An Application of Extreme Value Theory for Measuring Risk

Manfred Gilli, Evis Këllezi ⋆

Department of Econometrics, University of Geneva and FAME
CH–1211 Geneva 4, Switzerland

Abstract

Many fields of modern science and engineering have to deal with events which are rare but have significant consequences. Extreme value theory is considered to provide the basis for the statistical modelling of such extremes. The potential of extreme value theory applied to financial problems has only been recognized recently. This paper aims at introducing the fundamentals of extreme value theory as well as practical aspects for estimating and assessing statistical models for tail-related risk measures.

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 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 can we model the rare phenomena that lie outside the range of available observations. In such a situation

⋆ Supported by the Swiss National Science Foundation (projects 12–52481.97 and 1214–056900.99/1). We are grateful to an anonymous referee for corrections and comments and thank Elion Jani and Agim Xhaja for their suggestions.

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

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, more and more research has been undertaken to analyze 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 (1997b), Kuan and Webber (1998), Straetmans (1998), McNeil (1999), Jondeau and Rockinger (1999), Rootzèn and Klüppelberg (1999), Neftci (2000) and McNeil and Frey (2000). 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 assess tail related risk; it thus aims at providing a modelling tool for modern risk management.

Section 2 presents some definitions of common risk measures which provide the general background for practical applications. 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 the observations of thirty-one years of daily returns on an index representing the Swiss market are analyzed. In particular, the loss tail is modelled and point and interval estimates of the tail risk measures presented in Section 2 are computed. Finally, in Section 5, a brief analysis of out-of-sample performance of the model is presented which then suggests that the approach is robust and therefore useful.

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.

VaR Calculation

VaR is generally defined as the risk capital sufficient, in most instances, to cover 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. VaR 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* defined as the inverse of the distribution function F.

For internal risk control purposes, most of the financial firms compute a 5% VaR over a one-day holding period. The Basle accord proposed that VaR for the next 10 days and $p = 1\%$, based on a historical observation period of at least 1 year (220 days) of 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* (ES) or the *tail conditional expectation* which estimates the potential size of the loss exceeding VaR. The expected shortfall is defined as the expected size of a loss that exceeds VaR

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

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.

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 which may represent daily losses or returns. The first approach then considers the maximum (or minimum) 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 these block maxima for four periods with three observations.

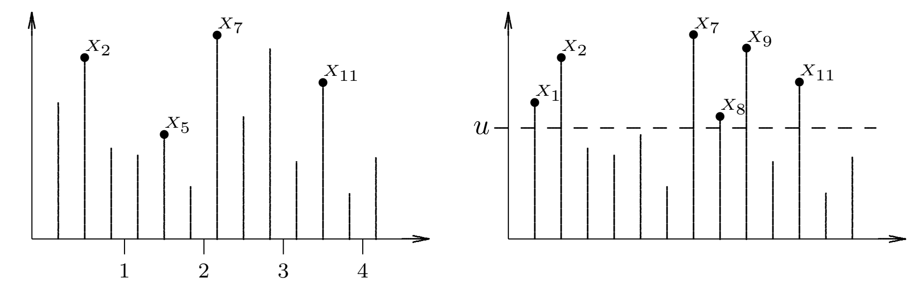


Fig. 1. Block-maxima (left panel) and excesses over a threshold u (right panel).

The second approach focuses on the realizations which exceed 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, threshold methods use data more efficiently and, for that reason, seem to become the choice method 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 (GEV)

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*

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

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.

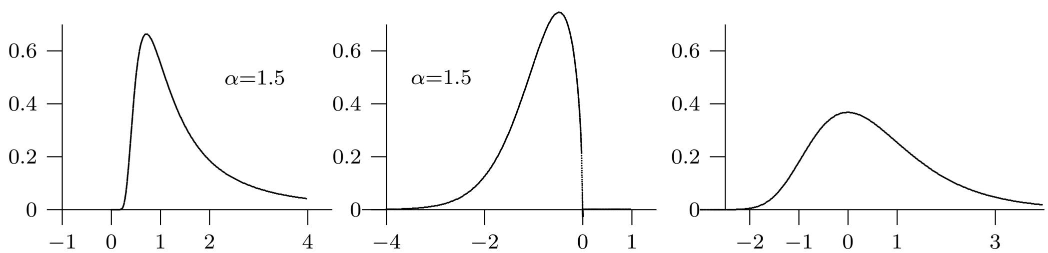


Fig. 2. Densities for the Fréchet, Weibull and Gumbel functions.

Jenkinson and von Mises 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.

The theory in theorem 1 strongly underlies the approach where we consider the distribution of exceedances. This method is also called the *peak over threshold* (POT) method.

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.

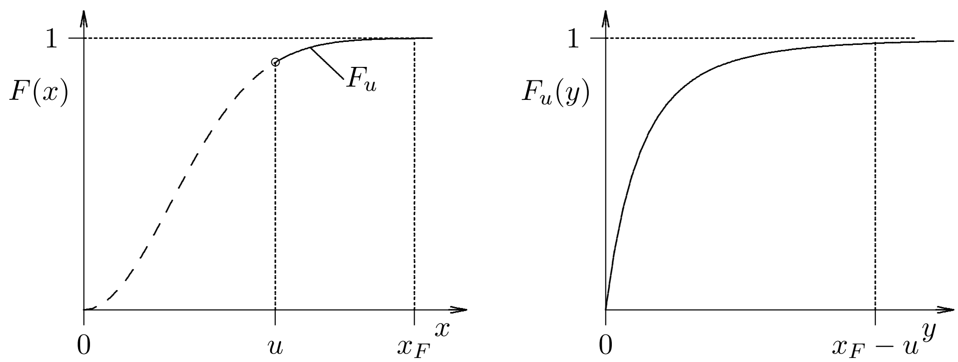


Fig. 3. Distribution function F and conditional distribution function F_u.

