Which Extreme Values are Really Extremes ?

Jesús Gonzalo *
Statistics and Econometrics
U. Carlos III de Madrid
jgonzalo@est-econ.uc3m.es

José Olmo
Statistics and Econometrics
U. Carlos III de Madrid
olmob@est-econ.uc3m.es

January 2003

Abstract

The aim of this paper is to give a formal definition and consistent estimates of the extremes of a population. This definition relies on a threshold value that delimits the extremes and on the uniform convergence of the distribution of these extremes to a Pareto type distribution. The tail parameter of this Pareto type distribution is the tail index of the data distribution. The estimator of the threshold is anchored in the Kolmogorov-Smirnov distance between consistent estimates of those two distributions. Our estimator is consistent and via the construction of confidence intervals for the tail index (derived from our threshold estimator) we overcome the bias problems of the usual tail index estimators (Hill or Pickands). The paper also explores the validity of our definition for standard sample sizes. For this purpose, a hypothesis test is designed in order to reject extremes estimates that are not really extremes. Applications for different stock returns are presented.

Keywords: Bootstrap, Goodness of fit test, Hill estimator, Kolmogorov-Smirnov distance, Balkema and De-Haan, Pickands Theorem, Tail index.

JEL: C12, C13, C14, C15, G10

*Corresponding Address: http://www.uc3m.es/uc3m/dpto/DEE/profesorado/personal/jgonzalo. Financial support DGCYT Grant (SEC01-0890) is gratefully acknowledged. We thank participants in the Conference on Extremal Events in Finance celebrated in Montreal.

1 Introduction

None doubts that Risk Management is one of the most important innovations of the 20th century. The question one would like to answer is: "If things go wrong, how wrong can they go?" The variance used as a risk measure is unable to answer this question, and therefore alternative measures regarding possible values out of the range of available information need to be defined. Extreme value theory (EVT) provides some tools to construct these new risk measures: Value at Risk (VaR), Expected Shortfall or the tail index of a distribution. All these measures need to start by identifying which values are extreme values. In practice this is done by graphical methods like QQ-plot, Sample Mean Excess Plot or by other ad-hoc methods that impose an arbitrary threshold $(5\%, 10\%, \ldots)$.

In this paper we propose a formal way of identifying which extreme values are really extremes. The goal is to estimate the lower bound of these extremes for finite samples, i.e. a threshold value. Our method is anchored in three key elements: Pickands, Balkema-De Haan theorem (BHP), a distance based on a Kolmogorov-Smirnov (KS) statistic and hypothesis testing via bootstrap methods. By Pickands, Balkema-De Haan theorem we know that the distribution of the exceedances of a random variable in the limit tends to a Pareto shape distribution. Therefore, extreme values considered as exceedances above certain threshold will asymptotically have this type of Pareto distribution. In order to estimate this threshold point we propose an alternative of Pickands estimator based on minimizing a Kolmogorov-Smirnov distance taking into account the length of the sample tail. One of the contributions of our threshold estimator is the obtention of confidence intervals for the tail index capturing the tail behavior of the data distribution. Moreover, the tail index estimators relying on our threshold estimator are consistent and allow to test our definition of extreme values (uniform convergence of the sample distribution of extremes to a Pareto type distribution). The paper concludes with some applications to extreme quantile estimation for simulated known distributions as well as for real financial series. For these series, extreme quantiles are the cornerstone of risk measures, as Value at Risk or Expected Shortfall.

The paper is structured as follows. In section 2 we present a summary of the existing methods to calculate the threshold value. Section 3 shows a brief review of the results from the Extreme Value Theory that we will be using in the core of the paper. Section 4 is devoted to define our concept of extreme value and to present a new estimation method for the threshold value that is used to define extreme observations. Section 5 introduces a bootstrap goodness of fit test to check the validity of our definition. The finite sample performance of our proposed method as well as some real applications are shown in section

6. The conclusions are given in section 7. All proofs as well as the notation used in the paper are gathered in the Appendices.

2 Existing ad-hoc Methods for Threshold Estimation

In the existing literature, there is not a clear definition of the threshold value ν that determines the extremes. There exist different popular estimation methods to select a threshold ($\hat{\nu}_n$) relying on the asymptotic Pareto distribution of the exceedances.

- QQ-plot
- Sample Mean Excess Plot
- Simulation Procedures

This estimation of the threshold has different challenges depending on how close $\hat{\nu}_n$ is to the right end point. A small $\hat{\nu}_n$ yields bias problems in the estimation of the parameters of the Pareto distribution. On the other hand, large $\hat{\nu}_n$ implies problems of great variance due to the abscence of points in the tail to estimate the Pareto distribution.

2.1 QQ-plot

The method is based on the following simple fact: if $U_{(1)} \leq U_{(2)} \leq \ldots U_{(n)}$ are the order statistics from n i.i.d. observations uniformly distributed on [0,1], then by symmetry $E(U_{(i+1)} - U_{(i)}) = \dfrac{1}{n+1}$ and hence $E(U_{(i)}) = \dfrac{i}{n+1}$. Since $U_{(i)}$ should be close to its mean $\frac{i}{(n+1)}$, the plot of $\{(\frac{i}{(n+1)}, U_{(i)}), 1 \leq i \leq n\}$ has to be linear. Suppose now, $X_{(1)} \leq X_{(2)} \leq \ldots X_{(n)}$ are the order statistics from an i.i.d. sample of size n which is suspected to come from a particular continuous distribution G. The plot of $\{(\frac{i}{(n+1)}, G(X_{(i)}), 1 \leq i \leq n\}$ should be approximately linear and hence also the plot of $\{G^{\leftarrow}(\frac{i}{(n+1)}), X_{(i)}), 1 \leq i \leq n\}$ should be linear.

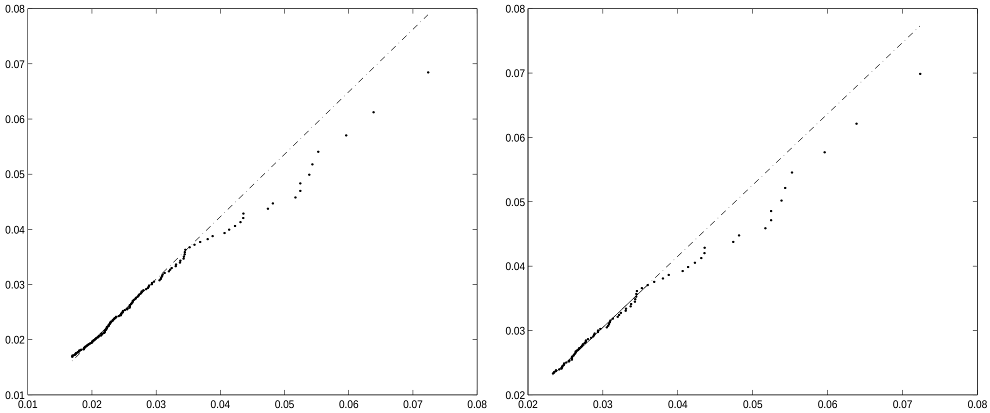

Figure 2.1. *QQ-plots of the negative tail of Nikkei Index returns over the period* 05/1997 − 05/2001 *with* $\hat{\nu} = x_{\lceil 0.90n \rceil}$ *and* $\hat{\nu} = x_{\lceil 0.95n \rceil}$.

It is not clear from Figure 2.1 which portion of the observations fits better to the Generalized Pareto distribution $GPD_{\hat{\Theta}}$, with parameters Θ estimated from the sample observations.

2.2 Sample Mean Excess Plot

Another standard tool for choosing suitable thresholds is the sample mean excess plot $(\nu, e_n(\nu))$ where $e_n(\nu)$ is the sample mean excess function defined by

$$e_n(\nu) = \frac{\sum_{i=1}^{n} (X_i - \nu)_+}{\sum_{i=1}^{n} 1_{\{X_i > \nu\}}},$$

with $x_+ = max(x, 0)$. The sample mean excess function $e_n(\nu)$ is the empirical counterpart of the mean excess function which is defined as $e(\nu) = E[X - \nu \mid X > \nu]$. If the empirical plot follows a reasonably straight line with positive gradient above a certain value of ν, then this is an indication that the exceedances over this threshold follow a Generalized Pareto distribution with positive tail index (ξ) parameter. This is derived from the fact that

$$e(\nu) = \frac{\sigma + \xi\nu}{1 - \xi},$$

where $\sigma + \xi\nu > 0$ and σ is the standard deviation of the GPD (see McNeil and Saladin, 2001).

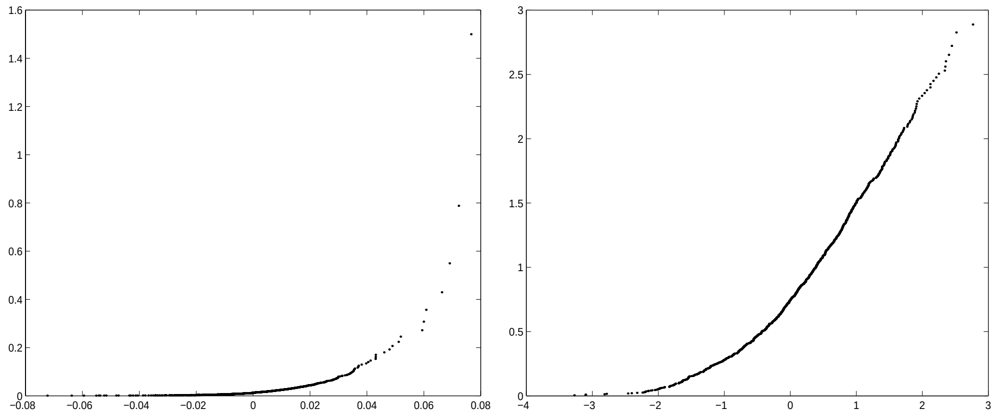

Figure 2.2. *Sample mean Excess plot for negative Nikkei Index returns over the period* $05/1997 - 05/2001$*, (left graph) and a sample from a normal distribution of size* $n = 1000$ *(right graph).*

Focusing on Figure 2.2 different candidates can be selected for the estimated threshold value.

Other methods in order to choose the threshold value ν take advantage of simulation procedures for known distributions. The idea is to determine a threshold ν from a sample of size n and consider the number of observations over this threshold N_ν. The goal is to obtain the necessary sample size to generate N_ν exceedances over the determined threshold ν. This sample size is employed to estimate an extreme quantile closer to the right end point than the threshold ν. In this way we can compare this extreme estimate with the actual extreme quantile of the known distribution and see the reliability of the ad-hoc threshold estimate (see McNeil, 1997).

3 Extreme Value Theory Results

The mathematical foundation of EVT is the class of extreme value limit laws, first derived heuristically by Fisher and Tippet (1928) and later from a rigorous standpoint by Gnedenko (1943). Suppose $X_1, \ldots, X_n$ are independent random variables with common distribution function $F(x) = P\{X \leq x\}$ and let $M_n = max(X_1, \ldots, X_n)$. Under some continuity conditions on F at its right end point, the maximum M_n properly centered and normalized has a limit law H_ξ with ξ the parameter of the limit distribution,

$$P\{\frac{M_n - d_n}{c_n} \leq x\} = F^n(c_n x + d_n) \quad \xrightarrow{d} \quad H_\xi(x). \tag{1}$$

The continuity on F is a sufficient condition but it is not necessary. It is only required some smoothness near the right end point.

