An Application of Extreme Value Theory for Measuring Financial Risk [1]

Manfred Gilli [a,*], Evis Këllezi [b,2],

[a] *Department of Econometrics, University of Geneva and FAME*

[b] *Mirabaud & Cie, Boulevard du Théâtre 3, 1204 Geneva, Switzerland*

Abstract

Assessing the probability of rare and extreme events is an important issue in the risk management of financial portfolios. Extreme value theory provides the solid fundamentals needed for the statistical modelling of such events and the computation of extreme risk measures. The focus of the paper is on the use of extreme value theory to compute tail risk measures and the related confidence intervals, applying it to several major stock market indices.

Key words: Extreme Value Theory, Generalized Pareto Distribution, Generalized Extreme Value Distribution, Quantile Estimation, Risk Measures, Maximum Likelihood Estimation, Profile Likelihood Confidence Intervals

---

1 Introduction

The last years have been characterized by significant instabilities in financial markets worldwide. This has led to numerous criticisms about the existing risk

* Corresponding author: Department of Econometrics, University of Geneva, Bd du Pont d'Arve 40, 1211 Geneva 4, Switzerland. Tel.: + 41 22 379 8222; fax: + 41 22 379 8299.

Email addresses: `Manfred.Gilli@metri.unige.ch` (Manfred Gilli), `Evis.Kellezi@Mirabaud.com` (Evis Këllezi).

[1] The original publication is available at `www.springerlink.com` (DOI: 10.1007/s10614-006-9025-7).

[2] Supported by the Swiss National Science Foundation (project 12–52481.97). We are grateful to three anonymous referees for corrections and comments and thank Elion Jani and Agim Xhaja for their suggestions.

Article published in Computational Economics 27(1), 2006, 1–23.

management systems and motivated the search for more appropriate methodologies able to cope with rare events that have heavy consequences.

The typical question one would like to answer is: "If things go wrong, how wrong can they go? " The problem is then how to model the rare phenomena that lie outside the range of available observations. In such a situation it seems essential to rely on a well founded methodology. Extreme value theory (EVT) provides a firm theoretical foundation on which we can build statistical models describing extreme events.

In many fields of modern science, engineering and insurance, extreme value theory is well established (see e.g. Embrechts et al. (1999), Reiss and Thomas (1997)). Recently, numerous research studies have analyzed the extreme variations that financial markets are subject to, mostly because of currency crises, stock market crashes and large credit defaults. The tail behaviour of financial series has, among others, been discussed in Koedijk et al. (1990), Dacorogna et al. (1995), Loretan and Phillips (1994), Longin (1996), Danielsson and de Vries (2000), Kuan and Webber (1998), Straetmans (1998), McNeil (1999), Jondeau and Rockinger (1999), Rootzèn and Klüppelberg (1999), Neftci (2000), McNeil and Frey (2000) and Gençay et al. (2003b). An interesting discussion about the potential of extreme value theory in risk management is given in Diebold et al. (1998).

This paper deals with the behavior of the tails of financial series. More specifically, the focus is on the use of extreme value theory to compute tail risk measures and the related confidence intervals.

Section 2 presents the definitions of the risk measures we consider in this paper. Section 3 reviews the fundamental results of extreme value theory used to model the distributions underlying the risk measures. In Section 4, a practical application is presented where six major developed market indices are analyzed. In particular, point and interval estimates of the tail risk measures are computed. Section 5 concludes.

2 Risk Measures

Some of the most frequent questions concerning risk management in finance involve extreme quantile estimation. This corresponds to the determination of the value a given variable exceeds with a given (low) probability. A typical example of such measures is the Value-at-Risk (`VaR`). Other less frequently used measures are the expected shortfall (`ES`) and the return level. Hereafter we define the risk measures we focus on in the following chapters.

Value-at-Risk

Value-at-Risk is generally defined as the capital sufficient to cover, in most instances, losses from a portfolio over a holding period of a fixed number of days. Suppose a random variable X with continuous distribution function F models losses or negative returns on a certain financial instrument over a certain time horizon. $\mathtt{VaR}_{\mathrm{p}}$ can then be defined as the p-th quantile of the distribution F

$$\mathtt{VaR}_{\mathrm{p}} = F^{-1}(1-p), \tag{1}$$

where F^{-1} is the so called *quantile function* [3] defined as the inverse of the distribution function F.

For internal risk control purposes, most of the financial firms compute a 5% $\mathtt{VaR}$ over a one-day holding period. The Basle accord proposed that $\mathtt{VaR}$ for the next 10 days and $p = 1\%$, based on a historical observation period of at least 1 year of daily data, should be computed and then multiplied by the 'safety factor' 3. The safety factor was introduced because the normal hypothesis for the profit and loss distribution is widely recognized as unrealistic.

Expected Shortfall

Another informative measure of risk is the *expected shortfall* ($\mathtt{ES}$) or the *tail conditional expectation* which estimates the potential size of the loss exceeding $\mathtt{VaR}$. The expected shortfall is defined as the expected size of a loss that exceeds $\mathtt{VaR}_{\mathrm{p}}$

$$\mathtt{ES}_{\mathrm{p}} = E(X \mid X > \mathtt{VaR}_{\mathrm{p}}). \tag{2}$$

Artzner et al. (1999) argue that expected shortfall, as opposed to Value-at-Risk, is a coherent risk measure.

Return Level

If H is the distribution of the maxima observed over successive non overlapping periods of equal length, the *return level* $R_n^k = H^{-1}(1 - \frac{1}{k})$ is the level expected to be exceeded in one out of k periods of length n. The return level can be used as a measure of the maximum loss of a portfolio, a rather more conservative measure than the Value-at-Risk.

[3] More generally a quantile function is defined as the generalized inverse of F: $F^{\leftarrow}(p) = \inf\{x \in \mathbb{R} : F(x) \geq p\}, \quad 0 < p < 1.$

3 Extreme Value Theory

When modelling the maxima of a random variable, extreme value theory plays the same fundamental role as the central limit theorem plays when modelling sums of random variables. In both cases, the theory tells us what the limiting distributions are.

Generally there are two related ways of identifying extremes in real data. Let us consider a random variable representing daily losses or returns. The first approach considers the maximum the variable takes in successive periods, for example months or years. These selected observations constitute the extreme events, also called block (or per period) maxima. In the left panel of Figure 1, the observations X_2, X_5, X_7 and X_{11} represent the block maxima for four periods of three observations each.

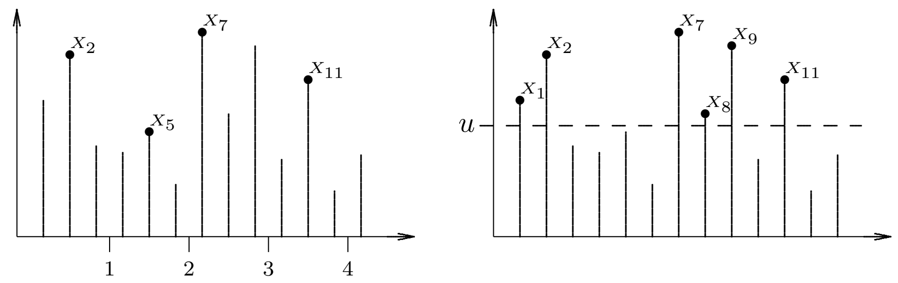


Fig. 1. Block-maxima (left panel) and excesses over a threshold u (right panel).

The second approach focuses on the realizations exceeding a given (high) threshold. The observations X_1, X_2, X_7, X_8, X_9 and X_{11} in the right panel of Figure 1, all exceed the threshold u and constitute extreme events.

The block maxima method is the traditional method used to analyze data with seasonality as for instance hydrological data. However, the threshold method uses data more efficiently and, for that reason, seems to become the method of choice in recent applications.

