# S Plus
# For Financial Engineers [1]
-
# PART III
-
# Beyond the Sample:
# Dealing with Extreme Values

Diethelm Würtz

Institut für Theoretische Physik
ETH Zürich

June 5, 2002

[1]This script collects material used in my lecture *Econophysics* held in SS 2002 at the "Insitute of Theoretical Physics" of ETH Zürich. These notes are thought for internal use only, please do not distribute! The software used in the script comes without any warranty!

.

# Overview

**Chapter 1 - Markets, Basic Statistics, Date and Time**
1.1 Introduction
1.2 Economic and Financial Markets
1.3 Brief Repetition from Probability Theory
1.4 Distribution Functions in Finance
1.5 Searching for Correlations and Dependencies
1.6 Hypothesis Testing
1.7 Calculating and Managing Calendar Dates

**Chapter 2 - The Dynamical Process Behind Financial Markets**
2.1 Introduction
2.2 ARIMA Modelling: Basic Concepts of Linear Processes
2.3 GARCH Modelling: Mastering Heteroskedastic Processes
2.4 Regression Modelling from the Time Series Point of View
2.5 Neural Networks: Feedforward Connectionist Networks
2.6 Design and Implementation of an Intra-Daily Trading System

**Chapter 3 - Beyond the Sample: Dealing With Extreme Values**
3.1 Introduction
3.2 Fluctuations of Maxima
3.3 Extremes of Point Processes
3.4 The Extremal Index

**Chapter 4 - Pricing and Hedging of Options**
4.1 Introduction
4.2 The Basics of Option Pricing
4.3 Pricing Formulas for Exotic Options
4.4 Heston Nandi Option Pricing
4.5 MC Simulation of Path Dependent Options

**Appendix**
An Introduction to Splus

.

# Contents

.

# Chapter 3

# Beyond the Sample: Dealing with Extreme Values

## 3.1 Introduction

From descriptive statistics of financial time series data in Chapter 1 and from the investigation of the dynamical process in Chapter 2 we have seen that the central point in the modelling of financial market data lies in a proper management of the extreme data values appearing almost in every financial market data set.

Two aspects are of major interest: (i) modelling the fat tails of heavy-tailed distributions and (ii) investigating the dependencies of extreme values. In this Chapter we introduce the `fExtremes` Library and describe the implemented algorithms together with the underlying concepts.

Section 2 is dedicated to the exploratory data analysis of extremes, EDA. We introduce probability and quantile plots and the mean excess function. We also consider records and the ratio of maximum and sum as exploratory tools.

Section 3 deals with fluctuations of maxima and the generalized extremal value distribution, GEV. We derive the limit law for maxima and discuss the maximum domain of attraction problem, MDA. Heavy tail estimation is the central application of this section. For Fréchet type of distributions ww estimate tail parameters from L-moments, under MDA conditions, and via the maximum log likelihood approach.

Section 4 is concerned with Point Processes. The point process is characterized and the generalized Pareto distribution derived, GPD.

Temporal dependencies and the Extremal Index are investigate in Section 5.

.

## 3.2 Exploratory Data Analysis of Extremes

### Introduction

Exploratory Data Analysis (EDA) is an approach for data analysis which employs a variety of techniques, mostly graphical, to maximize insight into a data set, and to uncover underlying structures. We use it here to explore and to analyze extreme values in data sets collected from financial markets.

The philosophy underlying EDA differs from the usual statistical graphics approaches; EDA encompasses a much larger venue. EDA describes a concept which postpones the usual assumptions about what kind of model the data follow with the more direct approach of allowing the data itself to reveal its underlying structure and model. Thus EDA is a kind philosophy as to how we disect a data set, what we look for, how we look, and how we interpret.

Usually EDA techniques are graphical in nature. The reason for the heavy reliance on graphics is that by its nature the main role of EDA is to open-mindedly explore, and graphics gives the analysts unparalleled power in terms of such open-mindedly exploration, enticing the data to reveal its structural secrets, and being ever-ready to gain some new, often unsuspected, insight into the data.

The graphical methods we discuss in the following to explore extreme values include probability and quantile plots, the mean excess function, records, and the ratio of maximum and sum.

### 3.2.1 Probability and Quantile Plots

Usually one starts with the discussion of the problem:

*Find a distribution function $F$ which is a good model for the iid data $X$, $X_1, ..., X_n$.*

The theoretical basis that underlies probability plots is a quantile transformation, which implies that for $F$ continuous, the random variable $U_i = F(X_i)$, for $i = 1, ..., n$, are *iid* uniform on (0,1). Moreover, defining the ordered sample $X_{n,n} \leq ... \leq X_{1,n}$,

$$(F(X_{k,n}))_{k=1,...,n} \;=\; (U_{k,n})_{k=1,...,n} \;. \tag{3.1}$$

From this it follows that

$$EF(X_{k,n}) \;=\; \frac{n-k+1}{n+1}\;,\quad k=1,\ldots,n\;. \tag{3.2}$$

Also note that $F_n(X_{k,n}) = (n-k+1)/n$, where $F_n$ stands for the empirical distribution function of $F$. The graph

$$\left\{(F(X_{k,n}), \frac{n-k+1}{n+1} : k=1,\ldots,n\right\} \tag{3.3}$$

is called the *probability plot*, PP-Plot. More common however is to plot the graph

$$\left\{(X_{k,n}, F^{\leftarrow}(\frac{n-k+1}{n+1})) : k=1,\ldots,n\right\}\;, \tag{3.4}$$

typically referred to as the *quantile plot*, QQ-Plot, where $F^{\leftarrow}$ denotes the quantile function of the distribution function $F$.

What are the main merits of QQ-Plots? They stem from the following properties:

- *Comparison of Distribution:* If the data were generated from a random sample of the reference distribution, the plot should look roughly linear. This remains true if the data come from a linear transformation of the distribution.
- *Outliers:* If one or a few of the data values contaminated by gross error or for any reason are markedly different in value from remaining values, the latter being more or less distributed like the reference distribution, the outlying points may easily identified on the plot.
- *Location and Scale:* Because a change of one of the distributions by a linear transformation simply transforms the plot by the same transformation, one may estimate graphically (through the intercept and slope) location and scale parameters for a sample of data, on the assumption that the data come from the reference distribution.
- *Shape:* Some difference in distributional shape may be deduced from the plot. For example if the reference distribution has heavier tails (tends to have more large values) the plot will curve down at the left and/or up to the right.

In the following we like to address these observations by some examples. In finance it is well known that usually the distributions of logarithmic returns behave leptokurtic, i.e. the distributions exhibit in comparison to a normal distribution function fatter tails and sharper peaks. This lepotkurtic behavior is clearly demonstrated in the following example where we show the QQ-Plot of the logarithmic returns of the NYSE Composite Index and of the BMW share prices.

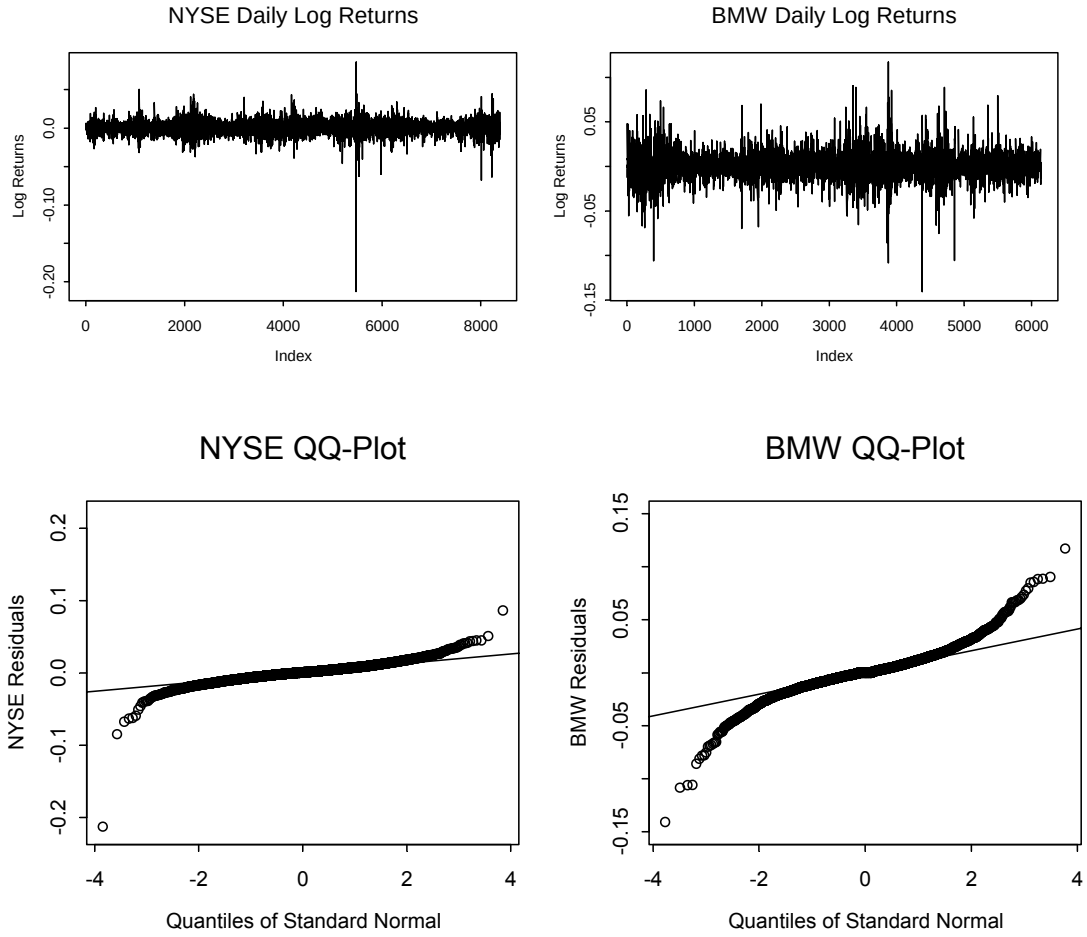
**Example: Quantile Plots -** `xmpEDAqqplot1`

Plot the time series and create QQ-plots of the logarithmic returns of the NYSE composite index and of the BMW share prices versus the normal distribution function.

Use the standard Splus functions `plot()`, `qqnorm()` and `qqline()` to produce the time series graphs and the QQ-Plots for the normal (Gaussian) distribution function:

```
> x <- scan("nyseres.csv")
> plot(x, type="l", ylab="Log Returns", main="NYSE Daily Log Returns")
> qqnorm(x, ylab="NYSE Residuals", main="NYSE QQ-Plot")
> qqline(x)
```

The last function draws a straight line through the normal QQ-Plots. Note, that a vector may be returned containing the intercept and slope of the line.




■ Figure 3.2.1: The upper two graphs show the daily logarithmic returns of the NYSE composite index (left) and the of the BMW share prices (right). In the lower row, the graph on the left side plots the daily logarithmic returns calculated from the NYSE composite index versus the normal distribution function. The graph to the right plots the same for logarithmic returns calculated from daily BMW stock prices.

Since the empirical distributions are far away from normal, the plotted points deviate significantly from the straight line. The extreme points have more variability than points toward the center. The typical "S" shape of the curve implies that one distribution has longer tails than the other. In the default configuration a plot from `qqnorm()` that is bent down on the left and bent up on the right means that the data have longer tails than the Gaussian.

### 3.2.2 The Mean Excess Function

The mean access function is a graphical tool, in particular for discrimination in the tails. Let $X$ be a random variable with right endpoint $x_F$, then

$$e(u) = E(X - u|X > u) \ , \quad 0 \leq u < x_F \ , \tag{3.5}$$

where $e(u)$ is called the mean excess function of $X$.

The quantity $e(u)$ is often referred to as the *mean excess over the threshold u*. In an insurance context, $e(u)$ can be interpreted as the expected claim size in the unlimited layer, over priority $u$. Here $e(u)$ is also called the *mean excess loss* function. In a reliability or medical context, $e(u)$ is referred to as the *mean residual life* function. In a financial risk management context, switching from the right tail to the left tail, $e(u)$ is referred to as the *shortfall*.

A graphical test for tail behavior can be based on the *empirical mean excess function* $e_n(u)$. Suppose that $X_1, \ldots, X_n$ are *iid* with distribution function $F$ and let $F_n$ denote the empirical distribution function and $\Delta_n(u) = \{i : i = 1, \ldots, n, X_i > 0\}$, then

$$e_n(u) = \frac{1}{\overline{F}_n(u)} \int_u^\infty \overline{F}(y)dy = \frac{1}{card\Delta_n(u)} \sum_{i \in \Delta_n(u)} (X_i - u) \ , \quad u \geq 0 \ , \tag{3.6}$$

with the convention that 0/0=0. A mean excess plot, ME-Plot, then consists of the graph

$$\{(X_{k,n} \ , \ e_n(X_{k,n})) : k = 1, \ldots, n\} \ . \tag{3.7}$$

**Example: Mean Excess Function Plot - `xmpEDAmxf`**

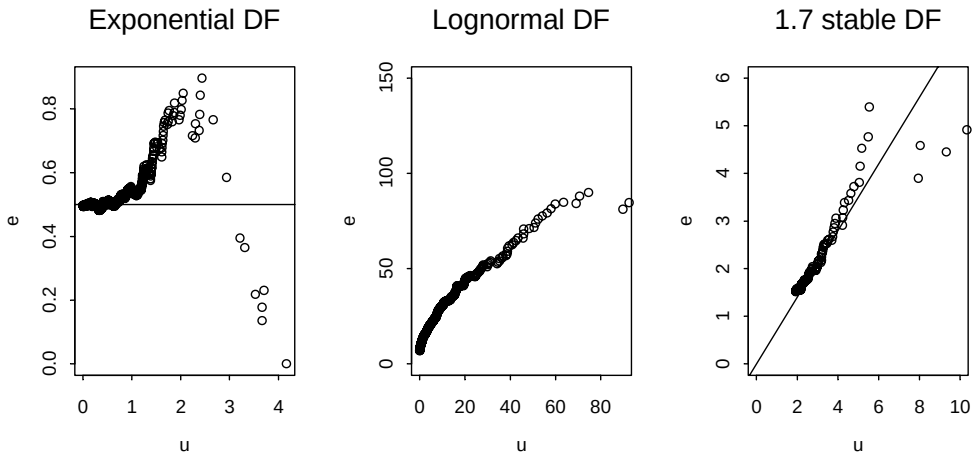
Compare the empirical mean excess function $e_n(u)$ of simulated data ($n = 1000$) with the corresponding theoretical mean excess function $e(u)$ for (i) standard exponential, (ii) lognormal with $lnX$ $N(0, 4)$, and (iii) alfa-stable distribution function with tail index 1.7.

Use the `fExtremes` library function `mxf.plot(data, tail=0.05, ...)` which simply evaluates $e(u)$ from:

```
u <- rev(sort(data))
n <- length(data)
e <- (cumsum(u)-(1:n)*u)/(1:n)
plot(u, e)
```

The variable `tail` lets restrict the data to a predefined tail depth, by default 5%:

```
> set.seed(7138)
> mxf.plot(rexp(1000, rate=2), tail=1, main="Exponential DF")
> abline(0.5,0)

> set.seed(6952)
> mxf.plot(rlnorm(1000, meanlog=0, sdlog=2), tail=1, main="Lognormal DF")

> set.seed(9835)
> mxf.plot(rstab(1000,index=1.7), tail=0.1, main="1.7 stable DF")
> abline(0,0.7)
```




■ Figure 3.2.2: The left figure shows the empirical excess function $e_n(u)$ for 1000 random variables from an exponential distribution function. The horizontal line gives the analytical result. The figure in the middle shows the same plot for lognormal distributed random variables. In the right figure the random variables are drawn from an $\alpha$-stable distribution function with index 1.7. We expect that $\exp(\lambda)$ distributed data follow a straight line with slope 0 and intercept $1/\lambda$ as shown in the upper graph.

Coles (1999) has written an Splus function, named mean residual life plot, which also allows to calculate confidence intervals. With this function we want to investigate 6 different small samples drawn from an $\alpha$-stable distribution function. The plots show very impressively how different the plots may appear. Due to the sparness of the data for calculating $e_n(u)$ for large values of $u$, the resulting plots are extremely sensitive to differences in the data towards the end of the range.

**Example: Mean Residual Life Plot -** `xmpEDAmxf2`

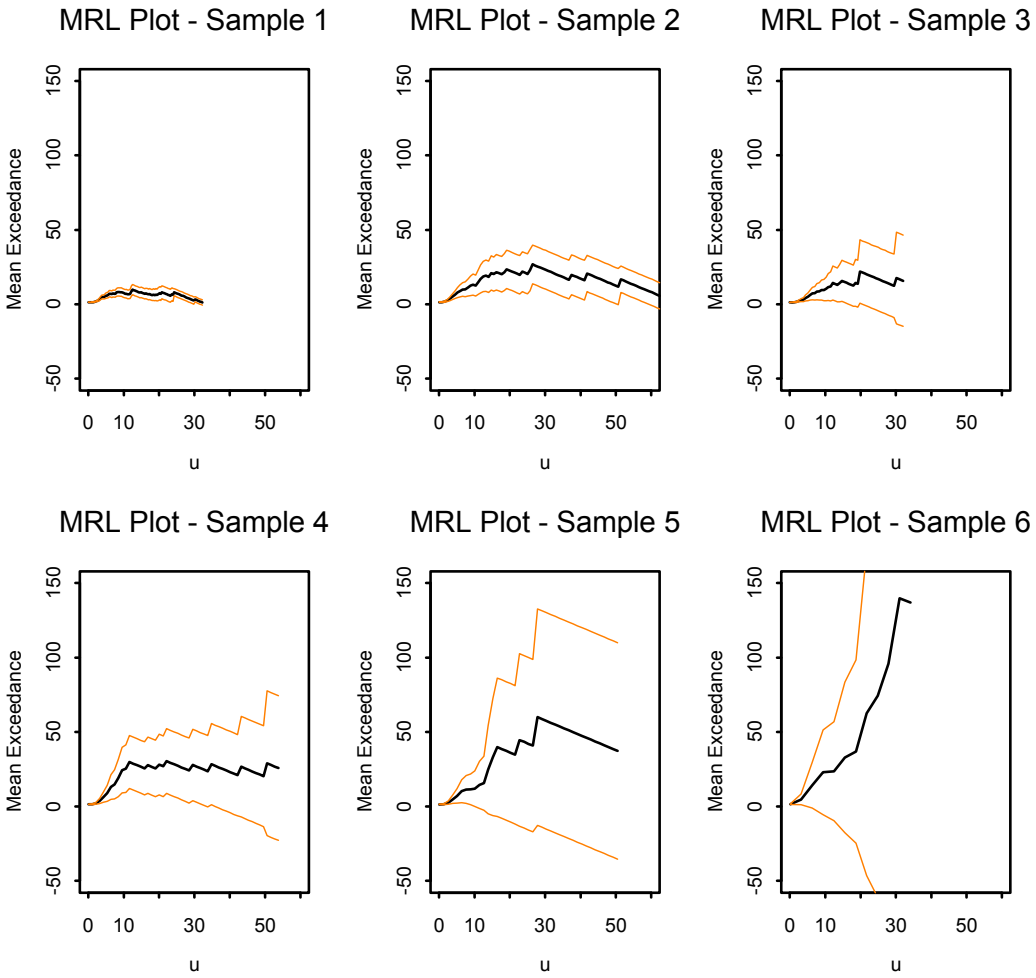
Compare on equal plotting scales the graphs (including 95% confidence intervals) of the empirical mean excess functions $e_n(u)$ for 6 different samples each of 5000 $\alpha$-stable distributed random variables with index $\alpha = 1.7$.

