

Which Extreme Values Are Really Extreme?

JESÚS GONZALO

Universidad Carlos III de Madrid

JOSÉ OLMO

Universidad Carlos III de Madrid

ABSTRACT

We define the extreme values of any random sample of size n from a distribution function F as the observations exceeding a threshold and following a type of generalized Pareto distribution (GPD) involving the tail index of F . The threshold is the order statistic that minimizes a Kolmogorov–Smirnov statistic between the empirical distribution of the corresponding largest observations and the corresponding GPD. To formalize the definition we use a semiparametric bootstrap to test the corresponding GPD approximation. Finally, we use our methodology to estimate the tail index and value at risk (VaR) of some financial indexes of major stock markets.

KEYWORDS: bootstrap, extreme values, goodness-of-fit test, Hill estimator, Pickands theorem, VaR

Risk management is one of the most important innovations of the 20th century in economics. During the last decade financial markets have realized the importance of monitoring risk. 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.

Alternative measures regarding possible values out of the range of available information need to be defined and calculated. Extreme value theory (EVT) provides the tools to model the asymptotic distribution of the maximum of a sequence of random variables $\{X_n\}$, and in this sense this theory can be very helpful in order to obtain a first impression about how wrong things

We thank participants at the Conference on Extremal Events in Finance, New Frontiers in Financial Volatility Modelling, Winter Meetings of the North American Econometric Society, and the Department of Statistics, University of North Carolina–Chapel Hill, especially to Ross Leadbetter and Richard Smith for excellent comments. We are also deeply grateful to Christopher Geczy, Alfonso Novales, Michael Wolf, and two anonymous referees for valuable suggestions and comments. Financial support provided by a DGICYT grant (SEC01-0890) is gratefully acknowledged. Address correspondence to Jesús Gonzalo, Department of Economics, Universidad Carlos III de Madrid, 28903, Getafe, Madrid, Spain, or e-mail: jesus.gonzalo@uc3m.es.

doi:10.1093/jfinec/nbh014

Journal of Financial Econometrics, Vol. 2, No. 3, © Oxford University Press 2004; all rights reserved.

can go. A deeper insight into EVT allows us to know not only the order of convergence of the maximum, but also the limiting distribution of the largest observations of the sequence. These observations are the main ingredients of more informative risk measures that have been recently introduced, like value at risk (VaR) or expected shortfall. These measures are functions of extreme quantiles of the data distribution. Attempting to model the tails of these distributions is troublesome and standard methodologies such as historical simulation or the gaussian distribution do not provide reliable approximations at very high quantiles.

On the other hand, the methodology derived from EVT covers this gap and produces a parametric framework to derive the VaR or any function of this extreme quantile. It is clear that the first task is to identify which values are really extreme values. In practice this is done by graphical methods such as the QQ plot, sample mean excess plot, or by other ad hoc methods that impose an arbitrary threshold (5%,10%,...) [see Embrechts, Klüppelberg, and Mikosch (1997)]. These methods do not propose any formal computable method, and moreover, they only give very rough estimates of the set of extreme values. In this article we propose a formal way of identifying and estimating the extreme values of any random sample of size n coming from a distribution function, say F . These values are going to be defined as the exceedances of a threshold sequence $\{u_n\}$ following a type of generalized Pareto distribution (GPD). The selection of this threshold plays a central role in this definition and in estimating the parameters of the GPD. The sequence of extreme values depends on the length of the data sequence by the choice of $\{u_n\}$. Therefore we need to introduce an appropriate test to assess statistically whether the distribution function of the set of extremes given by the threshold really satisfies the weak convergence to the GPD or not, with parameters driven by F . In order to achieve this task, we propose a semiparametric bootstrap test and study its asymptotic as well as its finite sample performance.

The final purpose of our methodology is to achieve a reliable approximation of F , paying special attention to its tails. Our tail estimate provides accurate approximations of the extreme quantiles of F , and from them it is straightforward to calculate the risk measures introduced in the financial literature.

The article is structured as follows. In Section 1 we present some general results of extreme value theory, focusing on the weak convergence of the largest observations of a random sequence. Section 2 introduces different approaches to select the threshold sequence and gives a brief review of estimation methods for the parameters of the GPD. Some simulations show the performance of our approach in terms of tail index estimation. The complete definition of the sequence of extreme values is given in Section 3 by means of a bootstrap hypothesis test. Monte Carlo simulations provide the finite sample performance of our proposed test. Section 4 presents an empirical application where the risk of financial indexes of major stock markets is analyzed via the tail index and VaR. Finally, Section 5 presents some concluding remarks. Proofs are presented in the appendix.

1 REVIEW OF EXTREME VALUE THEORY RESULTS

The purpose of this section is to briefly introduce the set of results of the so-called extreme value theory necessary to develop the theory of the article. The departing point is the study of the weak convergence for the sample maximum of a sequence of random variables $\{X_n\}$ with distribution function F . Our intention is to use the limiting distribution of this statistic to derive the weak convergence of the largest observations of a random sequence imposing a minimum set of assumptions on the distribution function F .

Let $M_n = \max\{X_1, \dots, X_n\}$ be the sample maximum of the sequence and let F be the common distribution function for $\{X_n\}$. Our first goal is to introduce the conditions under which M_n converges weakly to a nondegenerate distribution function.

Result 1 *Let $\{X_n\}$ be an independent and identically distributed (i.i.d.) sequence. Let $0 \leq \tau \leq \infty$ and suppose that $\{u_n\}$ is a sequence of real numbers such that*

$$n(1 - F(u_n)) \rightarrow \tau \text{ as } n \rightarrow \infty. \tag{1}$$

Then

$$P\{M_n \leq u_n\} \rightarrow e^{-\tau} \text{ as } n \rightarrow \infty. \tag{2}$$

Conversely, if Equation (2) holds for some $\tau, 0 \leq \tau \leq \infty$, then so does Equation (1).

The proof of this result is immediately derived from

$$P\{M_n \leq u_n\} = F^n(u_n) = \left(1 - \frac{n(1 - F(u_n))}{n}\right)^n. \tag{3}$$

However, this result does not guarantee the existence of a nondegenerate distribution for M_n . Define the right endpoint of a distribution function as $x_F = \sup\{x | F(x) < 1\} \leq +\infty$. It is clear that $M_n \rightarrow x_F$ with probability 1 as $n \rightarrow \infty$. Suppose now that F has a jump at x_F with $x_F < \infty$ (i.e., $F(x_F^-) < 1$ with $F(x_F^-) = \lim_{x \uparrow x_F} F(x)$), and consider a sequence $\{u_n\}$ satisfying Equation (2) with $0 \leq \tau \leq \infty$. Then either $u_n < x_F$ for infinitely many values of n and $n(1 - F(u_n)) \rightarrow \infty$, or $u_n > x_F$ and $n(1 - F(u_n)) = 0$. Therefore we also need some regularity condition on the tail of F to avoid the existence of such jumps.

Result 2 *Let F be a distribution function with right endpoint x_F such that*

$$\lim_{x \uparrow x_F} \frac{1 - F(x)}{1 - F(x^-)} = 1, \tag{4}$$

and let $\{u_n\}$ be a sequence with $u_n < x_F$ and $n(1 - F(u_n)) \rightarrow \tau$. Then $0 < \tau < \infty$.

We will assume hereafter these regularity conditions as our minimum set of assumptions on the distribution function F .

The choice of the sequence $\{u_n\}$ determines the value of τ . Suppose $v_n > u_n$ and Equation (2) holds, then $n(1 - F(v_n)) \rightarrow \tau'$ with $\tau' < \tau$. We can write Equation (2) as $P\{M_n \leq u_n(x)\} \rightarrow e^{-\tau(x)}$, with u_n depending on x . Moreover, there exist some

scaling sequences a_n, b_n varying according to F such that

$$P\{a_n^{-1}(M_n - b_n) \leq x\} \rightarrow G(x) \quad \text{as } n \rightarrow \infty, \tag{5}$$

with $u_n(x) = a_n x + b_n$ and $G(x) = e^{-\tau(x)}$ a distribution function. This function has been fully characterized by Gnedenko (1943) or de Haan (1976) via the analysis of domains of attraction for the maximum, and it can be summarized as follows:

Result 3 *The distribution function $G(x)$ derived in Equation (5) can only take three different forms,*

$$\begin{aligned} \text{Type I: (Gumbel)} \quad & G(x) = e^{-e^{-x}}, \quad -\infty < x < \infty, \\ \text{Type II: (Fréchet)} \quad & G(x) = \begin{cases} 0 & x \leq 0, \\ e^{-x^{-\xi}} & x > 0, \quad \xi > 0 \end{cases} \\ \text{Type III: (Weibull)} \quad & G(x) = \begin{cases} 1 & x \geq 0, \\ e^{-(x)^{-\xi}} & x < 0, \quad \xi < 0. \end{cases} \end{aligned}$$

The parameter ξ is the tail index of F and characterizes the tail behavior of the distribution function. The three types can be gathered in the so-called generalised extreme value distribution, first proposed by von Mises (1936),

$$G(x) = e^{-(1 + \xi \frac{x - \mu}{\sigma})^{\frac{1}{\xi}}}, \tag{6}$$

where μ is a location parameter, σ a scale parameter, and $\xi \neq 0$.