In the following subsections, the fundamental theoretical results underlying the block maxima and the threshold method are presented.

3.1 Distribution of Maxima

The limit law for the block maxima, which we denote by M_n, with n the size of the subsample (block), is given by the following theorem:

Theorem 1 (Fisher and Tippett (1928), Gnedenko (1943)) *Let* (X_n) *be a* *sequence of i.i.d. random variables. If there exist constants* $c_n > 0$, $d_n \in \mathbb{R}$ *and some non-degenerate distribution function* H *such that*

$$\frac{M_n - d_n}{c_n} \xrightarrow{\mathtt{d}} H,$$

then H *belongs to one of the three standard extreme value distributions:*

$$\text{Fréchet:} \qquad \Phi_\alpha(x) = \begin{cases} 0, & x \le 0 \\ e^{-x^{-\alpha}}, & x > 0 \end{cases} \quad \alpha > 0,$$

$$\text{Weibull:} \qquad \Psi_\alpha(x) = \begin{cases} e^{-(-x)^\alpha}, & x \le 0 \\ 1, & x > 0 \end{cases} \quad \alpha > 0,$$

$$\text{Gumbel:} \qquad \Lambda(x) = e^{-e^{-x}}, \ x \in \mathbb{R}.$$

The shape of the probability density functions for the standard Fréchet, Weibull and Gumbel distributions is given in Figure 2.

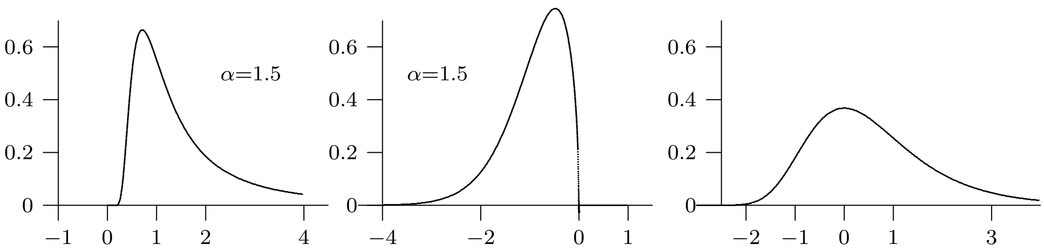


Fig. 2. Densities for the Fréchet, Weibull and Gumbel functions.

We observe that the Fréchet distribution has a polynomially decaying tail and therefore suits well heavy tailed distributions. The exponentially decaying tails of the Gumbel distribution characterize thin tailed distributions. Finally, the Weibull distribution is the asymptotic distribution of finite endpoint distributions.

Jenkinson (1955) and von Mises (1954) suggested the following one-parameter representation

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

of these three standard distributions, with x such that $1 + \xi x > 0$. This generalization, known as the *generalized extreme value* (GEV) distribution, is obtained by setting $\xi = \alpha^{-1}$ for the Fréchet distribution, $\xi = -\alpha^{-1}$ for the

Weibull distribution and by interpreting the Gumbel distribution as the limit case for $\xi = 0$.

As in general we do not know in advance the type of limiting distribution of the sample maxima, the generalized representation is particularly useful when maximum likelihood estimates have to be computed. Moreover the standard GEV defined in (3) is the limiting distribution of normalized extrema. Given that in practice we do not know the true distribution of the returns and, as a result, we do not have any idea about the norming constants c_n and d_n, we use the three parameter specification

$$H_{\xi,\sigma,\mu}(x) = H_\xi\left(\frac{x-\mu}{\sigma}\right) \quad x \in \mathcal{D}, \quad \mathcal{D} = \begin{cases}]-\infty,\ \mu - \frac{\sigma}{\xi}[& \xi < 0 \\ ]-\infty,\ \infty[& \xi = 0 \\ ]\mu - \frac{\sigma}{\xi},\ \infty[& \xi > 0 \end{cases} \tag{4}$$

of the GEV, which is the limiting distribution of the unnormalized maxima. The two additional parameters μ and σ are the location and the scale parameters representing the unknown norming constants.

The quantities of interest are not the parameters themselves, but the quantiles, also called return levels, of the estimated GEV:

$$R^k = H^{-1}_{\xi,\sigma,\mu}(1 - \tfrac{1}{k}) \,.$$

Substituting the parameters ξ, σ and μ by their estimates $\hat{\xi}$, $\hat{\sigma}$ and $\hat{\mu}$, we get

$$\hat{R}^k = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}\left(1 - \left(-\log(1-\frac{1}{k})\right)^{-\hat{\xi}}\right) & \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\sigma}\log\left(-\log(1-\frac{1}{k})\right) & \hat{\xi} = 0 \end{cases} \,. \tag{5}$$

A value of $\hat{R}^{10}$ of 7 means that the maximum loss observed during a period of one year will exceed 7% once in ten years on average.

3.2 Distribution of Exceedances

An alternative approach, called the *peak over threshold* (POT) method, is to consider the distribution of exceedances over a certain threshold. Our problem is illustrated in Figure 3 where we consider an (unknown) distribution function F of a random variable X. We are interested in estimating the distribution function F_u of values of x above a certain threshold u.

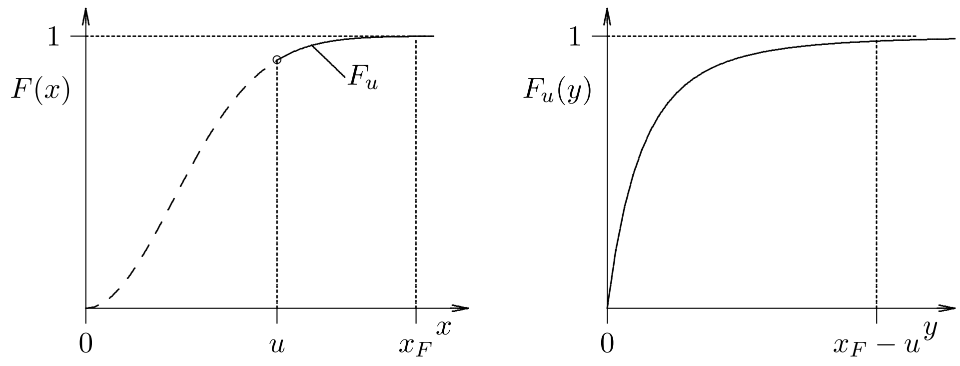


Fig. 3. Distribution function F and conditional distribution function F_u.

The distribution function F_u is called the *conditional excess distribution function* and is defined as

$$F_u(y) = P(X - u \leq y \mid X > u), \qquad 0 \leq y \leq x_F - u \tag{6}$$

where X is a random variable, u is a given threshold, $y = x - u$ are the excesses and $x_F \leq \infty$ is the right endpoint of F. We verify that F_u can be written in terms of F, i.e.

$$F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)} = \frac{F(x) - F(u)}{1 - F(u)}. \tag{7}$$

The realizations of the random variable X lie mainly between 0 and u and therefore the estimation of F in this interval generally poses no problems. The estimation of the portion F_u however might be difficult as we have in general very little observations in this area.

At this point EVT can prove very helpful as it provides us with a powerful result about the conditional excess distribution function which is stated in the following theorem:

Theorem 2 (Pickands (1975), Balkema and de Haan (1974)) *For a large class of underlying distribution functions F the conditional excess distribution function $F_u(y)$, for u large, is well approximated by*

$$F_u(y) \approx G_{\xi,\sigma}(y), \qquad u \to \infty,$$

where

$$G_{\xi,\sigma}(y) = \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma} y\right)^{-1/\xi} & \text{if} \quad \xi \neq 0 \\ 1 - e^{-y/\sigma} & \text{if} \quad \xi = 0 \end{cases} \tag{8}$$

for $y \in [0,\ (x_F - u)]$ if $\xi \geq 0$ and $y \in [0,\ -\frac{\sigma}{\xi}]$ if $\xi < 0$. $G_{\xi,\sigma}$ is the so called generalized Pareto distribution (GPD).