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

$$F_u(y) = P(X - u \leq y \mid X > u), \qquad 0 \leq y \leq x_F - u \tag{4}$$

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.

$$F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)} = \frac{F(x) - F(u)}{1 - F(u)}. \tag{5}$$

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 *cedf* 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

$$G_{\xi,\sigma}(y) = \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma} y\right)^{-1/\xi} & \text{if} \quad \xi \neq 0 \\ 1 - e^{-y/\sigma} & \text{if} \quad \xi = 0 \end{cases} \tag{6}$$

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.

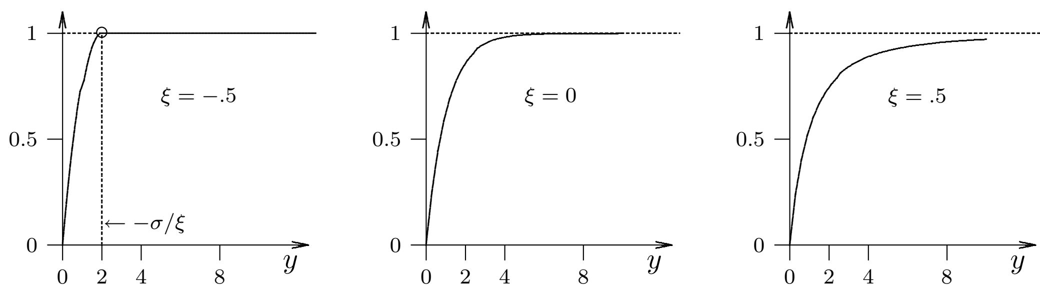


Fig. 4. Shape of the generalized Pareto distribution $G_{\xi,\sigma}$ for $\sigma = 1$.

As, in general, one cannot fix an upper bound for financial losses we see from Figure 4 that only distributions with shape parameter $\xi > 0$ are suited to model fat tailed distributions.

We will now derive analytical expressions for $\texttt{VaR}_\mathrm{p}$ and $\texttt{ES}_\mathrm{p}$ defined respectively in (1) and (2). First we isolate $F(x)$ from (5)

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

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 have

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

which simplifies to

$$F(x) = 1 - \tfrac{N_u}{n}\left(1 + \tfrac{\xi}{\sigma}(x-u)\right)^{-1/\xi}. \tag{7}$$

Inverting (7) for a given probability p gives

$$\mathtt{VaR}_{\mathrm{p}} = u + \tfrac{\sigma}{\xi}\left(\left(\tfrac{n}{N_u}p\right)^{-\xi} - 1\right). \tag{8}$$

Let us now rewrite the expected shortfall as

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

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

$$e(z) = E(X - z \mid X > z) = \frac{\sigma + \xi z}{1-\xi}, \qquad \sigma + \xi z > 0\,. \tag{9}$$

This function gives the average of the excesses of X over varying values of z. A general 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.[1]

Similarly, given the definition (2) for the expected shortfall and using expression (9), for $z = \mathtt{VaR}_{\mathrm{p}} - u$ and X representing the excesses y over u we obtain

$$\mathtt{ES}_{\mathrm{p}} = \mathtt{VaR}_{\mathrm{p}} + \frac{\sigma + \xi(\mathtt{VaR}_{\mathrm{p}} - u)}{1-\xi} = \frac{\mathtt{VaR}_{\mathrm{p}}}{1-\xi} + \frac{\sigma - \xi u}{1-\xi}. \tag{10}$$

4 Modelling the Fat Tails of Stock Returns

Our aim is to model the tail of the distribution of a market index negative movements in order to estimate extreme quantiles. We analyze the daily returns of the Swiss market represented by the "Credit Suisse General" index[2] for the period from 1–01–1969 to 1–12–1999. The application has been executed in a MATLAB 5.x programming environment. The files containing the data and the code can be downloaded from the URL `www.unige.ch/ses/metri/gilli/evtrm/evtrm_mm.html`. Figure 5 shows the plot of the $n = 8065$ observed daily returns.

[1] See Embrechts et al. (1999), page 165.

[2] This index has the longest history among indices representing the Swiss market. Data have been extracted from DataStream.

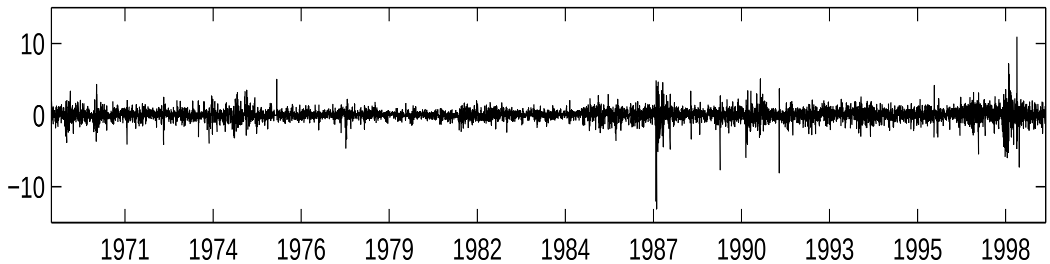


Fig. 5. Daily returns of the Credit Suisse General index.