Theorem 3.1. *Let F be a distribution function with right end point $x_F \leq \infty$ and let $\tau \in (0, \infty)$. There exists a sequence (u_n) satisfying $n\overline{F}(u_n) \to \tau$ if and only if*

$$\lim_{x \to x_F} \frac{\overline{F}(x)}{\overline{F}(x^-)} = 1 \tag{2}$$

(see Embrechts, Klüppelberg and Mikosch, 1997*, p.*117*).*

The condition $n\overline{F}(u_n) \to \tau$ is equivalent to say that the sample maximum has a non-degenerate distribution of exponential type $P(M_n \leq u_n) \to e^{-\tau}$. The asymptotic distribution of the maximum is called extreme value law. The key result of Fisher-Tippet and Gnedenko is that there are only three fundamental types of extreme value limit laws. These are

Type I: (Gumbel) $\quad \Lambda(x) = exp(-e^{-x}), \quad -\infty < x < \infty,$

Type II: (Frèchet) $\quad \Phi_\alpha(x) = \begin{cases} 0 & x \leq 0, \\ exp(-x^{-\alpha}) & x > 0 \end{cases}$

Type III: (Weibull) $\quad \Psi_\alpha(x) = \begin{cases} 1 & x \geq 0, \\ exp(-(-x)^{\alpha}) & x < 0 \end{cases}.$

In Types II and III α is a positive parameter. The three types may also be combined into a single generalised extreme value distribution, first proposed by Von Mises (1936), of the form

$$H_\xi(x) = \begin{cases} e^{-(1+\xi x)^{\frac{-1}{\xi}}} & \xi \neq 0 \\ e^{-e^{-x}} & \xi = 0 \end{cases} \tag{3}$$

with $1 + \xi x > 0$. The case $\xi > 0$ corresponds to Type II with $\alpha = \frac{1}{\xi}$, the case $\xi < 0$ to Type III with $\alpha = -1/\xi$, and the limit case $\xi \to 0$ to Type I.

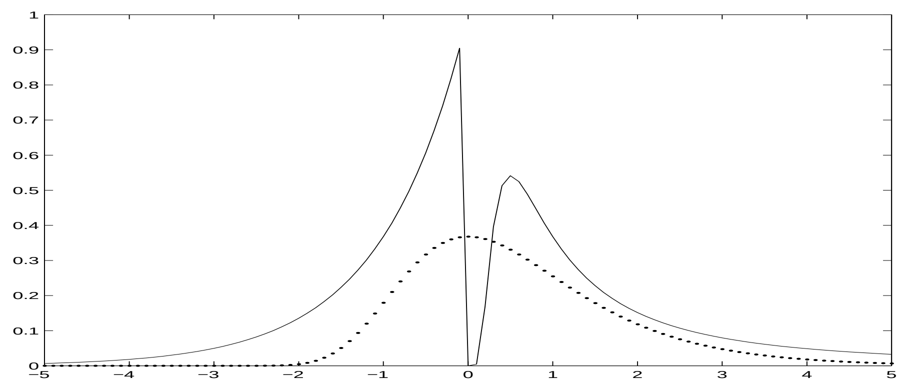

Figure 3.1. *The density function of the extreme value limit laws. The dot line is the Gumbel distribution. Frèchet and Weibull distributions are plotted with $\alpha = 1$.*

Corollary 3.1. *From expressions (1) and (3), the following relationships can be extracted depending on the value of the parameter* ξ, $n\overline{F}(c_n x + d_n) \xrightarrow{d} (1+\xi x)^{\frac{-1}{\xi}}$ *if* $\quad \xi \neq 0$ *and* $n\overline{F}(c_n x + d_n) \xrightarrow{d} e^{-x}$ *if* $\quad \xi = 0$.

These expressions can be considered as the survivor functions of a Generalized Pareto distribution. Moreover, the asymptotic distribution of the standardized tail of F depends on a parameter ξ (tail index), hence a distribution F verifying (2) can be classified according to this parameter.

Definition 3.1. *F belongs to the Maximum Domain of Attraction of an Extreme Value Distribution* H_ξ, $F \in MDA(H_\xi)$, *if and only if there exist constants* $c_n > 0$ *and* d_n, *such that* $c_n^{-1}(M_n - d_n) \xrightarrow{d} H_\xi$.

Notice that the commonly employed continuous distribution functions belong to the maximum domain of attraction of an extreme value limit law, $F \in MDA(H_\xi)$. From the results of Fisher-Tippet and Gnedenko it is derived that there are only three types of maximum domains of attraction in contrast with the number of domains of attraction of α-stable processes. This maximum domain of attraction of F depends on the sign of ξ.

Definition 3.2. *A df F such that the right tail satisfies*

$$\lim_{x\to\infty} \frac{1-F(tx)}{1-F(x)} = t^{-\alpha}, \qquad t > 0, \quad \alpha = \frac{1}{\xi} > 0 \tag{4}$$

is called regularly varying with index α *(*$F \in RV_{-\alpha}$*).*

The tail of a distribution F satisfying (4) decays polynomially (F is heavy tailed). This condition can be rewritten as

$$1 - F(x) = x^{-\frac{1}{\xi}} L(x), \qquad x \to \infty, \quad \xi > 0 \tag{5}$$

where $L(x)$ is a slowly varying function

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1, \qquad t > 0. \tag{6}$$

A distribution function F with positive tail index and verifying condition (2) indicates that the sample maximum should have a non degenerate distribution of Type II.

Proposition 3.1. $F \in RV_{-\alpha} \Leftrightarrow F \in MDA(\Phi_\alpha)$ *where* Φ_α *is the Fréchet EVD. The normalizing constants for this case are* $d_n = 0$ *and* $c_n = F^{\leftarrow}(1 - \frac{1}{n})$ *(see Embrechts, Klüppelberg and Mikosch,* 1997, *p.*132*).*

It is sufficient to know the tail index of a distribution F to know the asymptotic distribution of the standardized maximum. Moreover, this parameter ξ provides information about the behavior of the tail and therefore the tail index can help to give a formal definition for the tail avoiding the ad-hoc selection of arbitrary quantiles.

Definition 3.3. *The tail of a distribution is the set of extreme values; these extremes are the exceedances over a determined threshold (ν) with ν sufficiently large. The distribution of these large observations $F_\nu(x)$ is called the conditional excess distribution function (cedf) over ν and is defined as*

$$F_\nu(x) = P\{X \leq x | X > \nu\}, \qquad \nu \leq x \leq x_F, \tag{7}$$

where X is a random variable, ν is a given threshold and $x_F \leq \infty$ is the right endpoint of F. This distribution can be written in terms of F,

$$F_\nu(x) = \frac{F(x) - F(\nu)}{1 - F(\nu)} \qquad x > \nu. \tag{8}$$

From this expression it is deduced that $\overline{F}_\nu(x) = \frac{\overline{F}(x)}{\overline{F}(\nu)}$. The extremes of a population are determined by the threshold ν and by the tail index ξ of the distribution F. These parameters define the asymptotic distribution of standardized extremes.

Theorem 3.2. ***(Balkema and de Haan (**1974**), Pickands (**1975**)) (BHP)**. Let F be a distribution function such that $F \in MDA(H_\xi)$, the conditional excess distribution function $F_\nu(x)$ for ν large, is*

$$\lim_{\nu \to x_F} F_\nu(x) = GPD_{\xi\sigma}(x - \nu)$$

where

$$GPD_{\xi,\sigma}(x-\nu) = \begin{cases} 1 - (1 + \frac{\xi(x-\nu)}{\sigma})^{\frac{-1}{\xi}} & \text{if } \xi \neq 0 \\ 1 - exp\left(\frac{-(x-\nu)}{\sigma}\right) & \text{if } \xi = 0 \end{cases} \tag{9}$$

is the so-called Generalized Pareto distribution (GPD).

The Generalized Pareto is the asymptotic distribution of the extremes under some continuity conditions over F. If the distribution F has a regularly varying tail, then the distribution of the extremes as n goes to infinity can be reduced to a Pareto distribution.

Corollary 3.2. *For a distribution function F such that $F \in MDA(\Phi_\alpha)$, the conditional excess distribution function $F_\nu(x)$, for ν large, is*

$$\lim_{\nu \to x_F} F_\nu(x) = PD_\xi(\tfrac{x}{\nu}),$$

where

$$PD_\xi(\tfrac{x}{\nu}) = 1 - (\tfrac{x}{\nu})^{\frac{-1}{\xi}}, \qquad x > \nu$$

is the Pareto distribution.

We bring together the two approaches of the asymptotic Pareto type distribution of the tail in the notation G_Θ,

$$G_\Theta = \begin{cases} GPD_{\xi,\sigma}(x-\nu), & F \in MDA(H_\xi) \\ PD_\xi(\frac{x}{\nu}), & F \in MDA(\Phi_\alpha) \end{cases}$$

with $\Theta = \{\xi, \sigma\}$ for the GPD case and $\Theta = \{\xi\}$ for the Pareto distribution. It is important to notice that we are not contradicting BHP theorem. The Pareto distribution is included in the Generalized Pareto family: $PD_\xi(\frac{x}{\nu}) = GPD_{\xi,\sigma}(\frac{x-\nu}{\nu\xi})$.

4 Definition and Estimation of Extremes

Under definition (3.3) the tail of a distribution is the set of extreme values. We can extend BHP theorem to give a formal definition of the extremes based on uniform convergence between the two involved distribution functions.

Definition 4.1. *Let X be a random variable with a distribution $F \in MDA(H_\xi)$. Let $\nu \in$ support of $F = [x_0, x_F]$ with $x_F \leq \infty$, $F_\nu(x)$ be the conditional excess distribution function and $G_\Theta(x,\nu)$ be the Pareto type distribution. The extreme values of the distribution F are defined by a parameter $\nu < x_F$ such that $F_\nu(x)$ converges uniformly to $G_\Theta(x,\nu)$ as $\nu \to x_F$,*

$$\lim_{\nu \to x_F} \sup_{x \in \mathbb{R}} |F_\nu(x) - G_\Theta(x,\nu)| = 0 \tag{10}$$

with $\xi \in \Theta$ the tail index of the distribution F.

In order to obtain the extremes of a distribution F from a sample data we need to estimate the threshold parameter. Condition (10) is not possible to check for a given n, therefore we give a characterization of the definition for sample data.

Proposition 4.1. *Let $F(x)$ be a distribution function verifying condition (2) and consider $\hat{\nu}_n$ an estimator of ν. The extreme values estimates defined by $\hat{\nu}_n$ are extreme values if and only if*

- $\sup_{x \in \mathbb{R}} |F_{\hat{\nu}_n}(x) - G_\Theta(x,\hat{\nu}_n)| \xrightarrow{p} 0$.
- $F_{\hat{\nu}_n}(x) = G_\Theta(x,\hat{\nu}_n)$ *for almost every $x \in \mathbb{R}$ and given n.*

The first condition provides the consistency of the estimator $\hat{\nu}_n$. This is a necessary condition but not sufficient to define the set of extremes. We can obtain consistent estimators of ν that determine extremes estimates that do not follow Pareto type distributions for a given sample size n. This is the reason to impose the second condition. This condition is not usually possible to check because the parameters of the Pareto type distribution (Θ) are unknown. In consequence, we propose a goodness of fit test to circumvent this drawback.

4.1 Conditional Estimation of the Parameter Set Θ

The distribution of the extremes of a distribution depends on its tail behavior, i.e. on Θ. This set of parameters Θ must be estimated from the available data sample. Maximum Likelihood (ml) is the most conventional method and has very desirable properties; consistency, asymptotic efficiency and normality. The estimation of the tail parameters Θ is conditioned on the knowledge we have about the sample tail defined by the threshold ν. For the GPD approach, $\hat{\Theta}(\nu) = \{\hat{\xi}_{ml}(\nu), \hat{\sigma}_{ml}(\nu)\}$ and for the Pareto approach, $\hat{\Theta}(\nu) = \hat{\xi}_{ml}(\nu)$.