If x is defined as $x = u + y$, the GPD can also be expressed as a function of x, i.e. $G_{\xi,\sigma}(x) = 1 - (1 + \xi(x-u)/\sigma)^{-1/\xi}$.

Figure 4 illustrates the shape of the generalized Pareto distribution $G_{\xi,\sigma}(x)$ when ξ, called the *shape parameter* or *tail index*, takes a negative, a positive and a zero value. The scaling parameter σ is kept equal to one.

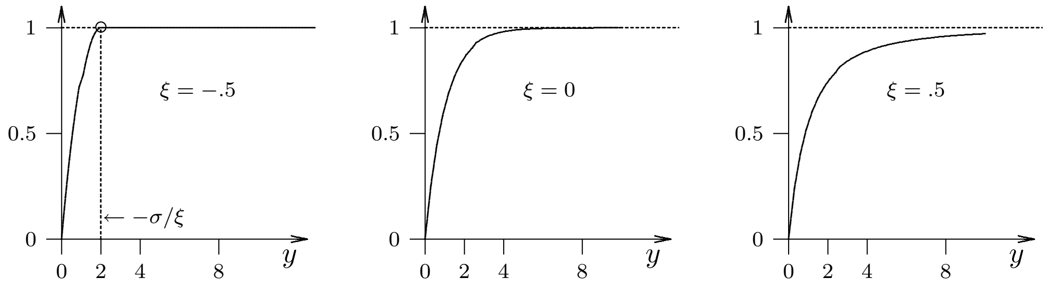


Fig. 4. Shape of the generalized Pareto distribution $G_{\xi,\sigma}$ for $\sigma = 1$.

The tail index ξ gives an indication of the heaviness of the tail, the larger ξ, the heavier the tail. As, in general, one cannot fix an upper bound for financial losses, only distributions with shape parameter $\xi \geq 0$ are suited to model financial return distributions.

Assuming a GPD function for the tail distribution, analytical expressions for $\mathtt{VaR}_\mathrm{p}$ and $\mathtt{ES}_\mathrm{p}$ can be defined as a function of GPD parameters. Isolating $F(x)$ from (7)

$$F(x) = (1 - F(u))\ F_u(y) + F(u)$$

and replacing F_u by the GPD and $F(u)$ by the estimate $(n - N_u)/n$, where n is the total number of observations and N_u the number of observations above the threshold u, we obtain

$$\widehat{F}(x) = \tfrac{N_u}{n}\left(1 - \left(1 + \tfrac{\hat{\xi}}{\hat{\sigma}}(x-u)\right)^{-1/\hat{\xi}}\right) + \left(1 - \tfrac{N_u}{n}\right)$$

which simplifies to

$$\widehat{F}(x) = 1 - \tfrac{N_u}{n}\left(1 + \tfrac{\hat{\xi}}{\hat{\sigma}}(x-u)\right)^{-1/\hat{\xi}}. \tag{9}$$

Inverting (9) for a given probability p gives

$$\widehat{\mathtt{VaR}}_\mathrm{p} = u + \tfrac{\hat{\sigma}}{\hat{\xi}}\left(\left(\tfrac{n}{N_u}p\right)^{-\hat{\xi}} - 1\right). \tag{10}$$

Let us rewrite the expected shortfall as

$$\widehat{\mathtt{ES}}_\mathrm{p} = \widehat{\mathtt{VaR}}_\mathrm{p} + E(X - \widehat{\mathtt{VaR}}_\mathrm{p} \mid X > \widehat{\mathtt{VaR}}_\mathrm{p})$$

where the second term on the right is the expected value of the exceedances over the threshold $\mathtt{VaR}_{\mathrm{p}}$. It is known that the *mean excess function* for the GPD with parameter $\xi < 1$ is

$$e(z) = E(X - z \mid X > z) = \frac{\sigma + \xi z}{1 - \xi}, \qquad \sigma + \xi z > 0\,. \tag{11}$$

This function gives the average of the excesses of X over varying values of a threshold z. Another important result concerning the existence of moments is that if X follows a GPD then, for all integers r such that $r < 1/\xi$, the r first moments exist.[4]

Similarly, given the definition (2) for the expected shortfall and using expression (11), for $z = \mathtt{VaR}_{\mathrm{p}} - u$ and X representing the excesses y over u we obtain

$$\widehat{\mathtt{ES}}_{\mathrm{p}} = \widehat{\mathtt{VaR}}_{\mathrm{p}} + \frac{\hat{\sigma} + \hat{\xi}(\widehat{\mathtt{VaR}}_{\mathrm{p}} - u)}{1 - \hat{\xi}} = \frac{\widehat{\mathtt{VaR}}_{\mathrm{p}}}{1 - \hat{\xi}} + \frac{\hat{\sigma} - \hat{\xi} u}{1 - \hat{\xi}}\,. \tag{12}$$

4 Application

Our aim is to illustrate the tail distribution estimation of a set of financial series of daily returns and use the results to quantify the market risk. Table 1 gives the list of the financial series considered in our analysis. The illustration focuses mainly on the S&P500 index, providing confidence intervals and graphical visualization of the estimates, whereas for the other series only point estimates are reported.

Table 1
Data analyzed.

Symbol	Index name	Start	End	Observations
`ES50`	Dow Jones Euro Stoxx 50	2–01–87	17–08–04	4555
`FTSE100`	FTSE 100	5–01–84	17–08–04	5215
`HS`	Hang Seng	9–01–81	17–08–04	5836
`Nikkei`	Nikkei 225	7–01–70	17–08–04	8567
`SMI`	Swiss Market Index	5–07–88	17–08–04	4050
`S&P500`	S&P 500	5–01–60	16–08–04	11270

The application has been executed in a Matlab 7.x programming environment.[5] The files with the data and the code can be downloaded from `www.`

[4] See Embrechts et al. (1999), page 165.
[5] Other software for extreme value analysis can be found at `www.math.ethz.ch/~mcneil/software.html` or in Gençay et al. (2003a). Standard numerical or statistical software, like for example Matlab, now also provide functions or routines that can be used for EVT applications.

unige.ch/ses/metri/gilli/evtrm/. Figure 5 shows the plot of the $n = 11270$ observed daily returns of the S&P500 index.

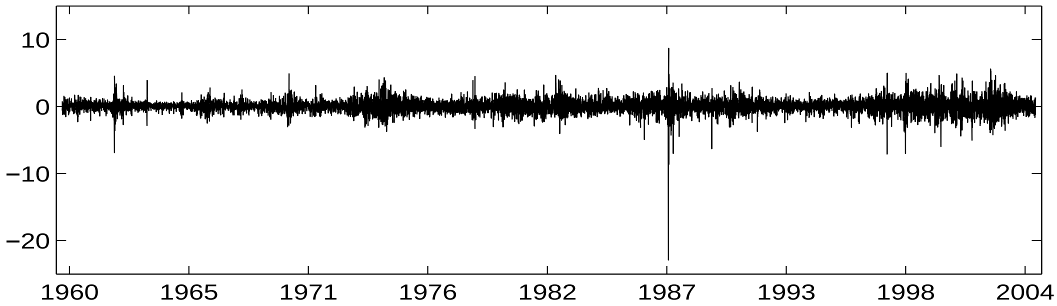


Fig. 5. Daily returns of the S&P500 index.

We consider both the left and the right tail of the return distribution. The reason is that the left tail represents losses for an investor with a long position on the index, whereas the right tail represents losses for an investor being short on the index.

