6}$$

The latter inequality implies that, on the set Ω_I, it holds

$$|S_{J'}| \leq \lambda V_{J'} / \left|J'\right| \leq \lambda \alpha_I (1+\Delta_I) \left|J'\right|^{-1/2}. \tag{7.7}$$

The decomposition $\widehat{\alpha}_{J'} = \alpha_{J'} + S_{J'}$ and the inequality (7.7) imply that, on the set Ω_I,

$$\left|\frac{\widehat{\alpha}_{J'}}{\alpha_{J'}} - 1\right| \leq \lambda (1+\Delta_I) \left|J'\right|^{-1/2}. \tag{7.8}$$

Note that $\left|\frac{\alpha_J - \alpha_{J^c}}{\alpha_I}\right| \leq 2\Delta_I$ and $|J'| \geq m_0$. Then, under the assumption $\sqrt{m_0} \geq 3\lambda(1+\Delta_I)$, the inequality (7.8) implies

$$\begin{aligned}
|R_{I,J}| &\leq \frac{2\Delta_I + \lambda(1+\Delta_I)\left(|J|^{-1/2} + |J^c|^{-1/2}\right)}{1 - \lambda(1+\Delta_I)|I|^{-1/2}} \\
&\leq \frac{2\Delta_I + 2\lambda(1+\Delta_I) m_0^{-1/2}}{1 - \lambda(1+\Delta_I) m_0^{-1/2}} \\
&\leq 3\Delta_I + 3\lambda(1+\Delta_I) m_0^{-1/2}.
\end{aligned} \tag{7.9}$$

We consider first the case of statistic T_I. The bounds (7.3) and (7.9) yield

$$T_{I,J} \leq \varepsilon(1-\varepsilon)|I| G(-|R_{I,J}|) \leq \frac{|I|}{4} G\left(-3\Delta_I - 3\lambda(1+\Delta_I) m_0^{-1/2}\right) \leq t_\gamma,$$

and the assertion of Theorem 6.1 concerning T_I follows.

In the same way we prove the assertion concerning $\overline{T}_I$. The inequality $|J| \cdot |J^c| \leq |I|^2/4$ implies, on the set Ω_I,

$$\overline{T}_{I,J} \leq \frac{|I|}{4} \frac{\left[3\Delta_I + 3\lambda\left(1+\Delta_I\right) m_0^{-1/2}\right]^2}{2} \leq \widehat{t}_\gamma.$$

Theorem 6.1 is proved.

7.3 Proof of Theorem 6.4

To keep the same notations as in Theorem 6.1 denote $J = I^*$, $J^c = I \backslash J = [k^*+1, k^*+m]$. Using Lemma 7.1, for any λ and m_0 satisfying $2\sqrt{\log \frac{1}{3\gamma}} \leq \lambda \leq \frac{2}{3}\sqrt{m_0}/(1+\Delta_I)$, one gets

$$P\left(\Omega_{I,J}\right) \geq 1 - 3e^{-\lambda^2/4} \geq 1 - \gamma.$$

It suffices to show that the event $\Omega_{I,J}$ implies $T_{I,J} \geq t_\gamma$. The lower bound in Lemma 7.4 implies

$$T_{I,J} \geq \varepsilon\left(1-\varepsilon\right)|I|\, G\left(\left|R_{I,J}\right|\right),$$

with $\varepsilon = |J|/|I|$ and $R_{I,J} = \frac{\widehat{\alpha}_J - \widehat{\alpha}_{J^c}}{\widehat{\alpha}_I}$. Since $k^* \geq m$ it follows that $\varepsilon = k^*/(k^*+m) \geq 1/2$. This and $1-\varepsilon = m/|I|$ imply

$$T_{I,J} \geq \frac{1}{2} mG\left(\left|R_{I,J}\right|\right), \tag{7.10}$$

Note that $V_J^2 = k^*\alpha^2$, $V_{J^c}^2 = m\beta^2$ and $V_I \leq V_J + V_{J^c}$. Then, similarly to the proof of Theorem 6.1, on the set $\Omega_{I,J}$, it holds

$$\left|R_{I,J}\right| \geq \frac{\left|\alpha_J - \alpha_{J^c}\right| - \lambda\left(\alpha/\sqrt{k^*} + \beta/\sqrt{m}\right)}{\alpha_I + \lambda\left(\alpha/\sqrt{k^*} + \beta/\sqrt{m}\right)}.$$