Use the Splus function `rstab(n=5000, index=1.7)` to generate the random variables. Select samples by changing the seed: `set.seed(n)`. Then use the `fExtremes` library function `mrl.plot(data, umin=mean(data), umax=max(data), conf=0.95, nint=100, ...)` which evaluates $e(u)$ from:

```
u <- seq(umin, umax, length=nint)
for(i in 1:nint) {
  data <- data[data > u[i]]
  n <- length(data)
  sdev <- sqrt(var(data))
  e[i] <- mean(data - u[i])
  xu[i] <- e[i] + (qnorm((1 + conf)/2) * sdev)/sqrt(n)
  xl[i] <- e[i] - (qnorm((1 + conf)/2) * sdev)/sqrt(n) }
plot(u, e)
lines(u[xl], xl); lines(u[xu], xu)
```

The variables `umin` and `umax` limit the range of data to be used (by default the minimal and maximal data values), `conf` denotes the size of the confidence level (by default 95%), and `nint` denotes the number of intervals used for the plot, by default 100, e.g.

```
> set.seed(seed(933))
> mrl.plot(rstab(n=5000, index=1.7), nint=100)
```




■ Figure 3.2.3: The graphs show the mean residual life plots for six different samples of $\alpha$-stable distributed random variables each of size 5000 with index 1.7.

### 3.2.3 Records as an Exploratory Tool

Suppose that the random variable $X_i$ are *iid* with distribution function $F$. A record $X_n$ occurs if $X_n > M_{n-1} = \max(X_1, \ldots, X_{n-1})$. By definition we take $X_1$ as a record. The record time $L_n$ are the random times at which the process $(M_n)$ jumps. Define the *record counting process* as

$$N_1 = 1 \ , \quad N_n = 1 + \sum_{k=2}^{n} I_{X_k > M_{k-1}} \ , \quad n \geq 2 \ . \tag{3.8}$$

The result on the expected mean $E(N_n)$ of $N_n$ may be surprising:

$$EN_n = \sum_{k=1}^{n} \frac{1}{k} \quad \textit{and} \quad \operatorname{var}(N_n) = \sum_{k=1}^{n} \left( \frac{1}{k} - \frac{1}{k^2} \right) \ . \tag{3.9}$$

Notice that $E(N_n)$ and $\mathrm{var}(N_n)$ are both of the order $\ln n$ as $n \to \infty$. More precisely $EN_n - \ln n \to \gamma$, where $\gamma = 0.5772\ldots$ denotes Euler's constant. As a consequence: *The number of records of iid data grows very slowly!*

**Example: Plot of Records -** `xmpEDArecords`

Plot the records, both on a linear and logarithmic scale for the daily BMW stock returns (the first 6000 data points partitioned in 6 subsamples) and the daily NYSE Composite Index (the first 8000 data points partitioned in 8 subsamples) returns. Compare the data with 8 iid random variables sets each of 1000 data points drawn from $\alpha$-stable distributions with index 1.7.

Use the `fExtremes` library function `records(x, cuts=1, plottype="lin", ...)` which evaluates the records from the following procedure:

```
u <- x[1]
v <- 1
x.records <- v
while (!is.na(v)) {
  u <- x[x>u][1]
  v <- y[x>u][1]
  if(!is.na(v)) x.records <- c(x.records,v) }
```

The variable `cuts` allows to partition the whole sample into subsamples. The default value is 1 taking all data records from the dataset. The variable `plottype="lin"` permits to plot on a linear scale, alternatively `"log"` creates a logarithmic plot. The graphical output from

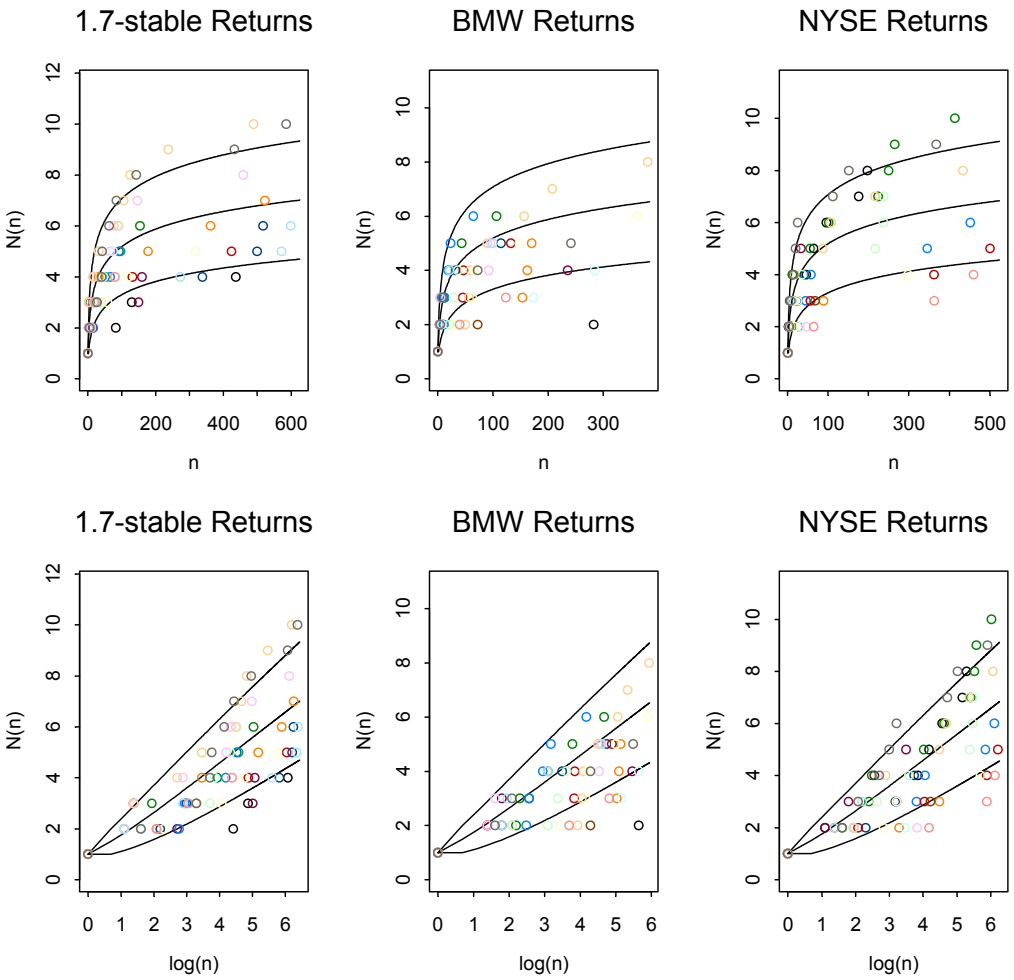
```
> set.seed(246)
> x1 <- rstab(n=8000, index=1.7)x1 <- scan(file="bmwres.csv")
> x2 <- scan(file="bmwres.csv")
> x3 <- scan(file="nyseres.csv")

> records.plot(x1, cuts=8, plottype="log", main="1.7-stable Returns")
> records.plot(x2, cuts=6, plottype="log", main="BMW Returns")
> records.plot(x3, cuts=8, plottype="log", main="NYSE Returns")

> records.plot(x1, cuts=8, plottype="lin", main="1.7-stable Returns")
> records.plot(x2, cuts=6, plottype="lin", main="BMW Returns")
> records.plot(x3, cuts=8, plottype="lin", main="NYSE Returns")
```

is shown in figure 3.2.4. Beside the records which are drawn as circles, both the expectation value and the $\pm 1$ standard deviations are drawn by straight lines. On a logarithmic scale these lines become straight lines in the limit $n \to \infty$.

The function returns as output for each subsample the number of records, e.g. in the case of the $\alpha$-stable distribution:

```
[1] 1  3
[1] 2 10
[1] 3  5
[1] 4  7
[1] 5  9
[1] 6  7
[1] 7  7
[1] 8  7
```

Notice, that the expectation value of the number of records is $EN_{1000}$, approximately 7.5, and its standard deviation is $\sqrt{var(N_{1000})}$, approximately 2.4.




■ Figure 3.2.4: Plot of records, both on a linear (upper row) and logarithmic scale (lower row) for the daily BMW stock returns and the daily NYSE composite index returns in comparison to *iid* random variables drawn from an alpha-stable distribution with index 1.7. Sixteen different subsamples are included in the plot to achieve better statistics.

### 3.2.4 The Ratio of Maximum and Sum

A simple tool for detecting heavy tails of a distribution and for giving a rough estimate of the order of its finite moments is the *Ratio of Maximum and Sum*. Suppose that the random variables $X, X_1, X_2, \ldots$ are iid and define for any positive $p$ the quantities

$$S_n(p) = |X_1|^p + \ldots + |X_n|^p \ , \quad M_n(p) = \max(|X_1|^p, \ldots, |X_n|^p) \ , \quad n \geq 1 \ . \tag{3.10}$$

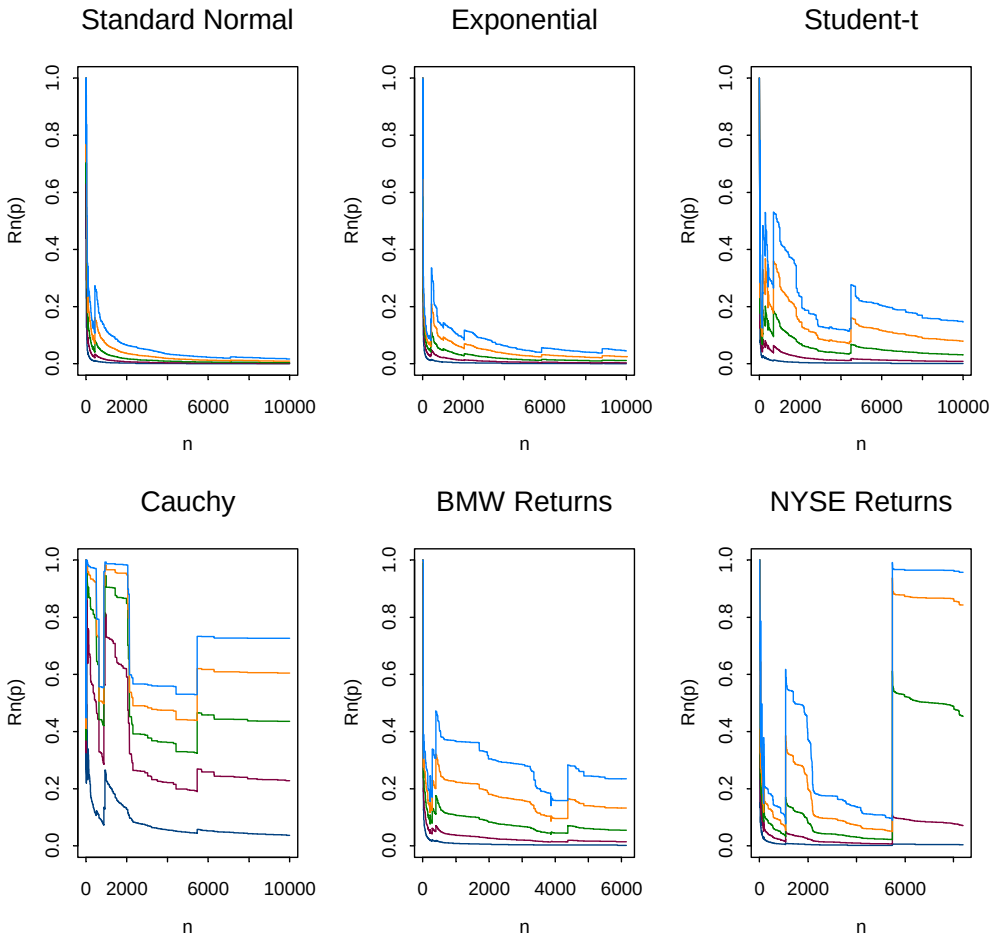
We also write $M_n = M_n(1)$ and $S_n = S_n(1)$ slightly abusing our usual notation. One way to study the underlying distribution is to look at the distributional or a.s. behavior of functionals $f(S_n(p), M_n(p))$.

Example: Ratio of Maximum and Sum Plot - xmpEDAmaxsum

Plot the Ratio of Maximum and Sum for exponents $p$ ranging from 1 to 4 for random variables drawn from a standard normal, and an $\alpha$-stable distribution with index 1.7. Compare the plots with those obtained from the daily log returns of BMW stock prices and the NYSE Composite index.

Use the `fExtremes` library function `maxsumratio.plot(x, p=seq(1,4,1), ...)`, where `p` denotes the power exponent, by default taking the values $1, 2, 3, 4$. The ratio is calculated as

```
max(abs(x)^p[i]) / cumsum(abs(x)^p[i])
```




■ Figure 3.2.5: The plots show the ratios for maximums and sums of random variables drawn from a standard normal (upper left) and an $\alpha$-stable distribution (upper right) with index 1.7 in comparison with the daily logarithmic returns of BMW stock prices (lower left) and the NYSE composite index (lower right).

The graphical output was produced from

```
> set.seed(876)
> nt <- 10000
> x1 <- rnorm(nt)
> x2 <- rexp(nt)
> x3 <- rt(nt,df=4)
> x4 <- rstab(nt,index=1.0)
> x5 <- scan(file="bmwres.csv")
```

```
> x6 <- scan(file="nyseres.csv")

> maxsumratio.plot (x1, main="Standard Normal")
> maxsumratio.plot (x2, main="Exponential")
> maxsumratio.plot (x3, main="Student-t")
> maxsumratio.plot (x4, main="Cauchy")
> maxsumratio.plot (x5, main="BMW Returns")
> maxsumratio.plot (x6, main="NYSE Returns")}
```

From the limit behavior of the ratio $M_n/S_n$ one can derive the following relationships:

$$\begin{array}{lll} R_n(p) \to 0 & \Leftrightarrow & E|X|^p < \infty \ , \\ R_n(p) \to 0 & \Leftrightarrow & E|X|^p \ I_{\{|X| \le x\}} \in \mathcal{R}_0 \ , \\ R_n(p) \to const(p) & \Leftrightarrow & P(|X| > x) \in \mathcal{R}_{-\alpha p} \text{ for some } \alpha \in (0,1) \ , \\ R_n(p) \to 1 & \Leftrightarrow & P(|X| > x) \in \mathcal{R}_0 \ . \end{array} \tag{3.11}$$

Now it is immediate how one can use these limit results to obtain some preliminary information about $P(|X| > x)$: Plot $R_n(p)$ against $n$ for a variety of $p$ values. Then $R_n(p)$ should be small for large $n$ provided that $E|X|^p < \infty$. On the other hand, if there are significant deviations of $R_n(p)$ from zero for large $n$, this is an indication for $E|X|^p$ being infinite.

Clearly, what has been said about the absolute values of the $X$, can be modified in the natural way to get information about the right or left distribution tail: Replace everywhere $|X_i|^p$ by the $p$th power of the positive or negative part of the $X_i$.

### Notes and Comments

The description of the material presented here follows closely some sections from the book of Embrechts, Klüppelberg and Mikosch (1997). Further references to this material can be found therein.

The `fExtremes` library implements several Splus functions for the exploratory data analysis of extremes. These include functions for quantile plots, mean excess function plots, for plots of records, and for plots of the ratio of maximum and sum. The Splus functions were written by D. Würtz implementing also some code written by S. Coles.

## 3.3 Fluctuations of Maxima: GEV Distribution

### Introduction

Starting an investigation of extremes it is useful to compare and contrast the Central Limit Theorem, CLT, with the main results of Extreme Value Theory, EVT. Remember CLT, where the general formulation goes as follows: Suppose $X_1, \ldots, X_n$ are *iid* random variables with distribution function $F$. Define the partial sum $S_n = X_1 + \ldots + X_n$. The CLT then solves the following problems:

- Given $F$, find constants $a_n > 0$ and $b_n \in \mathbb{R}$ so that
$$\frac{S_n - b_n}{a_n} \to Y \ , \quad n \to \infty \ , \tag{3.12}$$
where $Y$ is a non-degenerate random variable with distribution function $G$.
- Characterize a possible distribution function $G$ of $Y$.
- Given a possible distribution function $G$, find all distribution functions which satisfy (3.12) (the domain of attraction problem) and characterize the sequences $a_n$ and $b_n$.

Solutions to the above points are well known: e.g. consider the special case for distributions with second finite moment. Indeed then one gets $a_n = \sqrt{n}\sigma$, $b_n = n\mu$, $G = \mathcal{N}(0,1)$. Here $\mathcal{N}(0,1)$ is the standard normal distribution function, and all distribution functions with $\sigma^2 < \infty$, are attracted to $\mathcal{N}(0,1)$.