This expression boils down to $G(x) = e^{-e^{-\frac{x - \mu}{\sigma}}}$ when $\xi = 0$. Clearly $\tau(x) = (1 + \xi \frac{x - \mu}{\sigma})^{\frac{1}{\xi}}$ in Equation (5), and hence $n(1 - F(u_n(x))) \rightarrow (1 + \xi \frac{x - \mu}{\sigma})^{\frac{1}{\xi}}$ for all x , where a_n, b_n are suitable constants. This is the result we exploit in order to derive the weak convergence of the largest observations determined by a threshold sequence $u_{on} = a_n \mu + b_n$, with μ satisfying $-\log G(\mu) = 1$. By doing that

$$\frac{1 - F(u_n(x))}{1 - F(u_{on})} \rightarrow \left(1 + \xi \frac{x - \mu}{\sigma}\right)^{\frac{1}{\xi}}, \quad \text{as } n \rightarrow \infty. \tag{7}$$

This expression can be rewritten as

$$\frac{F(u_n(x)) - F(u_{on})}{1 - F(u_{on})} \rightarrow 1 - \left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-\frac{1}{\xi}}, \tag{8}$$

for all $x > \mu$ continuity points. The threshold sequence satisfies $u_n(x) = u_{on} + a_n(x - \mu)$, and we can define

$$F_{u_{on}}(a_n(x - \mu)) = \frac{F(u_{on} + a_n(x - \mu)) - F(u_{on})}{1 - F(u_{on})}, \tag{9}$$

as the conditional excess distribution function given u_{on} with $x > \mu$. This takes us directly to the following result:

Result 4 *Let $y = a_n(x - \mu)$, then*

$$\lim_{u_{on} \rightarrow x_F} \sup_{0 \leq y < \infty} |F_{u_{on}}(y) - GPD_{\xi, \sigma(u_{on})}(y)| = 0, \tag{10}$$

with

$$GPD_{\xi, \sigma(u_{on})}(y) = \begin{cases} 1 - \left(1 + \xi \frac{y}{\sigma(u_{on})}\right)^{-\frac{1}{\xi}} & \text{if } \xi \neq 0 \\ 1 - e^{-\frac{y}{\sigma(u_{on})}} & \text{if } \xi = 0 \end{cases}, \tag{11}$$

the generalized Pareto distribution and $\sigma(u_{on}) = \sigma a_n$.

This result is known as Pickands (1975) theorem. Pickands proposed a sequence u_{on} taken in the interval $[b_n, b_{n+1}]$, with b_n the suitable sequence in Equation (5). This approximation for the distribution of the largest observations regarded as the exceedances of a threshold sequence can be improved when the tail of F decays at a polynomial rate.

Suppose $1 - F(x) = x^{-\frac{1}{\xi}}L(x)$ with $L(tx)/L(x) \rightarrow 1$ as $x \rightarrow x_F$ and $\xi > 0$, then the distribution function F satisfies

$$\lim_{x \uparrow x_F} \frac{1 - F(tx)}{1 - F(x)} = t^{-\frac{1}{\xi}}, \quad t > 0. \tag{12}$$

This type of distribution function is regularly varying at a rate $\frac{1}{\xi}$ and the domain of attraction of the sample maximum is the Fréchet distribution [see Resnick (1987) or de Haan (1976)]. The function $L(x)$ is said to be slowly varying and is introduced to include the deviations of F from the Pareto probability law. When these departures from the polynomial law are small, $F_{u_{on}}(y)$ is better approximated by the Pareto distribution function. Consider a sequence $u_n(x) = u_{on}x$, where $u_{on} = u_n(1)$ is the threshold sequence that satisfies $1 - F(u_{on}) = u_{on}^{-\frac{1}{\xi}}L(u_{on})$. The conditional excess distribution function defined by u_{on} as $F_{u_{on}}(u_n(x)) = \frac{F(u_n(x)) - F(u_{on})}{1 - F(u_{on})}$ satisfies

$$F_{u_{on}}(u_n(x)) \rightarrow 1 - \left(\frac{u_n(x)}{u_{on}}\right)^{-\frac{1}{\xi}}, \quad \text{as } n \rightarrow \infty, \tag{13}$$

for $u_n(x) \geq u_{on}$ or equivalently for $x \geq 1$. This convergence holds for all continuity points of F and therefore for this case we can rewrite the previous result as

$$\lim_{u_{on} \rightarrow x_F} \sup_{[u_{on} \leq y < \infty]} |F_{u_{on}}(y) - PD_{\xi}(y)| = 0, \tag{14}$$

with $y = u_n(x)$ and $PD_{\xi}(y) = 1 - \left(\frac{y}{u_{on}}\right)^{-\frac{1}{\xi}}$.

Finally, the choice of the threshold sequence also has an effect on the error made by the approximations claimed in Pickands theorem. This error arises from the asymptotic relation $n(1 - F(u_n)) \rightarrow \tau$ and from the approximation of $F^n(u_n)$ by the exponential distribution. The latter approximation is of order $o(n^{-1})$ since

$$0 \leq e^{-x} - \left(1 - \frac{x}{n}\right)^n \leq 0.3 \frac{1}{n-1},$$

for $0 \leq x \leq n$ [see, e.g., Leadbetter, Lindgren, and Rootzén (1983)]. Nevertheless, if F is continuous one can always obtain an equality in Equation (2) by taking $u_n = F^{-1}(e^{-\frac{\tau}{n}})$ and making the approximation errors vanish. However, sequences of

type $u_n(x) = a_n x + b_n$, with a_n, b_n suitable constants are more appropriate to study the weak convergence of M_n . In these cases, the equality or any uniform bound for all x are not usually feasible in Equation (5).

2 THRESHOLD CHOICES TO DEFINE THE EXTREME VALUES

The last section has focused on finding the asymptotic laws that rule the largest observations of a random sequence from a distribution function F . This set of observations is defined by means of a threshold sequence and the tail index ξ that characterizes the corresponding generalized Pareto or Pareto. The choice of this sequence is troublesome since $u_{on} \rightarrow x_F$ when $n \rightarrow \infty$, but at an appropriate rate. This order of convergence depends on F represented by the sequences a_n and b_n when $u_n(x)$ is of the form $u_n(x) = a_n x + b_n$. Hence the threshold sequence u_{on} can be defined by the scaling sequences a_n, b_n and the value of x satisfying the condition $-\log G(x) = 1$, or equivalently $n(1 - F(u_{on})) \rightarrow 1$. For ease of notation we will use hereafter u_n instead of u_{on} to denote the threshold sequence satisfying these conditions. This sequence is immediately derived by direct calculations when F is known. Consider as an example the case $F(x) = 1 - e^{-x}$. By continuity of F we can choose $u_n(x) = F^{-1}(1 - \frac{\tau(x)}{n})$ with $\tau(x) > 0$, and hence $u_n(x) = -\log \tau(x) + \log n$. Equation (2) is written as

$$P\{M_n \leq -\log \tau(x) + \log n\} \rightarrow e^{-\tau(x)},$$

and then $P\{M_n - \log n \leq x\} \rightarrow e^{-e^{-x}}$, with $\tau(x) = e^{-x}$ for all $x \in \mathbb{R}$. The scaling constants are $a_n = 1, b_n = \log n$, and hence the threshold sequence is $u_n = \log n$, since $-\log G(0) = 1$. More examples can be found in Leadbetter, Lindgren, and Rootzén (1983).