A first information about the behaviour of the returns can be obtained by standardizing, i.e. centering and reducing them. In Figure 6, we reproduce the histogram of the standardized returns.

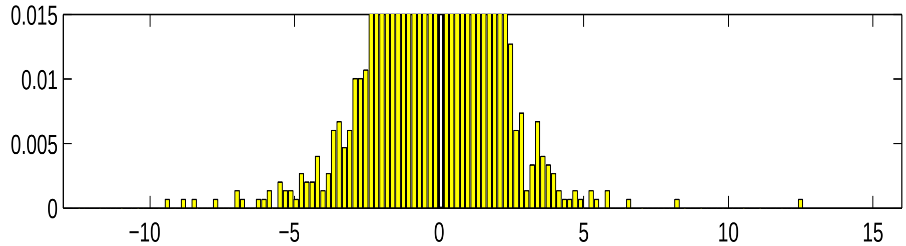


Fig. 6. Tails of the standardized returns (lower part of histogram).

We will now use the theory introduced in the previous chapter to analyze and model the losses. As, for convenience, we want the losses to be in the right tail, in the following we have changed the sign of the returns so that positive values correspond to losses.

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

4.1 The Peak Over Threshold (POT) Method

Despite the appealing theoretical framework EVT provides, small sample issues pose problems when it comes to statistical inference. The main problem is the 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. (1997), Danielsson and de Vries (1997a) and Dupuis (1998) among others. Tools from exploratory data analysis prove helpful in approaching this problem and we will present them together with our application.

Let us consider the *sample distribution function* $\widehat{F}_n(x_i^n)$ which, for a set of n observations, given in increasing order $x_1^n \leq \cdots \leq x_n^n$, is defined as

$$\widehat{F_n}(x_i^n) = \frac{i}{n} \qquad i = 1, \ldots, n. \tag{11}$$

The sample distribution function corresponding to the right tail of our data set is given in Figure 7. Our goal is to estimate the functional form of this portion of the distribution.

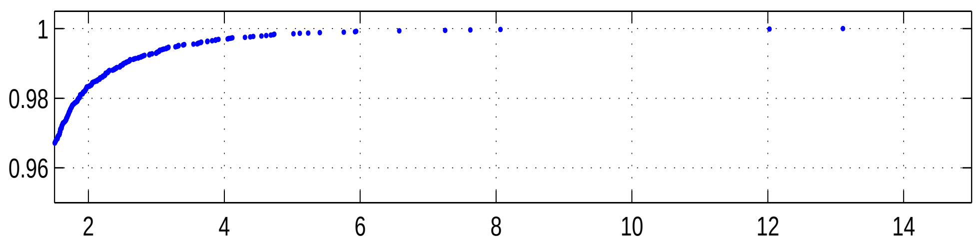


Fig. 7. Right portion of the sample distribution for our data.

Graphical Data Exploration Tools

Quantile plots (*QQ–plots*) can be used to distinguish visually between different distribution functions. Figure 8 shows the sample quantiles plotted against the GPD quantiles. Knowing that for financial data $\alpha = 1/\xi$ takes values in the range [3, 4] (see e.g. (Embrechts et al., 1999, p. 291)) the GPD quantiles are computed with $\xi = 0.3$.

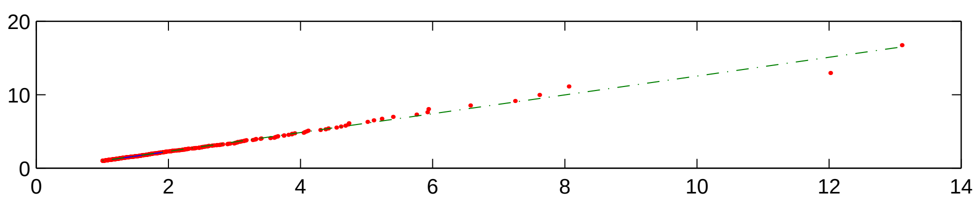


Fig. 8. QQ–plot of sample quantiles exceeding 1 against the corresponding quantiles of GPD distribution.

The picture strongly suggests that the hypotheses that our data follow a GPD distribution is acceptable.

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

$$(u,\ e_n(u)), \qquad x_1^n < u < x_n^n \tag{12}$$

and where the sample mean excess function is defined as

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

where $n-k+1$ is the number of observations exceeding the threshold u. The sample mean excess function is an estimate of the mean excess function $e(u)$. For the GPD it has been defined in (9) and is linear.

Figure 9 shows the sample mean excess plot corresponding to our data. From a closer inspection of the plot in the right panel, which zooms the function for a smaller range of values for u, we suggest trying the value $u = 1.56$ and $u = 3$ for the threshold as they are located at the beginning of a portion of the sample mean excess plot which is roughly linear.