### 3.3.1 Limit Laws for Maxima

Now we ask similar questions upon replacing $S_n$ by the maximum order statistic $M_n = X_{1,n} = \max(X_1, \ldots, X_n)$. We adopt the same approach in obtaining limits of the distribution of $(M_n - b_n)/a_n$. The solution to the range of possible limit distributions is given by the following

**Theorem (Fisher-Tippet):** *Suppose $(X_1, \ldots, X_n)$ are iid random variables with distribution function $F$. If there exist constants $a_n > 0$ and $b_n \in \mathbb{R}$ so that*

$$\frac{M_n - b_n}{a_n} \to Y \ , \quad n \to \infty \ , \tag{3.13}$$

*where $Y$ is a non-degenerate random variable with distribution function $G$ then $G$ is of one of the following types:*

$$
\begin{aligned}
\text{I} - \text{Gumbel}: \quad \Lambda(x) &= \exp\{-e^{-x}\}, && x \text{ real.} \\
\text{II} - \text{Frechet}: \quad \Phi_\alpha(x) &= \begin{cases} 0, & x \le 0 \\ \exp\{-x^{-\alpha}\}, & x > 0 \end{cases} && \alpha > 0. \\
\text{III} - \text{Weibull}: \quad \Psi_\alpha(x) &= \begin{cases} \exp\{-(-x)^{\alpha}\}, & x \le 0 \\ 1, & x > 0 \end{cases} && \alpha > 0.
\end{aligned} \tag{3.14}
$$

Collectively, the three classes of distribution are referred to as the extreme value distributions, with types I, II and III widely known as the Gumbel, Fréchet and Weibull types, respectively. For the Fréchet type, the extreme value distributions are heavy tailed functions whose tails decay powerlike. In contrast, the Gumbel and the Weibull functions are light tailed. Note that the Theorem does not guarantee the existence of a non-degenerate limit for $M_n$, nor specify which limit will arise when such a limit does exist. When such a limit does exist, however, by analogy with the CLT, we find that the limiting distribution of sample maxima follows one of the distributions specified in the Theorem whatever the parent distribution $F$.

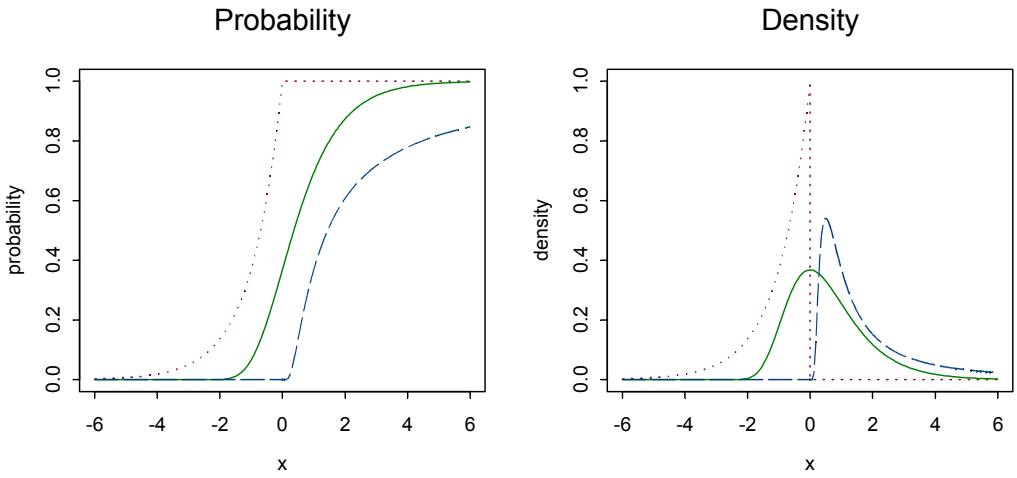
**Example: Gumbel, Fréchet and Weibull Distributions - `xmpGEVdistributions`**

Plot the density and probability functions for the Gumbel, Fréchet and Weibull distributions.

Use the `fExtremes` library functions

```
pfrechet(q, alpha=1), dfrechet(x, alpha=1),
pweibl(q, alpha=1), dweibl(x, alpha=1),
pgumbl(q), dgumbl(x),
```

where `x` and `q` are numeric vectors, and `alpha` the scale parameter.




■ Figure 3.3.1: Probabilities and densities for the standard ($\alpha = 1$) Weibull (dotted), standard Fréchet (dashed) and Gumbel (solid curve) distributions.

The limit distributions $\Lambda, \Phi, \Psi$ are determined up-to type: two distribution functions $F_1$, $F_2$ are of the same type if there exist constants $a_1 \in \mathbb{R}$, $a_2 > 0$ so that $F_2(x) = F_1((x - a_1)/a_2)$, i.e. $F_1$ and $F_2$ belong to the same location and scale family. Distribution functions of the Fréchet, Weibull and Gumbel types are called the *generalized extreme value distributions*. In many cases

it is inconvenient to work with three distinct classes of limiting distributions, so it is usual to adopt a parameterization which encompasses all three types

$$H_{\xi;\mu,\beta}(x) \;=\; \exp\left\{-\left(1+\xi\frac{x-\mu}{\beta}\right)^{\frac{1}{\xi}}\right\}\;, \tag{3.15}$$

defined on $\{x : 1+\xi(x-\mu)/\beta > 0\}$. The type II and III classes of extreme value distributions correspond respectively to the cases $\xi > 0$ and $\xi < 0$ in this parameterization, while the type I class arises in the limit $\xi \to 0$. The parameter $\xi$ is referred to as the *shape parameter*, while $\mu$ and $\beta$ are *location* and *scale* parameters, respectively. The standard case $\mu = 0$. $\beta = 1$ will be denoted by $H_\xi = H_{\xi;0,1}$.

The estimation of extreme quantiles $x_p$ make the importance of the shape parameter $\xi$ quite evident. Inverting equation (3.15) yields

$$x_p = \mu - \frac{\beta}{\xi}\left[1-\{-\log(1-p)\}^{-\xi}\right]\;, \tag{3.16}$$

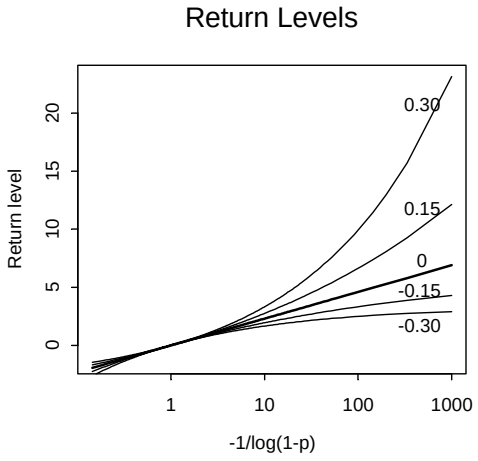
where $H(x_p) = 1-p$. In extreme value terminology, $x_p$ is the *return level* associated with the return period $1/p$, and it is common to extrapolate the relationship (3.16) to obtain estimates of return levels considerably beyond the range of the data to which the model has been fitted.

**Example: Return Levels: -** `xmpGEVreturnlevels`

Plot the return levels for different shape parameters of the GEV distribution.

Use the `fExtremes` library function `qgev(p, xi, mu=0, beta=1)` where `p` is a numeric vector, `xi` the shape parameter and `mu` and `beta` the location and scale parameters respectively, with default values `0` and `1`:

```
p <- seq(0.001, 0.999, length=500)
x <- -1/log(1-p)
plot (x, qgev(1-p, xi=+0.3), ...)
for (xi in (0.15, 0, -0.15, -0.3) lines(x, qgev(1-p, xi=xi))
```




■ Figure 3.3.2: Quantiles on log scale for GEV distribution with different shape parameters -0.3, -0.15, 0, 0.15 and 0.30.

If $\xi$ is negative, the quantiles are bounded above, so extrapolation to any level will lead to a finite limit. On the other hand, if $\xi$ is zero or positive, then the limiting distribution is unbounded, and extrapolation leads to an infinite limit, as can be seen from figure 3.3.2.

### 3.3.2 Maximum Domain of Attraction Problem

Similar to the CLT, one may try to find a solution to the *maximum domain of attraction* problem, i.e. given $H_\xi$ find conditions on $F$ so that equation (3.13) holds for appropriate sequences $(a_n)$ and $(b_n)$. In this case we say $F$ belongs to the maximum domain of attraction of $H_\xi$ with norming constants $a_n$ and $b_n$ denoted by $F \in MDA(H_\xi)$.

First let us consider two examples, the Exponential and Normal distributions, respectively.

**Example: Convergence of Exponential Maxima -** `xmpGEVconvergence`

Suppose $F$ is the standard exponential distribution, so that

$$F(x) = 1 - exp(-x) \; ; \quad \text{then} \quad F(x)^n = (1 - e^{-x})^n \quad .$$

Now let us inspect

$$F(x + \log(n))^n \;=\; (1 - e^{-x - log(n)})^n \;=$$
$$(1 - \frac{1}{n} e^{-x})^n \to \exp(-\exp(-x)) \quad as \quad n \to \infty \quad .$$

Thus with a normalization of

$$a_n \;=\; 1 \quad \text{and} \quad b_n = \log(n) \quad ,$$

$M_n$ has a limiting Gumbel distribution.

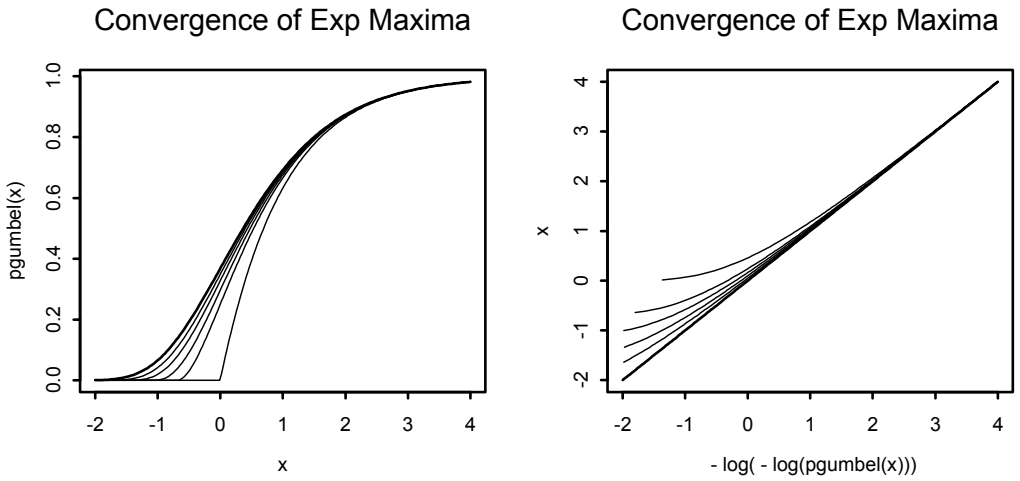
Illustrate this convergence graphically, showing the true distribution of $(M_n - b_n)/a_n$ for $n = 10, 5, 3, 2, 1$ together with the limit distribution $G(x) = exp(-exp(-x))$.

In order to examine more closely the convergence in the tail of the distribution, produce these plots on a different scale. In particular plot *quantiles*, $x_p$, of the distribution on a complementary log-log scale. That is, plot $x_p$ against $-\log(-\log(F(x_p))$. This choice of scale magnifies the tail region, and presents the Gumbel limit as a linear line.

The graphs below are produced in the following way:

```
a <- function(n) {1}

b <- function(n) {log(n)}

n <- c(10, 5, 3, 2, 1)

plot(x, pgumbel(x), ...)
for (i in n) {
  y <- (pexp(a(i)*x+b(i)))^i
  lines(x, y, ...) }

plot(-log(-log(pgumbel(x))), x, ...)
for (i in n) {
  y <- (pexp(a(i)*x+b(i)))^i
  lines(-log(-log(y)), x, ...)
```

First, define the functions `a(n)` and `b(n)`. Then plot the converging graphs as function of `n`. Here `pexp(x)` is the standard Splus function for the exponential probability function, and `pgumbel(x)` is the probability function of the Gumbel distribution from the `fExtremes` library.




■ Figure 3.3.3: Convergence of exponential maxima to Gumbul limit. Left probability plot, Right plot on alternative scale. Clearly the convergence is rapid, especially in the upper tail which is the region where greatest accuracy is required.

**Example: Normal Distribution: -** `xmpGEVconvergence`

An example where convergence is much slower is where $F$ is the standard normal distribution. Although by no means obvious $F$ converges with scalings

$$a_n = \frac{1}{\sqrt{2\log(n)}} ,$$

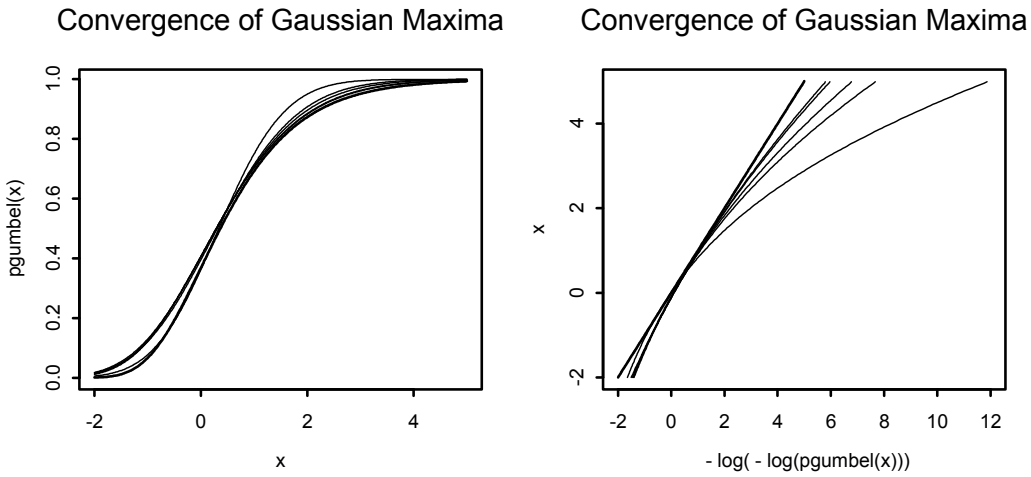
$$b_n = \sqrt{2/\log(n)} - \frac{\log\log n + log4\pi}{2\sqrt{2/\log(n)}} ,$$

$(M_n - b_n)/a_n$ converges again to the Gumbel distribution. In this case, even with $n = 100$, we will find considerable disagreement between the distribution of the normalized $M_n$ and the theoretical limit, especially at extreme levels. Show this behaviour graphically.

Again, first define the Splus function for `a(n)` and `b(n)` and further follow the code from the example above just replacing the function `pexp()` with `pnorm()`:

```
a <- function(n) {
       1/sqrt(2*log(n)) }

b <- function(n) {
       sqrt(2*log(n)) -
       (log(log(n))+log(4*pi))/sqrt(2*log(n))/2 }

n <- c(100, 50, 10, 5, 2)

plot(x, pgumbel(x), ...)
for (i in n) {
  y <- (pnorm(a(i)*x+b(i)))^i
  lines(x, y, ...) }
```

```
plot(-log(-log(pgumbel(x))), x, ...)
for (i in n) {
  y <- (pnorm(a(i)*x+b(i)))^i
  lines(-log(-log(y)), x, ...)
```




■ Figure 3.3.4: Convergence of normal maxima to Gumbul limit. Left probability plot, Right plot on alternative scale.

### 3.3.3 Heavy Tails: Fréchet Type Distribution

For applications in finance the Fréchet type turns out to be the most important model for extremal events. The domain of attraction condition for the Fréchet takes on a particularly easy form.

**Theorem (Gnedenko):** *Suppose $X_1, \dots X_2$ are iid with distribution function $F$. Then there exist sequences $(a_n)$ and $(b_n)$ so that (3.13) holds for $Y$ Fréchet distributed with parameter $\alpha > 0$ (i.e. $F \in MDA(H_{1/\alpha}) = MDA(\Phi_\alpha)$) if and only if*

$$1 - F(x) = x^{-\alpha} L(x) \quad x > 0 \tag{3.17}$$

*with $L$ slowly, i.e. $\lim_{x\to\infty} \frac{L(\lambda x)}{L(x)} = 1$ for all $\lambda > 0$.*

Thus the class of distribution functions with regularly varying tail (3.17) enter very naturally as standard models in the area of extreme value theory.

Now we turn to statistical issues, the estimation of the GEV distributional parameters. The problem is how to make inferences about the parameters $(\xi; \mu, \beta)$. First we present moment based estimators for the shape parameter $\xi$ and then the likelihood estimator for the distribution function.

### 3.3.4 Estimating $\xi$ from L-Moments

L-moments are summary statistics for probability distributions and data samples. They are analogous to ordinary moments in the sense, that they provide measures of location, dispersion, skewness, kurtosis, and other aspects of the shape of probability distributions or data samples, however, they are computed from linear combinations of the ordered data values. In the book of Hosking and Wallis (1997) one can find an excellent overview on the whole theory of L-moments.

Probability weighted moments, defined by Greenwood et al. (1979), are precursors of L-moments. *Sample probability weighted moments* are computed from data values $X_1, X_2, \ldots, X_n$ arranged in increasing order as

$$\begin{aligned} b_0 &= \frac{1}{n}\sum_{j=1}^{n} X_j \ , \\ b_1 &= \frac{1}{n}\sum_{j=r+1}^{n} \frac{(j-1)(j-2)\ldots(j-r)}{(n-1)(n-2)\ldots(n-r)} X_j \ . \end{aligned} \tag{3.18}$$

L-moments are derived as linear combinations of probability weighted moments that have simple interpretations as measures of the location, dispersion and shape of the data sample. The first few L-moments are defined by

$$\begin{aligned} l_1 &= b_0 \ , \\ l_2 &= 2b_1 - b_0 \ , \\ l_3 &= 6b_2 - 6b_1 + b_0 \ , \\ l_4 &= 20b_3 - 30b_2 + 12b_1 - b_0 \ , \end{aligned} \tag{3.19}$$

where the higher coefficients are those of shifted Legendre polynomials. The first L-moment is the sample mean, a measure of location. The second L-moment is (a multiple of) Gini's mean difference statistic, a measure of the dispersion of the data values about their mean. By dividing the higher-order L-moments by the dispersion measure, we obtain the *L-moment ratios*,

$$t_r = \frac{l_r}{l_2} \ . \tag{3.20}$$

These are dimensionless quantities, independent of the units of measurement of the data. $t_3$ is a measure of skewness and $t_4$ is a measure of kurtosis, these are respectively the L-skewness and L-kurtosis.