In general, F is unknown, and in this setting neither the theoretical derivation nor the direct comparison of different threshold choices is possible. This comparison is undertaken by analyzing the properties of the tail index estimator of F , as most of these estimators for ξ are tied to a threshold choice. Therefore their biases and variances are influenced by the effect of the selection of u_n . There is a large amount of literature in tail index estimation [chapter VI of Embrechts, Klüppelberg, and Mikosch (1997) gives an excellent review]. Among these estimators, the most popular are Hill's estimator (1975) and Pickands's estimator (1975). The former is given by

$$\hat{\xi}_n^{(H)}(u_n) = \frac{1}{k} \sum_{i=n-k+1}^n \log \frac{x_{(i)}}{x_{(n-k)}}, \tag{15}$$

with $u_n = x_{(n-k)}, x_{(n-k+1)} \leq \dots \leq x_{(n)}$ denoting the increasing order statistics and k an integer value in $[1, n]$. Pickands's estimator for the tail index is

$$\hat{\xi}_n^{(P)}(u_n) = \frac{1}{\log(2)} \log \left(\frac{x_{(n-k+1)} - x_{(n-2k+1)}}{x_{(n-2k+1)} - x_{(n-4k+1)}} \right), \tag{16}$$

and

$$\hat{\sigma}_n^{(P)}(u_n) = \frac{x_{(n-2k+1)} - x_{(n-4k+1)}}{\int_0^{\log 2} e^{\hat{\xi}_n^{(P)}(x_{(n-4k+1)})t} dt}, \tag{17}$$

for the variance, with $u_n = x_{(n-4k+1)}$ and $k = 1, \dots, n/4$. There are some features of both estimators that are worth mentioning. These estimators are heavily dependent on the threshold choice u_n , and both of them can be derived under the assumption that F_{u_n} is exactly Pareto with parameter ξ or generalized Pareto with parameters ξ and $\sigma(u_n)$. Moreover, if $F_{u_n} = PD_{\xi}$, Hill's estimator is the maximum-likelihood estimator of ξ inheriting the corresponding asymptotic properties: consistency and normal distribution. This approach is only valid for regularly varying distribution functions, that is, $\xi > 0$, otherwise the asymptotic properties of this estimator vary according to F [see Davis and Resnick (1984)].

Pickands's estimator for the tail index is obtained assuming $F_{u_n} = GPD_{\xi, \sigma(u_n)}$ and taking the inverse of the parametric GPD. This estimator is consistent and also converges to a normal distribution; but it is very sensitive to the choice of u_n . Alternatively, under the latter parametric assumption on F_{u_n} we can obtain the maximum-likelihood estimator for the parameter ξ and $\sigma(u_n)$ of the GPD. In this case there is not a closed expression for the maximum-likelihood estimators of these parameters, and we have to rely on numerical procedures [see Press (1992)]. The maximum-likelihood estimator for the tail index is consistent and asymptotically normal for $\xi > -\frac{1}{2}$, as is discussed in Smith (1985).

The threshold selection is carried out by studying the mean-squared error of these ξ estimators, as u_n varies. However, some explicit form is required for the distribution function F . Under the assumption

$$1 - F(x) = Cx^{-\frac{1}{\xi}}[1 + Dx^{-\beta} + o(x^{-\beta})], \tag{18}$$

where $\xi > 0, C > 0, \beta > 0$, and D is a real number, Hall (1982) proposed estimators for the tail index based on an optimal choice of intermediate order statistics as candidates for the threshold sequence. Nevertheless, the pioneering work for threshold selection is Pickands (1975), where F satisfies the regularity conditions of Result 2, but not necessarily Equation (18). The estimation of the tail index and the threshold selection are done in a single step. Pickands proposed as a candidate for the threshold the order statistic of a sample $\{x_n\}$ that minimizes the distance d^∞ involving the distribution functions $F_{u_n, n}$ and $GPD_{\hat{\xi}_n^{(P)}(u_n), \hat{\sigma}_n^{(P)}(u_n)}$. The empirical conditional excess distribution function $F_{u_n, n}(x)$ with $x > u_n$ is defined by

$$F_{u_n, n}(x) = \frac{\sum_{i=1}^n 1_{\{u_n < x_i \leq x\}}}{\sum_{j=1}^n 1_{\{x_j > u_n\}}}, \tag{19}$$

or equivalently, via the transformation $y = a_n(x - u_n) > 0$, by

$$F_{u_n, n}(y) = \frac{\sum_{i=1}^n 1_{\{0 < y_i \leq y\}}}{\sum_{j=1}^n 1_{\{y_j > 0\}}}. \tag{20}$$

The distance d^∞ can be written as a function of a variable θ , once n is given, as

$$d^\infty\left(F_{\theta,n}, GPD_{\hat{\xi}_n^{(P)}(\theta), \hat{\sigma}_n^{(P)}(\theta)}\right) = \sup_{0 \leq y < \infty} |F_{\theta,n}(y) - GPD_{\hat{\xi}_n^{(P)}(\theta), \hat{\sigma}_n^{(P)}(\theta)}(y)|. \tag{21}$$

The optimal threshold is then

$$u_n^{(P)} = \arg \min_{\theta} d^\infty\left(F_{\theta,n}, GPD_{\hat{\xi}_n^{(P)}(\theta), \hat{\sigma}_n^{(P)}(\theta)}\right), \tag{22}$$

with θ taking values along the ordered sample $x_{(3n/4)} \leq \dots \leq x_{(n)}$. More specifically, $u_n^{(P)} = x_{(n-k)}$ with $k \rightarrow \infty$, $n \rightarrow \infty$, and $k = o(n)$ to benefit of an increase in the sample size.

Alternatively we propose a version of the distance d^∞ where the number of tail observations is weighted differently. This new approach accounts for the estimation pitfalls that derive from the lack of observations when θ gets close to x_F .

Definition 1 Let $F_{\theta,n}$ be the empirical version of F_θ and $GPD_{\hat{\xi}_n^{(MI)}(\theta), \hat{\sigma}_n^{(MI)}(\theta)}$ the distribution function of the largest observations with parameters estimated by maximum likelihood (ML). Define the weighted Pickands distance d^{WP} as

$$d^{WP}\left(F_{\theta,n}, GPD_{\hat{\xi}_n^{(MI)}(\theta), \hat{\sigma}_n^{(MI)}(\theta)}\right) = k^\varepsilon \sup_{0 \leq y < \infty} |F_{\theta,n}(y) - GPD_{\hat{\xi}_n^{(MI)}(\theta), \hat{\sigma}_n^{(MI)}(\theta)}(y)|, \tag{23}$$

with $0 \leq \varepsilon \leq \frac{1}{2}$ and $k = \sum_{j=1}^n 1_{\{x_j > \theta\}}$.

The parameter ε determines the weight assigned by the distance d^{WP} to the tail observations defined by the corresponding θ . Notice that this distance is the one used by Pickands when $\varepsilon = 0$, and the Kolmogorov-Smirnov (KS) statistic [Kolmogorov (1933)] when $\varepsilon = \frac{1}{2}$. The corresponding threshold choice is the order statistic that minimizes the distance,

$$u_n^{(WP)} = \arg \min_{\theta} d^{WP}\left(F_{\theta,n}, GPD_{\hat{\xi}_n^{(MI)}(\theta), \hat{\sigma}_n^{(MI)}(\theta)}\right), \tag{24}$$

with θ taking values along the ordered sample $x_{(1)} \leq \dots \leq x_{(n)}$. The parameter ε can be useful to study the effect of different weighting schemes in the threshold selection; however, this is far beyond the scope of this article, where we will only focus on the value $\varepsilon = \frac{1}{2}$ (KS statistic).

It is clear that threshold values far from x_F produce biased estimates of the tail index. On the other hand, u_n close to the right endpoint will result in inefficient estimates of ξ . Goldie and Smith (1987) and Smith (1987) derive the asymptotic distribution functions of both the maximum-likelihood and Hill estimators of the tail index for a class of distribution functions such that $1 - F(x) = x^{-\frac{1}{\xi}}L(x)$, where $L(x)$ are slowly varying functions of different types. They also discuss in detail asymptotic bias and variance for these estimators and find that departures of F from a Pareto distribution function lead to biased and inefficient estimates of the tail index for both estimators. As a result, a right choice of the threshold sequence turns out to be of critical importance in order to minimize the mean-squared error (MSE).

Hall (1982) derives an analytical expression for the MSE of Hill's estimator when F satisfies Equation (18). All these results are achieved for determined

classes of distribution functions. In contrast, under the regularity conditions of Result 2 it is not possible to derive analytically the MSE expression for the tail index estimator. Therefore we propose bootstrap confidence intervals in order to measure the bias and uncertainty of the different tail index estimators we considered.