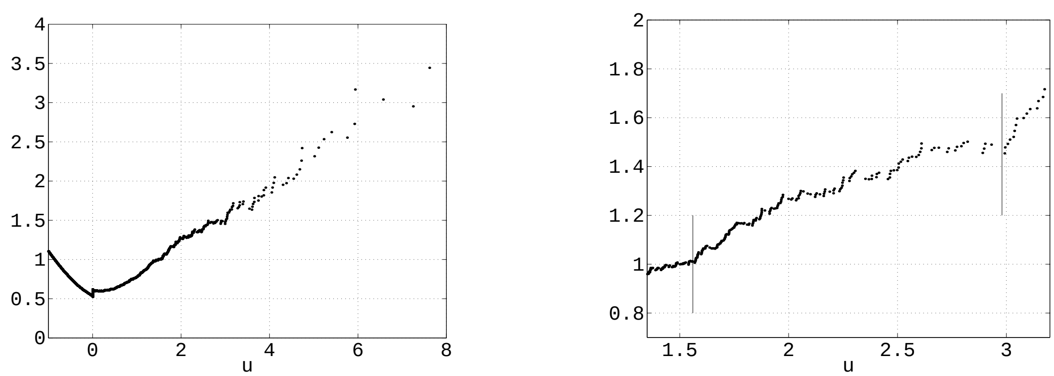


Fig. 9. Sample mean excess plot.

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). This is confirmed by the QQ-plot in Figure 8. Different methods can be used to estimate the parameters of the GPD. [3] We use the maximum likelihood estimation method.

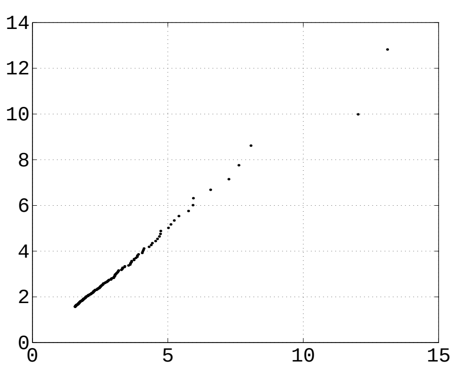
[3] These are the maximum likelihood estimation, the method of moments, the method of probability-weighted moments and the elemental percentile method. For comparisons and detailed discussions about their use for fitting the GPD to data, see Hosking and Wallis (1987), Grimshaw (1993), Tajvidi (1996a),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

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

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

According to our interpretation of the sample mean excess plot, we computed the values $\hat{\xi}$ and $\hat{\sigma}$ which maximize the log-likelihood function for the two samples corresponding to a threshold of $u = 1.56$ and $u = 3$. We obtained the estimates $\hat{\xi} = 0.35$ and $\hat{\sigma} = 0.67$ for $u = 1.56$ and $\hat{\xi} = 0.33$ and $\hat{\sigma} = 0.96$ for $u = 3$. We observe that the shape parameter varies very little between the two values of u and we therefore choose the threshold $u = 1.56$ leaving 247 observations in the tail instead of 56 for $u = 3$.

Again we may use a QQ–plot to visually check whether the data points satisfy the GPD assumption. The left panel in Figure 10 shows the plot of the sample quantiles against $G_{\hat{\xi},\hat{\sigma}}$ quantiles, from where we can conclude that the fit is satisfactory.



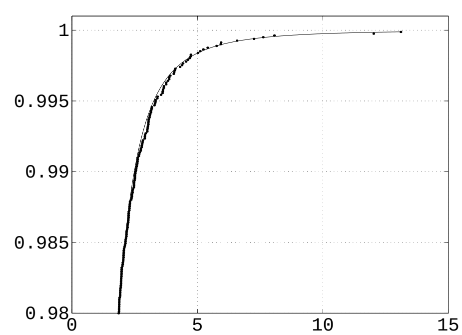


Fig. 10. QQ–plot of sample quantiles against $G_{\hat{\xi},\hat{\sigma}}$ quantiles (left panel) and GPD fitted to the 247 exceedances above the threshold $u = 1.56$ (right panel).

The GPD fitted to the $N_u = 247$ exceedances above the threshold $u = 1.56$ is plotted in the right panel in Figure 10. High quantiles may now be directly read in the plot or computed from equation (8) where we replace the parameters by their estimates

$$\widehat{\texttt{VaR}}_{\mathrm{p}} = u + \frac{\hat{\sigma}}{\hat{\xi}} \left(\left(\frac{n}{N_u} p \right)^{-\hat{\xi}} - 1 \right), \qquad \hat{\xi} \neq 0. \tag{13}$$

For instance, if we choose $p = 0.01$ we can compute $\widehat{\texttt{VaR}}_{0.01} = 2.48$ and $\widehat{\texttt{ES}}_{0.01} = 4.0$.

Confidence Intervals

If we admit that large-sample theory holds for our estimates, we can construct confidence intervals for the parameters ξ and σ by inverting the likelihood ratio test. [4] As only exceedances enter the estimation procedure, the estimates rely on very small data sets. For this reason one cannot always rely on the asymptotic optimality properties of the maximum likelihood estimators. Tajvidi (1996a) investigates the performance of several bootstrap and likelihood-based methods for constructing confidence intervals for the parameters and quantiles of the GPD. His conclusion is that the profile likelihood confidence intervals should be corrected for the small sample size using the methodology of Lawley (1956) and that they are comparable to the bias corrected and accelerated bootstrap confidence intervals [5] in terms of accuracy.

In addition to single confidence intervals on the parameters, we also consider joint confidence intervals. For example, a joint interval for the parameters ξ and σ is given by the contour at level $\chi^2_{\alpha,2}$ of the relative log-likelihood function defined as $L(\xi,\sigma) - L(\hat{\xi},\hat{\sigma})$.

To construct single confidence intervals, we need to compute the profile log-likelihood functions. For example, in the case we need an interval estimate for ξ, the profile log-likelihood function is

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

and the $1-\alpha$ confidence interval is given by values of ξ satisfying

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

where $\chi^2_{\alpha,1}$ is the $1-\alpha$ quantile of the χ^2 distribution with 1 degree of freedom.