For a probability distribution with cumulative distribution function $F(x)$, probability weighted moments are defined by

$$\beta_r = \int x\{F(x)\}^r dF(x) \ , \quad r = 0, 1, 2, \ldots \ . \tag{3.21}$$

L-moments are defined in terms of probability weighted moments, analogously to the sample L-moments.

$$\begin{aligned} \lambda_1 &= \beta_0 \ , \\ \lambda_2 &= 2\beta_1 - \beta_0 \ , \\ \lambda_3 &= 6\beta_2 - 6\beta_1 + \beta_0 \ , \\ \lambda_4 &= 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \ . \end{aligned} \tag{3.22}$$

L-moments ratios are defined by

$$\tau_r = \frac{\lambda_r}{\lambda_2} \ . \tag{3.23}$$

Estimating $\xi$ from L-moments now consists of equating model moments based on $H_{\xi;\mu,\beta}$ to the empirical moments based on the data. This approach yields

$$\begin{aligned} \frac{3b_2 - b_0}{2b_1 - b_0} &= \frac{3^\xi - 1}{2^\xi - 1} \ , \\ \beta &= \frac{((2b_1 - b_0)\xi}{\Gamma(1-\xi)(2^\xi - 1)} \ , \\ \mu &= b_0 + \frac{\beta}{\xi}(1 - \Gamma(1-\xi)) \ , \end{aligned} \tag{3.24}$$

where $b_0$, $b_1$ and $b_2$, are the probability weighted moments.

**Example: GEV simulated Data L-Moments Estimation -** `xmpGEVpwm1`

Create 8000 random variables, GEV distributed with parameters $\xi = 0.3$, $\mu = 0$, and $\beta = 1$. Estimate the parameters from the first three probability weighted moments. Plot the series of random variables and compare the empirical density with the estimated density and the true density.

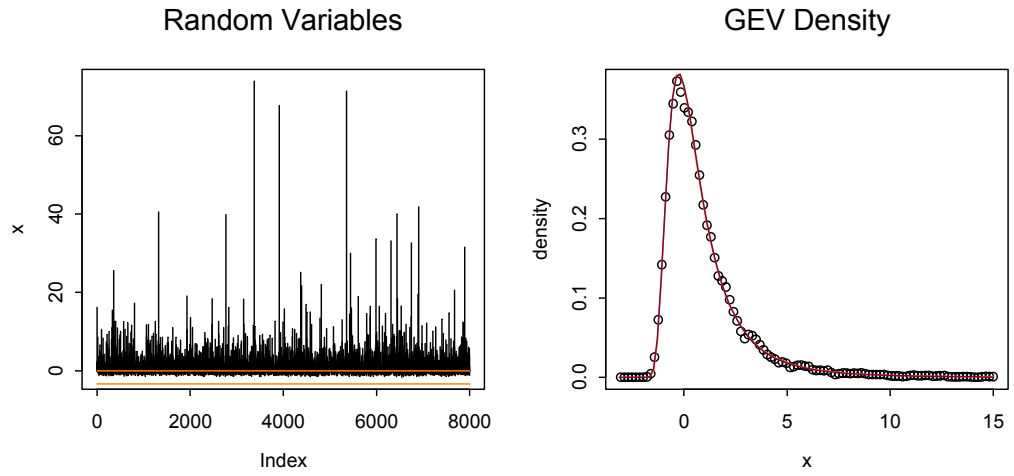
Use the function `gev.pmw()` from the `fExtremes` library to estimate `xi`, `mu` and `beta`. `gev.pwm()` evaluates the first three probability weighted moments implementing the FORTRAN routine "sampwm" from Hosking into the Splus function `sampwm()`. Having the weights `w0`, `w1`, and `w2` we can solve for `xi` with the help of the standard Splus function `uniroot()`. From `xi` and the weights we can derive `mu` and `beta` with the help of the standard Splus function `gamma()`. Here is a short outline of the Splus function code:

```
"sampwm" <- function(x, nmom=3, a=-0.35, b=0, kind=1) {
   result <- .Fortran( "sampwm",
     as.double(rev(sort(x))),
     as.integer(length(x)),
     as.double(rep(0,nmom)),
     as.integer(nmom),
     as.double(a),
     as.double(b),
     as.integer(kind))
   result[[3]] }
```

```
"y" <- function(xi, b0, b1, b2) {
   (3^xi-1)/(2^xi-1)-(3*b2-b0)/(2*b1-b0) }

"gev.pwm" <- function(x) {
   w <- sampwm(x, nmom=3, a=-0.35, b=0, kind=1)
   b0 <- w[1]; b1 <- w[2]; b2 <- w[3]
   xi <- uniroot(f=y, interval=c(-5,+5), b0=w[1], b1=w[2], b2=w[3])$root
   beta <- (2*b1-b0)*xi / gamma(1-xi) / (2^xi-1)
   mu <- b0 + beta*(1-gamma(1-xi))/xi
   list(xi=xi, mu=mu, beta=beta) }
```

First create and plot the GEV random variables, then plot their density, estimate the parameters and add to the density plot the graph for the estimated and true density:

```
> x <- rgev(8000, xi=0.3)
> plot(x)
> plot(density(x), n=100, width=0.5)
> parm <- gev.pwm(x)
> s <- seq(-1/parm$xi ,5, 100)
> lines(s, dgev(s, parm$xi, parm$mu, parm$beta))
> lines(s, dgev(s, xi=0.3))

> parm
$xi:
[1] 0.3251
$mu:
[1] 0.003107
$beta:
[1] 1.005
```




■ Figure 3.3.5: The graph to the left shows the simulated GEV variables. To the right the graph displays the density plot of the empirical data (circles) and the graphs of the estimated and true GEV. The latter two are so close, that they cannot be distinguished by visual inspection.

The result of this example shows us, that for a large sample the estimate is in a quite good agreement with the simulated empirical data. Now let us apply the same procedure to the monthly maxima of the negative logarithmic returns of the German BMW stock price data and of the NYSE Composite Index data.

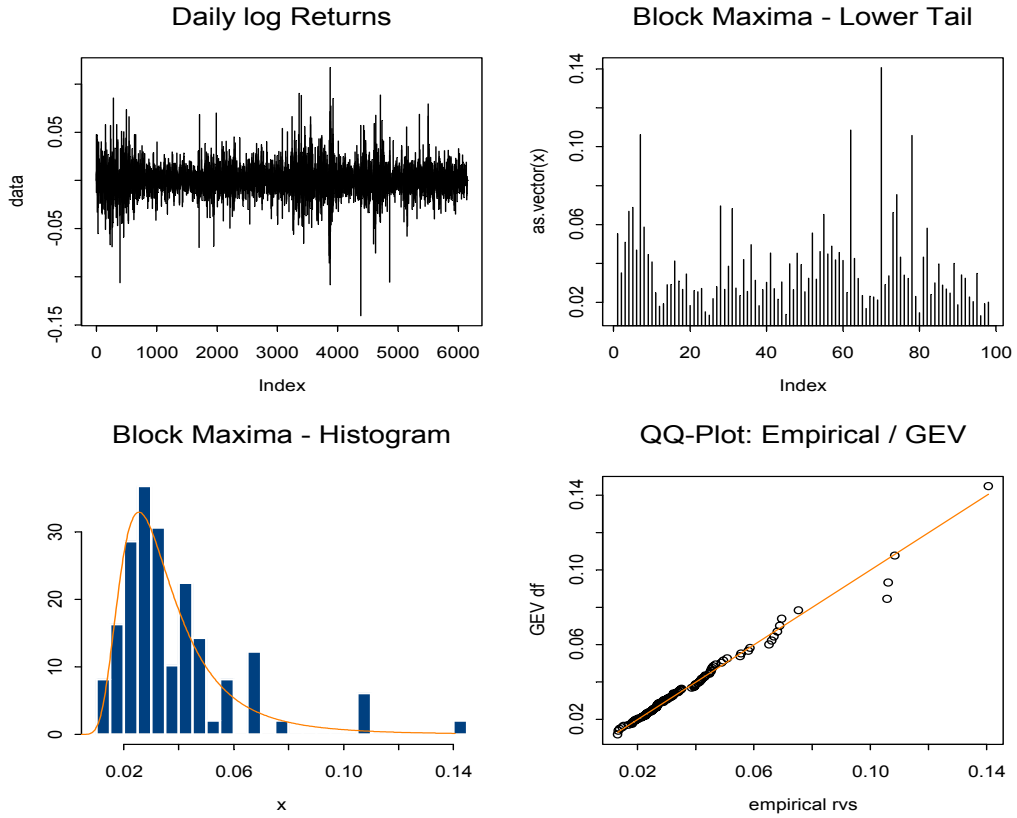
**Example: BMW Stock Data L-Moments Estimation - `xmpGEVpwm2`**

Plot the daily log returns of the BMW stock prices. Determine (approximately) quarterly block maxima by dividing the log-series into subseries of 63 business days. Estimate the parameters for the GEV distribution. Check the quality of the estimation process by comparing the block maxima histogram to the estimated GEV distribution and by plotting a QQ-Plot of the empircal data versus the estimated GEV distribution.

Use the `fExtremes` library function `block.maxima(x, blocklength=63)` to create the block maxima data and the standard Splus function `hist(x)` to display the histogram plot. The estimation yields the following values for the parameters $\xi$, $\mu$ and $\alpha$:

```
> x <- scan("bmwres.csv")
> gev.pwm(x)
$xi:
[1] 0.228
$mu:
[1] 0.028
$beta:
[1] 0.0115
> 1/parm$xi
[1] 4.39
```

The results for the NYSE Composite Index data are: $\xi = 0.406$, $\mu = 0.015$, $beta = 0.0052$, and $\alpha = 1/\xi = 2.46$.




■ Figure 3.3.6: The graphs in the upper row show the daily log-returns of the BMW stock prices (left) and the series of the 63-days block maxima data (right). The lower graphs show diagnostic plots: The histogram of block maxima together with the estimated GEV distribution (left) and the QQ-Plot of the empirical data versus the estimated GEV distribution(right).

### 3.3.5 Estimating $\xi$ under MDA Conditions

Let us start from a Theorem on an equivalence assertion for $MDA(H_\xi)$.

**Theorem: Characterization of** $MDA(H_\xi)$ For $\xi \in \mathbb{R}$ the assertion $F \in MDA(H_\xi)$ is equivalent to: For $x, y > 0, y \neq 1$,

$$\lim_{s\to\infty} \frac{U(sx) - U(s)}{U(sy) - U(s)} = \begin{cases} \frac{x^\xi - 1}{y^\xi - 1} & if\, \xi \neq 0, \\ \\ \frac{\ln x}{\ln y} & if\, \xi = 0. \end{cases} \tag{3.25}$$

where $U(t) = F^{\leftarrow}(1-t^{-1})$, $t > 0$ and $F^{\leftarrow}$ denotes the quantile function of a distribution function $F$.

#### Method 1: Pickands Estimator

A reformulation of relation (3.25) leads to Pickands' estimation procedure for quantiles outside the range of the data, using

$$\lim_{t\to\infty} \frac{U(c(t)t) - U(t)}{U(t) - U(t/c(t))} = 2^\xi \ . \tag{3.26}$$

whenever $\lim_{t\to\infty} c(t) = 2$ for a positive function $c$. From this expression an empirical estimator, the so called *Pickands' estimator*, can be constructed:

$$\widehat{\xi}^{(P)}_{k,n} = \frac{1}{\ln 2} \ln \frac{X_{k,n} - X_{2k,n}}{X_{2k,n} - X_{4k,n}} \to \xi \; n \to \infty \ . \tag{3.27}$$

Under further conditions on $k$ and $F$ the estimator is asymptotic normal,

$$\begin{aligned} \sqrt{k}(\widehat{\xi} - \xi) \ &\to \ \mathcal{N}(0, v(\xi)) \ , \quad n \to \infty \ , \\ v(\xi) \ &= \tfrac{\xi^2(2^{2\xi+1}+1)}{(2(2^\xi-1)\ln 2)^2} \ . \end{aligned}$$

This result forms the core of a whole series of results, Pickands (1975), and Decker and de Haan (1989), concerned with quantile and tail estimators and confidence interval constructions.

#### Method 2: Hill's Estimator

Suppose $X_1, \ldots, X_n$ are *iid* with distribution function $F \in MDA(\Phi_\alpha)$, $\alpha > 0$, thus $1 - F(x) = x^{-\alpha}L(x)$, $x > 0$, for slowly varying function $L$. The Hill Estimator of $\alpha$ takes on the following form:

$$\widehat{\alpha}^{(H)} = \widehat{\alpha}^{(H)}_{k,n} = \left( \frac{1}{k} \sum_{j=1}^{k} \ln X_{j,n} - \ln X_{k,n} \right)^{-1} \ . \tag{3.28}$$

Hill's Estimator should be viewed as a counterpart to Pickands' estimator. Note, under further conditions on $k$, $F$ and *iid* sequence $X_n$ the estimator is asymptotic normal,

$$\sqrt{k}(\widehat{\alpha}^{(H)} - \alpha) \to \mathcal{N}(0, \alpha^2) \ .$$

Here $k = k(n) \to \infty$ in an appropriate way, so that as in the case of the Pickands estimator, an increasing sequence of upper order statistics is used. One of the interesting facts concerning equation (3.28) is that various asymptotically equivalent versions of $\widehat{\alpha}^{(H)}$ can be derived through essentially different methods, see Embrechts et al. (1997).

**Example: Hill's Estimator - `xmpGEVhillplot`**

Plot the shape parameter $\alpha$ obtained from Hill's estimator as a function of the order statistic and threshold values together with its 95% confidence intervals.

Use the Splus functions `hill.plot(data, param="alpha", start=15, end=NA, reverse=F, ci=T, ci.p=0.95, auto.scale=T, labels=T, ...)` from the `fExtremes` library. The *required argument* `data` is a data vector or a time series. The *optional arguments* are: `param` allows to select whether "alpha" or "xi" (1/alpha) should be plotted. `start` is the lowest number of order statistics at which to plot a point, and `end` the highest number of order statistics at which to plot a point. `reverse` allows to select whether the plot is to be by increasing threshold (TRUE) or increasing number of order statistics (FALSE). `ci`, a logical flag, if TRUE, the probability for asymptotic confidence band is added to the plot; for no confidence band set to FALSE. `auto.scale` decides whether or not the plot should be automatically scaled; if not, `xlim` and `ylim` may be entered as graphical parameters. `labels`, a logical flag, if TRUE, the axes would be labeled.

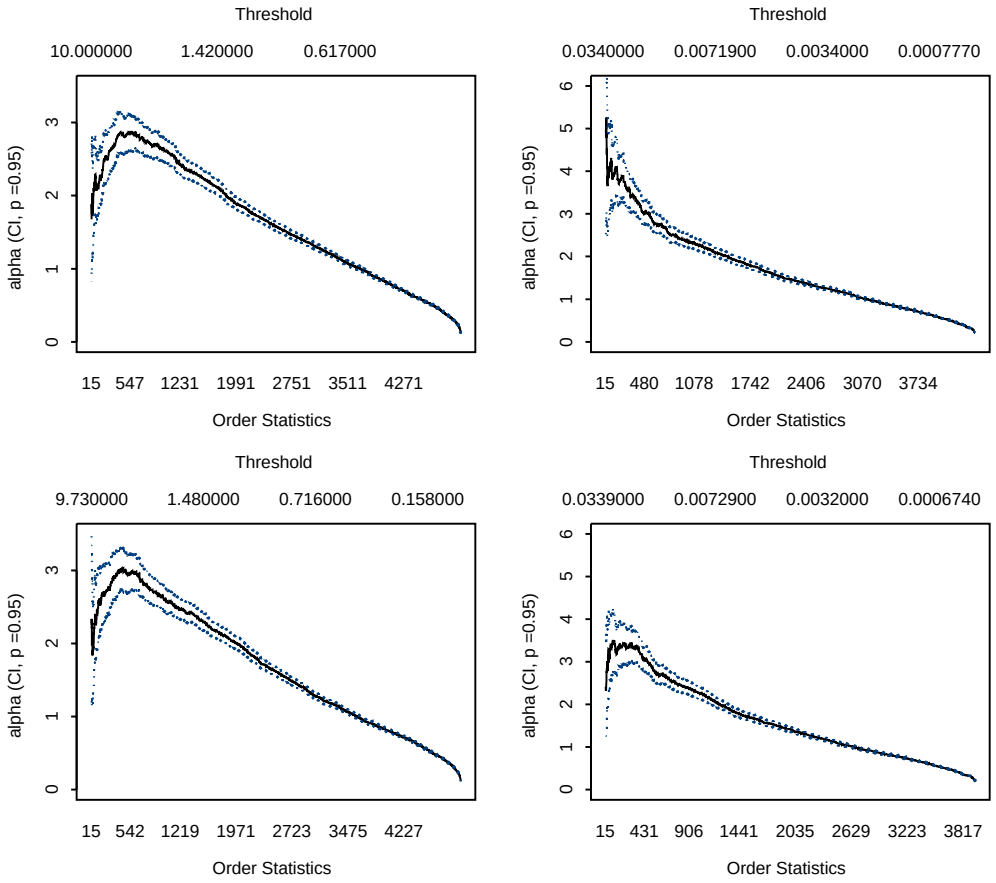
```
> set.seed(471)
> x1 <- rstab(10000, index=1.8)
> x2 <- scan(file="nyseres.csv")

> hill.plot( x1, auto.scale=F, ylim=c(0,3.5))
> hill.plot( x2, auto.scale=F, ylim=c(0,6.0))
> hill.plot(-x1, auto.scale=F, ylim=c(0,3.5))
> hill.plot(-x2, auto.scale=F, ylim=c(0,6.0))
```

Figure 3.3.7 shows the results for the log-return values of the NYSE composite index and a simulated series of 10000 alpha-stable random variables with index 1.8. Both, upper and lower tails are investigated. The first plot was created by the Splus command `hill.plot(rstab(1000,1.8)`.

### Method 3: Deckers-Einmahl-deHaan Estimator

A disadvantage of Hill's estimator is that is essentially designed for $F \in MDA(H_\xi)$, $\xi > 0$. In Deckers, Einmahl, de Haan (1989), Hill's estimator is extended to cover the whole class $H_\xi$, $\xi \in \mathbb{R}$. They came up with the following proposal:

$$\widehat{\xi}_{k,n} = 1 + H_{k,n}^{(1)} + \frac{1}{2}\left(\frac{(H_{k,n}^{(1)})^2}{H_{k,n}^{(2)}} - 1\right)^{-1} \ , \tag{3.29}$$




■ Figure 3.3.7: Hill plot for a simulated series of $\alpha$-stable random variables with index 1.8 (left) and the log-return values of the NYSE Composite Index (right). Both, upper and lower tails are investigated, shown in the upper and lowee figures, respectively.

where

$$H_{k,n}^{(1)} = \frac{1}{k}\sum_{j=1}^{k}(\ln X_{j,n} - \ln X_{k+1,n}) \tag{3.30}$$

is the reciprocal of Hill's estimator (modula an unimportant change from $k$ to $k+1$) and

$$H_{n}^{(2)} = \frac{1}{k}\sum_{j=1}^{k}(\ln X_{j,n} - \ln X_{k+1,n})^2. \tag{3.31}$$

Because $H_{k,n}^{(1)}$ and $H_{k,n}^{(2)}$ can be interpreted as empirical moments, $\widehat{\xi}_{k,n}$ is also referred to as a *moment estimator* of $\xi$.