The naïve nonparametric bootstrap is consistent since the empirical distribution function F_n is a consistent estimator of F and $\sqrt{k}(\hat{\xi}_n^{(i)}(u_n^{(l)}) - \xi)$, $i = H, MI, P$, and $l = P, WP, Ah$ (ad hoc) converges weakly to a normal distribution, with k being the number of exceedances over u_n . Then the bootstrap approximation $J_n(x, F_n)$ to the true sampling distribution function $J_n(x, F)$ of this statistic can be used to produce confidence regions, at the $1 - \alpha$ level, in the following way,

$$\xi \in \left[\hat{\xi}_n^{(i)}(u_n^{(l)}) - \frac{1}{\sqrt{k}} J_n^{-1} \left(1 - \frac{\alpha}{2}, F_n \right), \hat{\xi}_n^{(i)}(u_n^{(l)}) + \frac{1}{\sqrt{k}} J_n^{-1} \left(\frac{\alpha}{2}, F_n \right) \right], \tag{25}$$

where $J_n^{-1}(1 - \alpha, F_n)$ is the $1 - \alpha$ bootstrap quantile. To implement Equation (25), the bootstrap approximation is estimated by

$$\hat{J}_n(x, F_n) = \frac{1}{B} \sum_{j=1}^B \mathbf{1}_{\{\sqrt{k}(\hat{\xi}_{j,n}^{*(i)}(u_{j,n}^{*(l)}) - \hat{\xi}_n^{(i)}(u_n^{(l)})) \leq x\}}, \tag{26}$$

where B is the number of bootstrap iterations, $\hat{\xi}_{j,n}^{*(i)}(u_{j,n}^{*(l)})$ the corresponding estimator for the bootstrap sample j , and $u_{j,n}^{*(l)}$ the corresponding threshold choice.

The finite sample performance of the different estimators is analyzed in Table 1. The threshold u_n is chosen by both methods, Pickands and weighted Pickands with $\epsilon = \frac{1}{2}$. To emphasize the importance of the threshold selection to estimating the tail index, an ad hoc threshold ($u_n^{(Ah)} = x_{(\frac{95}{100}n)}$) is also included in the analysis.

The simulation experiment of Table 1 is done for different Student t -distributions, where the tail index ξ is well approximated by the inverse of the degrees of freedom [see chapter III of Embrechts, Klüppelberg, and Mikosch (1997)].

Before discussing the results of this Table 1 it is important to notice that although F is known, we replace it with F_n to calculate the bootstrap approximation $J_n(x, F_n)$. The reason for doing this is that the bootstrap procedure works

Table 1 Bootstrap confidence intervals I.

	$t_1 (\xi \sim 1)$	$t_5 (\xi \sim 0.2)$	$t_{10} (\xi \sim 0.1)$	$t_{30} (\xi \sim 0)$
$\hat{\xi}_n^{(MI)}(u_n^{(WP)})$	[0.70, 1.69]	[-0.17, 0.24]	[-0.28, 0.39]	[-0.43, 0.68]
$\hat{\xi}_n^{(P)}(u_n^{(P)})$	[0.29, 1.06]	[-0.39, 0.08]	[-0.63, -0.06]	[-0.64, -0.17]
$\hat{\xi}_n^{(MI)}(u_n^{(Ah)})$	[0.34, 1.75]	[0.19, 0.91]	[-0.26, 0.33]	[-0.28, 0.57]

Bootstrap confidence intervals at a significance level $\alpha = 0.05$ for different estimators of the tail index: $\hat{\xi}_n^{(MI)}(u_n)$ with u_n estimated by d^{WP} and by $u_n^{(Ah)} = x_{(\frac{95}{100}n)}$; and $\hat{\xi}_n^{(P)}(u_n^{(P)})$ with u_n estimated by d^∞ . $B = 1000$ bootstrap samples of size $n = 1000$ are drawn from a single sequence generated from t_ν , with $\nu = 1, 5, 10$ and 30 .

Table 2 Bootstrap confidence intervals II.

	$t_1 (\xi \sim 1)$	$t_5 (\xi \sim 0.2)$	$t_{10} (\xi \sim 0.1)$	$t_{30} (\xi \sim 0)$
$\hat{\xi}_n^{(MI)}(u_n^{(WP)})$	[0.70, 1.69]	[-0.17, 0.24]	[-0.28, 0.39]	[-0.43, 0.68]
$\hat{\xi}_n^{(H)}(u_n^{(WP)})$	[0.82, 1.23]	[0.08, 0.37]	[-0.42, 0.23]	[0.04, 0.20]

Bootstrap confidence intervals at a significance level $\alpha = 0.05$ for different estimators of the tail index when $u_n^{(WP)}$ is obtained from $GPD_{\xi, \sigma(u_n)}$ and from PD_{ξ} , respectively. Note $\hat{\xi}_n^{(MI)}(u_n^{(WP)})$ is $\hat{\xi}_n^{(H)}(u_n^{(WP)})$ for the PD_{ξ} case. $B = 1000$ bootstrap samples of size $n = 1000$ are drawn from a single sequence generated from t_{ν} , with $\nu = 1, 5, 10,$ and 30 .

even when F is unknown and we only have a realization from the random sequence $\{X_n\}$.

There are two clear results from Table 1: First, the confidence intervals for our estimator contain the true tail index, something that does not occur for Pickands’s method; and second, the confidence intervals estimated from the ad hoc threshold are wider than the ones derived from our method when ξ is significantly greater than zero.

Table 2 analyzes in more detail the advantages of the weighted Pickands method for selecting u_n when the data come from heavy-tailed distributions. In this case the $GPD_{\xi, \sigma(u_n)}$ is replaced by the PD_{ξ} in Definition 1 and Equation (24).

From Table 2 we conclude that when we are dealing with heavy-tailed distributions ($\xi > 0$), our method is more efficient with PD than with GPD . These simulation results are in line with the theoretical findings derived in Smith (1987).

3 HYPOTHESIS TESTING

Different threshold choices define different sets of possible extreme values of a particular sequence $\{X_n\}$. In this article the observations exceeding a certain threshold are considered extreme values only if they are distributed as a $GPD_{\xi, \sigma(u_n)}$, with ξ the tail index of F . In order to check this condition we propose a goodness-of-fit test for the following hypothesis:

$$H_{n,0}: \text{the sample } \{(x_1 - u_n)_+, \dots, (x_n - u_n)_+\} \text{ is distributed as } GPD_{\xi, \sigma(u_n)}$$

versus a general alternative of the form

$$H_{n,1}: \text{the sample } \{(x_1 - u_n)_+, \dots, (x_n - u_n)_+\} \text{ is not distributed as } GPD_{\xi, \sigma(u_n)}$$

with $u_n \in \mathbb{R}$, ξ the tail index of F and $(x)_+ = \max(x, 0)$.

A natural goodness-of-fit test statistic is the KS statistic [for other goodness-of-fit criteria see Anderson and Darling (1952)],

$$R_k(y; \xi, \sigma(u_n)) = \sqrt{k} \sup_{0 \leq y < \infty} |P_k(y) - GPD_{\xi, \sigma(u_n)}(y)|, \tag{27}$$

with $k = \sum_{j=1}^n 1_{\{x_j > u_n\}}$ and P_k the empirical distribution function of the observations exceeding u_n . When the parameters are known, the asymptotic distribution

of this test statistic is tabulated and the critical values can be derived. If the parameters are unknown, but consistently estimated, the bootstrap distribution function is a reliable approximation of the true sampling distribution of $R_\lambda(y; \xi, \sigma(u_n))$. In this case it can be proved [see Romano (1988)] that the bootstrap critical values are consistent estimates of the actual ones.

Our interest, however, does not lie in the definition of the extreme values of a particular sequence $\{X_n\}$, but in the definition of the extreme values of any sequence of length n with distribution function F . In this case a different hypothesis test is needed to determine whether the selected threshold is a good candidate to define the extremes of F given the sample size n . More formally, the testing problem under consideration is

$$H_0: F_{u_n} = GPD_{\xi, \sigma(u_n)}$$

versus a general alternative

$$H_1: F_{u_n} \neq GPD_{\xi, \sigma(u_n)},$$

with ξ being the tail index of F .

Now we can formally define the set of extreme values of any sequence with distribution function F .

Definition 2 Let $\{X_n\}$ be any sequence of a distribution function F . The extreme values of any sequence of length n from this distribution are given by the observations exceeding the threshold u_n , and satisfying $F_{u_n} = GPD_{\xi, \sigma(u_n)}$.

The test statistic in this case is a version of the family of KS test statistics,

$$T_n(y_n; \xi, \sigma(u_n)) = \sqrt{n} \sup_{0 \leq y < \infty} |F_{u_n, n}(y) - GPD_{\xi, \sigma(u_n)}(y)|, \tag{28}$$

with $y_i = (x_i - u_n)_+, i = 1, \dots, n$. This statistic depends on u_n, ξ , and $\sigma(u_n)$.