For the change point model $\alpha_J = \alpha$, $\alpha_{J^c} = \beta$ and $\alpha_I = \alpha k^*/\left(k^*+m\right) + \beta m/\left(k^*+m\right)$. This yields

$$\left|R_{I,J}\right| \geq \frac{b - \lambda\left(1/\sqrt{k^*} + \left(1+b\right)/\sqrt{m}\right)}{1 + b\frac{m}{k^*+m} + \lambda\left(1/\sqrt{k^*} + \left(1+b\right)/\sqrt{m}\right)},$$

where $b = \left|\frac{\beta}{\alpha} - 1\right|$. It is easy to see that, for a fixed m, the minimum over $k^* \geq m$ of the latter expression is attained for $k^* = m$. Therefore

$$\left|R_{I,J}\right| \geq \frac{b - \lambda\left(2+b\right)/\sqrt{m}}{1 + b/2 + \lambda\left(2+b\right)/\sqrt{m}} = \frac{d - \lambda/\sqrt{m}}{1/2 + \lambda/\sqrt{m}},$$

where $d = b/\left(2+b\right)$. Together with (7.10) this yields

$$T_{I,J} \geq \frac{1}{2} mG\left(\frac{d - \lambda/\sqrt{m}}{1/2 + \lambda/\sqrt{m}}\right).$$

Now the assertion of the theorem amounts to prove that the right hand side in the latter inequality is greater than t_γ. This is equivalent to

$$\frac{d-\lambda/\sqrt{m}}{1/2+\lambda/\sqrt{m}} \geq G_+^{-1}\left(\frac{2t_\gamma}{m}\right).$$

Since $G_+^{-1}(x) \leq 2\sqrt{x}$, for all $x \in [0,1/2]$ and $m > 4t_\gamma$, it suffices to show that

$$\frac{d-\lambda/\sqrt{m}}{1/2+\lambda/\sqrt{m}} \geq 2\sqrt{\frac{t_\gamma}{m}}.$$

The latter inequality is implied by the conditions (6.2) and $\lambda \leq \sqrt{m}$ of the theorem. This concludes the proof.

7.4 Proof of Theorem 6.5

To keep the same notations as in the proof of Theorem 6.1 let $J = I^*$, $J^c = I \setminus I^*$, $\varepsilon = |J| / |I|$ and $R_{I,J} = (\widehat{\alpha}_J - \widehat{\alpha}_{J^c}) / \widehat{\alpha}_I$. It is clear that $T_I \leq t_\gamma$ implies $T_{I,J} \leq t_\gamma$. The bounds (7.1) imply

$$|I|\, \varepsilon (1-\varepsilon)\, G(|R_{I,J}|) \leq T_{I,J} \leq t_\gamma,$$

from which it follows that

$$|R_{I,J}| \leq G_+^{-1}\left(\frac{t_\gamma}{\varepsilon(1-\varepsilon)|I|}\right),$$

where $G_+^{-1}(x)$, $x \geq 0$ is the inverse of the function $G(\cdot)$ on the interval $[0,\infty)$. Now by the definition of the set $\mathcal{J}_I$ one has $\varepsilon = |J| / |I| \geq 1/2$. Since $m_0 > 4t_\gamma$ it holds

$$\frac{t_\gamma}{\varepsilon(1-\varepsilon)|I|} \leq \frac{\frac{1}{4}m_0}{\frac{1}{2}|J|} \leq \frac{1}{2}.$$