**Example: Shape Parameter - `xmpGEVshaparm`**

The function `shaparm.plot(x, tails, doplot, alfa.range=c(-0.5,1.5), xi.range=c(0,10))` plots the shape parameter $\alpha$ or $\xi$ obtained from Pickands', Hill's and Decker-Einmahl-deHaan estimators. Individual plots can be displayed for predefined threshold values on a log-scale from which $\alpha$ or $\xi$ are estimated as mean values. Both, upper and lower tails are considered in the same plots. `x` is a data vector, in this example the logarithmic returns of the NYSE Composite Index, `tails` is a vector of threshold values for which $\alpha$ and $\xi$ are evaluated; here we select 10 different tail depths ranging from 1% to 10% of the data. `doplot` is a vector of logical values of the same length as `tails`. For those elements which are set TRUE a plot is generated, otherwise not. `alfa.range` and `xi.range` are optional arguments, which define the plot ranges for $\alpha$ and $\xi$ with default values set to `c(-0.5,1.5)` and `c(0,10)`, respectively. In the example we have overwritten the default values for the `xi.range`.

To become more definite, the plots of shape parameters in figure 3.3.8 for the NYSE Composite Index were generated by the following commands, producing this output:

```
> x <- scan(file="nyseres.csv")
> tails <- c(0.01,0.02,0.03,0.04,0.05,0.06,0.07,0.08,0.09,0.10)
> doplot <- c(  F,   F,   F,   F,   T,   F,   F,   F,   F,   F)
> shaparm.plot (x, tails, doplot, xi.range=c(0,8))

[1] 0.01
[1] 0.9020344 0.3838229 0.4609961
[1] 0.2176707 0.2192884 0.1760605
[1] 0.02
[1] 0.6415300 0.3462955 0.4669930
[1] 0.3016799 0.2322289 0.1862094
    ...
[1] 0.1
[1] 0.3072878 0.3082789 0.3896214
[1] 0.1344260 0.2693792 0.1807588
```

Here, for each tail value ranging from 1% to 10% of the data three lines are printed: The first line gives the tail depth (e.g. 0.02 2%) and the succeeding two lines give the $\alpha$ values for the upper and lower tails, respectively. The three numbers belong from left to right to Pickands', Hill's and Decker-Einmahl-deHaan's Estimator.

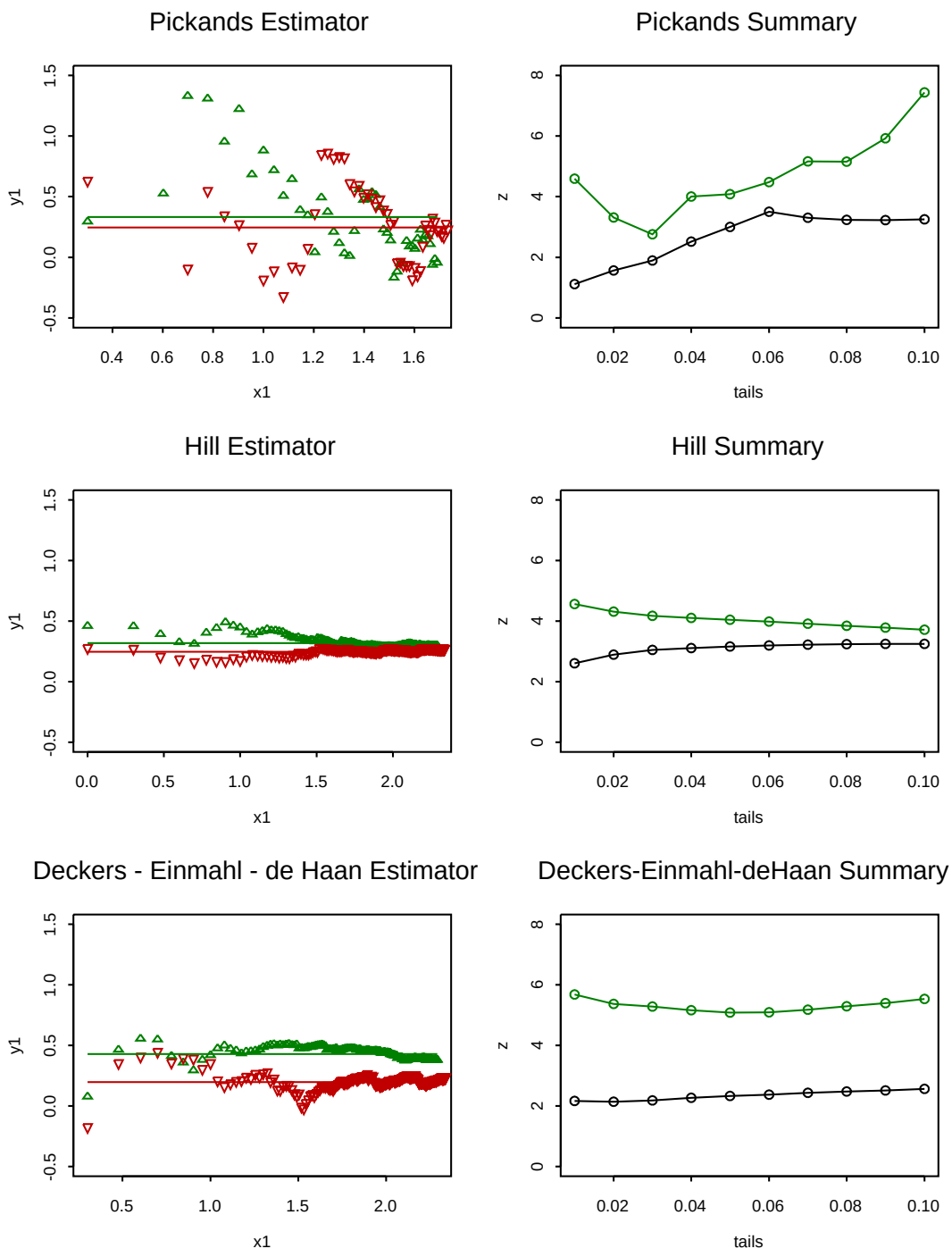
### 3.3.6 Maximum Likelihood Estimation of $\xi; \mu, \beta$

Suppose we have an observed sequence of maxima, $M_{n,1}, \ldots, M_{n,k}$, which for simplicity we rewrite as $X_1, X_2, \ldots, X_k$, and we assume that $n$ is sufficiently large for the GEV approximation to be valid. The inferential procedure would consist of the numerical maximization of the likelihood function

$$l(\xi, \mu, \beta) = \sum_{i=1} k \left\{ -\log \beta - (1 + 1/\xi) log[1 + \xi(\frac{x_i - \mu}{\beta})] - [1 + \xi(\frac{x_i - \mu}{\beta})]^{-1/\xi} \right\}. \qquad (3.32)$$

together with diagnostics checks.

- *Numerical maximization* of equation (3.32) presents generally no difficulties.
- The calculation of *standard errors* follows immediately from classical likelihood theory by inversion of the observed information matrix.




■ Figure 3.3.8: The figures give in the right column a summary of the inverse shape parameter as a function of tail depths for the Pickands', Hills' and Deckers-Einmahl-deHaan's estimators. The upper curves belong to the lower (loss) tail. The left column shows a snapshot for a tail depth of 5%. The shape parameter is derived from a l1-fit on a logarithmic scale.

- *Diagnostic checks* usually consists of a comparison of empirical and model-based estimates of the distribution of $M_n$. Of course, the model is required for extrapolation, but it should fit reasonably well to the most extreme of the observed data also. Comparison of quantiles based on the GEV and the empirical distribution function in a quantile-quantile plot give an assesment of fit. Similarly, the empirical and model-based distributions can be compared in a probability plot.
- In the case of *return level plots*, it is usual to plot estimated quantiles, $x_p$, obtained by substitution of the maximum likelihood estimates of $(\xi, \mu, \beta)$ into (**??**), on a complementary log-scale. That is, plot $x_p$ against $-\log(-\log F(x_p))$.
- The simplest method to calculate *confidence intervals for quantiles* is via the delta method, but asymmetry in the likelihood surface suggests that much more accurate intervals can be obtained using intervals based on the profile likelihood.

Alltogether it is worth to note, that in many cases likelihood based estimators are preferable compared to moment based estimators. Two R / S-plus packages for GEV fitting are available: One written by S. Coles and the other written by A. McNeil. They are slightly different in the produced output, we will give an example for both.

**Example: GEV Fitting -** `xmpGEVmle`

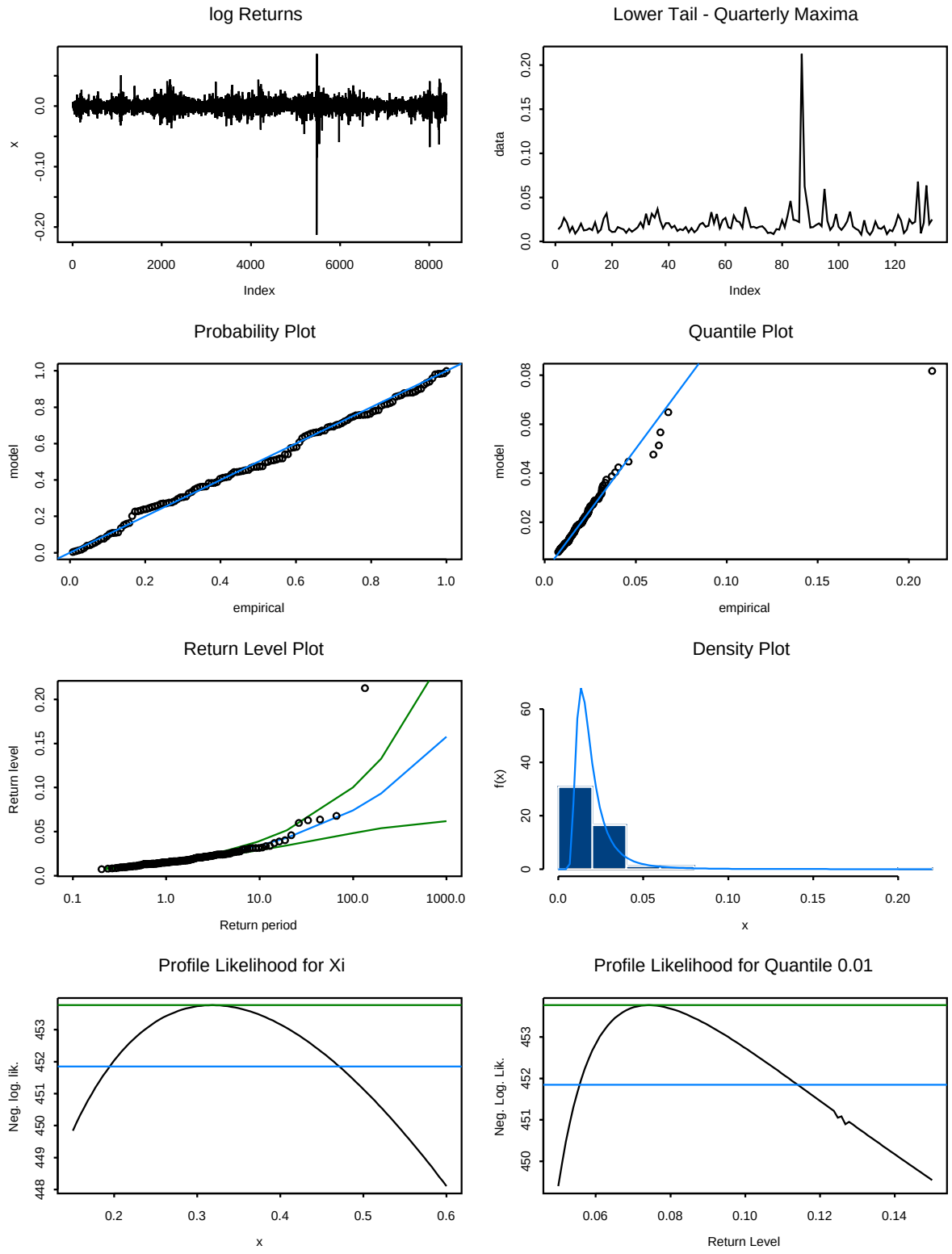
The Splus function `gev.mle(x)` fits through the likelihood approach the parameters $\xi$, $\mu$, and $\beta$ to a GEV distribution. The function returns a TRUE/FALSE indicator of whether the likelihood was successfully maximized or not, and if it was, the maximum likelihood estimates (($\xi$; $\mu$, $\beta$), respectively) and their standard errors. These items are just a summary of all the output items, such the covariance matrix of the estimates, which can be examined by storing the fit as an object. The covariance matrix and corresponding standard errors are obtained by inversion of the hessian matrix which in turn is obtained by numerical differencing.

Furthermore, the function `gev.diag(gevfit.object)` gives a graphical diagnostic of the GEV fit including probability, quantile, return level and density plots. The return level plot also includes 95% confidence intervals based on the delta method.

The function `gev.profxi(gevfit.object, xlow, xup, conf=0.95)` examines the profile likelihood for $\xi$ over the range c(xlow,xup). Also included are lines corresponding to the maximum likelihood, and a confidence profile likelihood interval for $\xi$, by default 95%.

Using the function `gev.prof(gevfit.object, p, xlow, xup, conf=0.95)` plots the profile likelihood for $q_p$ over the interval c(xlow,xup) and gives the corresponding 95% (by default) confidence interval. Thus, this function allows to study quantiles, particularly at extreme levels. For example the $p = 0.01\%$ quantile of the block maxima data, corresponding to the value which is exceeded in any given block with probability 0.01. For annual blocks this is the so called 100 year return level. By re-parametrizing the GEV likelihood in terms of $q_p$, the corresponding GEV quantile, and maximizing with respect to the other 2 parameters, yields the tprofile likelihood for $q_p$.
For example the plots in figure 3.3.9 for the NYSE Composite Index were generated by the following commands, producing this output:

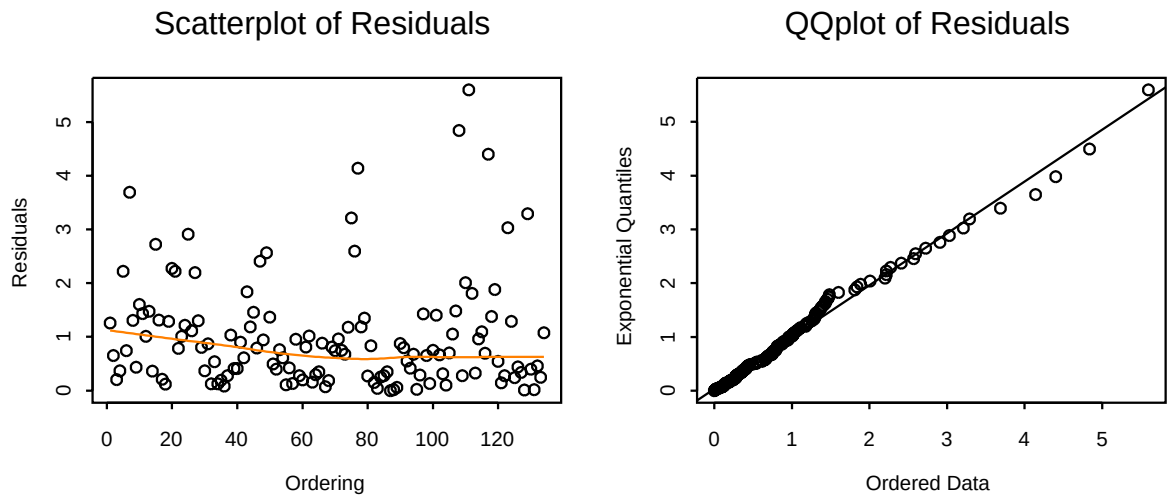
```
> x <- scan(file="nyseres.csv")
> xlength <- length(x)
> blocklength <- 63
> x <- x[1:(blocklength*floor(xlength/blocklength))]
> data <- apply(matrix(-x, byrow=T, ncol=blocklength), 1, max)
> plot(data, type="l", main="Lower Tail - Monthly Maxima")
> gevfit.object <- gev.fit(data)
> print(gevfit.object)
```




■ Figure 3.3.9: The plots show from upper left to lower right the daily log-Returns of the NYSE Composite Index, the quarterly maxima (minima) of the lower tail, the GEV probability, quantile, return level and density plots, the orofile likelihood for $\xi$ and quantile $q_{0.01}$.

```
$conv:
[1] T
$nllh:
[1] -453.7699
$mle:
[1] 0.014950939 0.005673978 0.318026031
$se:
[1] 0.0005531673 0.0004686979 0.0703838384
   ...
$cov:
             [,1]          [,2]          [,3]
[1,]  3.059941e-007 1.648557e-007 -9.640511e-006
[2,]  1.648557e-007 2.196777e-007  2.888646e-006
[3,] -9.640511e-006 2.888646e-006  4.953885e-003
> gev.diag(gevfit.bject)
> gev.profxi(gevfit.object, xlow=0.15, xup=0.60)
> gev.prof(gevfit.object, p=0.01, xlow=0.05, xup=0.15)
```

In addition he Splus function `gev.res()` calculates the residuals and shows a scatterplot.




■ Figure 3.17: The two plots show the residuals from GEV fitting.

## Notes and Comments

In this section we discussed fluctuations of maxima in the framework of the GEV distribution. Most of the material was taken from the textbook of Embrechts, Klüppelberg and Mikosch *Modelling Extremal Events* and from the review article of Coles and Tawn *Statistical Methods for Extreme Values*. For the L-moments approach we refer to the textbook of Hosking and Wallis *Regional Frequency Analysis*.

The `fExtremes` library for the investigation of fluctuations of maxima cllects a series of Splus functions implementing code written by Coles, McNeil and Würtz. Parameter estimation for the GEV distribution through maximum likelihood estimation implements source code written by McNeil and Coles. The L-Moments method accesses a Fortran routine written by Hosking. The Hill plot for estimating the shape parameter under MDA conditions includes functions written McNeil.