Proposition 4.2. *If $F \in MDA(\Phi_\alpha)$, $\hat{\xi}_{ml}(\nu)$ for PD_ξ is the Hill estimator (see Hill,* 1975*),*

$$\hat{\xi}_{Hill}(\nu) = \frac{1}{(n-k)} \sum_{i=k+1}^{n} log \frac{x_{(i)}}{\nu}, \tag{11}$$

with $\nu = x_{(k)}$ and $x_{(k+1)} \leq \ldots \leq x_{(n)}$ the increasing order statistics.

Hill estimator is gaining popularity in the EVT Literature because is easy to calculate and has good asymptotic properties, but in the financial literature is employed even for not heavy tailed distributions. Therefore, consistency and asymptotic normality may not hold any more. There exists some confusion about the conditions to use it.

Proposition 4.3. *Let $\hat{\xi}_{ml}$ and $\hat{\xi}_{Hill}$ be the maximum likelihood estimators of the parameter ξ of a Generalized Pareto and of a Pareto distribution respectively. These estimators are $\sqrt{n}$-consistent estimates of the tail index of a distribution function F verifying condition (2) if $\xi > -\frac{1}{2}$ for $\hat{\xi}_{ml}$ (see Smith,* 1984*) and if $\xi > 0$ for $\hat{\xi}_{Hill}$ (Goldie and Smith,* 1987*).*

The drawback of these estimators is their biases (see Guillou & Hall, 2000). This bias has two different sources: the distribution of data is not of a Pareto type and the choice of the number of order statistics used to construct the estimator. Let us concentrate on the Hill estimator and assume $F \in MDA(\Phi_\alpha)$. By BHP theorem the large observations $x_{(k'+1)} \leq \ldots \leq x_{(n)}$ greater than $\nu = x_{(k')}$ follow a PD_ξ with ν sufficiently large, therefore we eliminate the first source of bias. It is shown in Hill (1975) that the random variable $V_i = i\ [log\ x_{(n-i+1)} - log\ x_{(n-i)}]$ follows an exponential distribution with mean ξ. Consider $\hat{\nu}_n = x_{(k)}$ an estimate of ν such that $x_{(k+1)} \leq \ldots \leq x_{(k')} \leq \ldots \leq x_{(n)}$ are the exceedances over the estimate [1]. Hill estimator based on $x_{(k)}$ is $\hat{\xi}_{Hill}(\hat{\nu}_n) = \frac{1}{n-k}\sum_{i=k+1}^{n} log\frac{x_{(i)}}{x_{(k)}}$. This estimator can be decomposed as

$$\hat{\xi}_{Hill}(\hat{\nu}_n) = \frac{1}{n-k}\sum_{i=k'+1}^{n} log\frac{x_{(i)}}{\nu} + \frac{1}{n-k}\sum_{i=k+1}^{k'} log\frac{x_{(i)}}{\nu} + log\frac{\nu}{x_{(k)}} =$$

$$= \frac{n-k'}{n-k}\hat{\xi}_{Hill}(\nu) + \frac{1}{n-k}\sum_{i=k+1}^{k'} log\frac{x_{(i)}}{\nu} + log\frac{\nu}{x_{(k)}}.$$

On the other hand, Hill estimator based on the parameter ν can be expressed in terms of V_i, $\hat{\xi}_{Hill}(\nu) = \frac{1}{n-k'}\sum_{i=1}^{n-k'} V_i$. This estimator is unbiased ($E[\hat{\xi}_{Hill}(\nu)] = \xi$), however the expected value of the Hill estimator based on the estimate $\hat{\nu}$ is biased. This deviation from the parameter depends on the bias of the threshold estimator:

$$E[\hat{\xi}_{Hill}(\hat{\nu}_n)] = \frac{n-k'}{n-k}\xi + \frac{1}{n-k}\sum_{i=k+1}^{k'} Elogx_{(i)} - Elogx_{(k)} + \frac{n-k'}{n-k}log\nu.$$

Notice that the bias disappears if $k = k'$. Therefore, the bias of the Hill estimator of a distribution $F \in MDA(\Phi_\alpha)$ as n goes to infinity depends only on the bias of the threshold estimate. The problem is that the parameter $\nu = x_{(k')}$ is unknown. In order to minimize bias problems, confidence intervals are proposed as estimators of the tail index. It is well known that the random variable $S_k = \sqrt{n-k}(\hat{\xi}_{Hill} - \xi)$ has an asymptotic $N(0, \xi^2)$ distribution. We construct nonparametric bootstrap confidence intervals to approximate the exact confidence intervals for the tail index,

$$\xi \in [\hat{\xi}_{Hill}(\hat{\nu}_n) - \frac{1}{\sqrt{k}}J_k^{-1}(F_n, 1-\frac{\alpha}{2}), \hat{\xi}_{Hill}(\hat{\nu}_n) - \frac{1}{\sqrt{k}}J_k^{-1}(F_n, \frac{\alpha}{2})], \tag{12}$$

with F_n the empirical distribution function of the data, α the significance level and $J_k(x, F_n)$ the approximate bootstrap distribution of S_k. Note that the same procedure can be applied to calculate confidence intervals for the tail index based on the maximum likelihood estimator $\hat{\xi}_{ml}(\hat{\nu}_n)$ of a Generalized Pareto distribution.

[1] Notice that if $\hat{\nu} > \nu$ there is not a problem of bias, it is only a matter of efficiency of the estimator $\hat{\nu}$.

4.2 Estimation Method for the Threshold Value ν

Pickands (1975) proposed a method to estimate the threshold value ν based on uniform convergence (d_∞) between the Empirical Distribution associated to F_ν and a Generalized Pareto distribution estimated from data $GPD_{\Theta_\nu^{Pic}}$:

$$\nu_n^{Pic} = arg \quad \min_{\nu} \quad d_\infty^\nu(F_{\nu,n}, GPD_{\Theta_\nu^{Pic}}), \tag{13}$$

with Θ_ν^{Pic} the estimated parameters of the GPD. This estimator of ν is consistent in the sense that $P\{\sup_{x\in\mathbb{R}} \quad |F_{\nu_n^{Pic}} - GPD_{\Theta_{\nu_n^{Pic}}^{Pic}}| > \varepsilon\} \rightarrow 0$. The estimators for the parameters of the GPD proposed by Pickands depend on the different values of ν. Consider $\nu = X_{(n-4i+1)}$, $i = 1, \ldots, n/4$.

$$\hat{\xi}(\nu) = \frac{1}{log(2)} log(\frac{X_{(n-i+1)} - X_{(n-2i+1)}}{X_{(n-2i+1)} - \nu}),$$

for the tail index and

$$\hat{\sigma}(\nu) = \frac{X_{(n-2i+1)} - \nu}{\int_0^{log2} e^{\hat{\xi}u} du},$$

for the variance. This estimator for the tail index is consistent, but it is very sensitive to the choice of the order statistics and it is not efficient (Drees, 1995). For standard sample sizes the estimations of the tail index are biased and the confidence intervals for ξ do not give reliable information. Alternative statistics have been proposed for the tail index to overcome these drawbacks, (see Dekker, Einmahl and de Haan, 1989). Goldie and Smith (1987) or Dekker and de Haan (1993) establish the optimal number of order statistics for different estimators of the tail index.

On the other hand, Pickands estimator for the threshold (ν_n^{Pic}) does not take into account the length of the sample tails defined by ν to compute the distances in (13). As $\nu \rightarrow x_F$ the available samples of the tails are smaller yielding worse estimations of the tail parameters of the GPD. This implies worse goodness of fit of the conditional distributions F_ν to the theoretical asymptotic distribution. This is caused not only by the lack of fit of data to the theoretic GPD distribution but also by the estimation mechanism of the tail parameters. Hence, ν_n^{Pic} is not near the tail by its own construction. Consequently, the extremes estimates defined by Pickands estimator can be very misleading for standard sample sizes (see Table 6.1).

A natural distance to derive a good estimator to overcome Pickands drawbacks for finite samples is a distance based on Kolmogorov-Smirnov statistic.

Definition 4.2. ***(Kolmogorov-Smirnov distance)***
Let $F_{\nu,n}(x) = \frac{\sum_{i=1}^{n} 1_{\{\nu \le x_i \le x\}}}{\sum_{i=1}^{n} 1_{\{x_i > \nu\}}}$ *be the empirical distribution function associated to* F_ν *and* G_Θ *be a Pareto type distribution. The distance between* F_ν *and* G_Θ *is calculated by the following KS distance*

$$d^{\nu}_{ks}(F_{\nu,n}, G_{\hat{\Theta}_\nu}) = \sqrt{\sum_{i=1}^{n} 1_{\{x_i > \nu\}}} \quad \sup_{x \in \mathbb{R}} \left| \frac{\sum_{i=1}^{n} 1_{\{\nu \le x_i \le x\}}}{\sum_{i=1}^{n} 1_{\{x_i > \nu\}}} - G_{\hat{\Theta}_\nu}(x, \nu) \right|. \tag{14}$$

This statistic regards the number of observations of the available sample tails giving less weight to distances of samples with less data in order to compensate the estimation failure of the parameters of the theoretical distribution from small samples.

Definition 4.3. *Let* d^{ν}_{ks} *be the KS distance of (14) and* $\mathbf{x}_n = \{x_1 \ldots, x_n\}$ *be a sample of size n from a distribution F. The estimated threshold* $\hat{\nu}_n$ *is the order statistic* $x_{(k)}$ *that makes the distance* d^{ν}_{ks} *minimum.*

$$\hat{\nu}_n = arg \min_{\nu} \quad d^{\nu}_{ks}(F_{\nu,n}, G_{\hat{\Theta}_\nu}),$$

with $x_{(k)}$ *such that* $n - k \to \infty$, $\frac{n-k}{n} \to 0$.

The latter conditions are consequence of BHP theorem. As n becomes large, n-k should go to infinity to benefit of an increasing sample (more information as n increases and therefore smaller variance). At the same time, unless a portion of the upper tail follows exactly a Pareto type distribution we expect that $\frac{n-k}{n}$ tends to zero in order to improve the approximation to the theoretical distribution when $\nu \to x_F$ as BHP theorem states (smaller the bias).