As it can be seen from Figure 5, returns exhibit dependence in the second moment. McNeil and Frey (2000) propose a two stage method consisting in modelling the conditional distribution of asset returns against the current volatility and then fitting the GPD on the tails of residuals. On the other side, Danielsson and de Vries (2000) argue that for long time horizons an unconditional approach is better suited. Indeed, as Christoffersen and Diebold (2000) notice, conditional volatility forecasting is not indicated for multiple day predictions. For a detailed discussion on these issues, including the i.i.d. assumptions, we refer the reader to the above mentioned references, believing that the choice between conditional and unconditional approaches depends on the final use of the risk measures and the time horizon considered. For short time horizons of the order of several hours or days, and if an automatic updating of the parameters is feasible, a conditional approach may be indicated. For longer horizons, a non conditional approach might be justified by the fact that it provides stable estimates through time requiring less frequent updates.

The methodology applying to right tails, in the left tail case we change the sign of the returns so that positive values correspond to losses.

First, we consider the distribution of the block maxima, which allows the determination of the return level. Second, we model the exceedances over a given threshold which enables us to estimate high quantiles of the return distribution and the corresponding expected shortfall.

In both cases we use maximum likelihood estimation, which is one of the most common estimation procedures used in practice. We also compute likelihood-based interval estimates of the parameters and the quantities of interest which provide additional information related to the accuracy of the point estimates. These intervals, contrarily to those based on standard errors, do not rely on asymptotic theory results and restrictive assumptions. We expect them to be more accurate in the case of small sample size. Another advantage of the likelihood-based approach is the possibility to construct joint confidence intervals. The greater computational complexity of the likelihood-based approach is nowadays no longer an obstacle for its use.

4.1 Method of Block Maxima

The application of the method of block maxima goes through the following steps: divide the sample in n blocks of equal length, collect the maximum value in each block, fit the GEV distribution to the set of maxima and, finally, compute point and interval estimates for R_n^k.

The delicate point of this method is the appropriate choice of the periods defining the blocks. The calendar naturally suggests periods like months, quarters, etc. In order to avoid seasonal effects, we choose yearly periods which are likely to be sufficiently large for Theorem 1 to hold. The `S&P500` data sample has been divided into 45 non-overlapping sub-samples, each of them containing the daily returns of the successive calendar years. Therefore not all our blocks are of exactly the same length. The maximum return in each of the blocks constitute the data points for the sample of maxima M which is used to estimate the generalized extreme value distribution (GEV). Figure 6 plots the yearly maxima for the left and right tails of the `S&P500`.

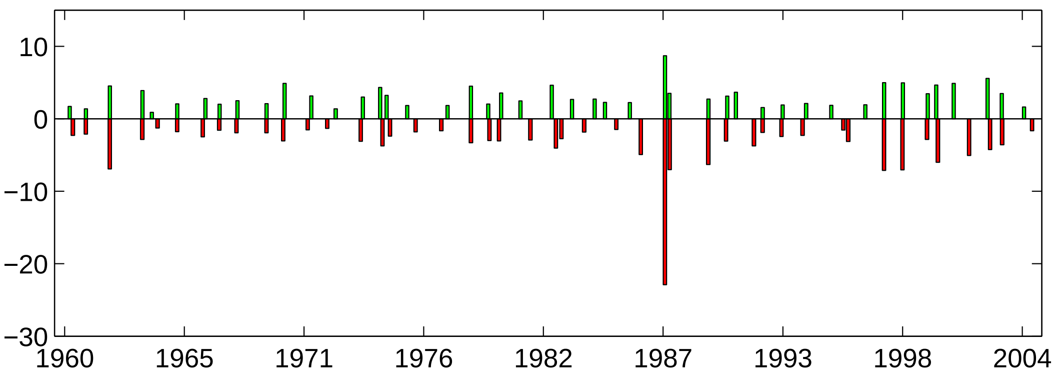


Fig. 6. Yearly minima and maxima of the daily returns of the `S&P500`.

The log-likelihood function that we maximize with respect to the three unknown parameters is

$$L(\xi, \mu, \sigma; x) = \sum_i \log\Big(h(x_i)\Big), \qquad x_i \in M \tag{13}$$

where

$$h(\xi, \mu, \sigma; x) = \frac{1}{\sigma}\left(1 + \xi \frac{x-\mu}{\sigma}\right)^{-1/\xi - 1} \exp\left(-\left(1 + \xi\,\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right)$$

is the probability density function if $\xi \neq 0$ and $1 + \xi\frac{x-\mu}{\sigma} > 0$. If $\xi = 0$ the function h is

$$h(\xi, \mu, \sigma; x) = \frac{1}{\sigma}\exp\left(-\frac{x-\mu}{\sigma}\right)\exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right).$$

In Figure 7, we give the plot of the sample distribution function [6] and the corresponding fitted GEV distribution. Point and interval estimates for the parameters are given in Table 2.

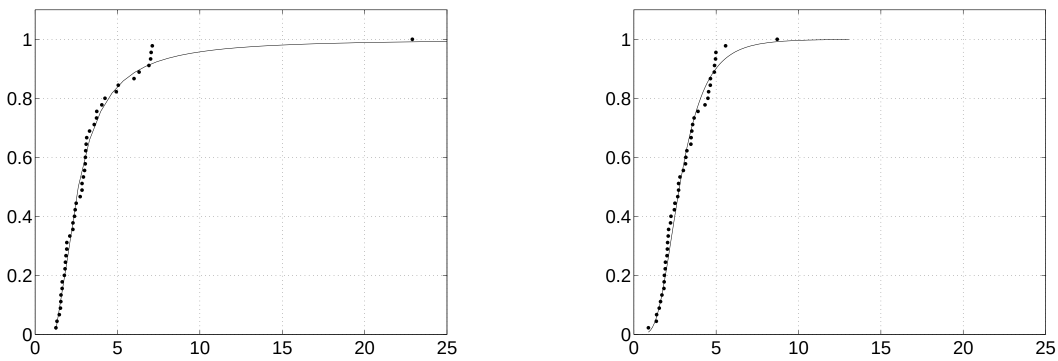


Fig. 7. Sample distribution (dots) of yearly minima (left panel) and maxima (right panel) and corresponding fitted GEV distribution for `S&P500`.

Interval Estimates

In order to be able to compute interval estimates, [7] it is useful to approach the quantile estimation problem by directly reparameterizing the GEV distribution as a function of the unknown return level R^k. To achieve this, we isolate μ from equation (5) and substitute it into $H_{\xi,\sigma,\mu}$ defined in (4). The

[6] The sample distribution function $\widehat{F}_n(x_i^n)$ for a set of n observations given in increasing order $x_1^n \leq \cdots \leq x_n^n$, is defined as $\widehat{F}_n(x_i^n) = \frac{i}{n}$, $i = 1, \ldots, n$.
[7] For a good introduction to likelihood-based statistical inference, see Azzalini (1996).