## 3.4 Extremes via Point Processes

### Introduction

As a procedure for statistical inference the technique of modelling block maxima with a GEV distribution is highly inefficient since only one data point within a block is used. This weakness has led to the development of other characterizations which enable extreme data other than just the block maxima to be incorporated into the analysis. A first class of such methods to be considered were "threshold methods" whereby all observations which exceed some specified threshold are modelled according to some distribution.

### 3.4.1 Point Process Characterization

Again we take $X_1, X_2, \ldots, X_n$ to be an *iid* series from an unknown distribution $F$. The basic idea is to form a 2-dimensional point process $\{(i, X_i); i = 1, \ldots, n\}$ and to characterize the behavior of this process in regions of the form $[t_1, t_2] \times (u, \infty)$, thus giving a representation for the behavior of the $X_i$ at large levels. More formally, the asymptotic argument is developed as follows. Suppose $F$ is in the domain of attraction of $G$. That is, there are sequences of constants $a_n$, $b_n$ such that equation (3.13) holds for some non-degenerate distribution $G$ given by

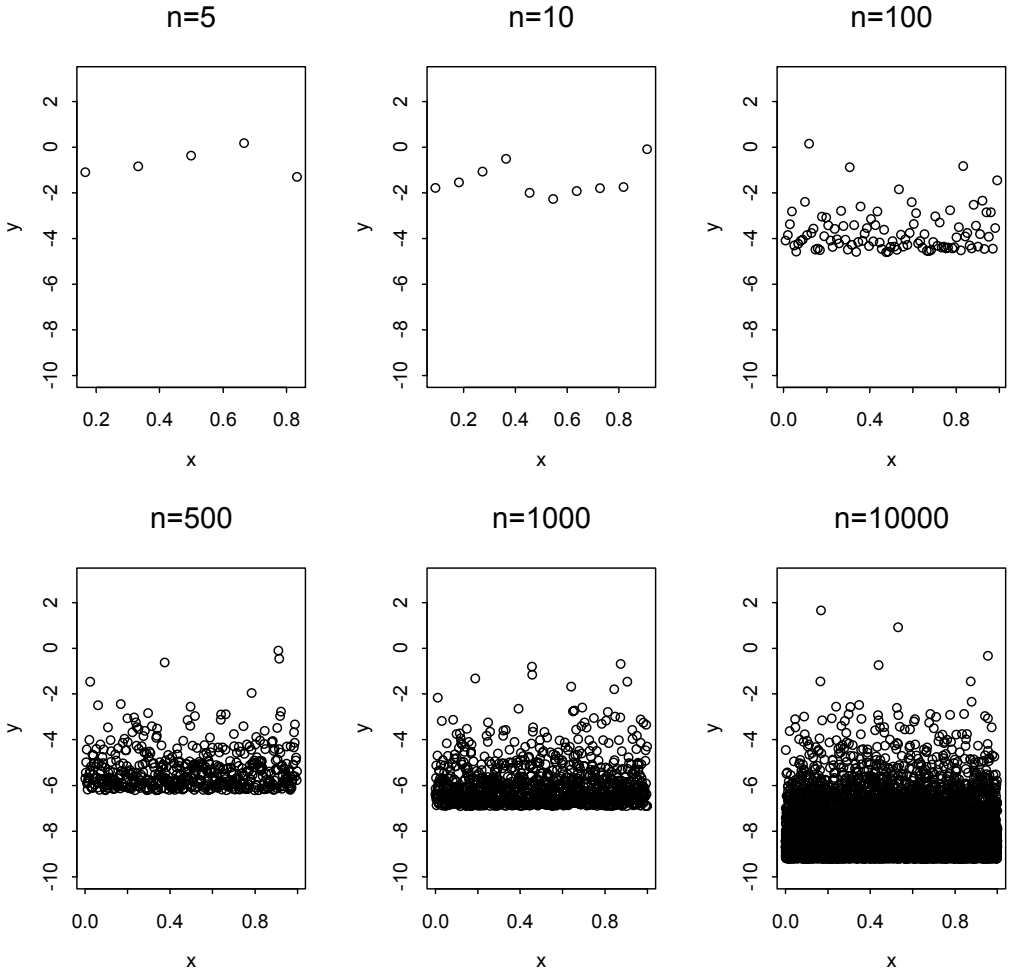
$$G(x) = \exp\left\{-[1 + \xi x]^{-1/\xi} \ .\right\} \tag{3.33}$$

We now construct a sequence of point processes on $\mathbb{R}^2$ by

$$P_n = \left\{ \left( \frac{i}{n+1}, (X_i - b_n)/a_n \right) \ : \ i = 1, \ldots, n \right\} \ . \tag{3.34}$$

The scaling is chosen to normalize the maximum over $P_n$; non-extreme observations become scaled towards the lower end-point. Figure 3.4.1 illustrates this procedure when the $X_i$ are exponentially distributed:

The key feature of this characterization is that away from the lower boundary the process behaves like a *non-homogenous Poisson process*. The simplest characterization of such a process is that the numbers of points in any non-overlapping sets are independent. The process can be uniquely specified by the *intensity measure* $\Lambda(A) = E(\text{number of points in } A)$. In this particular construction, the process $P_n$ converges weakly to a Poisson process on sets which exclude the




■ Figure 3.4.1: Point process $P_n$ for $n = 5, 10, 100, 500, 1000, 10000$ respectively with the $X_i$ exponentially distributed.

lower boundary. Denoting the intensity measure of this limiting process by $\Lambda$, it follows from the Poisson property that for the region $A = \{(t_1, t_2) \times (u, \infty)\}$:

$$\begin{aligned} exp\{-\Lambda(A)\} &= \mathrm{P}\{\text{number of points in A}\} \\ &= \mathrm{P}\{\mathrm{M_n} \leq \mathrm{x}\} \\ &\approx exp\left\{-[1+\xi x]^{-1/\xi}\right\} \ . \end{aligned} \tag{3.35}$$

Thus, at high levels, the process $P_n$ should approximate a Poisson process with intensity function given by

$$\Lambda\{(t_1, t_2) \times (x, \infty)\} = (t_1 - t_2)[1 + \xi x]^{-1/\xi} \ . \tag{3.36}$$

For statistical purposes it is useful to make two slight modifications to this representation. First, we can absorb the rescaling coefficients $a_n$ and $b_n$ into the GEV distribution and obtain that

$P_n = \{(\frac{i}{n+1}, X_i) : i = 1, \ldots, n\}$ may be approximated above high thresholds by a Poisson process with intensity function

$$\Lambda\{(t_1, t_2) \times (x, \infty)\} = (t_1 - t_2)[1 + \xi(x - \mu)/\beta]^{-1/\xi} \ . \tag{3.37}$$

Consequently, the distribution of the maximum *over the entire observation period* is GEV with distribution function

$$G(x) = \exp\left\{ - \left[ 1 + \xi \frac{(x - \mu)}{\beta} \right]^{-1/\xi} \right\} \ . \tag{3.38}$$

In applications it is more usual to parameterize the model so that the maximum over *one block* of observations has this distribution. This is easily achieved. If the $n$ observations correspond to $n_b$ blocks of data then we write the intensity function as

$$\Lambda\{(t_1, t_2) \times (x, \infty)\} = n_b(t_1 - t_2)[1 + \xi(x - \mu)/\beta]^{-1/\xi} \ . \tag{3.39}$$

This does not alter the structure of the model, but ensures that the parameters $(\xi; \mu, \beta)$ correspond to the GEV distribution of the block maximum. To summarize, the overall procedure becomes:

- Build the point process

$$P_n = \left\{ (\frac{i}{n+1}, X_i) \ : \ i = 1, \ldots, n \right\} \ .$$

- Above a high threshold $u$, $P_n$ approximates a non-homogeneous Poisson process.
- The intensity function of the process $P$ is

$$\Lambda\{(t_1, t_2) \times (x, \infty)\} = n_b(t_1 - t_2)[1 + \xi(x - \mu)/\beta]^{-1/\xi} \ .$$

The essential point is that this result gives us a characterization for *all* observations which are extreme in the sense of having exceeded a high threshold $u$. Note, that what determines a high threshold is yet to be made clear however. Thus we can incorporate all such extreme observations into the inference. The most natural approach is based on the point process likelihood. Suppose we make $n$ observations in total over $n_b$ periods. Then with a region of the form $A_v = [0, 1] \times (v, \infty)$ for $v > u$ the likelihood is given by

$$\begin{aligned} L(A_v; \xi, \mu, \beta) &= \exp\{-\Lambda(A_v)\} \prod_{i=1}^{N_{A_v}} d\Lambda(t_i, x_i) \\ &= \exp\left\{ -n_b \left( 1 + \xi \tfrac{v - \mu}{\beta} \right)^{-1/\xi} \right\} \prod_{i=1}^{N_{A_v}} \beta^{-1} \left( 1 + \xi \tfrac{x_i - \mu}{\beta} \right)^{-1/\xi - 1} \ , \end{aligned} \tag{3.40}$$

where $x_1, \ldots, x_{N_A}$ is an enumeration of the $N_A$ points that exceed the threshold $v$.

**Example: Fitting Point Processes - `xmpPOTppfit`**

The Splus function `pp.fit(xdat, ufun, nobs)` estimates the parameters $(\xi, \mu, \beta)$ for a data vector `xdat` from the point process over a threshold `ufun`.

In addition it is possible to produce a range of diagnostic plots to check the quality of the fit. Simple probability and quantile-quantile plots are calculated using the function `pp.diag(pp.object)`.

There is also the issue how to make the choice of threshold, and having made the choice, deciding whether it was satisfactory or not. There is a trade-off in threshold choice: thresholds which are too low incur bias due to the invalidity of the asymptotic argument; thresholds which are too high have few exceedances and so sampling variability is high. One (computationally intensive and consequently very slow) method of threshold verification is based on the likelihood analysis itself. The asymptotic argument presented above suggests that the model with parameters $(\xi, \mu, \beta)$ is valid at all levels above some suitably high threshold. Thus, if we fit the model at a range of thresholds, then we should observe some stability in the parameter estimates (relative to their sampling variability) at a threshold where the asymptotics are valid. The function `pp.fitrange()` plots each of the MLE's against threshold, together with confidence bands corresponding to 1.96 standard errors.

For example the plots in figure 3.19 for the NYSE Composite Index were generated by the following commands, producing this output:

```
> data <- scan(file="nyseres.csv")
> plot(data, type="l", main="log-Returns")
> u <- 0.02
> y <- data[data < -u]
> x <- (1:length(data))[data < -u]
> points(x, y, col=5)
> pp.object <- pp.fit(-data, ufun=u, nobs=252)
> print(pp.object)
    ...
> pp.diag(pp.object)
> pp.fitrange(data, umin=0.01, umax=0.04, nobs=252, nint=15)
    ...
```

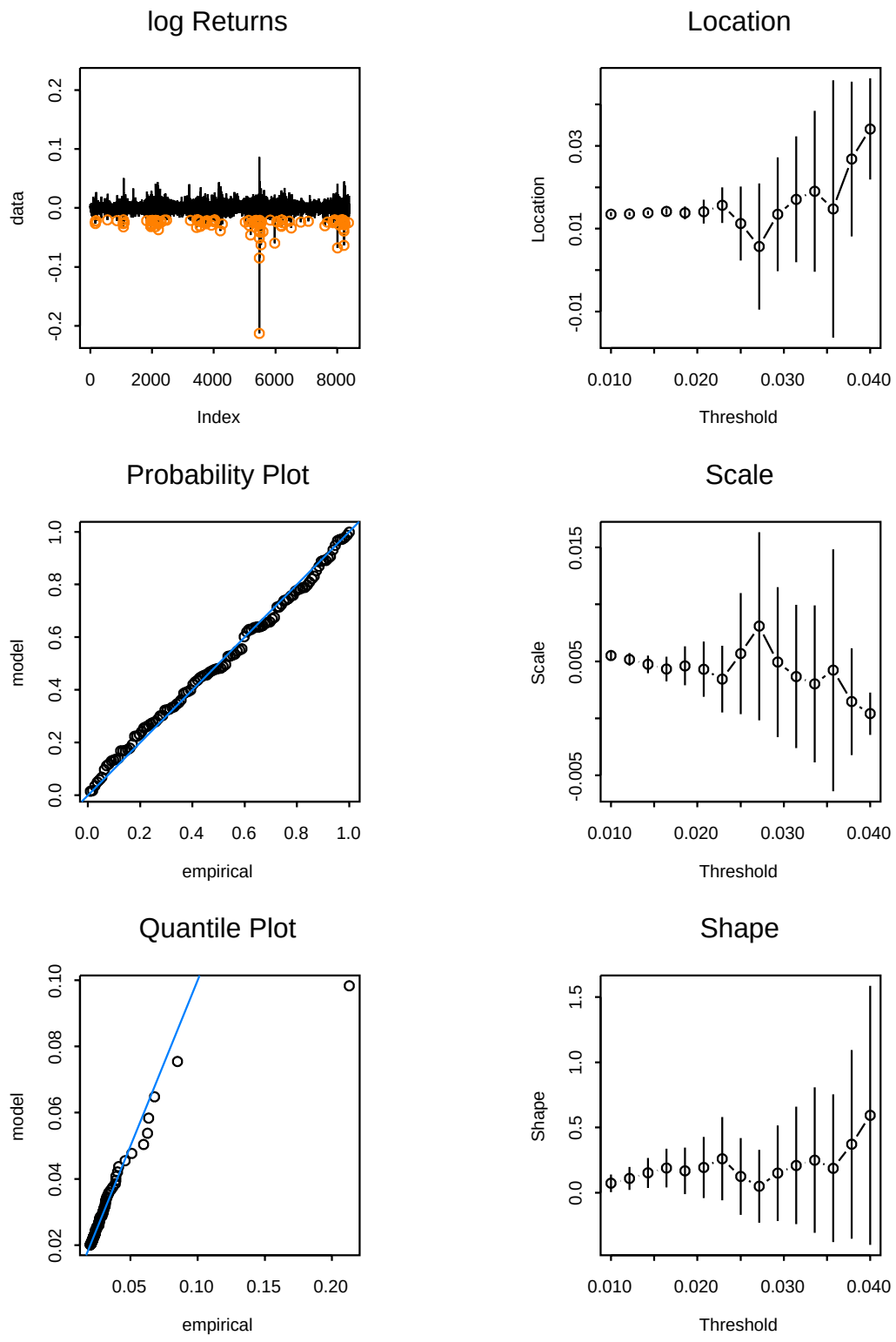
### 3.4.2 Generalized Pareto Distribution

This type of analysis derives as a simple special case of the point process approach. The idea is to obtain an explicit approximation for the conditional distribution. With the same notation as before, let $X^{\star}_{n,i} = (X_i - b_n)/a_n$ for $i = 1, \ldots, n$. Then, for $u$ sufficiently large for the Poisson limit with intensity (3.36) to be a valid approximation on $[0,1] \times (u, \infty)$,

$$\begin{aligned} P\{X^{\star}_{n,i} > u + x | X^{\star}_{n,i}\} &\approx \frac{\Lambda\{(0,1)\times(u+x,\infty)\}}{\Lambda\{(0,1)\times(u,\infty)\}} \\ &= [1 + \xi x]^{-1/\xi} \ . \end{aligned} \tag{3.41}$$

In applications, therefore, exceedances of a high threshold, $u$, can be modelled as having distribution specified by equation (3.41). Again, the unknown coefficients $a_n$ and $b_n$ can be absorbed into the distribution to give

$$P\{X_{n,i} > u + x | X_{n,i}\} = [1 + \xi \frac{x}{\beta}]^{-1/\xi} \ . \tag{3.42}$$




■ Figure 3.19: Point process analysis of the lower tail of the daily log-Returns of the NYSE Composite Index. The plots to the left show the returns together with data points exceeding from above the threshold value $u = 0.02$, the probability plot and quantile-quantile plot comparing empirical with model-based results. To the right we have plots showing the MLE's for the three parameters against threshold values.

This is referred to as the *Generalized Pareto distribution* (GPD). Taking the limit $\xi \to 0$ gives the exponential distribution as a special case. Note also that there is no location parameter in this model, as the effect of the re-location by $a_n$ is lost through the conditioning.

The role of this distribution in the context of extremes was first established by Pickands (1975) who showed that convergence of the sample maxima to a GEV distribution is equivalent to a Poisson rate of exceedances of asymptotically high thresholds, together with a GPD distribution for the excesses themselves. From a practical viewpoint, the GPD distribution gives us a model which has some asymptotic justification for the distribution of excesses above a high threshold, conditional on being above that threshold see Figure 3.X. Inference can be based directly on this model, but of course, this amounts to exactly the point process likelihood discussed above. There is, however, one characteristic of the GPD distribution which acts as a useful diagnostic in threshold selection. That is, if the conditional distribution of the X follow the GPD model (2.13), then

$$E(X - u|X > u) = \frac{\beta + \xi u}{1 - \xi} \tag{3.43}$$

provided $\xi < 1$. Moreover, if the GPD model is valid at a threshold $u$, then it should be valid at all thresholds $v$ greater than $u$. Hence, a plot of the sample mean of exceedances of $u$ against $u$ should be linear above a level at which the asymptotic argument is valid. Furthermore, the slope and intercept of that line give simple estimates of $\xi/(1 - \xi)$ and $\beta/(1 - \xi)$, respectively. This is a mean residual life plot.