Figure 11 shows the single and joint confidence intervals for the estimated parameters.

In order to further explore the reliability of the confidence intervals we applied the bootstrap method to generate 1000 samples. For each sample, we estimated the parameters and plotted the pairs $(\hat{\xi}_i,\, \hat{\sigma}_i)$, $i = 1,\ldots,1000$. These points are plotted in Figure 11 together with the the log-likelihood based confidence intervals. We can verify that about 5% lie outside the 95% joint confidence region computed with the likelihood ratio test. For the marginal distributions we observe that about 13.5% of the estimates of ξ and 18% of the estimates of σ lie outside the maximum likelihood confidence intervals for a single parameter. We can also observe that the density of the bootstrap estimates differs from

[4] See for instance Azzalini (1996) for an introduction.
[5] See Efron and Tibshirani (1993) or Shao and Tu (1995).

the log-likelihood based confidence interval and further investigation about the statistical properties of the estimates would be needed.

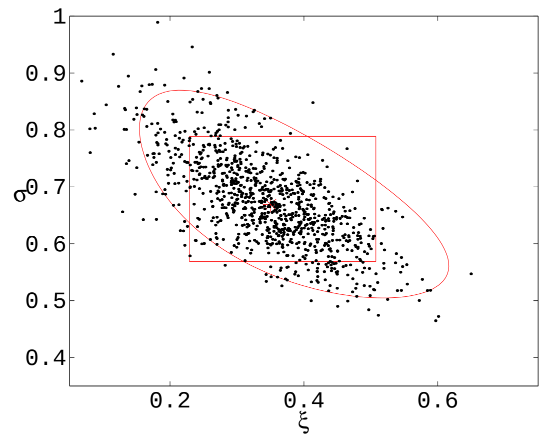


Fig. 11. Single and joint 95% confidence intervals for ξ and σ for the POT method. Dots represent 1000 bootstrap estimates.

The empirical marginal distributions of the bootstrap values for ξ and σ as well as the corresponding empirical distribution of $\mathtt{VaR}_{0.01}$ and $\mathtt{ES}_{0.01}$ are reproduced in the plots in Figure 12.

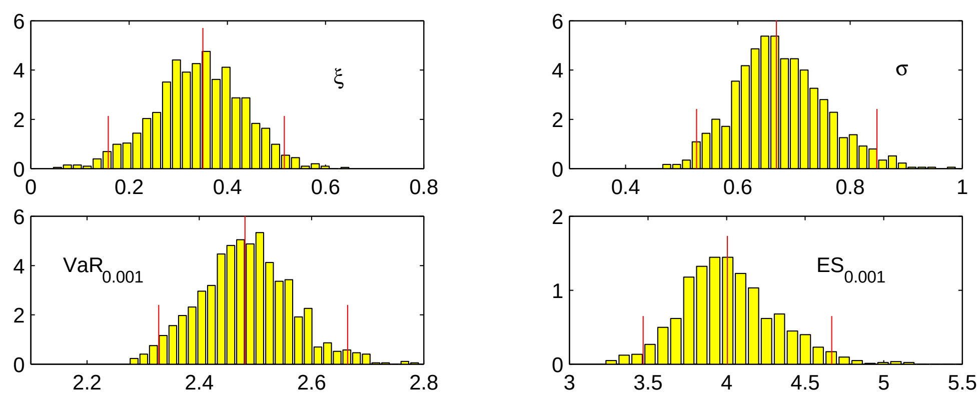


Fig. 12. Empirical bootstrap marginal distributions for $\hat{\xi}$ (upper left), $\hat{\sigma}$ (upper right), $\mathtt{VaR}_{0.01}$ (lower left) and $\mathtt{ES}_{0.01}$ (lower right) for the POT method.

Table 1 summarizes the point estimates, the maximum likelihood (ML) and the bootstrap (BS) confidence intervals of the marginal distributions.

Table 1
Point estimates and 95% maximum likelihood (ML) and bootstrap (BS) confidence intervals for the POT method.

	Lower bound		Point estimate	Upper bound	
	BS	ML	ML	ML	BS
$\hat{\xi}$	0.16	0.23	0.35	0.51	0.52
$\hat{\sigma}$	0.52	0.57	0.67	0.79	0.85
$\widehat{\mathtt{VaR}}_{0.01}$	2.33	2.34	2.48	2.65	2.66
$\widehat{\mathtt{ES}}_{0.01}$	3.47	3.53	4.00	5.03	4.67

The results in Table 1 indicate that with probability 0.01 the tomorrow's loss will exceed the value 2.48% and that the corresponding expected loss, that is the average loss in situations where the losses exceed 2.48%, is 4.00%. These point estimates are completed with 95% confidence intervals. Thus the expected loss will, in 95 out of 100 cases, lie between 3.47% and 4.67%.

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. This guarantees that the estimated expected shortfall, which is a conditional first moment, exists.

Log-likelihood based confidence intervals for $\mathtt{VaR}_\mathrm{p}$ can be obtained by using a reparameterized version of GPD defined as a function of ξ and $\mathtt{VaR}_\mathrm{p}$:

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

The corresponding probability density function is

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

Figure 13 these likelihood based confidence regions obtained by using this reparameterized version of GPD.