In order to derive the asymptotic distribution of Equation (28) and to assess the bootstrap approximation, the following results are required. Let

$$U_\lambda(t) = \frac{P\{\lambda < T \leq t\}}{P\{T > \lambda\}} \tag{29}$$

be the conditional excess distribution function, with parameter λ on $[0, 1]$, of a uniform $[0, 1]$ random variable T . Its empirical counterpart

$$U_{\lambda, n}(t) = \frac{1}{n} \sum_{i=1}^n \frac{1_{\{\lambda < t_i \leq t\}}}{\frac{1}{n} \sum_{j=1}^n 1_{\{t_j > \lambda\}}}, \tag{30}$$

with t_1, \dots, t_n and $t \in [0, 1]$, defines an empirical process $B_n(t) = \sqrt{n}(U_{\lambda, n}(t) - U_\lambda(t))$ similar to the uniform empirical process $\sqrt{n}(U_n(t) - U(t))$. It is well known that the latter converges weakly to the distribution of a mean-zero gaussian process $Z_U(\cdot)$ [see chapter V of Pollard (1984)]. By an analogue reasoning, it is immediate to derive the probability law of the process $S_n(y) = \sqrt{n}(F_{u_n, n}(y) - F_{u_n}(y))$, where the threshold u_n plays the role of the parameter λ .

Theorem 1 Consider a continuous and strictly increasing distribution function F and a threshold u_n , with $u_n < x_F$. The empirical process $S_n(y)$ converges weakly to the distribution of a mean-zero gaussian process $Z_{F_{u_n}}(\cdot)$ with covariance function

$$\begin{aligned} & \text{cov}(Z_{F_{u_n}}(y_1), Z_{F_{u_n}}(y_2)) \\ &= \frac{(F(\min(y_1, y_2)) - F(u_n)) - (F(y_1) - F(u_n))(F(y_2) - F(u_n))}{(1 - F(u_n))^2}, \end{aligned} \tag{31}$$

with $y_1, y_2 \in \mathbb{R}$. Moreover, under the null hypothesis H_0 , this empirical process takes the form $\sqrt{n}(F_{u_n, n}(y) - \text{GPD}_{\xi, \sigma(u_n)}(y))$ and the covariance function becomes

$$\begin{aligned} \text{cov}(Z_{F_{u_n}}(y_1), Z_{F_{u_n}}(y_2)) &= \frac{\text{GPD}_{\xi, \sigma(u)}(\min(y_1, y_2))}{1 - F(u_n)} \\ &\quad - \text{GPD}_{\xi, \sigma(u_n)}(y_1)\text{GPD}_{\xi, \sigma(u_n)}(y_2). \end{aligned} \tag{32}$$

By the continuous mapping theorem, the limiting distribution function, denoted by $L(x, F)$, of the test statistic T_n is the distribution of the supremum of a mean-zero gaussian process with the covariance function of Equation (32). The proof is in the appendix.

In order to test H_0 , we should be using the following rejection criteria:

$$\{T_n(y_n; \xi, \sigma(u_n)) > L_n^{-1}(1 - \alpha, F)\}, \tag{33}$$

where $L_n^{-1}(1 - \alpha, F)$ is the $1 - \alpha$ quantile of the exact finite sample distribution $L_n(x, F)$ of the statistic T_n . This distribution L_n is clearly unknown, and in practice has to be approximated by the asymptotic distribution $L(x, F)$. This limiting distribution takes a complicated form and depends on the knowledge of F , on the parameters of the GPD, as well as on the threshold u_n . The nuisance parameters dependency forces us to look for an alternative method to approximate the distribution $L_n(x, F)$.

3.1 Bootstrap Approximation

Let $L_n(x, Q_n)$ be the bootstrap distribution that approximates $L_n(x, F)$, and $L_n^{-1}(1 - \alpha, Q_n)$ the bootstrap quantile that approximates the corresponding finite sample distribution quantile $L_n^{-1}(1 - \alpha, F)$. In order for the bootstrap to be consistent, Q_n has to satisfy certain conditions.

Lemma 1 Let Q_n be an estimator of F based on $\{x_1, \dots, x_n\}$ that satisfies $\sup_{x \in \mathbb{R}} |Q_n(x) - F(x)| \xrightarrow{P} 0$ whenever $F \in H_0$, and let $L(x, F)$, the limiting distribution of the test statistic T_n , be continuous and strictly increasing. Then

$$P\{T_n > L_n^{-1}(1 - \alpha, Q_n)\} \rightarrow \alpha, \text{ as } n \rightarrow \infty. \tag{34}$$

The naïve nonparametric bootstrap from $Q_n = F_n$ fails to produce consistent estimates of a distribution function under H_0 if F does not belong to the null. On the other hand, the parametric bootstrap from the $\text{GPD}_{\xi, \sigma(u_n)}$ [see Equation (27)] fails to capture the structure of F for the observations smaller than the threshold u_n .

To fulfill the conditions of Lemma 1 corresponding to Q_n and therefore to solve the two previously mentioned problems, a semiparametric bootstrap methodology is introduced. Define

$$Q_n(x) = \begin{cases} F_n(x) & x \leq u_n \\ GPD_{\xi, \sigma(u_n)}(x - u_n) + F_n(u_n)(1 - GPD_{\xi, \sigma(u_n)}(x - u_n)) & x > u_n. \end{cases} \tag{35}$$

This distribution function is derived from the conditional probability theorem, since

$$P\{X \leq x\} = P\{X \leq u_n\}P\{X \leq x | X \leq u_n\} + P\{X > u_n\}P\{X \leq x | X > u_n\}, \tag{36}$$

where $P\{X \leq u_n\}$ is consistently approximated by $F_n(u_n)$, and under the null $P\{X \leq x | X > u_n\} = GPD_{\xi, \sigma(u_n)}(y)$ with $y = x - u_n$.

Denote $\{x_n^*\}$ a bootstrap sample obtained from Q_n and consider the transformed bootstrap sample $y_i^* = x_i^* - u_n$ with $i = 1, \dots, n$. The value of the test statistic is $t_n(y_1^*, \dots, y_n^*; \xi, \sigma(u_n))$ and for the sake of notation is denoted as $t_n^*(y_n; \xi, \sigma(u_n))$. The bootstrap approximation $L_n(x, Q_n)$ is then estimated by the empirical distribution of the B (number of bootstrap samples) values of T_n ,

$$\hat{L}_n(x, Q_n) = \frac{1}{B} \sum_{j=1}^B \mathbf{1}_{\{t_{n,j}^*(y_n; \xi, \sigma(u_n)) \leq x\}}. \tag{37}$$

The $1 - \alpha$ quantile of $\hat{L}_n(x, Q_n)$ is the order statistic $t_{n, \lceil (1-\alpha)B \rceil}^*(y_n; \xi, \sigma(u_n))$ of the sequence $\{t_{n,j}^*(y_n; \xi, \sigma(u_n))\}$ of B elements, where $\lceil x \rceil$ is the upper integer part of x . The rejection criteria Equation (33) is replaced now by

$$\{T_n(y_n; \xi, \sigma(u_n)) > t_{n, \lceil (1-\alpha)B \rceil}^*(y_n; \xi, \sigma(u_n))\}, \tag{38}$$

and hence for a sample $\{x_n\}$, the null hypothesis is rejected if $t_n(y_1, \dots, y_n; \xi, \sigma(u_n))$ is in this rejection region. This means that the conditional excess distribution function defined by u_n is not a $GPD_{\xi, \sigma(u_n)}$, and according to our definition these candidates for extreme observations are not really extreme.

Recall that until now we have assumed the parameters to be known. Nevertheless this condition is rarely satisfied in practice. To make our test operational, we replace these parameters with their maximum-likelihood estimators, and instead of Q_n , we define its counterpart distribution function \hat{Q}_n :

$$\hat{Q}_n(x) = \begin{cases} F_n(x) & x \leq u_n \\ GPD_{\hat{\xi}_n^{(M)}(u_n), \hat{\sigma}_n^{(M)}(u_n)}(x - u_n) + F_n(u_n)(1 - GPD_{\hat{\xi}_n^{(M)}(u_n), \hat{\sigma}_n^{(M)}(u_n)}(x - u_n)) & x > u_n. \end{cases} \tag{39}$$

Notice that the new bootstrap distribution function $L_n(x, \hat{Q}_n)$ boils down to $L_n(x, Q_n)$ for $x \leq u_n$, and for $x > u_n$, the former \sqrt{k} converges to the latter, where k is the number of observations of the tail defined by u_n . Moreover, if F belongs to the

null hypothesis defined by u_n , the conditions in Lemma 1 still hold and the rejection region of Equation (38) becomes

$$\{\hat{T}_n(y_n; \hat{\xi}_n^{(ML)}(u_n), \hat{\sigma}_n^{(ML)}(u_n)) > t_{n, ((1-\alpha)B)}^* (y_n; \hat{\xi}_n^{*,(ML)}(u_n), \hat{\sigma}_n^{*,(ML)}(u_n))\}, \tag{40}$$

where \hat{T}_n and $\hat{\xi}_n^{(ML)}(u_n), \hat{\sigma}_n^{(ML)}(u_n)$ are calculated from the original sample $\{x_n\}$, and $\hat{\xi}_n^{*,(ML)}(u_n), \hat{\sigma}_n^{*,(ML)}(u_n)$ are estimated from the corresponding bootstrap sequences.