**Example: Mean Residual Life Plot - `xmpPOTmrlplot`**

The Splus function `mrl.plot(data)` produces a mean residual life plot. Approximate confidence intervals are also included, 95% being the default level. Usually, the interpretation of mrl plots is not easy, since the variability in the plotted data is high.
For example the following plots were generated as follows:

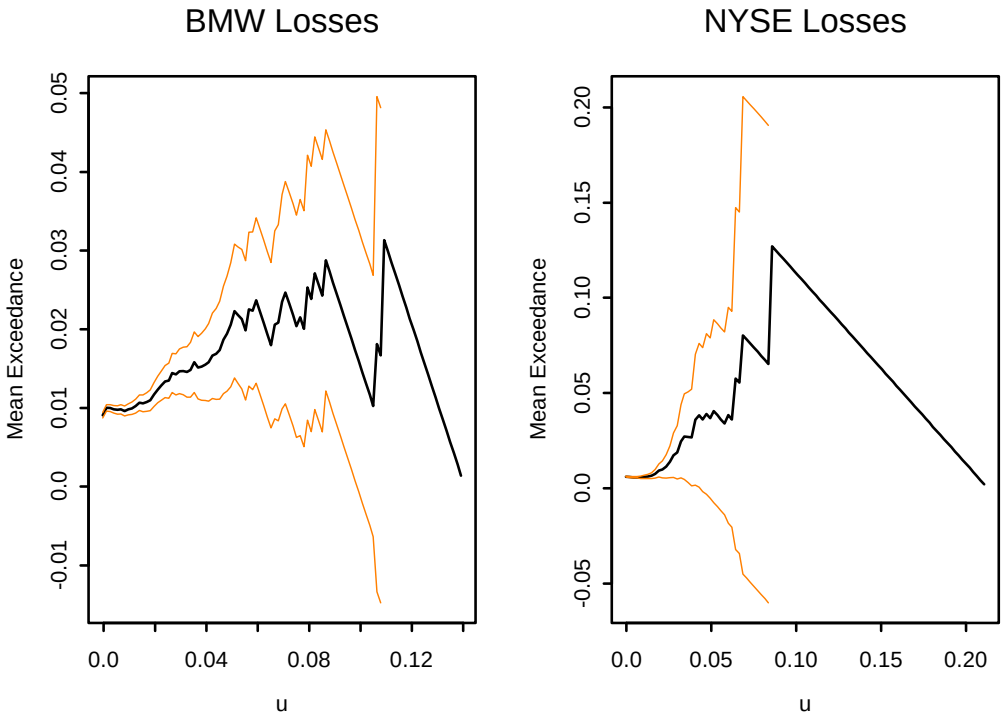
```
> x <- scan(file="nyseres.csv")
> mrl.plot(-x, main="NYSE Losses")
```

Thus from plotting the exceedances we can get the idea of a suitable threshold to fit the GPD.

McNeil (1997) has also shown how to find a closely related approximation to the distribution of the ground-up exceedances of a threshold, i.e. the excesses plus $u$, and to the tail of the original unknown underlying distribution. The distribution of the exceedances we define as $F_u(x - u)$ for $x \geq u$ and this may be approximated by $G_{\xi;\beta}(x - u) = G_{\xi;u,\beta}(x)$. For points in the tail of the distribution ($x \geq u$) we note that

$$F(x) = P\{X \leq X\} = (1 - P\{X \leq u\})F_u(x - u) + P\{X \leq u\} \ .$$

We know that we can estimate $F_u(x-u)$ by $G_{\xi;u,\beta}(x)$ for $u$ large. We can also estimate $P\{X \leq u\}$ from the data by $F_n(u)$, the empirical distribution function evaluated at $u$. Thus for $X \geq u$ we can use the tail estimate




■ Figure 3.4.3: These are the mean residual life plots for the lower (losses) tails of the log returns of the BMW stock and the NYSE Composite index values.

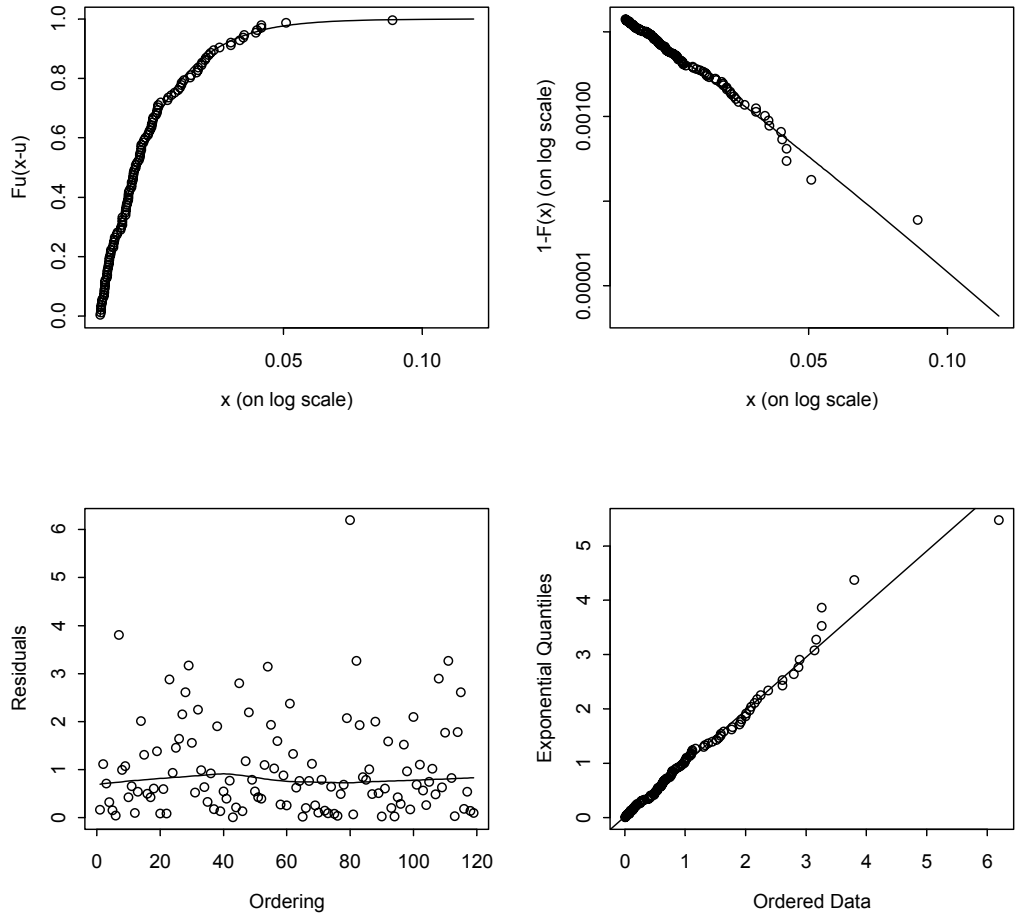
$$\widehat{F(x)} = (1 - F_n(u))G_{\xi;u,\beta}(x) + F_n(u)$$

to approximate the distribution function $F(x)$. It can be shown that $\widehat{F(x)}$ is also a generalized Pareto distribution, with the same shape parameter $xi$, but with scale parameter $\widetilde{\beta} = \beta(1 - F_n(u))^x i$ and location parameter $\widetilde{\mu} = u - \widetilde{\beta}((1 - F_n(u))^{-\xi} - 1)/\xi$.

**Example: Fitting the GPD and the Tail - `xmpPOTgpd`**

The Splus function `gpd.fit()` fits the GPD. Additionally `gpd.diag()` plots the GPD and the tail in comparison to the estimated distributions, and displays a residual and QQ-plot

```
> x <- scan(file="nyseres.csv")
> result <- gpd.mle(x, u=0.02)
> qpd.diag(result)
[1] "Excess Distribution"
[1] "threshold = 0.02   xi = 0.202   scale = 0.00537   location = 0.02"
[1] "Tail of Underlying Distribution"
[1] "threshold = 0.02   xi = 0.202   scale = 0.00228   location = 0.00466"
```




■ Figure 3.4.4: Fitting the GPD (upper left) and tail (upper right) above threshold $u = 0.02$ for the NYSE Composite Index log-returns. The lower graphs show residual and QQ-plot.

### 3.4.3 The $r$-Largest Order Statistics Method

This technique can be derived as a special case of the point process approach. Given a sequence of *iid* variables $X_i, X_2, \ldots$ the block maximum method is based on the limiting distribution of $(M_n - b_N)/a_n$, the normalized maximum order statistic.

Defining $M_n^{(i)}$ to be the $i$-th largest order statistic, the technique here is to obtain the limiting joint distribution of

$$\left( \frac{M_n^{(1)} - b_n)}{a_n} \ , \quad \frac{M_n^{(2)} - b_n)}{a_n} \ , \quad \ldots \ , \quad \frac{M_n^{(r)} - b_n)}{a_n} \right) \tag{3.44}$$

for sone choice of $r$. Setting $u = M_n^{(r)}$ in the likelihood ([**?**]) gives the likelihood

$$L = \exp\left\{ -\left(1 + \xi \frac{M_n^{(r)} - \mu}{\beta}\right)^{-1/\xi} \right\} \prod_{i=1}^{r} \beta^{-1} \left(1 + \xi \frac{M_n^{(i)} - \mu}{\beta}\right)^{-1/\xi - 1} . \tag{3.45}$$

Again, because this result derives directly from the point process representation, as an inferential procedure we may just as well use the point process likelihood (**??**) directly, in this case setting the threshold at the $r$-th order statistic.

Finally, note that using the $r$-largest method wit $r = 1$ is equivalent to the GEV block maxima method.

### Notes and Comments

This section discusses the point process characterization as described in Chapter 2 *Alternative Characterizations - Improved Inferences* of Coles' and Tawn's review article *Statistical Methods for Extreme Values*.

The Splus functions in the `fExtremes` library were implemented from Coles' and McNeil's software packages.

## 3.5 The Extremal Index

*Assume a dyke has to be build at the seashore to protect against floods with 95% certainty for the next 100 years. Suppose it has been established that the 99.9 and 99.95 percentiles of the annual wave-height are 10m and 11m, respectively. If the annual maxima are believed to be iid then the dyke should be 11m high,* $0.9995^{10} \approx 0.95$. *But if the annual maxima are stationary (temporal dependent with extremal index* $\theta = 0.5$*, then one can show that a height of 10m is sufficient,* $0.999^{0.5*100} \approx 0.095$. *How do we can understand this result?*

*Source: Embrechts et al. (1997) taken from Weismann.*

### Introduction

The study of extremes in the presence of temporal dependence has focused on two types of dependence: (i) Long-range dependence and (ii) Short-term dependence. In applications, short-term dependence at extreme levels is common, with a tendency for extreme events to cluster. On the other hand long-range dependence is in general sufficiently weak not to affect the asymptotics of extreme value analysis.

Until now independence among observations was assumed. In practice however time series are often characterized by serial correlations. The maxima tend to cluster. In order to quantify the effect of serial dependence the extremal index $\theta$ is defined as

$$\theta = lim_{n\to\infty} P\{max(X_2, \ldots X_{p_n} \leq u_n | X_1 \geq U_n\} \ ,$$

where $p_n = o(n)$ and the sequence $u_n$ is such that $P\{M_n \leq u_n\}$ converges. The correct interpretation is that $\theta$ is the reciprocal of the mean cluster size at asymptotically extreme levels, but more loosely it is just the reciprocal of the mean cluster size. The following theorem states the effect of the extremal index $\theta$ on the asymptotics of maxima. In this theorem the quantity $M_n^*$ denotes the maximum of a sequence of independent variables $X_1^*, \ldots, X_n^*$ with the same marginal properties as the original $X_i$, $(i = 1, \ldots, n)$, series.

**Theorem:**

If $X_1, X_2, \ldots$ is a stationary process with no long range dependence and with extremal index $\theta > 0$, then

$$P\left\{\frac{M_n - b_n}{a_n} < x\right\} \to G_1(x)$$

if and only if

$$P\left\{\frac{M_n^* - b_n}{a_n} < x\right\} \to G_2(x)$$

with $G_1$ and $G_2$ non degenerate and

$$G_1(x) = G_2^{\theta}(x) \ .$$

From the theorem it is concluded that the asymptotic distribution of the maxima of a serial correlated sequence is still a GEV distribution. The extremal index $\theta$ only affects the location and scale parameters but not the shape parameter. Hence the type of the limiting distribution is unaffected.

### 3.5.1 The Extremal Index: Blocks Method

Starting from the definition of the extremal index $\theta$, we have

$$P(M_n \leq u_n) \ \approx \ P^{\theta}(\widetilde{M_n} \leq u_n) \ = \ F^{\theta n}(u_n) \ , \tag{3.46}$$

provided $n\overline{F}(u_n) \to \tau > 0$. Hence

$$\lim_{n\to\infty} \frac{P(M_n \leq u_n)}{n \ln F(u_n)} \ = \ \theta \ . \tag{3.47}$$

This simple limit relation suggests constructing naive estimators of $\theta$. Since we do not know $F(u_n)$ and $P(M_n \leq u_n)$ these quantities have to be replaced by estimators. A candidate for estimating the tail $n\overline{F}(u_n)$ is its empirical version

$$\frac{N}{n} \ = \ \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i > u_n\}} \ . \tag{3.48}$$

To find an empirical estimator for $P(M_n \leq u_n)$ is not straightforward. However, assuming a specific type of asymptotic independence (see Embrechts et al. 1997, Chapter 4.4) implies

$$P(M_n \leq u_n) \ \approx \ P^k(M_{[n/k]} \leq u_n) \tag{3.49}$$

for constant $k$ or slowly increasing $k = k(n)$. This approximation forms the basis for the *blocks method*. For the sake of arguments assume that $n = rk$ for integers $r = r(n) \to \infty$ and $k = k(n) \to \infty$. Otherwise, let $r = [n/k]$. This divides the sample $X_1, \ldots, X_n$ into $k$ blocks of size $r$:

$$X_1, \ldots, X_r; \ldots; X_{(k-1)r+1}, \ldots, X_{kr} \ . \tag{3.50}$$

For each block we calculate the maximum

$$M_r^{(i)} \ = \ \max(X_{(i-1)r+1}, \ldots, X_{ir}) \ , \quad i = 1, \ldots, k \ . \tag{3.51}$$

Relation (3.49) then suggests the approximation

$$P(M_n \leq u_n) \;=\; P\left(\max_{1\leq i\leq k} M_r^{(i)} \leq u_n\right) \;\approx\; P^k(M_r \leq u_n) \;\approx\; \left(\frac{1}{k}\sum_{i=1}^{k} I_{M_r^{(i)}\leq u_n}\right)^k \;=\; \left(1-\frac{K}{k}\right)^k \;. \tag{3.52}$$

A combination of these heuristic arguments with (.) leads to the following estimator of $\theta$:

$$\hat{\theta}_n^{(1)} \;=\; \frac{k}{n}\frac{\ln(1-K/k)}{\ln(1-N/n)} \;=\; \frac{1}{r}\frac{\ln(1-K/k)}{\ln(1-N/n)} \;. \tag{3.53}$$

Here $N$ is the number of exceedences of $u_n$ by $X_1, \ldots, X_n$ and $K$ is the number of blocks with one or more exceedences. A Taylor expansion argument yields the estimator

$$\hat{\theta}_n^{(2)} \;=\; K/N \;=\; \frac{1}{r}\frac{K/k}{N/n} \;\approx\; \hat{\theta}_n^{(1)} \;. \tag{3.54}$$

The blocks method accounts for *clustering* in the data. If the event

happens one says that a cluster occured in the $i$th block. These events characterize the extremal behavior of $X_n$ if we assume that the size $r(n)$ of the blocks increases slowly with $n$. This gives us some feeling for the dependence structure in the sequence $X_n$. In this case, the extremal index is a measure of the clustering tendency of high threshold exceedances in a stationary sequence.

In practice, it turns out (as in the case of the estimation of the index $xi$) that it is not easy to obtain accurate estimates of $\theta$. Especially, given that we have $n$ observation, how do we choose the values of $r$ (or $k$) and $u_n$?

### 3.5.2 The Extremal Index: Reciprocal Mean Cluster Size

An alternative approach yields the interpretation of the extremal index as the reciprocal of the mean cluster size []. This interpretation of $\theta$ suggests an estimator based on the blocks method:

$$\hat{\theta}_n^{(2)} \;=\; \frac{\Sigma_{i=1}^{k} I_{M_r^i > u_n}}{\Sigma_{i=1}^{n} I_{\{X_i > u_n\}}}\frac{K}{N} \;, \tag{3.55}$$

i.e. number $K$ of clusters of exceedances divided by the total number $N$ of exceedances. The same estimator has already been suggested as an approximation to $\hat{\theta}_n^{(1)}$.

### 3.5.3 The Extremal Index: Runs Method

...

$$\hat{\theta}_n^{(3)} \;=\; \frac{\Sigma_{i=1}^{n-r} I_{A_{i,n}}}{\Sigma_{i=1}^{n} I_{\{X_i > u_n\}}} \;=\; \frac{\Sigma_{i=1}^{n-r} I_{A_{i,n}}}{N} \;, \tag{3.56}$$

where

$$A_{i,n} \;=\; \{X_i > u_n, X_{i+1} \le u_n, \ldots, X_{i+r} \le u_n\}\ . \tag{3.57}$$

This means we take any sequence of $r = r(n)$ consecutive observations below the threshold as separating two clusters.

**Example: Extremal Index:** `xmpEXindex1`

Plot the values of the extremal index estimates $\hat{\theta}_n^{(1,2,3)}$ as a function of quantiles ranging between 99% and 99.9%. Select as data samples a large (80'000 points) and a short (8'000 points) series of random variables drawn from Student's distribution with four degrees of freedom together with their exponentially smoothed averages (EMA created dependencies with $\lambda = 0.2$). As real world examples consider the BMW stock and NYSE composite index returns.