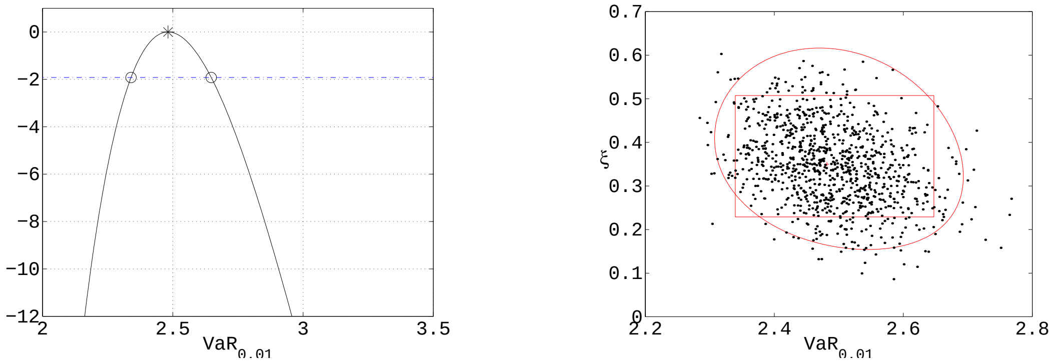


Fig. 13. Left panel: Relative profile log-likelihood function and confidence interval for $\mathtt{VaR}_{\mathrm{p}}$. Right panel: Single and joint confidence intervals at level 95% for $\hat{\xi}$ and $\mathtt{VaR}_{\mathrm{p}}$. Dots represent 1000 bootstrap estimates for the parameters of the POT method.

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

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

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

we compute a log-likelihood based confidence interval for the expected shortfall $\mathtt{ES}_{\mathrm{p}}$. In Figure 14 we see that the log-likelihood confidence interval for $\mathtt{ES}_{\mathrm{p}}$ is not symmetric with respect to $\widehat{\mathtt{ES}}_{0.01}$.

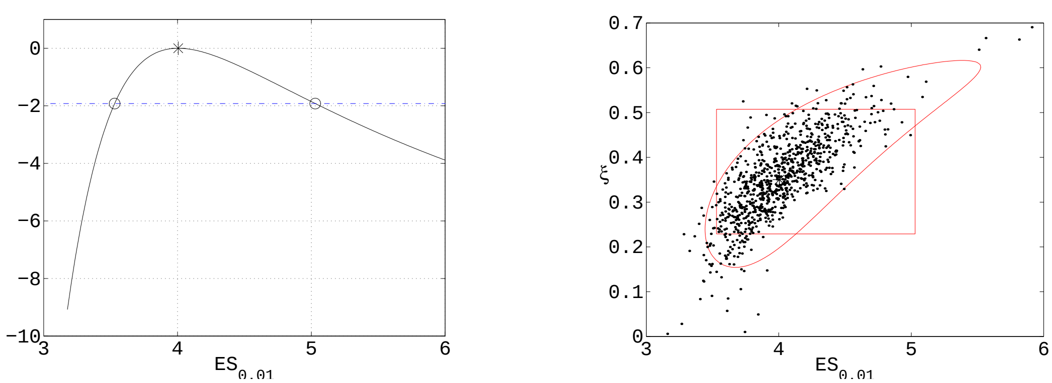


Fig. 14. Left panel: Relative profile log-likelihood function and confidence interval for $\mathtt{ES}_{\mathrm{p}}$. Right panel: Single and joint confidence intervals at level 95% for $\hat{\xi}$ and $\mathtt{ES}_{\mathrm{p}}$. Dots represent 1000 bootstrap estimates for the parameters of the POT method.

We observe that in both Figures 13 and 14 the number of bootstrap estimates lying outside the 95% likelihood joint confidence intervals is 49 out of 1000.

We now apply the block maxima method to our daily return data. For this method the delicate point 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.

Thus our sample has been divided into 31 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 absolute value of the minimum return in each of the blocks constitute the data points in the sample of minima M which is used to estimate the generalized extreme value distribution (GEV). Figure 15 plots the yearly minima and maxima of our daily returns.

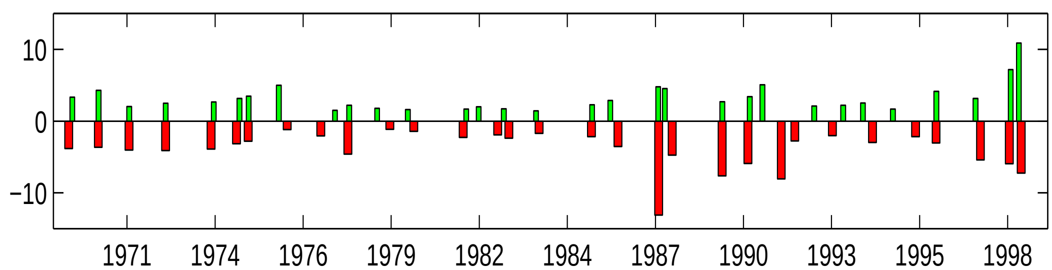


Fig. 15. Yearly minima and maxima of the daily returns of the Credit Suisse General index.

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

$$H_{\xi,\sigma,\mu}(x) = H_\xi\left(\frac{x-\mu}{\sigma}\right) \quad x \in \mathcal{D}, \quad \mathcal{D} = \begin{cases}]-\infty,\ \mu - \frac{\sigma}{\xi}[& \xi < 0 \\ ]-\infty,\ \infty[& \xi = 0 \\ ]\mu - \frac{\sigma}{\xi}\ \infty[& \xi > 0 \end{cases} \tag{14}$$

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

$$L(\xi,\mu,\sigma;x) = \prod_i \log(h(x_i)), \qquad x_i \in M \tag{15}$$

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

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

The log-likelihood estimates we obtain are $\hat{\xi} = 0.28$, $\hat{\sigma} = 1.37$ and $\hat{\mu} = 2.65$. In Figure 16, we give the plot of the sample distribution and the corresponding fitted GEV distribution.