GEV distribution function then becomes

$$H_{\xi,\sigma,R^k}(x) = \begin{cases} \exp\left(-\left(\frac{\xi}{\sigma}(x - R^k) + \left(-\log(1-\frac{1}{k})\right)^{-\xi}\right)^{-1/\xi}\right) & \xi \neq 0 \\ (1-\frac{1}{k})^{\exp\left(-\frac{x-R^k}{\sigma}\right)} & \xi = 0 \end{cases}$$

for $x \in \mathcal{D}$ defined as

$$\mathcal{D} = \begin{cases}]-\infty,\ \left(R^k - \frac{\xi}{\sigma}\left(-\log(1-\frac{1}{k})\right)^{-\xi}\right)[& \xi < 0 \\ ]-\infty,\ \infty[& \xi = 0 \\ ]\left(R^k - \frac{\xi}{\sigma}\left(-\log(1-\frac{1}{k})\right)^{-\xi}\right),\ \infty[& \xi > 0 \end{cases}$$

and we can directly obtain maximum likelihood estimates for R^k. The profile log-likelihood function can then be used to compute separate or joint confidence intervals for each of the parameters. For example, in the case where the parameter of interest is R^k, the profile log-likelihood function will be defined as

$$L^*(R^k) = \max_{\xi,\sigma} L(\xi, \sigma, R^k)\,.$$

The confidence interval we then derive includes all values of R^k satisfying the condition

$$L^*(R^k) - L(\hat{\xi}, \hat{\sigma}, \hat{R}^k) > -\tfrac{1}{2}\,\chi^2_{\alpha,1}$$

where $\chi^2_{\alpha,1}$ refers to the $(1-\alpha)$–level quantile of the χ^2 distribution with 1 degree of freedom. The function $L^*(R^k) - L^*(\hat{\xi}, \hat{\sigma}, \hat{R}^k)$ is called the *relative profile log-likelihood function* and is plotted in the left panel of Figure 8. The point estimate of 6.41% of R^{10} is included in the rather large interval (4.74, 11). As less observations are available for higher quantiles, the interval is asymmetric, indicating more uncertainty for the upper bound of maximum losses.

Sometimes we are also interested in the value of ξ, which characterizes the tail heaviness of the underlying distribution. In this case, a joint confidence region on both ξ and R^{10} is needed. The profile log-likelihood function is then defined as

$$L^*(\xi, R^k) = \max_{\sigma} L(\xi, \sigma, R^k),$$

with the confidence region defined as the contour at the level $-\frac{1}{2}\,\chi^2_{\alpha,2}$ of the relative profile log-likelihood function

$$L^*(\xi, R^k) - L(\hat{\xi}, \hat{\sigma}, \hat{R}^k).$$

In the right panel of Figure 8, we reproduce single and joint confidence regions at level 95% for $\hat{\xi}$ and $\hat{R}^{10}$ of the `S&P500`. In the same graph we also plot the pairs $(\hat{R}^{10}, \hat{\xi})$ estimated on 1000 bootstrap samples. The joint confidence region covers approximately 95% of the bootstrap pairs, indicating that computing the joint interval region gives a good idea about the likely values of the parameters. Moreover, we notice that the joint region is significantly different from the one defined by the single confidence intervals.

In order to account for the small sample size, single confidence intervals are also computed with a bias-corrected and accelerated (BCa) bootstrap method. [8] As a result, for $\hat{R}^{10}$, the BCa interval narrows to (5.01, 9.19). Regarding the shape parameter ξ, the difference is less pronounced. However, in both cases, the intervals clearly indicate a positive value for ξ, which implies that the limiting distribution of maxima belongs to the Fréchet family.

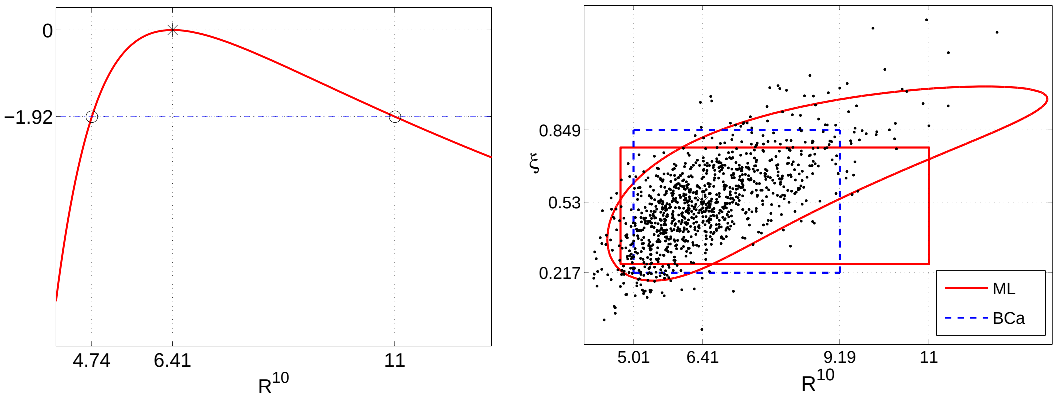


Fig. 8. Left panel: Relative profile log-likelihood and 95% confidence interval for $\hat{R}^{10}$ of the left tail. Right panel: Single and joint confidence regions for $\hat{\xi}$ and $\hat{R}^{10}$ at level 95%. Maximum likelihood estimates are marked with the symbol $*$.

The point estimates and the single confidence intervals for the reparameterized GEV distribution for `S&P500` are summarized in Table 2.

Table 2
Point estimates and 95% maximum likelihood (ML) and bootstrap (BCa) confidence intervals for the GEV method applied to `S&P500`.

	Lower bound		Point estimate	Upper bound	
	BCa	ML	ML	ML	BCa
			Left tail		
$\hat{\xi}$	0.217	0.256	0.530	0.771	0.849
$\hat{\sigma}$	0.802	0.815	0.964	1.213	1.188
$\hat{R}^{10}$	5.006	4.741	6.411	11.001	9.190
			Right tail		
$\hat{\xi}$	−.288	−.076	0.100	0.341	0.392
$\hat{\sigma}$	0.815	0.836	1.024	1.302	1.312
$\hat{R}^{10}$	4.368	4.230	4.981	6.485	5.869

[8] For an introduction to bootstrap methods see Efron and Tibshirani (1993) or Shao and Tu (1995).

For $k = 10$, we obtain for our data $\hat{R}^{10} = 6.41$, meaning that the maximum loss observed during a period of one year exceeds 6.41% in one out of ten years on average. In the same way we can derive that a loss of $R^{100} = 21.27\,\%$ is exceeded on average only once in a century. Notice that this is very close to the 87 crash daily loss of 22.90 %.

Table 3 summarizes point estimates for GEV for all six indices. Because of the low number of observations for `ES50` and `SMI` the corresponding point estimates are less reliable. We notice the high value of R^{10} for the left tail of Hang Seng (`HS`), twice as big as the next riskiest index.

Table 3
Point estimates for the GEV method for six market indices.

	`ES50`	`FTSE100`	`HS`	`Nikkei`	`SMI`	`S&P500`
# maxima	18	21	24	35	17	45
			Left tail			
$\hat{\xi}$	−.301	0.679	0.512	0.251	0.172	0.530
$\hat{\sigma}$	1.773	0.705	2.707	1.616	1.563	0.964
$\hat{R}^{10}$	7.217	6.323	16.950	8.243	8.341	6.411
			Right tail			
$\hat{\xi}$	0.185	0.309	0.179	0.096	−.032	0.100
$\hat{\sigma}$	1.252	0.919	1.754	1.518	1.773	1.024
$\hat{R}^{10}$	6.366	5.501	8.439	7.159	7.611	4.981

One way to better exploit information about extremes in the data sample is to use the POT method. Coles (2001, p. 81) suggests the estimation of return levels using GPD. However, if the data set is large enough, GEV may still prove useful as it can avoid dealing with data clustering issues, provided that blocks are sufficiently large. Furthermore, the estimation is simplified as the selection of a threshold u is not needed.

4.2 The Peak Over Threshold Method

The implementation of the peak over threshold method involves the following steps: select the threshold u, fit the GPD function to the exceedances over u and then compute point and interval estimates for Value-at-Risk and the expected shortfall.

Selection of the threshold u

Theory tells us that u should be high in order to satisfy Theorem 2, but the higher the threshold the less observations are left for the estimation of the parameters of the tail distribution function.

So far, no automatic algorithm with satisfactory performance for the selection of the threshold u is available. The issue of determining the fraction of data belonging to the tail is treated by Danielsson et al. (2001), Danielsson and de Vries (1997) and Dupuis (1998) among others. However these references do not provide a clear answer to the question of which method should be used.