An applications of the upper bound in Lemma 7.3 yields

$$|R_{I,J}| \leq 2\sqrt{\frac{t_\gamma}{\varepsilon(1-\varepsilon)|I|}}.$$

From the identities (7.4) it follows that $R_{I,J} = \left(\frac{\widehat{\alpha}_J}{\widehat{\alpha}_I} - 1\right) / (1-\varepsilon)$, which together with the previous inequality gives

$$\left|\frac{\widehat{\alpha}_J}{\widehat{\alpha}_I} - 1\right| \leq \frac{2\sqrt{(1-\varepsilon)\, t_\gamma}}{\sqrt{\varepsilon\,|I|}} \leq \frac{2\sqrt{t_\gamma}}{\sqrt{|J|}}.$$

This implies

$$\left|\frac{\delta}{1-\delta}\right| \leq 2\sqrt{t_\gamma\,|J|^{-1}},$$

where $\delta = (\widehat{\alpha}_J - \widehat{\alpha}_I) / \widehat{\alpha}_J$, which in turn implies $|\delta| \leq \rho / (1-\rho)$, where $\rho = 2\sqrt{t_\gamma\,|J|^{-1}}$, and the assertion concerning T_I follows. The case of the statistic $\overline{T}_I$ can be handled in the same way.

7.5 Proof of Corollary 6.6

Since $\Omega_{I'} \subset \Omega_I$, for any $I' \subset I$, Theorem 6.5 implies that on the set Ω_I,

$$|\widehat{\alpha}_I - \widehat{\alpha}_{I^*}| \leq \widehat{\alpha}_{I^*} \frac{\rho}{1-\rho}.$$

From this it follows that, on the set Ω_I,

$$|\widehat{\alpha} - \alpha_{I^*}| \leq |\widehat{\alpha} - \widehat{\alpha}_{I^*}| + |\widehat{\alpha}_{I^*} - \alpha_{I^*}| \leq \frac{\rho}{1-\rho}\alpha_{I^*} + \frac{1}{1-\rho}\left|\widehat{\alpha}_{I^*} - \alpha_{I^*}\right|.$$

Since, on the set Ω_I,

$$|\widehat{\alpha}_{I^*} - \alpha_{I^*}| = |S_{I^*}| \leq \frac{\lambda V_{I^*}}{|I^*|},$$

one gets

$$\frac{|\widehat{\alpha} - \alpha_{I^*}|}{\alpha_{I^*}} \leq \frac{1}{1-\rho}\frac{\lambda V_{I^*}}{\alpha_{I^*}\,|I^*|} + \frac{\rho}{1-\rho}.$$

The inequality $V_{I^*}^2 \leq (1+\Delta_{I^*})^2\,\alpha_{I^*}^2\,|I^*|$ (see (7.6)) implies

$$\frac{|\widehat{\alpha} - \alpha_{I^*}|}{\alpha_{I^*}} \leq \frac{1}{1-\rho}\frac{\lambda\,(1+\Delta_{I^*})}{\sqrt{|I^*|}} + \frac{\rho}{1-\rho}.$$

References

[1] Barron, A., Birgé, L. and Massart, P. (1999). Risk bounds for model selection via penalization. *Probability Theory and Related Fields,* **113**, 301-413.

[2] Beirlant, J., Teugels, J. and Vinysaker, P. (1996). Practical analysis of extreme values, Leuven University Press.

[3] Beirlant, J., Dierskx, G., Goegebeur, Y. et Matthys, G. (1999). Tail index estimation and an exponential regression model. *Extremes,* **2:2**, 177-200.

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

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

[6] Deheuvels, J., Haeusler, E. and Masson, D. (1988). Almost sure convergence of the Hill estimator. *Mathematical Proceedings of the Cambridge Philosophical society,* **104**, 371-384.