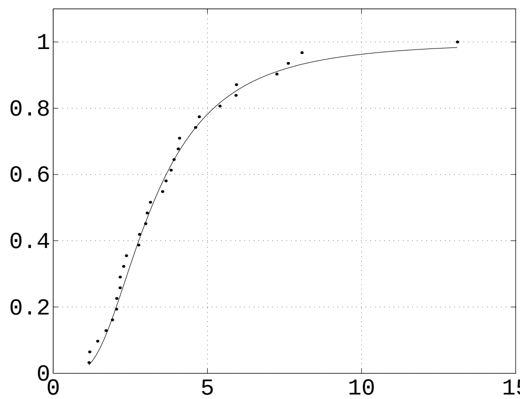


Fig. 16. Sample distribution of yearly maxima and corresponding fitted GEV distribution.

In practice the quantities of interest are not the parameters themselves, but the quantiles, also called return levels, of the estimated GEV. The return level R^k is the level we expect to be exceeded in one out of k one year periods:

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

Substituting the parameters ξ, σ and μ by their estimates we get

$$\hat{R}^k = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}\left(1-\left(-\log(1-\frac{1}{k})\right)^{-\hat{\xi}}\right) & \xi \neq 0 \\ \hat{\mu} - \hat{\sigma}\log\left(-\log(1-\frac{1}{k})\right) & \xi = 0 \end{cases}. \tag{16}$$

Taking for example $k = 10$, we obtain for our data $\hat{R}^{10} = 6.94$, which means that the maximum loss observed during a period of one year will exceed 6.94% in one out of ten years on average.

Confidence Intervals

As already discussed in the case of the GPD estimation, it may prove 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 $\hat{\mu}$ from equation (16) and substitute it into $H_{\xi,\sigma,\mu}$ defined in (14). The GEV distribution function then becomes

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

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 and is plotted in the left panel of Figure 17. For $\alpha = 0.05$ the interval estimate for R^{10} is [5.29, 11.63].

In the case where a joint confidence region on two parameters, say ξ and R^k is needed, the profile log-likelihood function, which in this case is a surface, is defined as

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

In this case the confidence region is 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 17, we reproduce the confidence regions at level 95% for the parameters $\hat{\xi}$ and $\hat{R}^k$. We also generated 1000 bootstrap samples and computed the bootstrap confidence intervals for ξ, μ, σ and R^{10}. The pairs (R^{10}, ξ) are plotted in the right panel of Figure 17.

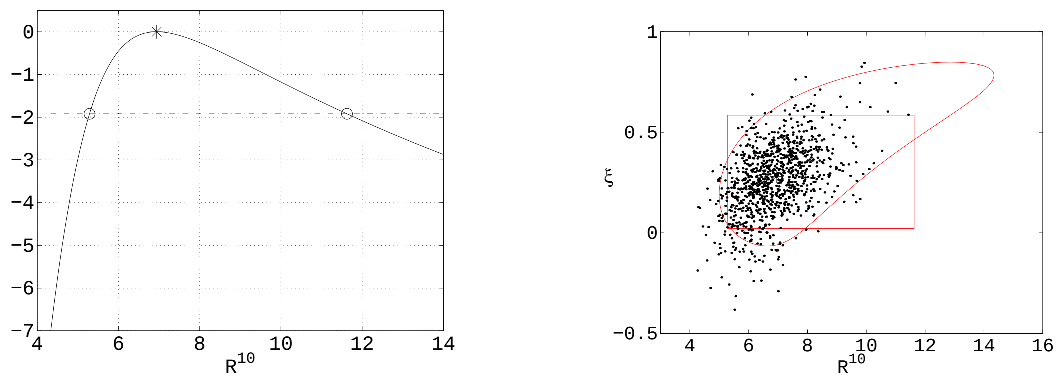


Fig. 17. Left panel: Relative profile log-likelihood and 95% confidence interval for $\hat{R}^{10}$ estimated with the method of block maxima. Right panel: Joint confidence region for $\hat{\xi}$ and $\hat{R}^{10}$ at level 95%. In both panels the maximum likelihood estimates are marked with the symbol $*$.

The empirical marginal distributions of the bootstrap values for ξ, σ, μ and R^{10} are reproduced in the plots in Figure 18.

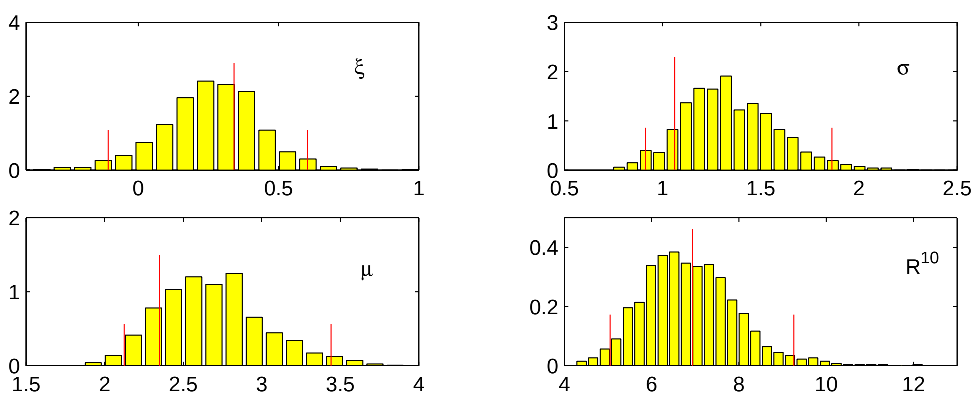


Fig. 18. Empirical marginal distributions of the bootstrap estimates.