A graphical tool that is very helpful for the selection of the threshold u is the *sample mean excess plot* defined by the points

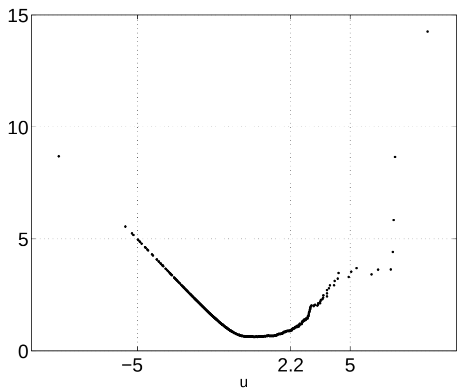
$$\Big(u,\ e_n(u)\Big), \qquad x_1^n < u < x_n^n\,, \tag{14}$$

where $e_n(u)$ is the sample mean excess function defined as

$$e_n(u) = \frac{\sum_{i=k}^{n}(x_i^n - u)}{n-k+1}, \qquad k = \min\{i \mid x_i^n > u\},$$

and $n-k+1$ is the number of observations exceeding the threshold u.

The sample mean excess function, which is an estimate of the mean excess function $e(u)$ defined in (11), should be linear. This property can be used as a criterion for the selection of u. Figure 9 shows the sample mean excess plots corresponding to the S&P500 data. From a closer inspection of the plots we suggest the values $u = 2.2$ for the threshold of the left tail and $u = 1.4$ for the threshold of the right tail. These values are located at the beginning of a portion of the sample mean excess plot that is roughly linear, leaving respectively 158 and 614 observations in the tails (see Figure 10).



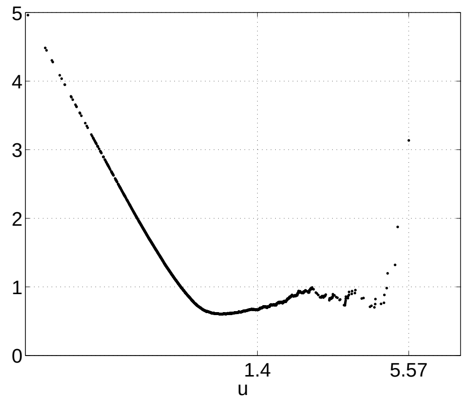


Fig. 9. Sample mean excess plot for the left and right tail determination for S&P 500 data.

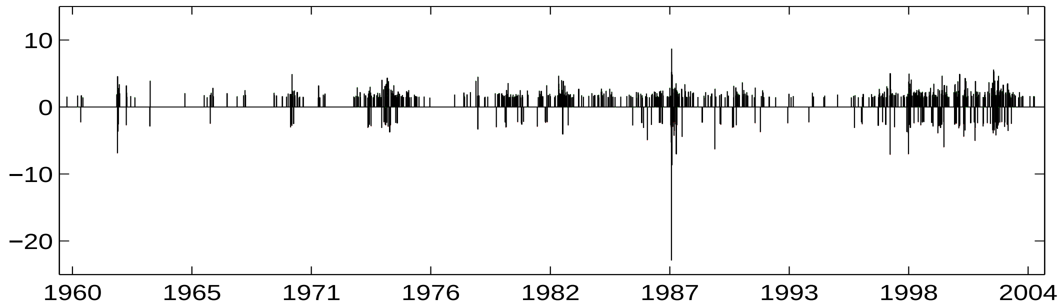


Fig. 10. Exceedances of daily returns of the S&P500 index.

Maximum Likelihood Estimation

Given the theoretical results presented in the previous section, we know that the distribution of the observations above the threshold in the tail should be a generalized Pareto distribution (GPD). Different methods can be used to estimate the parameters of the GPD.[9] In the following we describe the maximum likelihood estimation method.

For a sample $y = \{y_1, \ldots, y_n\}$ the log-likelihood function $L(\xi, \sigma|y)$ for the GPD is the logarithm of the joint density of the n observations

$$L(\xi, \sigma \mid y) = \begin{cases} -n \log \sigma - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^{n} \log \left(1 + \frac{\xi}{\sigma}\, y_i\right) & \text{if} \quad \xi \neq 0 \\ -n \log \sigma - \frac{1}{\sigma} \sum_{i=1}^{n} y_i & \text{if} \quad \xi = 0. \end{cases}$$

We compute the values $\hat{\xi}$ and $\hat{\sigma}$ that maximize the log-likelihood function for the sample defined by the observations exceeding the threshold u. We obtain the estimates $\hat{\xi} = 0.388$ and $\hat{\sigma} = 0.545$ for the left tail exceedances and $\hat{\xi} = 0.137$ and $\hat{\sigma} = 0.579$ for the right tail. Figure 11 shows how GPD fits to exceedances of the left and right tails of the S&P500. Clearly the left tail is heavier than the right one. This can also be seen from the estimated value of the shape parameter ξ which is positive in both cases, but higher in the left tail case.

High quantiles and expected shortfall may be directly read in the plot or computed from equations (10) and (12) where we replace the parameters by their estimates. For instance, for $p = 0.01$ we can compute $\widehat{\mathtt{VaR}}_{0.01} = 2.397$ and

[9] These are the maximum likelihood estimation, the method of moments, the method of probability-weighted moments and the elemental percentile method. For comparisons and detailed discussions about their use for fitting the GPD to data, see Hosking and Wallis (1987), Grimshaw (1993), Tajvidi (1996a), Tajvidi (1996b) and Castillo and Hadi (1997).

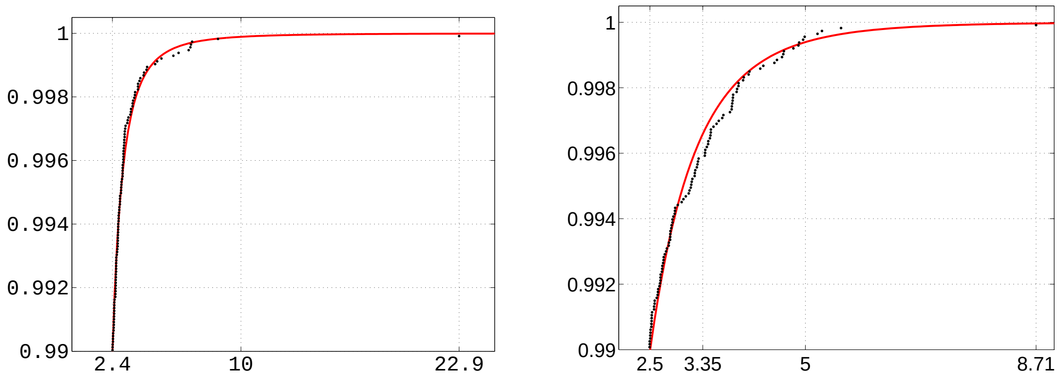


Fig. 11. Left panel: GPD fitted to the 158 left tail exceedances above the threshold $u = 2.2$. Right panel: GPD fitted to the 614 right tail exceedances above the threshold $u = 1.4$.

$\widehat{\texttt{ES}}_{0.01} = 3.412$ ($\widehat{\texttt{VaR}}_{0.01} = 2.505$, $\widehat{\texttt{ES}}_{0.01} = 3.351$ for the right tail). We observe that, with respect to the right tail, the left tail has a lower VaR but a higher ES which illustrates the importance to go beyond a simple VaR calculation.