3.2 Finite Sample Performance: Empirical Power

The power of our test,

$$P\{\hat{T}_n > L_n^{-1}(1 - \alpha, \hat{Q}_n)\}, \tag{41}$$

depends on three key parameters: the threshold choice, the distribution function F , and the length of the sequence. To calculate this power it is important to realize that the maximum-likelihood estimates $\hat{\xi}_n^{(ML)}(u_n), \hat{\sigma}_n^{(ML)}(u_n)$ that are entered in the expression of \hat{T}_n are the ones used to define the null distribution \hat{Q}_n .

This test lies in constructing a distribution function \hat{Q}_n , such that its conditional excess distribution is a $GPD_{\hat{\xi}_n^{(ML)}(u_n), \hat{\sigma}_n^{(ML)}(u_n)}$. In that way the observations coming from the null hypothesis are drawn from \hat{Q}_n and not from F . The empirical size of the test is calculated from the former distribution. For a deeper insight into how to calculate the power via bootstrap [see Beran (1986) and Romano (1988)].

The following algorithms are devoted to describing the simulation experiment. Algorithm 1 generates bootstrap samples $\{x_n^*\}$ from the distribution function \hat{Q}_n and calculates the empirical bootstrap approximation of $L_n(x, F)$. The threshold value u_n and the maximum-likelihood estimates are obtained from a particular sample $\{x_n\}$ from F and are used to construct \hat{Q}_n .

Algorithm 1 (Bootstrap Procedure)

1. $l = 1$.
2. Generate $x_{1,l}^*, \dots, x_{n,l}^*$ drawn from \hat{Q}_n .
3. Calculate $\hat{\xi}_n^{*,(ML)}(u_n)$ and $\hat{\sigma}_n^{*,(ML)}(u_n)$ from the bootstrap sample.
4. $t_{n,l}^*(y_n; \hat{\xi}_n^{*,(ML)}(u_n), \hat{\sigma}_n^{*,(ML)}(u_n)) = \sqrt{n} \sup_{0 \leq y < \infty} |F_{u_n, n}(y) - GPD_{\hat{\xi}_n^{*,(ML)}(u_n), \hat{\sigma}_n^{*,(ML)}(u_n)}(y)|$ with $y = x - u_n$.
5. $l++$. Go to step 2 while $l \leq B$.
6. $\hat{L}_n(x, \hat{Q}_n) = \frac{1}{B} \sum_{j=1}^B 1_{\{t_{n,j}^*(y; \hat{\xi}_n^{*,(ML)}(u_n), \hat{\sigma}_n^{*,(ML)}(u_n)) \leq x\}}$

In practice, the p -value replaces the rejection criteria given in Equation (40). The empirical p -value is

$$p = \frac{1}{B} \sum_{j=1}^B 1_{\{t_{n,j}^* > \hat{t}_n\}}, \tag{42}$$

with \hat{t}_n obtained from the sample $\{x_n\}$.

The probability of Equation (41) cannot be directly derived, and we have to rely on Monte Carlo simulations to calculate it. The following algorithm describes how to implement this procedure.

Algorithm 2 (Empirical Power)

1. $j = 1$.
2. Let $\{x_{1,j}, \dots, x_{n,j}\}$ be a sample from F and obtain $u_n, \hat{\xi}_n^{(ML)}(u_n)$ and $\hat{\sigma}_n^{(ML)}(u_n)$.
3. Construct \hat{Q}_n and $\hat{L}_n(x, \hat{Q}_n)$ as in Algorithm 1.
4. Generate $\{x'_1, \dots, x'_n\}$ from a distribution function F_1 .
5. Calculate $\hat{t}_n(x'_n; \hat{\xi}_n^{(ML)}(u_n), \hat{\sigma}_n^{(ML)}(u_n))$ if $F_1 \neq F$. Otherwise $\hat{t}_n(x'_n; \hat{\xi}_n^{*(ML)}(u_n), \hat{\sigma}_n^{*(ML)}(u_n))$ with $\hat{\xi}_n^{*(ML)}(u_n), \hat{\sigma}_n^{*(ML)}(u_n)$ from $\{x'_n\}$.
6. Calculate the p -value as in Equation (42).
7. $\delta_j = \begin{cases} 1 & \text{if } p < \alpha \\ 0 & \text{otherwise.} \end{cases}$
8. $j++$. Repeat while $j \leq m$.
9. $\hat{\alpha} = \frac{1}{m} \sum_{j=1}^m \delta_j$.

As $n \rightarrow \infty$, the estimate $\hat{\alpha}$ approaches the size of the test if the threshold u_n is really defining the extremes of F for a given length n . On the other hand, when the conditional distribution function defined by the threshold is not a $GPD_{\xi, \sigma(u_n)}$, or the sequence of data does not come from F , the estimate $\hat{\alpha}$ tends to one.

Table 3 gives the simulation results of the empirical power for a family of Student t -distribution functions with the threshold u_n obtained by our weighted Pickands method.

Table 3 points out two clear results. First, the fact that the diagonal is very close to the nominal size reveals that our procedure performs very well in capturing the extremes of sequences of length n coming from F_0 (distribution function under H_0). Second, extreme value candidates coming from F_1 (distribution function under H_1) are rejected as extreme values of F_0 . A by-product of this table is that our test can be considered a goodness-of-fit test via the tails. In principle our test is more sensitive than standard KS statistics in detecting deviations in the tails [see Mason and Schuenemeyer (1983)].

Another alternative to selecting the threshold is to choose a fixed order statistic. In this case, the set of extreme values is defined by a fixed number of observations given the sample size n .

Table 3 Empirical power for different distributions.

F_0	F_1			
	$t_{30} (\xi \sim 0)$	$t_{10} (\xi \sim 0.1)$	$t_5 (\xi \sim 0.2)$	$t_1 (\xi \sim 1)$
t_{30}	0.08	0.62	0.93	0.97
t_{10}	0.65	0.09	0.76	0.98
t_5	0.95	0.72	0.09	0.99
t_1	0.94	0.94	0.92	0.09

Empirical power of T_n for a family of Student t -distribution functions, with u_n from d^{WP} . F_0 denotes the data-generating process and F_1 the distribution under the alternative hypothesis. Bootstrap replications $B = 1000$, Monte Carlo simulations $m = 1000$, $n = 1000$, significance level $\alpha = 0.05$.

Table 4 Empirical power for ad hoc thresholds.

F_0	$x_{(700)}$	$x_{(800)}$	$x_{(900)}$	$x_{(950)}$
t_{30}	0.49	0.48	0.45	0.44
t_{10}	0.48	0.48	0.46	0.46
t_5	0.54	0.50	0.49	0.47
t_1	0.64	0.58	0.52	0.49

Empirical power for a family of Student t -distribution functions, with different ad hoc threshold choices for a sample size $n = 1000$. F_0 denotes the data-generating process. Bootstrap replications $B = 1000$, Monte Carlo simulations $m = 1000$, significance level $\alpha = 0.05$.

The message from Table 4 is clear: These ad hoc selections of the set of extreme values can be valid for particular sequences of F_0 , but in general are rejected to define the extremes of any sequence of F_0 with the same length n .

4 EMPIRICAL APPLICATION: VaR ESTIMATION IN FINANCIAL INDEXES

An important application of the semiparametric approximation \hat{Q}_n of F is quantile estimation in the tail region, where there is usually a lack of observations because we are dealing with extremal events. This question is becoming of primary importance in a wide variety of research fields, including finance, climatology, and hydrology.

The goal of this section is to obtain a deeper insight into risk management for financial indexes of different major markets. Market risk management is inherently related to the probability of occurrence of extreme events, that is, very large negative or positive returns. We focus on a particular measure of this market risk: value at risk (VaR), the amount of money necessary to provide the institution with coverage against losses that can occur with a p probability over some holding period. It is not our intention to get into details of the VaR methodology; we only pursue it to present some results about tail index estimation (tail behavior) and a naïve calculus of VaR under i.i.d. assumptions for financial data. Of course, we know this assumption is unrealistic and we should go a step further regarding heteroskedastic conditional volatility models, but this is left for future research.

General practitioners calculate VaRs in two different ways: (i) complete parametric, where it is assumed an underlying distribution (normal, student's t , etc.), and (ii) fully nonparametric, where the main actor is the empirical distribution F_n . Our approach can be considered as something in the middle, because we use a semiparametric approximation \hat{Q}_n .