[7] Dress, H. (2001). Minimax risk bounds in extreme value theory. *Ann. Statist.,* **29**, 266-294.

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

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

[10] de Haan, L. and Resnik, S. I. (1980). A simple asymptotic estimate for the index of a stable distribution. *J. Roy. Statist. Soc.,* Ser. B **42**, 83-87.

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

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

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

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

[15] Pickands, J. (1975). III, Statistical inference using extreme order statistics. *Ann. Statist.,* **3**, 119-131.

[16] Picard, D. and Tribulieu, K. (2002). Evolutionary Pareto distributions. *Ann. I. H. Poincaré - PR* **38**, 1009-1022.

[17] Reiss, R.-D. (1989). Approximate distributions of order statistics: with applications to nonparametric statistics. Springer.

[18] Resnik, S. and Starica, C. (1997). Smoothing the Hill estimator. *Adv. in Appl. Probab.* **29**, 271-293.

[19] Resnik, S. and Starica, C. (1998). Tail index estimation for dependent data. *Ann. Appl. Probab.,* **8**, 1156-1183.

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

[21] Seneta, E. (1976). Regularly varying functions. *Lecture Notes in Mathematics,* Vol. **508**. Springer.

[22] Spokoiny, V. (1998). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *Ann. Statist.,* **26** no. 4, 1356–1378.

8 Appendix

Table 3: The list of distribution functions used in the simulations.

	$F(x)$	Parameters
Pareto	$1 - x^{-1/\alpha},\ x \geq 1$	$\alpha = 1$
Pareto-log	$F(x) = 1 - (x/e)^{-1/\alpha} \log x,\ x \geq e$	$\alpha = 1$
Pareto-CP	$1 - \left(\frac{x}{x_1}\right)^{-1/\alpha_1},\ \text{if } x_1 \leq x < x_2$ $1 - \left(\frac{x_2}{x_1}\right)^{-1/\alpha_1} \left(\frac{x}{x_2}\right)^{-1/\alpha_2},\ \text{if } x > x_2$	$\alpha_1 = 1/2,\ \alpha_2 = 1$ $x_1 = 1,\ x_2 = 5$
Cauchy-plus	$F(x) = \frac{2}{\pi} \arctan x,\ x \geq 0$	
GPD	$1 - (1 + \alpha \frac{x-a}{\sigma})^{-1/\alpha},\ x \geq a$	$a = 0, \sigma = 1, \alpha = 1$
Hall model	$1 - cx^{-1/\alpha}(1 + x^{-1/\beta}),\ x \geq 1$	$\alpha = 1, \beta = 1$

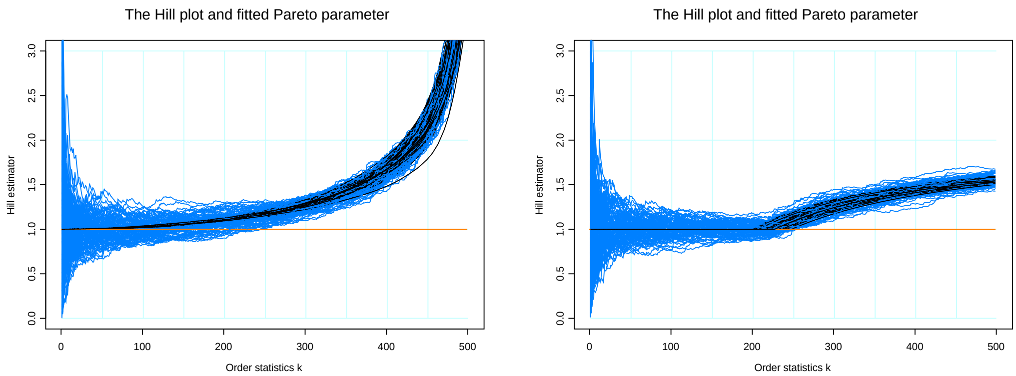

Figure 4: 100 realizations of the Hill estimator for Cauchy-plus (left) and Pareto-CP (right) d.f.'s and the corresponding fitted Pareto parameters. Here the dark lines represent the fitted Pareto parameter computed from the approximation formula (3.5) and the light ones are the corresponding Hill plots.