Interval Estimates

Again, we consider single and joint confidence intervals, based on the profile log-likelihood functions. Log-likelihood based confidence intervals for $\texttt{VaR}_\text{p}$ can be obtained by using a reparameterized version of GPD defined as a function of ξ and $\texttt{VaR}_\text{p}$:

$$G_{\xi,\texttt{VaR}_\text{p}}(y) = \begin{cases} 1 - \left(1 + \frac{\left(\frac{n}{N_u}p\right)^{-\xi} - 1}{\texttt{VaR}_\text{p} - u}\, y\right)^{-\frac{1}{\xi}} & \xi \neq 0 \\ 1 - \frac{n}{N_u}\, p \exp\left(\frac{y}{\texttt{VaR}_\text{p} - u}\right) & \xi = 0 \end{cases}.$$

The corresponding probability density function is

$$g_{\xi,\texttt{VaR}_\text{p}}(y) = \begin{cases} \frac{\left(\frac{n}{N_u}p\right)^{-\xi} - 1}{\xi(\texttt{VaR}_\text{p} - u)} \left(1 + \frac{\left(\frac{n}{N_u}p\right)^{-\xi} - 1}{\texttt{VaR}_\text{p} - u}\, y\right)^{-\frac{1}{\xi} - 1} & \xi \neq 0 \\ -\frac{\frac{n}{N_u}\, p \exp\left(\frac{y}{\texttt{VaR}_\text{p} - u}\right)}{\texttt{VaR}_\text{p} - u} & \xi = 0 \end{cases}.$$

Similarly, using the following reparameterization for $\xi \neq 0$

$$G_{\xi,\texttt{ES}_\text{p}} = 1 - \left(1 + \frac{\xi + \left(\frac{n}{N_u}p\right)^{-\xi} - 1}{(\texttt{ES}_\text{p} - u)(1 - \xi)} y\right)^{-\frac{1}{\xi}},$$

$$g_{\xi,\texttt{ES}_\text{p}} = \frac{\xi + \left(\frac{n}{N_u}p\right)^{-\xi} - 1}{\xi(1 - \xi)(\texttt{ES}_\text{p} - u)} \left(1 + \frac{\xi + \left(\frac{n}{N_u}p\right)^{-\xi} - 1}{(\texttt{ES}_\text{p} - u)(1 - \xi)} y\right)^{-\frac{1}{\xi} - 1},$$

we compute a log-likelihood based confidence interval for the expected shortfall $\mathtt{ES}_p$. Figures 12–13 show the likelihood based confidence regions for the left tail `VaR` and `ES` of `S&P500` obtained by using these reparameterized versions of GPD.

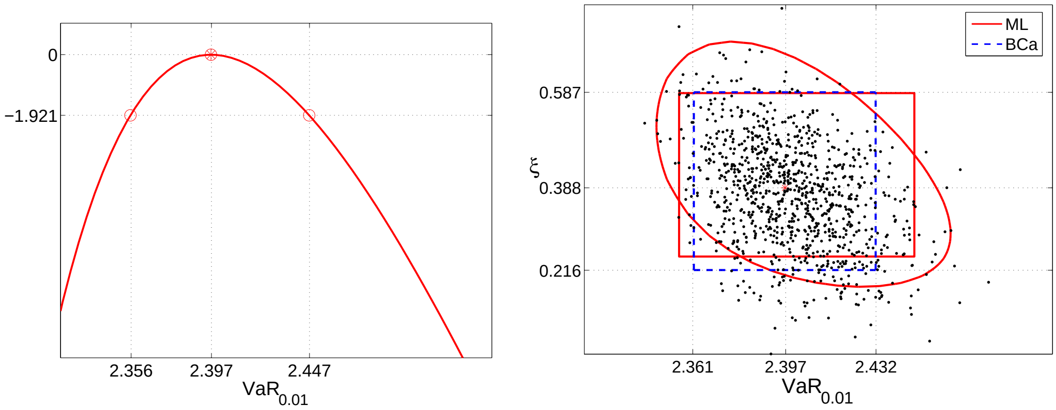


Fig. 12. Left panel: Relative profile log-likelihood function and confidence interval for $\mathtt{VaR}_{0.01}$. Right panel: Single and joint confidence intervals at level 95% for $\hat{\xi}$ and $\mathtt{VaR}_{0.01}$. Dots represent 1000 bootstrap estimates from `S&P500` data.

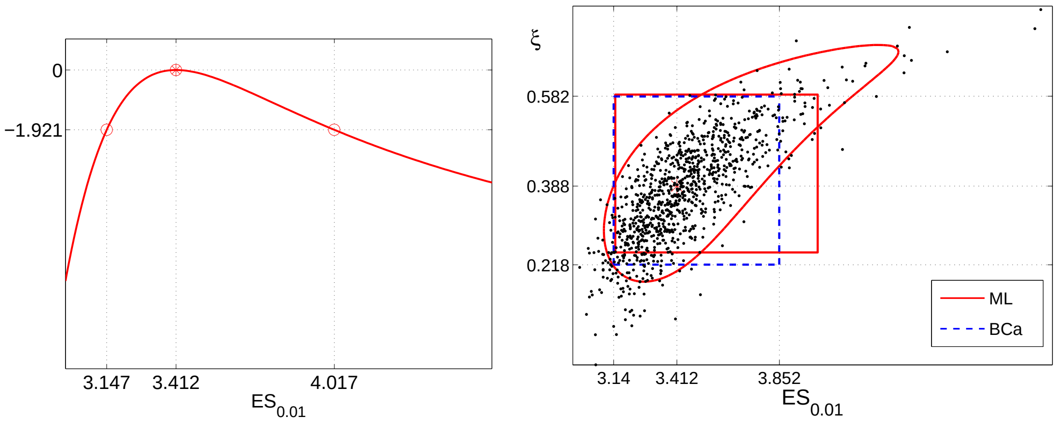


Fig. 13. Left panel: Relative profile log-likelihood function and confidence interval for $\mathtt{ES}_{0.01}$. Right panel: Single and joint confidence intervals at level 95% for $\hat{\xi}$ and $\mathtt{ES}_{0.01}$. Dots represent 1000 bootstrap estimates from `S&P500` data.

In the same figures we show the single bias-corrected and accelerated bootstrap confidence intervals. We also plot the pairs $(\hat{\xi}_i, \widehat{\mathtt{VaR}}_{0.01,i})$, $(\hat{\xi}_i, \widehat{\mathtt{ES}}_{0.01,i})$, $i = 1, \ldots, 1000$ estimated from 1000 resampled data sets. We observe that about 5% lie outside the 95% joint confidence region based on likelihood (which is not the case for the single intervals). Again this shows the interest of considering joint confidence intervals.

The maximum likelihood (ML) point estimates, the maximum likelihood and the BCa bootstrap confidence intervals for $\hat{\xi}$, $\hat{\sigma}$, $\widehat{\mathtt{VaR}}_{0.01}$ and $\widehat{\mathtt{ES}}_{0.01}$ for both tails of the S&P500 are summarized in Table 4.

Table 4
Point estimates and 95% maximum likelihood (ML) and bootstrap (BCa) confidence intervals for the S&P500.

	Lower bound		Point estimate	Upper bound	
	BCa	ML	ML	ML	BCa
			Left tail		
$\hat{\xi}$	0.221	0.245	0.388	0.585	0.588
$\hat{\sigma}$	0.442	0.444	0.545	0.671	0.651
$\widehat{\mathtt{VaR}}_{0.01}$	2.361	2.356	2.397	2.447	2.432
$\widehat{\mathtt{ES}}_{0.01}$	3.140	3.147	3.412	4.017	3.852
			Right tail		
$\hat{\xi}$	0.068	0.075	0.137	0.212	0.220
$\hat{\sigma}$	0.520	0.530	0.579	0.634	0.638
$\widehat{\mathtt{VaR}}_{0.01}$	2.427	2.411	2.505	2.609	2.591
$\widehat{\mathtt{ES}}_{0.01}$	3.182	3.151	3.351	3.634	3.569

The results in Table 4 indicate that, with probability 0.01, the tomorrow's loss on a long position will exceed the value 2.397% and that the corresponding expected loss, that is the average loss in situations where the losses exceed 2.397%, is 3.412%.