Theorem 4.1. *Let* $\hat{\nu}_n$ *be the threshold estimator derived from the KS distance* (d^{ν}_{ks}) *and let* $\hat{\xi}(\hat{\nu}_n)$ *be a consistent estimator of the tail index based on* $\mathbf{x}_n$ *with* $\xi \in \Theta$. *Then,* $\hat{\nu}_n$ *is a consistent estimator of the threshold parameter* ν *in the sense that*

$$P\{\sup_{x \in \mathbb{R}} \quad |F_{\hat{\nu}_n}(x) - G_\Theta(x, \hat{\nu}_n)| > \varepsilon\} \to 0, \, \forall \, \varepsilon > 0.$$

The concept of consistency can be puzzling in this context because the parameter ν according to BHP theorem must go to the right end point. The uniqueness of the threshold makes no sense, because as ν goes to x_F the approximation of the conditional distribution is better. In consequence, we prove the consistency of our estimator in the sense that mimics the properties of the parameter ν. However, other estimators can mimic as well the behavior of the parameter; $\nu \to x_F$, $\hat{\xi}(\nu) \xrightarrow{p} \xi$ and $F_\nu = G_\Theta(x, \nu)$. In order to check the performance of these other estimators we propose a hypothesis test in the next section. In practice our estimator of the threshold is obtained in the following way,

Algorithm 4.1. *:*

1. *Fix a threshold,* $\nu = x_{(k)}$, $(k = k_0 = n/2)$ [2]
2. *Estimation* [3] *of* $\hat{\Theta}_\nu = \begin{cases} \hat{\xi}_{ml}(\nu), \hat{\sigma}_{ml}(\nu) & \textit{GPD approach} \\ \hat{\xi}_{Hill}(\nu) & \textit{Pareto approach} \end{cases}$
3. *Compute* $F_{\nu,n}(x) = \frac{\sum_{i=1}^{n} 1_{\{\nu \le x_i \le x\}}}{\sum_{i=1}^{n} 1_{\{x_i > \nu\}}}$.
4. *Compute* $G_{\hat{\Theta}_\nu} = \begin{cases} GPD_{\hat{\xi},\hat{\sigma}}(x_i - \nu) & \textit{GPD approach} \\ PD_{\hat{\xi}_{Hill}}(\frac{x_i}{\nu}) & \textit{Pareto approach} \end{cases}$
5. *Calculate the distance defined by*
   $d^{\nu}_{ks}(F_{\nu,n}, G_{\hat{\Theta}}) = \sqrt{\sum_{i=1}^{n} 1_{\{x_i > \nu\}}} \quad \underset{x \in \mathbb{R}}{sup} \quad \left| \frac{\sum_{i=1}^{n} 1_{\{\nu \le x_i \le x\}}}{\sum_{i=1}^{n} 1_{\{x_i > \nu\}}} - G_{\hat{\Theta}_\nu}(x, \nu) \right|$
6. $k++$
   Repeat the process until $k = n - 1$.
7. *At the end of the day, we estimate* $\hat{\nu}_n = x_{(\hat{k})}$ *such that*

$$\hat{\nu}_n = arg \min_{\nu} \quad d^{\nu}_{ks}(F_{\nu,n}, G_{\hat{\Theta}_\nu}).$$

Alternative distance measures can be proposed for this threshold selection. For instance the ones based on Cramér-von Mises or Anderson-Darling Statistics,

- $W_n^2 = n \int_{-\infty}^{\infty} (F_{\nu,n}(x) - G_{\hat{\Theta}_\nu}(x))^2 dG_{\hat{\Theta}_\nu}(x)$
- $A_n^2 = n \int_{-\infty}^{\infty} \frac{(F_{\nu,n}(x) - G_{\hat{\Theta}_\nu}(x))^2}{G_{\hat{\Theta}_\nu}(x)(1 - G_{\hat{\Theta}_\nu}(x))} dG_{\hat{\Theta}_\nu}(x)$.

These statistics rely on the euclidean distance. The drawback of these measures with respect to KS type statistics for threshold selection is that these first ones are less sensitive to large deviations from the Pareto type distribution due to isolate observations (outlier observations).

5 Hypothesis Testing

The threshold estimate $\hat{\nu}_n$ provides the lower limit of the estimation of the extreme values in finite samples. Our threshold estimator $\hat{\nu}_n$ is such that as the sample size increases,

[2] Consider $k = 1, \ldots, n-1$ is computationally very costly. The method is implemented taking fractions of the sample. $x_{(k)}$ s.t. $k = n * \frac{i}{100}$, $i = 50, 60, 70, 80, 90, 91, \ldots, 99$.

[3] The algorithm to estimate the threshold depends on the maximum domain of attraction of the distribution F.

condition (10) asymptotically holds ($F_{\hat{\nu}_n} = G_\Theta$). The key question to answer is whether this condition can be rejected or not for the extremes estimates produced from the threshold value estimation. In other words, are these estimates really extreme values according to our definition of extremes? The answer boils down to test

$$H_0 : F_\nu = G_\Theta \tag{15}$$

with $G_\Theta = \begin{cases} GPD_{\xi,\sigma}(x-\nu) & \text{GPD approach} \\ PD_\xi(\frac{x_i}{\nu}) & \text{Pareto approach.} \end{cases}$

The statistic proposed to test H_0 is the following goodness of fit test

$$T_n(\mathbf{x}_n, \Theta) = \sqrt{n} \sup_{x\in\mathbb{R}} |F_{\hat{\nu}_n,n}(x) - G_\Theta(x, \hat{\nu}_n)|. \tag{16}$$

Although there are alternatives that are more sensitive to the deviations from the null distribution that occur in both tails (Modified KS tests, see Mason and Schuenemeyer, 1983) we concentrate on the standard KS test because our concern is the distribution of the largest observations exceeding the threshold value ν. The sampling distribution of this test statistic $J_n(x, F; \Theta) = P\{T_n(\mathbf{x}_n, \Theta) \leq x\}$ is not known and the asymptotic null distribution $J(x, F)$ is parameter free (see Kolmogorov, 1933) but it is not possible to obtain a value of the estimator based on a sample $\mathbf{x}_n$ because the set of parameters Θ is unknown. Therefore, the test statistic needed to test the null hypothesis is $T_n(\mathbf{x}_n, \hat{\Theta})$, where $\hat{\Theta}$ is an estimate of the true Θ. This statistic follows asymptotically a functional of a centered gaussian process that depends on Θ, see Durbin (1973). The asymptotic critical values vary with H_0 and the estimation of this set of parameters. Bootstrap methodology can be applied to calculate the sampling quantiles of the Bootstrap distribution $J_n(x, \hat{F}_n; \hat{\Theta}^*)$ with $\hat{\Theta}^*$ the estimated set of parameters from the bootstrap sample $\mathbf{x}_n^* = \{x_1^*, \ldots, x_n^*\}$ and with $\hat{F}_n$ an estimate of F. These quantiles will be close to the exact quantiles of the distribution of the statistic $J_n(x, F; \Theta)$ if the Bootstrap is consistent ($J_n(x, F; \hat{\Theta}) \simeq J_n(x, \hat{F}_n; \hat{\Theta}^*)$) and if $\hat{\Theta}$ is a $\sqrt{n}$-consistent estimator of Θ ($J_n(x, F; \Theta) \simeq J_n(x, F; \hat{\Theta})$), see Babu and Rao (2002) for details.

Proposition 5.1. *Let $\mathbf{x}_n$ be a sample of size n from F. Assume that $\hat{F}_n$ is an estimate of F based on $\mathbf{x}_n$ and let $J_n(x, F; \hat{\Theta})$ be the true sampling distribution of the statistic $T_n(\mathbf{x}_n, \hat{\Theta})$. If the following two conditions hold*

- $\sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)| \xrightarrow{p} 0.$
- $J_n(x, F; \hat{\Theta}) \to J(x, F; \Theta)$ *with* $J(x, F; \Theta)$ *being a strictly increasing continuous function in x.*

Then, the Bootstrap approximation $J_n(x, \hat{F}_n; \hat{\Theta}^*)$ is consistent ($J_n(x, \hat{F}_n; \hat{\Theta}^*) \simeq J_n(x, F; \hat{\Theta})$).

5.1 Methodology

The statistic $T_n(\mathbf{x}_n, \hat{\Theta})$ follows asymptotically a functional of a centered gaussian process. Therefore, in order to obtain a consistent bootstrap approximation ($J_n(x, \hat{F}_n; \hat{\Theta}^*)$) of the true sampling distribution of $T_n(\mathbf{x}_n, \hat{\Theta})$ we need to construct $\hat{F}_n$ verifying uniform convergence in probability to F.

Definition 5.1. *Let $\hat{F}_n(x)$ be a mixture of $F_n(x)$ for values smaller than the estimated threshold $\hat{\nu}_n$ and of a Pareto type distribution for values above it:*

$$\hat{F}_n(x) = \begin{cases} \frac{1}{n}\sum_{i=1}^{n} 1_{\{x_i \le x\}} & x \le \hat{\nu}_n \\ G_{\hat{\Theta}_{\hat{\nu}_n}}(x) + \frac{1}{n}\sum_{i=1}^{n} 1_{\{x_i \le \hat{\nu}_n\}} \overline{G}_{\hat{\Theta}_{\hat{\nu}_n}}(x) & x > \hat{\nu}_n. \end{cases} \tag{17}$$

It is obvious to check that $\hat{F}_n(x)$ in expression (17) is a distribution function.

Proposition 5.2. *Let $\mathbf{x}_n$ be a sample of size n with distribution function $F(x)$. Then, the distribution function $\hat{F}_n$ is such that $\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \xrightarrow{p} 0$.*

The first task is to generate a bootstrap sample $\mathbf{x}_n^*$ of size n from the distribution $\hat{F}_n$.

Algorithm 5.1. ***(Generating Process of Data):***

1. *Let $\hat{\nu}_n = x_{(k)}$ be the estimated threshold and $\hat{\Theta}$ be the estimated parameter space.*
2. *Generate $0 \le j \le 1$ and calculate $\lceil nj \rceil$*
3. $x_i^* = \begin{cases} x_{(\lceil nj \rceil)} & \text{if } \lceil jn \rceil \le k \\ z & \text{if } \lceil jn \rceil > k \end{cases}$

   $$z = G^{\leftarrow}_{\hat{\Theta}_{\hat{\nu}_n}}\left(\frac{nj - \sum_{i=1}^{n} 1_{\{x_i \le \hat{\nu}_n\}}}{\sum_{i=1}^{n} 1_{\{x_i > \hat{\nu}_n\}}}\right)$$
4. $i++$

   Go to step 2

Once a bootstrap sample is generated it is immediate to calculate $J_n(x, \hat{F}_n; \hat{\Theta}^*)$ under H_0.

Algorithm 5.2. ***(Bootstrap Distribution of T_n):***

1. $l = 1$.
2. *Generate $\mathbf{x}_n^*$ a bootstrap sample coming from $\hat{F}_n$.*

3. *Compute* $\hat{\Theta}^*$ *from the exceedances of* $\mathbf{x}_n^*$ *over the fixed threshold* $\hat{\nu}_n$.

4. *Compute* $T_l^*(\mathbf{x}_n^*, \hat{\Theta}^*) = \sqrt{n} \quad \sup_{x \in \mathbb{R}} \quad |\frac{\sum_{i=1}^{n} 1_{\{\hat{\nu}_n \leq x_i^* \leq x\}}}{\sum_{i=1}^{n} 1_{\{x_i^* > \hat{\nu}_n\}}} - G_{\hat{\Theta}^*_{\hat{\nu}_n}}(x, \hat{\nu}_n)|$.

5. $l = 1, \ldots, B$

6. $J_n(x, \hat{F}_n; \hat{\Theta}^*) = \frac{1}{B} \sum_{i=1}^{B} 1_{\{T_i^* \leq x\}}$.

Notice that the set of parameters Θ is consistently estimated by $\{\hat{s}^2, \hat{\xi}_{ml}\}$ for the Generalized Pareto distribution, and by $\hat{\xi}_{Hill}$ for the Pareto distribution. Both estimators of the tail index are $\sqrt{n}$ -consistent for some values of the tail index (see proposition (4.3)). Therefore, the knowledge of $J_n(x, \hat{F}_n; \hat{\Theta}^*)$ allows us to estimate the p-value of the test (16):

$$p = P\{J_n(x, F; \Theta) > T_n(\mathbf{x}_n, \Theta)\} \simeq P\{J_n(x, \hat{F}_n; \hat{\Theta}^*) > T_n(\mathbf{x}_n, \hat{\Theta})\} = \frac{1}{B} \sum_{i=1}^{B} 1_{\{T_l^* > T_n\}} = \hat{p}.$$

Large values of the test statistic imply rejection of the null hypothesis. In other words, it is rejected if $\hat{p} < \alpha$ for a given significance level α.

5.2 Size of the Test

Theorem 5.1. *Let* $\hat{Q}_n$ *be an estimator of* F *based on a sample* $\mathbf{x}_n$ *of size* n *that satisfies* $\sup_{x \in \mathbb{R}} \quad |\hat{Q}_n - F(x)| \xrightarrow{p} 0$ *whenever* $F \in F_{H_0}$. *Then,* $P\{T_n(x, \hat{\Theta}) > j_n(1 - \alpha, \hat{Q}_n; \hat{\Theta}^*)\} \to \alpha$, *with* $j_n(1 - \alpha, \hat{Q}_n; \hat{\Theta}^*)$ *the* $1 - \alpha$ *quantile of the Bootstrap distribution* $J_n(x, \hat{Q}_n; \hat{\Theta}^*)$ *of* $T_n(x, \hat{\Theta})$.