The inverse of \hat{Q}_n provides a consistent estimator of VaR for the distribution function F . In this case,

$$\widehat{VaR}_p = \begin{cases} \inf\{x | F_n(x) \geq 1 - p\}, & 1 - p \leq F_n(u_n) \\ u_n + \frac{\hat{\sigma}_n^{(MI)}(u_n)}{\hat{\xi}_n^{(MI)}(u_n)} \left(\left(\frac{p}{1 - F_n(u_n)} \right)^{-\hat{\xi}_n^{(MI)}(u_n)} - 1 \right), & 1 - p > F_n(u_n) \end{cases} \quad (43)$$

When the distribution function is regularly varying ($\xi > 0$), the tail of \hat{Q}_n is modeled as a Pareto distribution and the inverse of F is consistently estimated by

$$\widehat{VaR}_p = \begin{cases} \inf \{x | F_n(x) \geq 1 - p\}, & 1 - p \leq F_n(u_n) \\ u_n \left(\frac{1 - F_n(u_n)}{p} \right)^{\hat{\xi}_n^{(M)}(u_n)}, & 1 - p > F_n(u_n) \end{cases} \quad (44)$$

The uncertainty of these estimates can be measured by bootstrap confidence intervals, since the exact finite sample distribution function of $V_n = \sqrt{n}(\widehat{VaR}_p - VaR_p)$ is not known and its asymptotic distribution depends on nuisance parameters. Let $J_n(x, \hat{Q}_n)$ be the bootstrap approximation of the exact distribution of V_n . A two-sided, equal-tailed confidence interval for VaR_p , at a significance level α , is therefore given by

$$CI_\alpha(VaR_p) = \left[\widehat{VaR}_p - \frac{1}{\sqrt{n}} J_n^{-1} \left(1 - \frac{\alpha}{2}, \hat{Q}_n \right), \widehat{VaR}_p - \frac{1}{\sqrt{n}} J_n^{-1} \left(\frac{\alpha}{2}, \hat{Q}_n \right) \right], \quad (45)$$

where $J_n^{-1}(1 - \alpha, \hat{Q}_n)$ is the $1 - \alpha$ bootstrap quantile.

4.1 Data Features

The data we use to illustrate how the methodology proposed in this article can be applied consist of five financial indexes of major stock markets over the period December 19, to April 20, Frankfurt (Dax), London (FTSE-100), Madrid (Ibex), Tokyo (Nikkei), and New York (Dow Jones). These data have been collected from <http://www.freelunch.com>. The observations considered for the analysis are the logarithmic returns measured in percentage terms and denoted as r_t :

$$r_t = 100 (\log P_t - \log P_{t-1}),$$

where P_t is the original price at time t . For calculating ease, the negative observations (losses) are depicted in the positive tail.

A first glance to the standard statistic for kurtosis shows that most of these series are leptokurtic. For instance, the Dax index has a coefficient of corrected kurtosis of 5.70; FTSE, 1.34; Ibex, 3.88; Nikkei, 2.77, and the Dow Jones has a coefficient of 3.25. Traditionally this measure has been considered an indicator of heavy tails. Nevertheless, the coefficient of kurtosis does not provide us with adequate information about the source of the heaviness. The tail index, however, provides this kind of information, focusing on a particular tail. For instance, $\xi > 0$ corresponds to distributions where that tail has a polynomial decay [a more detailed discussion can be found in Shiryaev (2001)].

Table 5 presents nonparametric bootstrap confidence intervals for the tail index [see Equation (25)] obtained by the different approaches investigated throughout the article.

From Table 5, it appears that the tail index ξ is greater than zero, indicating the existence of heavy right-hand side tails (corresponding to losses). The only

Table 5 Bootstrap confidence intervals for tail index.

	$\hat{\xi}_n^{(MI)}(u_n^{(WP)})$	$\hat{\xi}_n^{(H)}(u_n^{(WP)})$	$\hat{\xi}_n^{(P)}(u_n^{(P)})$	$\hat{\xi}_n^{(MI)}(x_{(\frac{0.05}{100}n)})$
Dax	[-0.02; 0.24; 0.84]	[0.30; 0.31; 0.36]	[-0.50; -0.37; -0.20]	[-0.13; 0.22; 0.65]
Ftse	[-0.57; -0.26; 0.04]	[0.07; 0.11; 0.12]	[-0.44; -0.28; -0.08]	[-0.54; -0.29; 0.13]
Ibex	[-0.12; 0.28; 0.87]	[0.32; 0.37; 0.38]	[-0.43; -0.21; -0.04]	[-0.04; 0.46; 0.90]
Nikkei	[-0.13; 0.11; 0.55]	[0.33; 0.34; 0.39]	[-0.34; -0.19; -0.03]	[-0.25; 0.07; 0.50]
Dow Jones	[-0.11; 0.63; 1.52]	[0.33; 0.41; 0.44]	[-0.24; -0.22; -0.03]	[0.05; 0.76; 1.72]

Bootstrap confidence intervals ($\alpha = 0.05$) and pointwise estimation of the tail index ξ for stock returns over the period December 19, 1994, to April 20, 2001. Bootstrap samples $B = 1000$.

Table 6 Bootstrap confidence intervals for VaR.

VaR	GPD	PD	F_n	Gaussian
Dax	[3.57; 4.16; 7.83]	[3.48; 4.25; 4.93]	[2.96; 4.33; 5.04]	[3.52; 3.62; 3.71]
Ftse	[2.81; 3.04; 3.40]	[2.83; 3.05; 3.31]	[2.83; 3.08; 3.32]	[2.65; 2.78; 2.85]
Ibex	[3.25; 3.92; 4.69]	[2.94; 3.91; 4.62]	[3.02; 4.50; 5.80]	[3.08; 3.19; 3.32]
Nikkei	[3.69; 4.24; 8.30]	[3.33; 4.31; 5.00]	[4.09; 4.73; 5.95]	[3.75; 3.79; 3.83]
Dow Jones	[1.47; 2.09; 2.60]	[1.56; 2.09; 2.49]	[1.36; 1.90; 2.15]	[1.55; 1.73; 1.97]

Confidence intervals ($\alpha = 0.05$) and pointwise estimation of the VaR for the different financial returns calculated with different methodologies: our GPD and PD approaches, nonparametric approach F_n , and a parametric approach based on a gaussian assumption. The VaR indicates the percentage of return losses with $p = 0.01$ and a holding period of 1 day. The data covers the period December 19, 1994 to April 20, 2001. Bootstrap samples $B = 1000$.

exception is the Ftse index, where there are some reasonable doubts. For that reason, in the next table the VaR is calculated under both the GPD and PD methodologies.

In Table 6 we provide pointwise estimates and confidence intervals for VaR under four different approaches. The first two correspond to the methods developed in this article, and the last two correspond to the standard empirical methodologies that will be used here as a benchmark.

From Table 6, three conclusions can be obtained: (i) Comparing our two approaches and taking into account the results of the previous table, the PD method outperforms the GPD from an efficiency point of view, given that the point estimates are very similar. This is the expected result under the presence of heavy tails. (ii) The approach based on the empirical distribution is less efficient compared to the PD method. The main reason is the lack of observations coming from the tail, something that our PD method overcomes by properly parameterizing the tail. (iii) The approach based on gaussianity, as expected, is very conservative in the sense of requiring a lesser amount of capital (smaller VaR).

5 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 to 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 extreme?” This is the main goal of this article, where, following Pickands (1975) methodology, we not only formally define the set of extreme observations of a particular sequence, but also, by means of a hypothesis test, we define the extreme values of any sequence of the same length and with the same distribution function. Identification of the extreme observations allows us to estimate risk measures such as VaR very accurately, as well as to make inferences on different tail parameters of interest. Extensions to dependent data and to multivariate extremes constitute current research by the authors.