Table 2 summarizes the point estimates, the maximum likelihood (ML) and the bootstrap (BS) confidence intervals for the reparameterized GEV distribution.

Table 2
Point estimates and 95% maximum likelihood (ML) and bootstrap (BS) confidence intervals.

	Lower bound		Point estimate	Upper bound	
	BS	ML	ML	ML	BS
$\hat{\xi}$	-.11	0.02	0.28	0.58	0.60
$\hat{\sigma}$	0.91	1.09	1.37	1.84	1.86
$\hat{R^{10}}$	5.05	5.29	6.94	11.63	9.26

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. In order to assess the reliability of the methods presented in the previous sections, we did some out-of-sample analysis by repeating all computations for a subsample of the data covering the period from 1969 to 1989.

For the POT method, Table 3 reproduces the point estimates and the bootstrap confidence intervals obtained from the 5477 first observations defining the subsample.

Table 3
Point estimates for the POT method and 95% maximum likelihood (ML) and bootstrap (BS) confidence intervals corresponding to the period 1969–1989.

	Lower bound		Point estimate	Upper bound	
	BS	ML	ML	ML	BS
$\hat{\xi}$	0.12	0.24	0.41	0.65	0.68
$\hat{\sigma}$	0.44	0.48	0.60	0.77	0.83
$\widehat{\mathtt{VaR}}_{0.01}$	2.02	2.03	2.17	2.33	2.34
$\widehat{\mathtt{ES}}_{0.01}$	3.01	3.07	3.60	5.41	4.78

We observe that the estimated values differ very little from values reported in Table 1 which correspond to the estimates obtained from the whole sample.

We also compared `VaR` estimated with the POT method with the `VaR` proposed by the Basle accord. Assuming the normal distribution for the observations until 1989, the 1% lower quantile is 1.95. Multiplying this value by 3 gives 5.86, whereas in our calculation the upper bound for the expected shortfall is 4.78. Clearly the POT method provides more accurate information.

As far as the estimation of GEV is concerned, the estimated parameters over the 21 first yearly maximum losses are $\hat{\xi} = 0.33$, $\hat{\sigma} = 1.2$ and $\hat{\mu} = 2.36$. The corresponding return level R^{10} is 6.4, which means that yearly maximum losses will exceed the value 6.4% once in ten years on average. In Figure 19, we verify that in the out-of-sample period this value is exceeded twice, which is not in contradiction with our model.

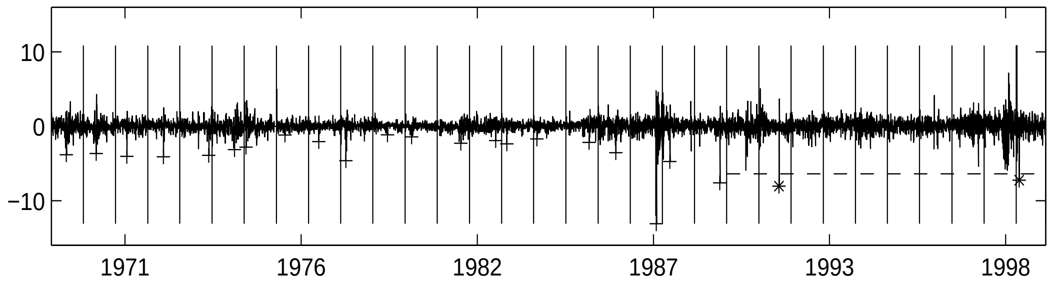


Fig. 19. Out of sample comparisons of the return level R^{10} (horizontal dashed line). Observed minima exceeding R^{10} are marked with a star. GEV distribution is estimated over yearly minima from 1969 to 1989 (marked with +).

Finally, it might be interesting to show how the model allows extrapolation beyond the sample. Using the entire sample, we computed $R^{100} = 15.4\%$, i.e. the level we expect to be exceeded only in one year every century. This value varies only insignificantly whether the full or the subsample is considered. The models also tells us that the 1987 crash is likely to happen once in 60 years. This is obtained by computing $H(13.1)$, where H is the GEV distribution with parameters $\hat{\xi} = 0.33$, $\hat{\sigma} = 1.2$ and $\hat{\mu} = 2.36$.

From the POT model, we conclude that the "once a century loss" of 15.4% predicted by GEV occurs once every 60 years. We obtain this result by computing the probability $p = 1 - F(15.4)$ where F is the estimated GPD. Such an event will happen on average every $1/p$ days which in our case gives 60 years. Similarly using the POT method we can compute that the 1987 crash loss is likely to be exceeded every 37 years. We should mention that the events forecasted by the POT and block maxima method are not exactly the same, which in part explains the difference concerning the results of 60 and 100 years.

References

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.

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., de Haan, L., Peng, L., and de Vries, C. (1997). Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation. Mimeo.

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

Danielsson, J. and de Vries, C. (1997b). Value-at-Risk and Extreme Returns. Working Paper.

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.

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.

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.

Lawley, D. N. (1956). A General Method for Approximating to the Distribution of Likeligood Ratio Criteria. *Biometrika*, 43:295–303.

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/`.