The distribution function $\hat{F}_n(x)$ of expression (17) verifies the condition of theorem 5.1, therefore $j_n(1 - \alpha, \hat{F}_n; \hat{\Theta}^*) \simeq j_n(1 - \alpha, F; \hat{\Theta})$. In consequence, $\hat{F}_n$ is a good candidate to estimate the size of the proposed test.

Algorithm 5.3. *:*

1. $j = 1$.

2. *Estimate* $\hat{\nu}_n = x_{(k)}$ *and* $G_{\hat{\Theta}_{\hat{\nu}_n}}$ *by KS method from a sample* $\mathbf{x}_{j,n}$ *that follows* F.

   (a) $i = 1$

   (b) *Generate a sample* $\mathbf{x}_{i,n}^* \sim \hat{F}_n$ *from* $\mathbf{x}_{j,n}$.

   (c) *Calculate* $T_i^*(\mathbf{x}_{i,n}^*, \hat{\Theta}^*)$

   (d) $i++$. *Go to step (b) while* $i \leq B$.

   (e) *Construct* $J_n(x, \hat{F}_n) = \frac{1}{B} \sum_{i=1}^{B} 1_{\{T_i^* \leq x\}}$.

3. *Generate a sample* $\mathbf{x}'_n$ *under* H_0.

4. *Calculate* $T'_n(\mathbf{x}'_n, \hat{\Theta}')$.

5. $\hat{p} = \frac{1}{B}\sum_{i=1}^{B} 1_{\{T_i^* > T_n'\}}$.

6. *Reject* H_0 *if* $\hat{p} < \alpha$ *with* α *the significance level.*

7. $\delta_j = \begin{cases} 1 & \text{if } H_0 \text{ is rejected} \\ 0 & \text{if } H_0 \text{ is accepted.} \end{cases}$

8. $j++$. *Go to step* 2 *while* $j \leq m$.

9. $\hat{\alpha} = \frac{1}{m}\sum_{i=1}^{m} \delta_i$, *where* $\hat{\alpha}$ *is the estimation of the type I error.*

$\hat{\alpha}$ should be close to the significance level α.

5.3 Power of the Test

The choice of $\hat{Q}_n$ can bring some problems under the alternative hypothesis ($F \in F_{H_1}$). $\hat{Q}_n$ should satisfy three conditions under the alternative hypothesis in order to avoid that the critical values of $J_n(x, \hat{Q}_n; \hat{\Theta}^*)$ go to infinity as n increases.

- $T_n(\mathbf{x}_n, \hat{\Theta}) \to \infty$ under $F \in F_{H_1}$.
- $\hat{Q}_n$ with $F \in F_{H_1}$ such that $\hat{Q}_n \not\simeq F$, but some F_0 under (F_{H_0}).
- The critical value should satisfy $j_n(1-\alpha, \hat{Q}_n; \hat{\Theta}^*) \simeq j_n(1-\alpha, F_0; \Theta) \to$
  $\to j(1-\alpha, F_0) < \infty$.

If these conditions hold, then by Slutsky's theorem,
$P\{T_n(x, \hat{\Theta}) > j_n(1-\alpha, \hat{Q}_n; \hat{\Theta}^*)\} \simeq P\{T_n(x, \Theta) > j_n(1-\alpha, F_0)\} \to 1$ as $n \to \infty$.

Proposition 5.3. *Let* $\mathbf{x}_n$ *be a sample of size n from a distribution F under the alternative hypothesis* F_{H_1} *and let* $T_n(x, \hat{\Theta})$ *be the test statistic of (16) with* $\hat{\nu}_n$ *and* $G_{\hat{\Theta}_{\hat{\nu}_n}}$ *estimated under the null hypothesis. Then,* $T_n(x, \hat{\Theta}) \to \infty$.

The problem is how to construct $\hat{Q}_n$ such that does not approach the distribution F, but F_{H_0} when the sample $\mathbf{x}_n$ comes from F_{H_1}. $\hat{F}_n$ is not valid in this case because $F \in F_{H_1}$ ($\mathbf{x}_n \sim F_{H_1}$). At least, a sample $\mathbf{x}_{o,n}$ of size n under F_{H_0} is required to construct $\hat{Q}_n$, a consistent estimate of F_{H_0}.

$$\hat{Q}_n(x) = \begin{cases} \frac{1}{n}\sum_{i=1}^{n} 1_{\{x_{0,i} \leq x\}} & x \leq \hat{\nu}_n \\ G_{\hat{\Theta}_{\hat{\nu}_n}}(x) + \frac{1}{n}\sum_{i=1}^{n} 1_{\{x_{0,i} \leq \hat{\nu}_n\}} \overline{G}_{\hat{\Theta}_{\hat{\nu}_n}}(x) & x > \hat{\nu}_n \end{cases} \tag{18}$$

with $\hat{\Theta}_{\hat{\nu}_n}$ a consistent estimate of Θ under the null hypothesis.

The algorithm to estimate the power is equivalent to the algorithm proposed for the size, but in step 3 the sample is generated from $F \in F_{H_1}$. Therefore, $\hat{\alpha}$ is an estimate of the power of the test. The objective of this hypothesis test is to reject extremes estimates defined by $\hat{\nu}_n$ which are not really extremes. This situation can occur for small sample sizes where $\hat{\nu}_n$ can be not near the right end point x_F defining more extremes estimates than there really exist. We can also test if the extremes estimates defined by other $\tilde{\nu}_n$ are really extremes.

6 Simulations and Some Financial Applications

In this section we present how our estimation and testing methodology perform in finite samples, with simulated data from different distributions as well as with real data. Under our methodology the extremes of the distribution are well estimated by the observations exceeding a determined threshold value once the null hypothesis (15) is not rejected. The extreme quantile estimates and their bootstrap confidence intervals rely upon the construction of $\hat{F}_n(x)$. We distinguish two cases: if F has heavy tails, G_Θ is a PD_ξ and a consistent estimator is given by

$$\hat{F}_n(x) = \begin{cases} \frac{1}{n}\sum_{i=1}^{n} 1_{\{x_i \le x\}} & x \le \hat{\nu}_n \\ 1 - \frac{\sum_{i=1}^{n} 1_{\{x_i > \hat{\nu}_n\}}}{n} (\frac{x}{\hat{\nu}_n})^{\frac{-1}{\hat{\xi}}} & x > \hat{\nu}_n \end{cases}$$

otherwise, G_Θ is a $GPD_{\xi,\sigma}$ and a consistent estimate of F is

$$\hat{F}_n(x) = \begin{cases} \frac{1}{n}\sum_{i=1}^{n} 1_{\{x_i \le x\}} & x \le \hat{\nu}_n \\ 1 - \frac{\sum_{i=1}^{n} 1_{\{x_i > \hat{\nu}_n\}}}{n} (1 + \hat{\xi}\frac{(x-\hat{\nu}_n)}{\hat{\sigma}})^{\frac{-1}{\hat{\xi}}} & x > \hat{\nu}_n \end{cases}$$

By the conditional probability theorem,

$$P\{X \le x\} = P\{X \le \nu\}P\{X \le x \mid X \le \nu\} + P\{X > \nu\}P\{X \le x \mid X > \nu\} \qquad (19)$$

with $P\{X \le x \mid X \le \nu\} = 1$ for $x > \nu$. The conditional probability
$P\{X \le x \mid X > \nu\} = F_\nu(x)$ can be well approximated by a Pareto type distribution G_Θ for ν large (BHP theorem). Consider x_p such that $P\{X \le x_p\} = 1 - p$, $0 < p < 1$ and $\hat{\nu}_n = x_{(k)}$ estimated by our KS distance estimator. Then, converting expression (19) into its empirical counterpart and approximating $F_\nu(x)$ by $G_{\hat{\Theta}_{\hat{\nu}_n}}$ we obtain

$$1 - p = \frac{1}{n}\sum_{i=1}^{n} 1_{\{x_i \le \hat{\nu}_n\}} + \frac{1}{n}\sum_{i=1}^{n} 1_{\{x_i > \hat{\nu}_n\}} G_{\hat{\Theta}_{\hat{\nu}_n}}.$$

For $F \in MDA(\Phi_\alpha)$, $G_{\hat{\Theta}} = PD_{\hat{\xi}}$,

$$\hat{x}_p = \hat{\nu}_n (\frac{\sum_{i=1}^{n} 1_{\{x_i > \hat{\nu}_n\}}}{pn})^{\hat{\xi}}. \qquad (20)$$

For $F \in MDA(H_\xi)$, $G_{\hat{\Theta}} = GPD_{\hat{\xi},\hat{\sigma}}$,

$$\hat{x_p} = \hat{\nu}_n + \frac{\hat{\sigma}}{\hat{\xi}}((\frac{n}{\sum_{i=1}^{n} 1_{\{x_i > \hat{\nu}_n\}}} p)^{-\hat{\xi}} - 1). \tag{21}$$

Quantile estimation is very important as a risk measure in many fields. In Finance is used as a risk indicator (Value at Risk) and in Hydrology or Meteorology to determine security levels of rainfalls or floods. Another application of $\hat{F}_n$ is to measure the uncertainty of the tail parameter estimates. There are two challenges to make inference about these parameters. First, F and the true sampling distribution of the statistic $h_n(\mathbf{x}_n, \theta)$ of the extreme parameter θ are not known, and second, the asymptotic distribution of h_n depends on nuisance parameters. $\hat{F}_n$ defined from $\hat{\nu}_n$ allows to generate bootstrap samples $\mathbf{x}_n^*$ in order to calculate the Bootstrap sampling distribution $L_n(x, \hat{F}_n) = P(h_n(\mathbf{x}_n^*, \hat{\theta}) \leq x)$ of the statistic.

Proposition 6.1. *Let $h_n(\mathbf{x}_n, \theta(F))$ be a statistic such that depends on the sample $\mathbf{x}_n$ and on the parameter $\theta(F)$. Let $L_n(x, F)$ the true sampling distribution of the statistic and $L_n(x, \hat{F}_n)$ be the bootstrap approximation. Consider $\hat{\theta}(\hat{F}_n)$ an estimator of $\theta(F)$. Then, if the Bootstrap approximation is consistent ($L_n(x, F) \simeq L_n(x, \hat{F}_n)$),*

$$P\{L_n^{-1}(\frac{\alpha}{2}, \hat{F}_n) \leq h_n(\mathbf{x}_n, \theta(F)) \leq L_n^{-1}(1 - \frac{\alpha}{2}, \hat{F}_n)\} \simeq 1 - \alpha.$$

Suppose $h_n(\mathbf{x}_n, \theta(F)) = n^\gamma(\hat{\theta}(\hat{F}_n) - \theta(F))$, $\gamma > 0$. Then, a confidence interval for $\theta(F)$ at significance level α is

$$I.C(\alpha) = [\hat{\theta}(\hat{F}_n) - n^{-\gamma} L_n^{-1}(1 - \frac{\alpha}{2}, \hat{F}_n), \hat{\theta}(\hat{F}_n) - n^{-\gamma} L_n^{-1}(\frac{\alpha}{2}, \hat{F}_n)]. \tag{22}$$

Confidence intervals for the tail index parameter proposed in (12) are calculated with this methodology but with no information about the tail behavior, i.e. $\hat{F}_n$ is the empirical distribution. Once the null hypothesis of (15) is not rejected, confidence intervals from expression (12) can be improved approximating F by our semi-parametric distribution $\hat{F}_n$ because we are counting with crucial information about the tail of F.

6.1 Finite Sample Performance

The scope of this section is to give simulated evidence about the finite sample properties of the different estimators of the threshold, as well as the impact of these estimators in the tail index estimators.