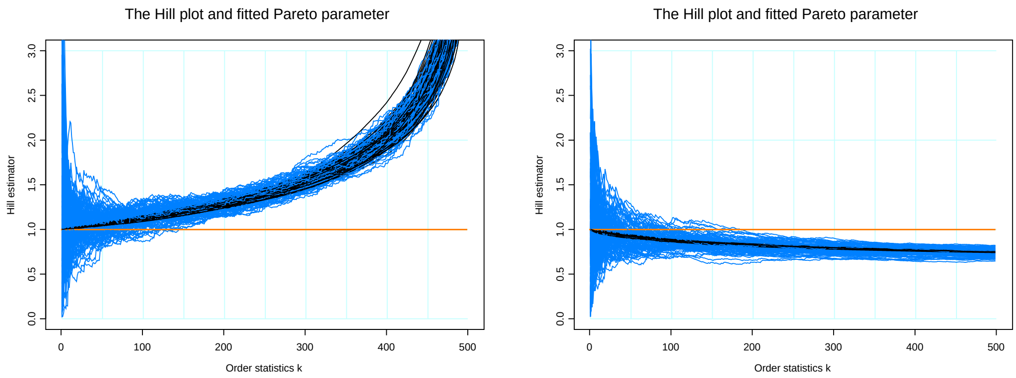

Figure 5: 100 realizations of the Hill estimator for GPD (left) d.f. and for the Hall model (right) and the corresponding fitted Pareto parameters. Here the dark lines represent the fitted Pareto parameter computed from the approximation formula (3.5) and the light ones are the corresponding Hill plots.

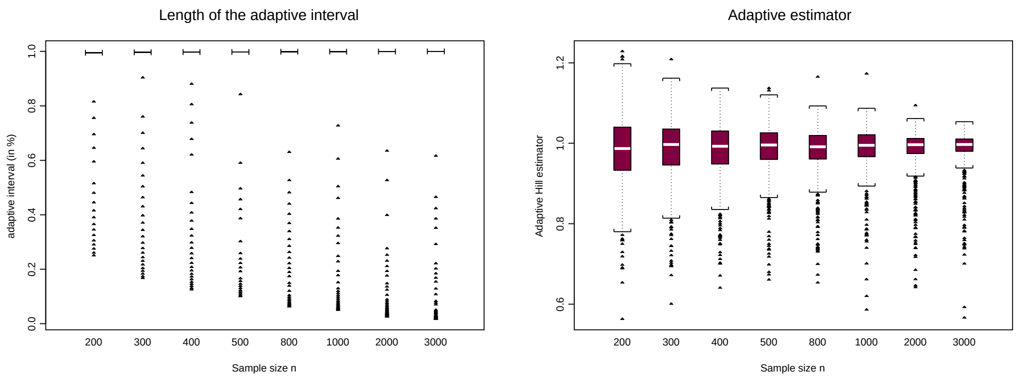

Figure 6: Box-plots of selected intervals (in %) and the adaptive estimators for Pareto d.f. from 500 realization for different sample sizes.