APPENDIX

Proof of Theorem 1 Let $\{U_n\}$ be a sequence of independent and identically distributed (i.i.d.) uniform random variables on $[0, 1]$ and let λ be a parameter in $0 < \lambda < 1$. Define the empirical process $B_n(t) = \sqrt{n}(U_{\lambda,n}(t) - U_\lambda(t))$ with $U_{\lambda,n}(t) = \frac{1}{n} \sum_{i=1}^n \left(\frac{1_{\{\lambda < t_i \leq t\}}}{\frac{1}{n} \sum_{j=1}^n 1_{\{t_j > \lambda\}}} \right)$. This process has a binomial distribution $Bin(n, U_\lambda(t))$. By the Donsker theorem or empirical central limit theorem, $B_n(t)$ converges weakly to $N(0, U_\lambda(t)(1 - U_\lambda(t)))$, therefore the finite dimensional distributions are normal for any fixed $t \in [0, 1]$. In addition, the process is tight due to the uniform continuity of the distribution function U and of $U_\lambda(t)$. This implies that $B_n(t)$ converges weakly to a mean-zero gaussian process $Z_{U_\lambda}(t)$. It only remains to find the asymptotic covariance function,

$$\text{cov}(B_n(s), B_n(t)) = \text{cov}[\sqrt{n}(U_{\lambda,n}(s) - U_\lambda(s)), \sqrt{n}(U_{\lambda,n}(t) - U_\lambda(t))],$$

with $0 < s, t < 1$. As $U_\lambda(t)$ is constant given $t \in (0, 1)$, the covariance function boils down to

$$\text{cov}(B_n(s), B_n(t)) = \frac{n}{(1 - U_n(\lambda))^2} \text{cov} \left(\frac{1}{n} \sum_{i=1}^n 1_{\{\lambda < t_i \leq s\}}, \frac{1}{n} \sum_{i=1}^n 1_{\{\lambda < t_i \leq t\}} \right).$$

The observations $\{t_1, \dots, t_n\}$ are i.i.d., and therefore $\text{cov}(1_{\{\lambda < t_i \leq s\}}, 1_{\{\lambda < t_j \leq t\}}) = 0$ with $i \neq j$. The covariance function is in this case

$$\text{cov}(B_n(s), B_n(t)) = \frac{1}{(1 - U_n(\lambda))^2} \text{cov}(1_{\{\lambda < t_i \leq s\}}, 1_{\{\lambda < t_i \leq t\}}) \tag{46}$$

$$\begin{aligned} &= \frac{1}{(1 - U_n(\lambda))^2} [E(1_{\{\lambda < t_i \leq \min(s,t)\}}) - E(1_{\{\lambda < t_i \leq s\}})E(1_{\{\lambda < t_i \leq t\}})] \\ &= \frac{(U(\min(s,t)) - U(\lambda)) - (U(s) - U(\lambda))(U(t) - U(\lambda))}{(1 - U_n(\lambda))^2}, \tag{47} \end{aligned}$$

with $0 < s, t < 1$. Therefore $B_n(t)$ converges weakly to the distribution of a mean-zero gaussian process $Z_{U_\lambda}(t)$ with covariance function given by

$$\text{cov}(Z_{U_\lambda}(s), Z_{U_\lambda}(t)) = \frac{(\min(s, t) - \lambda) - (s - \lambda)(t - \lambda)}{(1 - \lambda)^2}. \tag{48}$$

For F continuous and strictly increasing, we can define $u_n = F^{-1}(\lambda)$. Construct x_1, \dots, x_n i.i.d. from F via $x_i = F^{-1}(t_i)$ and let $F_n(x)$ denote the empirical distribution function based on x_1, \dots, x_n . By the monotonicity of F , $\sum_{i=1}^n 1_{\{u_n < x_i \leq x\}} = \sum_{i=1}^n 1_{\{F(u_n) < F(x_i) \leq F(x)\}}$ and therefore $F_{u_n, n}(x)$ defined in Equation (19) satisfies $F_{u_n, n}(x) = U_{\lambda, n}(t)$ with $x = F^{-1}(t)$. Then the process $B_n(t)$ becomes equal to the process $\sqrt{n}(F_{u_n, n}(y) - F_{u_n}(y))$ with $y = x - u_n$ [see Equations (19) and (20)] and the covariance function is

$$\text{cov}(Z_{F_{u_n}}(y_1), Z_{F_{u_n}}(y_2)) = \frac{(F(\min(y_1, y_2)) - F(u_n)) - (F(y_1) - F(u_n))(F(y_2) - F(u_n))}{(1 - F(u_n))^2}, \tag{49}$$

with $y_1 = F^{-1}(s)$ and $y_2 = F^{-1}(t)$.

Under the null hypothesis $F_{u_n} = \text{GPD}_{\xi, \sigma(u_n)}$, the empirical process $S_n(y)$ amounts to $\sqrt{n}(F_{u_n, n}(y) - \text{GPD}_{\xi, \sigma(u_n)}(y))$ and the covariance function of the limiting process is

$$\text{cov}(Z_{F_{u_n}}(y_1), Z_{F_{u_n}}(y_2)) = \frac{\text{GPD}_{\xi, \sigma}(\min(y_1, y_2))}{1 - F(u_n)} - \text{GPD}_{\xi, \sigma}(y_1)\text{GPD}_{\xi, \sigma}(y_2). \tag{50}$$

Proof of Lemma 1 Let $0 < \alpha < 1$ be the significance level of the test and consider $L(x, F)$ continuous and strictly increasing. By definition

$$P\{T_n > L^{-1}(1 - \alpha, F)\} = \alpha,$$

with $L^{-1}(1 - \alpha, F)$ the $1 - \alpha$ asymptotic quantile. ■

Consider $L_n(x, Q_n)$ the bootstrap approximation of $L_n(x; F)$ and $L_n^{-1}(1 - \alpha, Q_n)$ its $1 - \alpha$ quantile. Therefore if $\sup_{x \in \mathbb{R}} |Q_n(x) - F(x)| \xrightarrow{p} 0$, then $L_n^{-1}(1 - \alpha, Q_n) \rightarrow L^{-1}(1 - \alpha, F)$ with probability one and by Slutsky's theorem

$$P\{T_n > L_n^{-1}(1 - \alpha, Q_n)\} \rightarrow P\{T_n > L^{-1}(1 - \alpha, F)\} = \alpha. \tag{51}$$

Received March 5, 2003; revised February 19, 2004; accepted April 21, 2004

REFERENCES

Anderson, T. W., and D. A. Darling. (1952). "Asymptotic Theory of Certain Goodness of Fit Criteria Based on Stochastic Processes." *Annals of Mathematical Statistics* 23, 193–212.

Beran, R. (1986). "Simulated Power Functions." *Annals of Statistics* 14, 151–173.

- Davis, R. A., and S. I. Resnick. (1984). "Tail Estimates Motivated by Extreme Value Theory." *Annals of Statistics* 12, 1467–1487.
- de Haan, L. (1976). "Sample Extremes: An Elementary Introduction." *Statistica Neerlandica* 30, 161–172.
- Embrechts, P., C. Klüppelberg, and T. Mikosch. (1997). *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer.
- Gnedenko, B. V. (1943). "Sur la distribution limit du terme d'une série aléatoire." *Annals of Mathematics* 44, 423–453.
- Goldie, C. M., and R. L. Smith. (1987). "Slow Variation with Remainder: Theory and Applications." *Quarterly Journal of Mathematics* 38(2), 45–71.
- Granger, C. W. J. (2002). "Some Comments on Risk." *Journal of Applied Econometrics* 17, 447–456.
- Hall, P. (1982). "On Some Simple Estimates of an Exponent of Regular Variation." *Journal of Royal Statistical Society London B* 44, 37–42.
- Hill, B. M. (1975). "A Simple General Approach to Inference About the Tail of a Distribution." *Annals of Statistics* 5, 1163–1174.
- Kolmogorov, A. N. (1933). "Sulla Determinazione Empirica di una Legge di Distribuzione." *Giornale del Istituto Italiano degli Attuari* 4, 83–91.
- Leadbetter, M. R., G. Lindgren, and H. Rootzén. (1983). *Extremes and Related Properties of Random Sequences and Processes*. New York: Springer-Verlag.
- Mason, D., and J. H. Schuenemeyer. (1983). "A Modified Kolmogorov-Smirnov Test Sensitive to Tail Alternatives." *Annals of Statistics* 3, 933–946.
- Pickands, J. III (1975). "Statistical Inference Using Extreme Order Statistics." *Annals of Statistics* 1, 119–131.
- Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer-Verlag.
- Press, W. H. (1992). *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge: Cambridge University Press.
- Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. New York: Springer-Verlag.
- Romano, J. P. (1988). "A Bootstrap Revival of Some Nonparametric Distance Tests." *Journal of the American Statistical Association* 83, 698–708.
- Shiryaev, A. N. (2001). *Essentials of Stochastic Finance: Facts, Models, Theory*. Singapore: World Scientific Publishing.
- Smith, R. L. (1985). "Maximum Likelihood Estimation in a Class of Non-Regular Cases." *Biometrika* 72, 67–90.
- Smith, R. L. (1987). "Estimating Tails of Probability Distributions." *Annals of Statistics* 15, 1174–1207.
- von Mises, R. (1936). "La Distribution de la Plus Grande de n Valeurs." In *Selected Papers of Richard von Mises*, Vol. 2. Providence, RI: American Mathematical Society.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.