Extremes are characterized by a threshold parameter ν such that satisfies: $\hat{\xi}(\nu) \xrightarrow{p} \xi$, $\nu \to x_F$ and $F_\nu = G_\Theta(x,\nu)$. Let us start with the tail index estimator. We consider three alternative estimators: $\hat{\xi}_{ml}(\hat{\nu})$ based on a GPD with the threshold estimated by KS distance, $\hat{\xi}_{Hill}(\hat{\nu}_n)$ with $\hat{\nu}_n$ also estimated by KS distance and $\hat{\xi}_{Pic}(\nu^{Pic})$ Pickands estimator with the threshold estimated by Pickands method (see Section 4.2). These statistics depend on the threshold, therefore the method to select $\hat{\nu}_n$ is crucial to minimize possible bias effects and to get consistency. We have constructed bootstrap confidence intervals for the tail index yielded from these three different approaches. KS(GPD) and KS(PD) are the methods anchored in a Generalized Pareto and a Pareto distribution respectively. Pickands method is constructed with the estimates of the Pickands estimator obtained from the values over the estimated threshold proposed by Pickands (1975).

F	ξ	KS (GPD)	KS (PD)	Pickands
$N(0,1)$	$\xi = 0$	$[-0.41, 0.18]$	$[0.08, 0.19]$	$[-0.80, -0.35]$
$Exp(1)$	$\xi = 0$	$[-0.23, 1.22]$	$[-0.29, 0.25]$	$[-0.34, -0.05]$
t_{60}	$\xi \sim 0$	$[-0.39, 0.27]$	$[0, 0.24]$	$[-0.6, -0.31]$
t_{10}	$\xi \sim 0.1$	$[-0.28, 0.48]$	$[0.16, 0.30]$	$[-0.67, -0.09]$
$PD_{1/4,1}$	$\xi = 0.25$	$[0.02, 0.59]$	$[0.16, 0.37]$	$[0.13, 0.43]$
$PD_{1/2,1}$	$\xi = 0.5$	$[-0.13, 1.41]$	$[0.23, 0.81]$	$[0.46, 0.79]$

Table 6.1. *Confidence intervals at* $\alpha = 0.05$ *for the tail index* ξ *yielded from the three proposed estimators,* $\hat{\xi}_{ml}(\hat{\nu}_n)$, $\hat{\xi}_{Hill}(\hat{\nu}_n)$ *and* $\hat{\xi}_{Pic}(\nu^{Pic})$ *with* ν *estimated by the KS distance method and Pickands estimator, respectively.* $B = 1000$ *bootstrap samples of size* $n = 1000$ *have been generated from a sample of the distribution* F.

It can be observed that KS(GPD) confidence intervals always contain the parameter, although they are longer than the other ones. KS(PD) method outperforms the GPD method when F has heavy tails, in other cases, this estimator can produce biased confidence intervals. Pickands method only performs well for distributions with heavy tails. It is important to notice that these bootstrap intervals rely on the empirical distribution function, F_n. For large sample sizes it is not relevant the bootstrap approximation of F, however, for as the sample size decreases it is better to use $\hat{F}_n$ of expression (17), because it provides us with information about the tail when there is no sufficient available data of F.

F	KS (GPD)	KS (PD)		
	F_n	$\hat{F}_n$	F_n	$\hat{F}_n$
$N(0,1)$	$[-0.48, 1.45]$	$[-0.67, -0.11]$	$[-1.38, 0.08]$	$[0.04, 0.38]$
$Exp(1)$	$[-0.35, 1.39]$	$[-0.48, 1.56]$	$[0.02, 0.42]$	$[-2.32, 0.13]$
t_{60}	$[-1.49, 1.50]$	$[-0.62, -0.01]$	$[-0.89, 0.32]$	$[-0.03, 0.30]$
t_{10}	$[-0.39, -0.29]$	$[-0.43, 0.31]$	$[0.20, 0.59]$	$[-0.25, 0.29]$
$PD_{1/4,1}$	$[-0.78, -0.66]$	$[-0.14, 0.70]$	$[0.10, 0.42]$	$[0.19, 0.30]$
$PD_{1/2,1}$	$[0.06, 0.95]$	$[0.11, 1.11]$	$[0.18, 1.70]$	$[0.37, 0.67]$

Table 6.2. *Confidence intervals at $\alpha = 0.05$ for the tail index ξ yielded from $\hat{\xi}_{ml}(\hat{\nu}_n)$ and $\hat{\xi}_{Hill}(\hat{\nu}_n)$ with ν estimated by the KS distance method. $B = 1000$ bootstrap samples of size $n = 250$ have been generated from a sample of the distribution F.*

In the rest of the section we will be using F_n to construct the confidence intervals for the tail index, because in order to employ $\hat{F}_n$ we have first to accept the null hypothesis $F_\nu = G_\Theta$.

To check in more detail the performance of these estimators for heavy tails, in Table 6.3 we analyze t-student distributions with different degrees of freedom.

	$t_1(\xi \sim 1)$	$t_3(\xi \sim 0.33)$	$t_5(\xi \sim 0.2)$	$t_{10}(\xi \sim 0.1)$	$t_{30}(\xi \sim 0)$
KS (GPD)	$[0.37, 1.11]$	$[0.10, 1.53]$	$[-0.17, 0.33]$	$[-0.48, 0.14]$	$[-1.31, 0.50]$
KS (PD)	$[0.67, 1.24]$	$[0.09, 0.42]$	$[0.15, 0.39]$	$[0.16, 0.30]$	$[-0.03, 0.24]$
Pickands	$[0.61, 1.36]$	$[-0.44, 0.14]$	$[0.01, 0.90]$	$[-0.67, -0.09]$	$[-0.83, -0.36]$

Table 6.3. *Confidence intervals at $\alpha = 0.05$ for the tail index ξ from the three proposed estimators, $\hat{\xi}_{ml}(\hat{\nu}_n)$, $\hat{\xi}_{Hill}(\hat{\nu}_n)$ and $\hat{\xi}_{Pic}(\nu^{Pic})$ with ν estimated by the KS distance method and Pickands estimator, respectively. $B = 1000$ bootstrap samples of size $n = 1000$ have been generated from a sample of the different t-student distributions.*

In practice, the problem arises when the generating process of data is unknown and there is no information about the ratio of decay of the tail. The tail index can be estimated by both methods (KS(GPD) and KS(PD)) and depending on the results we should apply an adequate estimator for the threshold parameters, $\hat{\nu}_{n,ks}^{GPD}$ or $\hat{\nu}_{n,ks}^{PD}$, to achieve more accurate and reliable estimations of the extremes. Some financial indexes are considered in Table 6.4.

	KS (GPD)	KS (PD)	C.I. Pickands
Dax	$[-0.18, 0.89]$	$[0.23, 0.37]$	$[-0.49, -0.15]$
Ftse	$[-0.25, 0.07]$	$[-0.31, 0.15]$	$[-0.46, -0.06]$
Ibex	$[-0.11, 0.87]$	$[0.25, 0.47]$	$[-0.46, 0.04]$
Nikkei	$[-0.11, 0.56]$	$[0.27, 0.41]$	$[-0.36, 0.03]$
Dow-Jones	$[-0.15, 1.55]$	$[0.039, 0.53]$	$[-0.43, -0.03]$

Table 6.4. *Confidence intervals at* $\alpha = 0.05$ *for the tail index* ξ *for real data over roughly the period* $05/1997 - 05/2001$. $B = 1000$ *bootstrap samples of size* $n = 1000$ *have been generated for the bootstrap intervals.*

Almost all financial indexes analyzed in this Table can be considered to be fat tailed and the extremes of these distributions are well defined by $\hat{\nu}_n$ yielded from the KS estimator and the Pareto distribution (PD_ξ) with ξ contained in a precise confidence interval. Some doubts can exist with respect *Ftse* index. In this case we conclude that the extremes follow a $GPD_{\xi,\sigma}$. Consider now the second property of the threshold parameter: $\nu \to x_F$. By consistency, the threshold estimators should go to the right end point as the sample size increases.

Distribution	$n = 500$	$n = 1000$	$n = 1500$	$n = 2000$	$n = 5000$
$N(0,1)$					
$\hat{\nu}_{n,ks}^{GPD}$	1.19	1.37	1.45	1.51	1.67
	(0.57)	(0.49)	(0.47)	(0.46)	(0.42)
$\hat{\nu}_n^{Pic}$	0.44	0.52	0.59	0.64	0.88
	(0.26)	(0.29)	(0.32)	(0.33)	(0.36)
t_{10}					
$\hat{\nu}_{n,ks}^{PD}$	2.18	2.28	2.33	2.39	2.49
	(0.47)	(0.43)	(0.41)	(0.38)	(0.32)
$\hat{\nu}_n^{Pic}$	0.47	0.56	0.63	0.69	0.96
	(0.27)	(0.31)	(0.34)	(0.36)	(0.39)
$PD_{\frac{1}{4},1}$					
$\hat{\nu}_{n,ks}^{PD}$	2.14	2.13	2.11	2.07	2.07
	(0.64)	(0.62)	(0.62)	(0.61)	(0.61)
$\hat{\nu}_n^{Pic}$	1.29	1.29	1.29	1.29	1.29
	(0.07)	(0.07)	(0.08)	(0.08)	(0.08)

Table 6.5. *Threshold estimation with KS distance and Pickands estimators as n increases.* 5000 *samples of size n of different distributions are generated. The unbiased estimated standard deviation from simulations of* $\hat{\nu}_n$ *is displayed in brackets.*

As n increases, the two estimators go to the right end point of the distribution. Pickands estimator provides estimates far from the right end point and the variance slowly increases. This result points out that extremes estimates produced by Pickands method may be not very reliable. On the other hand, the estimators anchored in KS distance have decreasing variance and approach to x_F as $n \to \infty$. Notice that for $PD_{\frac{1}{4},1}$ distribution, $\hat{\nu}^{PD}_{n,ks}$ estimator has a greater variance as before and $\frac{n-k}{n} \nrightarrow 0$. This is because this distribution is exactly of Pareto type but the term of the KS statistic accounting for the sample length of the tails produces this uncertainty in the threshold estimates from the bootstrap samples. Pickands estimator detects the shape of the distribution from the beginning.

One of the goals of this paper is to propose a test to check if the extreme estimates yielded from a proposed threshold estimator verify the third property: $F_\nu = G_\Theta(x,\nu)$. The rejection of the null hypothesis means the extremes estimates defined by $\hat{\nu}_n$ are not really extremes. Tables 6.6 and 6.7 show size and power of the goodness of fit test proposed in (16). The proposed alternatives to measure the power of this test are constructed as deviations from the theoretical distribution of the extremes. Table 6.6 shows the empirical rejection rates of our test for $F \in MDA(H_\xi)$.

$n = 1000$	Size		Power (5%)		
	0.01	0.05	$Exp(1)$	$GPD_{-1/4,1}$	$GPD_{1/4,1}$
$N(0,1)$	0.014	0.07	0.98	0.96	0.96
$Exp(1)$	0.014	0.04	0.5	0.72	0.75
t_{60}	0.02	0.05	0.97	0.95	0.96
FTSE	0.006	0.048	1	1	1

Table 6.6. *B=1000 Bootstrap samples of length $n = 1000$ of the different distributions with tail exponentially decaying. m=500 simulations are generated for the bootstrap test.*

Notice that the results from the exponential distribution reflect certain lack of power of the test. This is because the tail of an exponential with mean 1 is a GPD with $\xi = 0$. Thus, our proposed alternatives are very close to the null hypothesis.

Next table displays the the empirical rejection rates of our test for $F \in MDA(\Phi_\alpha)$.