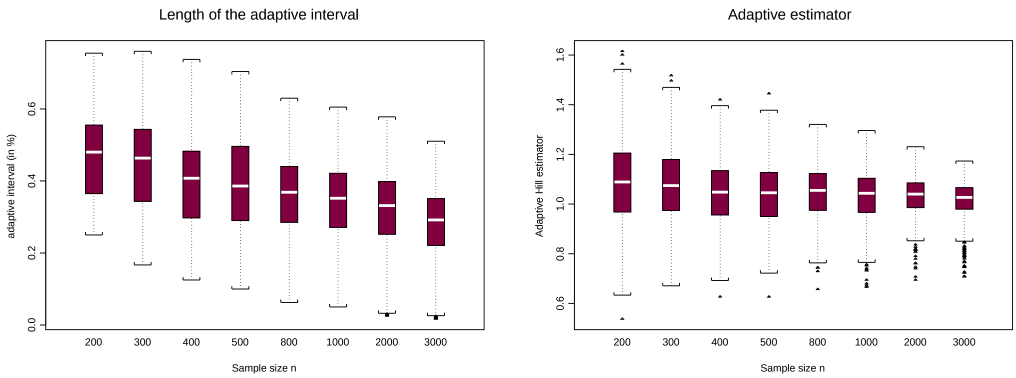

Figure 7: Box-plots of selected intervals (in %) and the adaptive estimators for Cauchy-plus d.f. from 500 realization for different sample sizes.

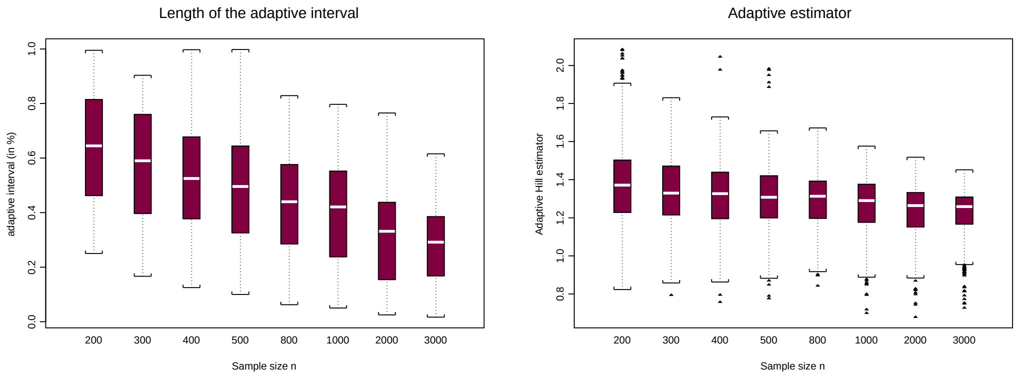

Figure 8: Box-plots of selected intervals (in %) and the adaptive estimators for Pareto-log d.f. from 500 realization for different sample sizes.

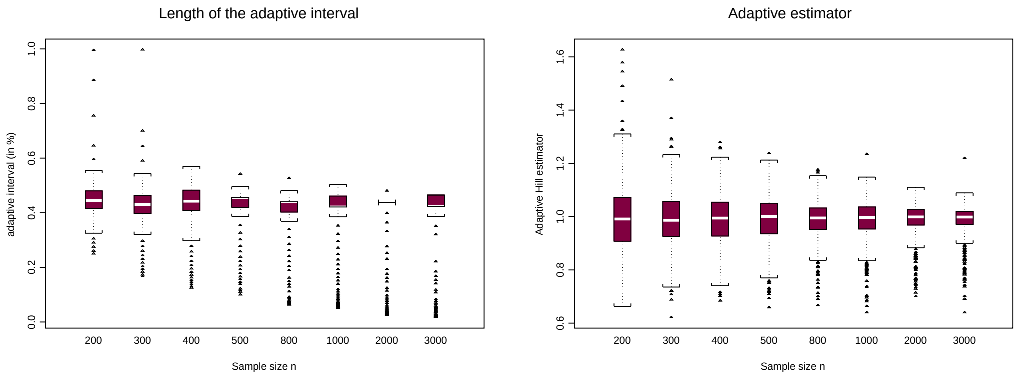

Figure 9: Box-plots of selected intervals (in %) and the adaptive estimators for Pareto-CP d.f. from 500 realization for different sample sizes.