It is interesting to note that the upper bound of the confidence interval for the parameter ξ is such that the first order moment is finite $(1/0.671 > 1)$. This guarantees that the estimated expected shortfall, which is a conditional first moment, exists for both tails.

The point estimates for all the six market indices are reported in Table 5. Similarly to the S&P500 case the left tail is heavier than the right one for all indices. Looking at estimated VaR and ES values, we observe that Hang Seng (HS) and DJ Euro Stoxx 50 (ES50) are the most exposed to extreme losses, followed by Nikkei and the Swiss Market Index (SMI). The less exposed indices are S&P500 and FTSE 100.

Regarding the right tail Hang Seng is again the most exposed to daily extreme moves and S&P500 and FTSE100 are the least exposed.

Table 5
Point estimates for the POT method for six market indices.

	ES50	FTSE100	HS	Nikkei	SMI	S&P500
			Left tail			
$\hat{\xi}$	0.045	0.232	0.298	0.181	0.264	0.388
$\hat{\sigma}$	1.102	0.656	1.395	0.810	0.772	0.545
$\widehat{\mathtt{VaR}}_{0.01}$	3.819	2.862	5.146	3.435	3.347	2.397
$\widehat{\mathtt{ES}}_{0.01}$	5.057	4.103	8.046	4.697	4.880	3.412
			Right tail			
$\hat{\xi}$	0.113	0.093	0.156	0.165	0.185	0.137
$\hat{\sigma}$	0.953	0.711	1.102	0.912	0.764	0.579
$\widehat{\mathtt{VaR}}_{0.01}$	3.517	2.707	4.581	3.316	3.078	2.505
$\widehat{\mathtt{ES}}_{0.01}$	4.785	3.562	6.258	4.470	4.259	3.351

5 Concluding Remarks

We have illustrated how extreme value theory can be used to model tail-related risk measures such as Value-at-Risk, expected shortfall and return level, applying it to daily log-returns on six market indices.

Our conclusion is that EVT can be useful for assessing the size of extreme events. From a practical point of view this problem can be approached in different ways, depending on data availability and frequency, the desired time horizon and the level of complexity one is willing to introduce in the model. One can choose to use a conditional or an unconditional approach, the BMM or the POT method, and finally rely on point or interval estimates.

In our application, the POT method proved superior as it better exploits the information in the data sample. Being interested in long term behavior rather than in short term forecasting, we favored an unconditional approach. Finally, we find it is worthwhile computing interval estimates as they provide additional information about the quality of the model fit.

References

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.

Azzalini, A. (1996). *Statistical Inference Based on the Likelihood.* Chapman and Hall, London.

Balkema, A. A. and de Haan, L. (1974). Residual life time at great age. *Annals of Probability*, 2:792–804.

Castillo, E. and Hadi, A. (1997). Fitting the Generalized Pareto Distribution to Data. *Journal of the American Statistical Association*, 92(440):1609–1620.

Christoffersen, P. and Diebold, F. (2000). How relevant is volatility forecasting for financial risk management. *Review of Economics and Statistics*, 82:1–11.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values.* Springer.

Dacorogna, M. M., Müller, U. A., Pictet, O. V., and de Vries, C. G. (1995). The distribution of extremal foreign exchange rate returns in extremely large data sets. Preprint, O&A Research Group.

Danielsson, J. and de Vries, C. (1997). Beyond the Sample: Extreme Quantile and Probability Estimation. Mimeo.

Danielsson, J. and de Vries, C. (2000). Value-at-Risk and Extreme Returns. *Annales d'Economie et de Statistique*, 60:239–270.

Danielsson, J., de Vries, C., de Haan, L., and Peng, L. (2001). Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation. *Journal of Multivariate Analysis*, 76(2):226–248.

Diebold, F. X., Schuermann, T., and Stroughair, J. D. (1998). Pitfalls and opportunities in the use of extreme value theory in risk management. In Refenes, A.-P., Burgess, A., and Moody, J., editors, *Decision Technologies for Computational Finance*, pages 3–12. Kluwer Academic Publishers.

Dupuis, D. J. (1998). Exceedances over high thresholds: A guide to threshold selection. *Extremes*, 1(3):251–261.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* Chapman & Hall, New York.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1999). *Modelling Extremal Events for Insurance and Finance.* Applications of Mathematics. Springer. 2nd ed.(1st ed., 1997).

Fisher, R. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of largest or smallest member of a sample. *Proceedings of the Cambridge philosophical society*, 24:180–190.

Gençay, R., Selçuk, F., and Ulugülyağci, A. (2003a). EVIM: a software package for extreme value analysis in Matlab. *Studies in Nonlinear Dynamics and Econometrics*, 5:213–239.

Gençay, R., Selçuk, F., and Ulugülyağci, A. (2003b). High volatility, thick tails and extreme value theory in value-at-risk estimation. *Insurance: Mathematics and Economics*, 33:337–356.

Gnedenko, B. V. (1943). Sur la distribution limite du terme d'une série aléatoire. *Annals of Mathematics*, 44:423–453.

Grimshaw, S. (1993). Computing the Maximum Likelihood Estimates for the Generalized Pareto Distribution to Data. *Technometrics*, 35(2):185–191.

Hosking, J. R. M. and Wallis, J. R. (1987). Parameter and quantile estimation for the generalised Pareto distribution. *Technometrics*, 29:339–349.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (minimum) values of meteorological events. *Quarterly Journal of the Royal*

Meteorological Society, 81:158–172.

Jondeau, E. and Rockinger, M. (1999). The tail behavior of stock returns: Emerging versus mature markets. Mimeo, HEC and Banque de France.

Koedijk, K. G., Schafgans, M., and de Vries, C. (1990). The Tail Index of Exchange Rate Returns. *Journal of International Economics*, 29:93–108.

Kuan, C. H. and Webber, N. (1998). Valuing Interest Rate Derivatives Consistent with a Volatility Smile. Working Paper, University of Warwick.

Longin, F. M. (1996). The Assymptotic Distribution of Extreme Stock Market Returns. *Journal of Business*, 69:383–408.

Loretan, M. and Phillips, P. (1994). Testing the covariance stationarity of heavy-tailed time series. *Journal of Empirical Finance*, 1(2):211–248.

McNeil, A. J. (1999). Extreme value theory for risk managers. In *Internal Modelling and CAD II*, pages 93–113. RISK Books.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7(3–4):271–300.

Neftci, S. N. (2000). Value at risk calculations, extreme events, and tail estimation. *Journal of Derivatives*, pages 23–37.

Pickands, J. I. (1975). Statistical inference using extreme value order statistics. *Annals of Statististics*, 3:119–131.

Reiss, R. D. and Thomas, M. (1997). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields.* Birkhäuser Verlag, Basel.

Rootzèn, H. and Klüppelberg, C. (1999). A single number can't hedge against economic catastrophes. *Ambio*, 28(6):550–555. Preprint.

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap.* Springer Series in Statistics. Springer.

Straetmans, S. (1998). Extreme financial returns and their comovements. Ph.D. Thesis, Tinbergen Institute Research Series, Erasmus University Rotterdam.

Tajvidi, N. (1996a). Confidence Intervals and Accuracy Estimation for Heavy-tailed Generalized Pareto Distribution. Thesis article, Chalmers University of Technology. `www.maths.lth.se/matstat/staff/nader/`.

Tajvidi, N. (1996b). Design and Implementation of Statistical Computations for Generalized Pareto Distributions. Technical Report, Chalmers University of Technology. `www.maths.lth.se/matstat/staff/nader/`.

von Mises, R. (1954). La distribution de la plus grande de n valeurs. In *Selected Papers, Volume II*, pages 271–294. American Mathematical Society, Providence, RI.