$n = 1000$	Size		Power (5%)		
	0.01	0.05	$Exp(1)$	$PD_{0.1,1}$	$PD_{0.65,1}$
t_{10}	0.012	0.038	0.79	0.74	0.97
$PD_{1/4,1}$	0.012	0.056	0.75	0.92	0.95
$PD_{1/2,1}$	0.01	0.046	0.98	0.99	0.67
Nikkei	0.014	0.042	1	1	1

Table 6.7. *B=1000 Bootstrap samples of length $n = 1000$ of the different heavy tailed distributions. m=500 simulations are generated for the bootstrap test.*

Another possibility for the alternative hypothesis is to consider more extremes than with our definition of extremes, i.e. $\tilde{\nu}_n < \hat{\nu}_n$. Let us concentrate on distributions with heavy tails. We should test $F_\nu = PD_\xi$ fixing the threshold $\tilde{\nu}_n$ in order to check if there are more data in the population that follow a Pareto distribution with tail index ξ. In addition, the opposite case can be tested as well. Consider a smaller set of extremes than the ones produced with our definition of extremes. In this case the null hypothesis should be accepted because $F_{\tilde{\nu}_n} = PD_\xi$ implies $F_{\hat{\nu}_n} = PD_\xi$ with $\hat{\nu}_n < \tilde{\nu}_n$.

Data	$\hat{\nu}_n$	$\tilde{\nu}_n = x_{(950)}$	$\tilde{\nu}_n = x_{(900)}$	$\tilde{\nu}_n = x_{(800)}$	$\tilde{\nu}_n = x_{(700)}$
t_{10}	$\gamma_{0.97} = 2.27$	0.19	0.01	0.00	0.00
	$\hat{s} = (0.42)$	(0.29)	(0.07)	(0.00)	(0.00)
t_3	$\gamma_{0.97} = 2.97$	0.29	0.13	0.0001	0.00
	$\hat{s} = (0.97)$	(0.33)	(0.26)	(0.002)	(0.00)
DaX	$x_{(910)} = 0.025$	0.69	0.20	0.00	0.00
Nikkei	$x_{(920)} = 0.021$	0.97	0.05	0.00	0.00

Table 6.8. *p-values of the bootstrap hypothesis tests $H_0 : F_{\tilde{\nu}} = PD_\xi$ for samples of $n =$ 1000 observations. For the t-student distributions $m = 500$ iterations are generated. γ_p is the extreme quantile $\hat{\nu}_n$ of the distribution. The unbiased estimated standard deviation of the p-values is displayed in brackets.*

7 Conclusion

Risk and uncertainty are not the same thing (see Granger, 2002) and therefore they need to be characterized by different measures. It is accepted that variance is well designed to capture the latter but not the former. To measure risk, in other words, to respond the question *if things go wrong how wrong they can go?* it is first necessary to find an answer to the question *which extreme values are really extremes?* This is the main goal of this paper, where following Pickands (1975) methodology we do not only define formally and analytically the set of extreme observations of a given population, but we propose a simple estimator of them and construct a test to answer the previous question. Identification of the extreme observations allows to estimate very accurately risk measures as Value at Risk or Expected Shortfall, as well as to make inference on different tail parameters of interest. Boths issues are extensions of this paper and constitute undergoing research by the authors.

A Appendix: Proofs

Corollary 3.1: Taking logs in expression (1), we have $n\ log(1-\overline{F}(c_n x+d_n)) \xrightarrow{d} logH_\xi(x)$. Therefore, $log\ (1-\frac{n\overline{F}(c_n x+d_n)}{n})^n \xrightarrow{d} logH_\xi(x)$. This is equivalent to $n\ \overline{F}(c_n x+d_n) \xrightarrow{d} -logH_\xi(x)$, with $H_\xi = e^{-(1+\xi x)^{\frac{-1}{\xi}}}$ if $\xi \neq 0$ and $H_\xi = e^{-e^{-x}}$ if $\xi = 0$. We obtain $n\overline{F}(c_n x+d_n) \xrightarrow{d} (1+\xi x)^{\frac{-1}{\xi}}$ if $\xi \neq 0$ and $n\overline{F}(c_n x+d_n) \xrightarrow{d} e^{-x}$. □

Corollary 3.2: Let $F \in MDA(\Phi_\alpha)$ and $M_n = max(x_1,\ldots,x_n)$. By definition, there exist constants, $c_n = F^{\leftarrow}(1-\frac{1}{n})$ and $d_n = 0$ such that $c_n^{-1}(M_n - d_n) \xrightarrow{d} \Phi_\alpha$ with $\Phi_\alpha = e^{-x^{-\alpha}}, x > 0$, and $\alpha > 0$. By proposition (3.1), $F \in RV_{-\alpha}$. Consider $\nu, x \in$ support(F) with $x_F = \infty$ and $x = \nu t$ with $t > 1$. Notice that for $0 < t \leq 1$, $F_\nu(x) = 0$. Operating in expression (4),

$$1 - \lim_{\nu\to\infty} \frac{1-F(x)}{1-F(\nu)} = \lim_{\nu\to\infty} \frac{F(x)-F(\nu)}{1-F(\nu)} = \lim_{\nu\to\infty} F_\nu(x) = 1-(\frac{x}{\nu})^{-\alpha} = PD_\xi(\frac{x}{\nu}). \qquad \square$$

Proposition 4.1: First the if part. Consider $\hat{\nu}_n$ a threshold estimator such that the values above it are extreme values. Therefore expression (10) can be written, replacing the parameter by the estimator, as

$$\lim_{\hat{\nu}_n\to x_F} \sup_{x\in\mathbb{R}} |F_{\hat{\nu}_n}(x) - G_\Theta(x,\hat{\nu}_n)| = 0.$$

This implies

$$P\{\sup_{x\in\mathbb{R}} |F_{\hat{\nu}_n}(x) - G_\Theta(x,\hat{\nu}_n)| > \varepsilon\} \to 0.$$

In addition, if $\hat{\nu}_n$ defines the set of extreme values there may exist a subset $A_n \subseteq \mathbb{R}$ such that $|F_{\hat{\nu}_n}(A_n) - G_\Theta(A_n,\hat{\nu}_n)| > \varepsilon$, although from (10) $\sup_{x\in A_n} |F_{\hat{\nu}_n}(x) - G_\Theta(x,\hat{\nu}_n)| \to 0$. Then, it is derived that $F_{\hat{\nu}_n}(x) = G_\Theta(x,\hat{\nu}_n)\ \forall x \in \mathbb{R}\backslash A_n$.

With respect to the only if part, this result follows from condition (2). The continuity near the right end point x_F and the consistency of the estimator $\hat{\nu}_n$ imply that

$$\lim_{\hat{\nu}_n\to x_F} \sup_{x\in\mathbb{R}} |F_{\hat{\nu}_n}(x) - G_\Theta(x,\hat{\nu}_n)| = 0. \qquad \square$$

Proposition 4.2: Let $x_1, \ldots, x_k \sim PD_\xi$ with $PD_\xi(\frac{x}{\nu}) = 1 - (\frac{x}{\nu})^{-\alpha}, x > \nu$. The density function is $pd(x) = \alpha(\frac{x}{\nu})^{-(\alpha+1)}\frac{1}{\nu}$. Then, the likelihood function is $l(x_1, \ldots, x_k; \nu, \alpha) = (\frac{\alpha}{\nu})^k \prod_{i=1}^{k} (\frac{x_i}{\nu})^{-(\alpha+1)}$. Let $\xi = \frac{1}{\alpha}$, then from the first order conditions, it is easy to obtain $\hat{\xi} = \frac{1}{k}\sum_{i=1}^{k} log\frac{x_i}{\nu}$. $\square$

Theorem 4.1: Let $\hat{\nu}_n$ be the threshold estimator derived from the KS distance and let $\hat{\xi}(\hat{\nu}_n)$ be a consistent estimator of the tail index based on $\mathbf{x}_n$ with $\xi \in \Theta$.
$\sup_{x\in\mathbb{R}} |F_{\hat{\nu}_n}(x) - G_\Theta(x, \hat{\nu}_n)| \leq \sup_{x\in\mathbb{R}} |F_{\hat{\nu}_n}(x) - G_{\hat{\Theta}}(x, \hat{\nu}_n)| + \sup_{x\in\mathbb{R}} |G_\Theta(x, \hat{\nu}_n) - G_{\hat{\Theta}}(x, \hat{\nu}_n)|$.
$\hat{\Theta}$ is a consistent estimator of Θ, therefore, $\sup_{x\in\mathbb{R}} |G_\Theta(x, \hat{\nu}_n) - G_{\hat{\Theta}}(x, \hat{\nu}_n)| \xrightarrow{p} 0$.
Let $X_n(\nu) = \sqrt{\sum_{i=1}^{n} 1_{\{x_i > \nu\}}} \sup_{x\in\mathbb{R}} |F_\nu(x) - G_{\hat{\Theta}_\nu}(x, \nu)|$ such that for values of ν sufficiently large $X_n(\nu)$ is a random variable that follows a functional of a centered gaussian process depending on the parameter Θ (see Durbin, 1973). Consider now, $X_n(\hat{\nu}_n) = min\{X_{n,1}(\nu_1), \ldots, X_{n,k}(\nu_k)\}$ with $\{\nu_1, \ldots, \nu_k\}$ greater than a ν_0 verifying BHP theorem and $X_{n,i}(\nu_i)$ random variables. $\hat{\nu}_n$ is the argument of the minimum of this finite set; $\hat{\nu}_n = arg \min_\nu X_n(\nu)$. Then,
$P(X_n(\hat{\nu}_n) > \varepsilon) = P(min\{X_{n,1}(\nu_1), \ldots, X_{n,k}(\nu_k)\} > \varepsilon) = P(X_{n,i}(\nu_i) > \varepsilon)^k$.
As n goes to infinity k increases as well. In addition, $P(X_n(\nu) > \varepsilon) < 1$, therefore, $P(X_n(\hat{\nu}_n) > \varepsilon) \to 0$ as $n, k \to \infty$. This expression is equivalent to

$$P\{\sqrt{\sum_{i=1}^{n} 1_{\{x_i > \hat{\nu}_n\}}} \sup_{x\in\mathbb{R}} |F_{\hat{\nu}_n}(x) - G_{\hat{\Theta}_{\hat{\nu}_n}}(x, \hat{\nu}_n)| > \varepsilon\} \to 0.$$

Then,

$$P\{\sup_{x\in\mathbb{R}} |F_{\hat{\nu}_n}(x) - G_{\hat{\Theta}_{\hat{\nu}_n}}(x, \hat{\nu}_n)| > \varepsilon^*\} \to 0 \text{ with } 0 < \varepsilon^* < \varepsilon. \quad \square$$

Proposition 5.1: Let $\mathbf{x}_n$ be a sample of size n from F. Assume that $\hat{F}_n$ is an estimate of F based on $\mathbf{x}_n$ verifying $\sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)| \xrightarrow{p} 0$ and let $J_n(x, F; \hat{\Theta})$ be the true sampling distribution of the statistic $T_n(\mathbf{x}_n, \hat{\Theta})$. This distribution is such that $J_n(x, F; \hat{\Theta}) \to J(x, F; \Theta)$ with $J(x, F; \Theta)$ being a strictly increasing continuous function in x. Then,
$P\{T_n(x, \hat{\Theta}) \leq J_n^{\leftarrow}(1-\alpha, F; \hat{\Theta})\} \to P\{T_n(x, \hat{\Theta}) \leq J^{\leftarrow}(1-\alpha, F; \Theta)\} = 1 - \alpha$. In addition, $J(x, F; \Theta)$ is continuous and strictly increasing, therefore
$J_n^{\leftarrow}(1-\alpha, F; \hat{\Theta}) \to J^{\leftarrow}(1-\alpha, F; \Theta)$. Then, as $n \to \infty$, $J_n^{\leftarrow}(1-\alpha, \hat{F}_n; \hat{\Theta}^*) \to J^{\leftarrow}(1-\alpha, F; \Theta)$ because $\sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)| \xrightarrow{p} 0$. Consequently, $P\{T_n(x, \hat{\Theta}) \leq J_n^{\leftarrow}(1-\alpha, \hat{F}_n; \hat{\Theta}^*)\} \to P\{T_n(x, \hat{\Theta}) \leq J^{\leftarrow}(1-\alpha, F; \Theta)\} = 1 - \alpha$. Then, $\sup_{x\in\mathbb{R}} |J_n(x, F; \hat{\Theta}) - J_n(x, \hat{F}_n; \hat{\Theta}^*)| \leq \sup_{x\in\mathbb{R}} |J_n(x, F; \hat{\Theta}) - J(x, F; \Theta)| + \sup_{x\in\mathbb{R}} |J_n(x, \hat{F}_n; \hat{\Theta}^*) - J(x, F; \Theta)| \to 0$. $\square$

Proposition 5.2: For $x \leq \hat{\nu}_n$, $\hat{F}_n(x)$ is the Empirical distribution function. By Glivenko-Cantelli theorem, $\sup_{x \leq \hat{\nu}_n} |\hat{F}_n(x) - F(x)| = \sup_{x \leq \hat{\nu}_n} |F_n(x) - F(x)| \xrightarrow{a.s.} 0$. For $x > \hat{\nu}_n$, under the null hypothesis, $F_{\hat{\nu}_n} = G_\Theta$ and $G_{\hat{\Theta}_{\hat{\nu}_n}}$ is a consistent estimate of G_Θ.

$P\{X \leq x\} = P\{X \leq \hat{\nu}_n\}P\{X \leq x \mid X \leq \hat{\nu}_n\} + P\{X > \hat{\nu}_n\}P\{X \leq x \mid X > \hat{\nu}_n\}$. Therefore, $\hat{F}_n(x) = F_n(\hat{\nu}_n) + \overline{F}_n(\hat{\nu}_n)G_{\hat{\Theta}_{\hat{\nu}_n}}(x)$.

$\underset{x>\hat{\nu}_n}{sup} \quad |\hat{F}_n(x) - F(x)| = \underset{x>\hat{\nu}_n}{sup} \quad |1 - \overline{F}_n(x)\overline{G}_{\hat{\Theta}_{\hat{\nu}_n}}(x) - F(x)| = \underset{x>\hat{\nu}_n}{sup} \quad |\overline{F}(x) - \overline{F}_n(\hat{\nu}_n)\overline{G}_{\hat{\Theta}_{\hat{\nu}_n}}(x)|$.

The distribution of data $F(x)$ can be written in terms of the tail distribution as $\overline{F}(x) = \overline{F}(\hat{\nu}_n)\overline{F}_{\hat{\nu}_n}(x)$. Then, $\underset{x>\hat{\nu}_n}{sup} \quad |\overline{F}(\hat{\nu}_n)\overline{F}_{\hat{\nu}_n}(x) - \overline{F}_n(\hat{\nu}_n)\overline{G}_{\hat{\Theta}_{\hat{\nu}_n}}(x)| \xrightarrow{p} 0$ by consistency of the threshold estimator (see Theorem 4.1) and by Glivenko-Cantelli theorem. $\square$

Theorem 5.1: Let $\hat{Q}_n$ be an estimator of F based on a sample $\mathbf{x}_n$ of size n that satisfies $\underset{x\in\mathbb{R}}{sup} \quad |\hat{Q}_n(x) - F(x)| \rightarrow 0$ in probability whenever $F \in F_{H_0}$. Let $j(1-\alpha, F)$ be the $1-\alpha$ asymptotic quantile of the distribution $J_n(x, F)$ of a statistic $T_n(x, \Theta)$. Then, $P\{T_n(x, \Theta) > j(1-\alpha, F)\} \rightarrow \alpha$. The asymptotic distribution of the statistic is continuous, then, $j_n(1-\alpha, \hat{Q}_n) \rightarrow j(1-\alpha, F)$ with $j_n(1-\alpha, \hat{Q}_n)$ the $1-\alpha$ quantile of the Bootstrap distribution $J_n(x, \hat{Q}_n)$. By Slutsky's theorem, $P\{T_n(x, \Theta) > j_n(1-\alpha, \hat{Q}_n\} \rightarrow \alpha$ as $n \rightarrow \infty$. $\square$

Proposition 5.3: Let $\mathbf{x}_n$ be a sample of size n from a distribution F under the alternative hypothesis, $H_1 : F_{\hat{\nu}_n} \ncong G_\Theta$. $\exists A_n \in \mathbb{R}$ such that $\forall\ \delta > 0$, $|F_{\hat{\nu}_n,n}(A_n) - G_{\hat{\Theta}}(A_n, \hat{\nu}_n)| > \delta$ with $\hat{\Theta}$ a consistent estimate of Θ under H_0. Then, $\sqrt{n}\underset{x\in\mathbb{R}}{sup} \quad |F_{\hat{\nu}_n,n}(x) - G_{\hat{\Theta}_{\hat{\nu}_n}}(x)| \geq \sqrt{n}\underset{x\in\mathbb{A}_n}{sup} \quad |F_{\hat{\nu}_n,n}(x) - G_{\hat{\Theta}_{\hat{\nu}_n}}(x)| > \sqrt{n}\delta$. $\square$

Proposition 6.1: Let $\hat{\theta}(\hat{F}_n)$ be an estimator of $\theta(F)$ and $h_n(\mathbf{x}_n, \theta(F))$ be the statistic with sampling distribution $L_n(x, F)$. Let α be the significance level.

$P\{L_n^{-1}(\frac{\alpha}{2}, F) \leq h_n(\mathbf{x}_n, \theta(F)) \leq L_n^{-1}(1-\frac{\alpha}{2}, F)\}$ can be approximated by

$P\{L_n^{-1}(\dfrac{\alpha}{2}, \hat{F}_n) \leq h_n(\mathbf{x}_n, \theta(F)) \leq L_n^{-1}(1-\dfrac{\alpha}{2}, \hat{F}_n)\} \simeq 1-\alpha$, if the bootstrap is consistent ($L_n(x, F) \simeq L_n(x, \hat{F}_n)$).

Let $h_n(\mathbf{x}_n, \theta(F)) = n^\gamma(\hat{\theta}(\hat{F}_n) - \theta(F))$, $\gamma > 0$

$$P\{L_n^{-1}(\tfrac{\alpha}{2}, \hat{F}_n) \leq n^\gamma(\hat{\theta}(\hat{F}_n) - \theta(F)) \leq L_n^{-1}(1-\tfrac{\alpha}{2}, \hat{F}_n)\} \rightarrow 1-\alpha$$

$$I.C_{1-\alpha}(\theta)(F) = [\hat{\theta}(\hat{F}_n) - n^{-\gamma}L_n^{-1}(1-\tfrac{\alpha}{2}, \hat{F}_n), \hat{\theta}(\hat{F}_n) - n^{-\gamma}L_n^{-1}(\tfrac{\alpha}{2}, \hat{F}_n)]. \quad \square$$

B Appendix: Notation List

$x_{(i)}$: i^{th} order statistic.

$\mathbf{x}_n$: $x_1, \ldots, x_n$.

$x_+ : max(x, 0)$.

$x \rightarrow a^+$: x approaches to a by the right side.

$\lceil x \rceil$: integer greater or equal than x.

$\rightarrow$: convergence as n goes to ∞.

$\stackrel{p}{\to}$: convergence in probability.

$\stackrel{d}{\to}$: weak convergence.

$\stackrel{a.s.}{\to}$: almost sure convergence.

$F^{\leftarrow}$: inverse of the distribution function.

$\overline{F}$: tail of F

$\sim$: follows a distribution.

$\simeq$: approximates.

$\not\simeq$: does not approximate.

Θ_{Pic}: estimated parameters by Pickands method.

References

Balkema, A.A., and Haan, L.de, (1974): "Residual lifetime at great age," *Annals of Probability*, 2, $792-804$.

Babu, G.J., and Rao, C.R., (2002): "Goodness-of-fit tests when parameters are estimated," Pennsylvania State University, (Working Paper).

Danielsson, J., Haan, L.de, Peng, L., de Vries, C.G., (2001): "Using a Bootstrap Method to choose the sample fraction in the tail index estimation," *Journal of Multivariate Analysis*, Vol 76, $226-248$.

Dekkers, A.L.M., and Haan, L.de, (1993): "Optimal choice of sample fraction in extreme value estimation," *Journal of Multivariate Analysis*, Vol 47, $173-195$.

Dekkers, A.L.M., Einmahl, J.H., and Haan, L.de, (1989): "A moment estimator for the index of an extreme value distribution," *Annals of Statistics*, Vol 17, $1833-1855$.

Drees, H., (1995): "Refined Pickands estimators of the extreme value index," *Annals of Statistics*, Vol 23, $2059-2080$.

Drees, H. and Kaufmann, E. (1998): "Selection of the optimal sample fraction in univariate extreme value estimation," *Stochastic Processes and their applications*, Vol 75, $149-172$.

Durbin, J., (1973): "Weak convergence of the sample distribution function when parameters are estimated," *Annals of Statistics*, Vol 1, $279-290$.

Embrechts, P., Klüppelberg, C., and T.Mikosch, (1997): *Modelling extremal events for insurance and finance*, Ed. Springer, Berlin.

Fisher, R.A., and Tippet, L.H.C., (1928): "Limiting forms of the frecuency distribution of the largest and smallest member of a sample," *Proc. Cambridge Philos. Soc.*, 24, $180-190$.

Gnedenko, B.V., (1943): "Sur la distribution limit du terme d'une série aléatorie," *Annals of Mathematics*, 44, 423 − 453.

Goldie, C.M., and Smith R.L., (1987): "Slow variation with remainder: theory and applications," *Quart. J. Math. Oxford Ser.*(2) 38, 45 − 71.

Granger, C.W.J., (2002): "Some comments on risk," *Journal of Applied Econometrics.*, 17, 447 − 456.

Guillou, A., and Hall, P., (2001): "A diagnostic for selecting the threshold in extreme value analysis," *Journal of Royal Statistical Society B*, [**63**], Part 2, 3(2), 293 − 305.

Hill, B.M., (1975): "A simple general approach to inference about the tail of a distribution," *Annals of statistics*, Vol 3, 5, 1163 − 1174.

Kolmogorov, A.N., (1933): "Sulla determinazione empirica di una legge di distribuzione," *Giornale del Instituto Italiano degli Attuari*, 4, 83 − 91.

Mason, D., and Schuenemeyer, J.H., (1983): "A modified Kolmogorov-Smirnov test sensitive to tail alternatives," *Annals of Statistics*, Vol 11, 3, 933 − 946.

McNeil, J., (1997): "Estimating the Tails of Loss Severity Distributions using Extreme Value Theory," *ASTIN Bulletin*, 27, 117 − 137.

McNeil, J., and Saladin, T., (2001): "The Peaks over Thresholds Method for Estimating High Quantiles of Loss Distributions," in Proceedings of XXVIIth International ASTIN Colloquium, 23 − 43, Cairns, Australia.

Pickands III, J., (1975): "Statistical Inference Using Extreme Order Statistics," *Annals of Statistics*, Volume 3, Issue 1, 119 − 131.

Smith, R.L., (1984): "Threshold methods for Sample Extremes," *Statistical Extremes and Applications*, ed. J. Tiago de Oliveira, Dordrecht: D. Reidel, 621 − 638.

von Mises, R., (1936): "La distribution de la plus grande de n valeurs", Selected papers of Richard von Mises, 2, 271 − 294. *American Mathematical Society*, Providence.