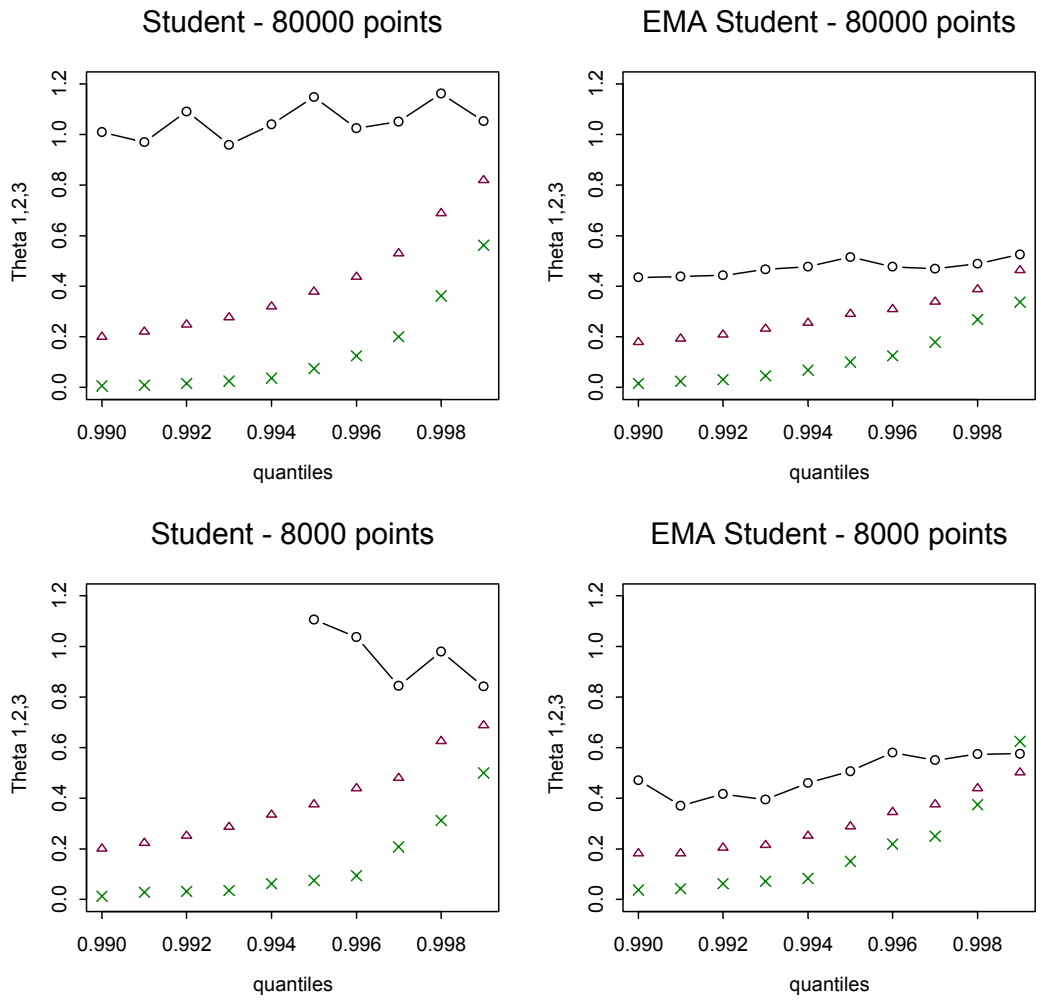
Use the fExtreme library function

```
exindexes(x, blocklength, quantiles=seq(0.990,0.999,0.001),
  doprint=T, doplot=T, ...},
```

where `x` is a numeric vector of input data $X$ (in finance usually the log-returns or residuals of a time series), `blocklength` is a numeric value specifying the length of the blocks (from which the number of blocks $r$ can be determined), `quantiles` is a numeric vector characterizing the desired tail depths, by default ranging in ten equal steps from 99.0% to 99.9% (from which the thresholds $u_n$ can be determined). Note, setting the `blocklength` to 1, allows block maxima data as input. Furthermore, replacing the data vector `x` by `-x` allows for the investigation of the lower tail of the distribution.

A look into the Splus function shows us how the estimates are evaluated:

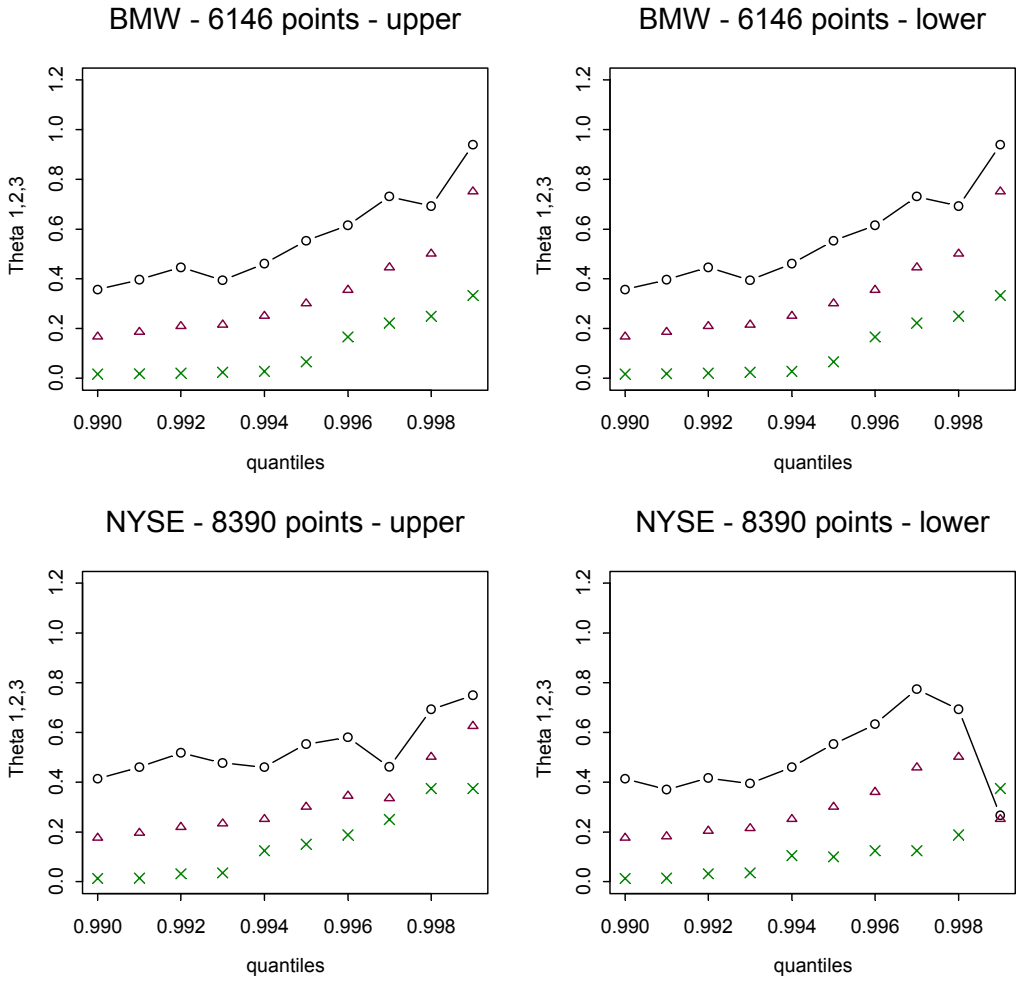
```
...
k <- floor(length(x)/blocklength)            # Number of blocks
n <- k*blocklength                           # Number of data points
x <- x[1:(k*blocklength)]                    # Truncate rest of data
x <- matrix(x, ncol=blocklength, byrow=T)    # Arrange in matrix form
ordered <- sort(x)                           # Sort in reverse order
u <- ordered[floor(quantiles*length(x))]     # Determine thresholds
...
# Now calculate extremal index for each threshold u:
    ...
    N <- length(x[x>z])                      # Number of exceedances
    K <- sum(sign(apply(x, 1, max)-z)+1)/2   # Blocks with exceedances
    ...
    theta1 <- (k/n)*log(1-K/k)/log(1-N/n)    # Theta (1)
    theta2 <- K/N                            # Theta (2)
      t <- diff((1:n)[x>z])
    theta3 <- length(t[t>blocklength])/N     # Theta (3)
    ...
```




■ The four plots show the extremal indexes for a long (80'000 points) and a short (8'000 points) series of Student t-distributed (4 degrees of freedom) ranom variables (left graphs), together with the EMA smoothed series ($\lambda = 0.2$), graphs on the right side. This example demonstrates the influence of dependencies on the extremal index created by smoothed data records. We also see that it becomes rather difficult to estimate the extremal index for shorter time series. One also observes the fact that usually $\theta^1$ is larger than $\theta^2$, and the latter is larger than $\theta^3$.

Remark: The EMA filtered series $\overline{x}_i = \lambda x_i + (1-\lambda)\overline{x_{i-1}}$ were calculated with the help of the standard Splus function `filter()` as:

```
lambda <- 0.2
xbar <- x * lambda
xbar[1] <- x[1]
xbar <- filter(xbar, filter=(1-lambda), method="rec")
```




■ The four plots show the extremal index for the logarithmic returns of the BMW stock prices (upper row) and the NYSE composite index (lower row) for both, the upper (to the left) and lower tail (to the right).

A different point of view in plotting the extremal index takes the Splus function `exindex()` from the EVIS software package of A. McNeil []:

**Example: Extremal Index:** `xmpEXindex2`

Plot the values of the extremal index estimate $\hat{\theta}_n^{(1)}$ as a function of thresholds $u$ and $K$ for the BMW stock and NYSE composite index returns.

Use the fExtreme library function

```
exindex(data, block, start = 5, end = NA, reverse = F,
  auto.scale = T, labels = T, table = T, ...),
```

where `data` is a numeric vector of input data or a daily time series, `block` is either an integer value specifying the number of blocks $r$ (from which the blocklengths can be determined) or a string variable taking one of the values `month|quarter|year` if `data` is a time series. In this case the Splus function

```
blocks(data, block = "quarter", func = max)
```

allows to determine the block maxima.

The output produced by the function `exindex()` for the lower tail in the case of the NYSE data is

```
        N  K         un     theta2    theta1
 [1,]  15  5 0.03386695 0.33333333 0.3892907
 [2,]  16  6 0.03330663 0.37500000 0.4561044
 [3,]  18  7 0.03194501 0.38888889 0.4941326
 [4,]  19  8 0.03149290 0.42105263 0.5610423
 [5,]  21  9 0.03128161 0.42857143 0.6015455
 [6,]  22 10 0.03113046 0.45454545 0.6758827
 [7,]  25 11 0.03015246 0.44000000 0.6979819
 [8,]  39 12 0.02667477 0.30769231 0.5253125
 [9,]  72 13 0.02259791 0.18055556 0.3357635
[10,]  87 14 0.02168161 0.16091954 0.3328211
[11,] 162 15 0.01794776 0.09259259 0.2195216
[12,] 181 16 0.01727691 0.08839779 0.2598157
```

## Notes and Comments

The presentation of the theoretical background of the extremal index was taken from Chapter 8.1 from the book of Embrechts et al. (1997). Further references to this material ... .

The Splus function `exindexes()` was written by D. Würtz and the functions exindex() and blocks() were implemented from the EVIS software package of A. McNeil.

## 3.6 Risk Management

### Introduction

Sometimes later ...

### 3.6.1 Value at Risk

A good introduction to value at risk is given by the *Risk Grades Technical Document* written by Kim and Mina (2000) document. More detailed material can be found in the original Resmetrics Documentations from JP Morgan. A further source is the PhD thesis of ...

### 3.6.2 Stress Testing

A brief introduction to stress testing provides the article *Stress Testing* written by Simozar (1998). To get insight into into current practice we refer to the paper *Stress Testing by Large Financial Institutions: Current Practice and aggregation Issues* published by the Bank for International Settlemets, BIS. For further reading we mention the reports *A Coherent Framework for Stress Testing* written by Berkowitz from the US Federal Reserve and *Stress Testing* edited by Grau from the Austrian National Bank.

### 3.6.3 Extremes and Implied Risk Management

This is also the title of book edited by Embrechts (2000) with 18 contributions from several authors. The book covers in four chapters basic extreme value theory, risk measures and extreme value theory, applications to finance, and applications to insurance. A detailed description of ... can be found in the diploma thesis of Frahm ().

...

## 3.7 The fExtremes Library

### 3.7.1 Summary of Splus Function

The following section gives an overview over the Splus functions available in the fExtremes Library. The programs are grouped by their functionalities. A short description follows each Splus function name.

#### Exploratory Data Analysis Functions

Standard Splus offers the qq.plot() function for the investigation of extremal values. The fExtremes library adds the following functions:

```
mxf.plot            Mean excess function plot
mrl.plot            Mean residual life plot
records.plot        Plot of Records
maxsumratio.plot    Ratio of Maximum and Sum Plot
```

#### Distribution Functions

Several additional density, cumulative probability, and quantile functions as well as random number generators which play important roles in the extremal value theory can be accessed through the fExtremes library. These functions include:

```
dfrechet            Density function for the Frechet distribution
pfrechet            Probability function
qfrechet            Quantile function
rfrechet            Random variables

dweibl              Density function for the Weibull distribution
pweibl              Probability function
qweibl              Quantile function
rweibl              Random variables

dgumbel             Density function for the Gumbel distribution
pgumbel             Probability
qgumbel             Quantile function
rgumbel             Random variables

dgev                Density function for the GEV distribution
```

```
pgev                Probability
qgev                Quantile function
rgev                Random variables

dgpd                Density function for the GPD distribution
pgpd                Probability
qgpd                Quantile function
rgpd                Random variables
```

### Fluctutations of Maxima: GEV

For an investigation of the fluctuations of maxima the `fExtreme` library implements tools to fit the GEV distribution via the l-moments approach and via the maximum likelihood estimation procedure. Additionally functions are available to find the shape parameter from Pickands', Hill's and Decker-Einmahl-deHaan's estimators.

```
gev.pwm             PWM parameter fit to a GEV distribution

gev.fit             MLE parameter fit to a GEV distribution
gev.diag            Graphical diagnosis of the GEV fit
gev.profxi          Examination of the profile likelihood for xi
gev.profq           Examination of the profile likelihood for q_p
gev.res

hill.plot           Plot Hill's Estimator
shaparm.plot        Pickands, Hill and Decker-Einmahl-deHaan Estimator

block.maxima        Utility to calculate block maxima
```

### Extremes via Point Processes: GPD

To explore extremes via point processes the `fExtreme` library has functions for the peak over threshold parameter fit and for the mean excess function fit.

```
pp.fit              Peak over threshold parameter fit
pp.diag             Diagnostic Analysis - Probability and QQ-plots
pp.fitrange         Plot each of the MLE's against threshold

mxf.fit             Mean Excess function fit to xi
```

### Extremal Index

For the calculation of the extremal index two Splus functions are offered by the `fExtremes` library:

```
exindex             Extremal Index
exindexes           Extremal Indexes
```

## 3.7.2 List of Splus Datasets

```
nyseres.csv         log returns of NYSE Composite index values
bmwres.csv          log returns of BMW stock prices
```

### 3.7.3 List of Splus Examples

The material presented is illustrated by several Splus examples In the following we give a full list of all examples together with a short description:

```
xmpEDAqqplot[123]      3.2  Quantile-quantile plots
xmpEDAmxf[123]         3.2  Mean excess function plots
xmpEDArecords          3.2  Records plot
xmpEDAmaxsum           3.2  Ratio of maximum and sum plot

xmpGEVdistributions    3.3  Gumbel, Frechet, Weibull Ddistributions
xmpGEVreturnlevels     3.3  Return levels
xmpGEVconvergence      3.3  Convergence of exponential and Gaussian maxima
xmpGEVpwm[12]          3.3  GEV probability weighted moments fit
xmpGEVhillplot         3.3  Hill plot
xmpGEVshaparm          3.3  Shape parameter summary plots
xmpGEVmle.ssc          3.3  GEV maximum likelihood estimation

xmpPOTppfit            3.4  Fitting Ppoint processes
xmpPOTmrlplot          3.4  Mean residual life plot
xmpPOTgpd              3.4  Fitting the FPD and the tail

xmpEXindex             3.5  Extremal Index
xmpEXindexes           3.5  Extremal Index
```

### 3.7.4 Software Packages

The `fExtremes` was written by the author implementing beside his own functions additional functions and routines written by others. This includes McNeils' EVIS Splus package, a collection of Splus functions written by S. Coles and the Fortran package on L-moments from Hosking.

#### Mc Neil's Splus Package EVIS:

Alexander McNeil, ETH Zurich, EVIS Software for Extreme Values in S-Plus: Working with Paul Embrechts at ETH Zurich A. McNeil has developed SPlus functions for extreme value data analysis. The central activities are: fitting the generalized extreme value distribution (GEV) to block maxima/minima; fitting the generalized Pareto distribution (GPD) to excesses over high thresholds; fitting a Poisson point process with GPD marks to the point process of threshold exceedances. Also included are a number of useful exploratory data analysis functions.

Source available from:
*http://www.math.ethz.ch/ mcneil/*

#### Coles' Splus Package:

Stuart Coles Home Page: Extreme value course notes. This is a link to a manuscript on extreme value theory and modelling. Both the notes and associated Splus code have been revised (as of 9/6/99) to include the fitting of arbitrary covariates in extreme value models. The text includes a recently revised chapter on multivariate extremes, written also with Jan Currie and Jonathan Tawn.

Source available from:
*http://www.stats.bris.ac.uk/ masgc*

**Hosking's FORTRAN Library L-moments:**

An archive of 73 FORTRAN subroutines for statistical analysis using L-moments, and some auxiliary routines used by the L-moment routines. Submitted by M. Handcock, handcoc@ibm.com) [2/Feb/90] to the Stat Library.

Source available from:
*http://www.stat.cmu.edu/general/lmoments1.0*

# Bibliography


[1] Sorry, not yet complete ...

[2] Bassi F., Embrechts P., Kafetzaki M., (1999), *A survival kit on quantile estimation*, Preprint.

[3] Bauer C., (2000), *Value at Risk Using Hyperbolic Distributions*, Journal of Economics and Business 52, 455-467.

[4] Bensalah Y., (2000), *Steps in Applying Extreme Value Theory to Finance: A Review*, Preprint, Bank of Canada Working Paper, 22 pages.

[5] Casson E., Coles S. (2000) *Spatial Regression Models for Extremes*, Preprint 2000, 21 pages.

[6] Coles S., (1999), *Extreme Value Theory and Applications*, Preprint 1999, 78 pages.

[7] Coles S., Tawn J., (1998), *Statistical Methods for Extreme Values*, Preprint 1999, 78 pages.

[8] Coles S., Currie J., Tawn J., (1999), *Dependence Measures for Extreme Value Analysis*, reprint 1999, 25 pages.

[9] Dacorogna M.M., Müller U.A., Pictet O.V., Vries C.G. de, (2000), *Extremal Forex Returns in Extremely Large Data Sets*, Preprint, 49 pages.

[10] Diebold F.X., Schuermann T., Stroughair J.D, (2000), *Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management*, Preprint, 10 pages.

[11] Dixon M.J, Ledford A.W., Marriott P.K., (1998) Finite Sample Inference for Extreme Value Distributions, Preprint, 16 pages.

[12] Ebnoether S., Vanini P., McNeil A.J., Antolinez-Fehr P., (2001), *Modelling Operational Risk*, Preprint.

[13] Embrechts P., Klüppelberg C., Mikosch Th. (1997), *Modelling Extremal Events*, Springer Verlag, Berlin.

[14] Embrechts P., McNeil A.J., Straumann D., (1999), *Correlation: pitfalls and alternatives*, RISK, May 1999, pages 69-71.

[15] Embrechts P., Lindskog F., McNeil A.J., (2001), *Modelling dependence with copulas and applications to risk management*, Preprint.

[16] Embrechts P., McNeil A.J. and Straumann D., (2002), *Correlation and dependence in risk management: properties and pitfalls*, In Risk management: value at risk and beyond, edt. by M. Dempster, publ. by Cambridge University Press, Cambridge.

[17] Fisher R.A., Tippett L.H.C. (1928), *Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample*, Proceedings of Cambridge Philosophical Society, 24, 180-190.

[18] Frey R., McNeil A.J., (2001), *Modelling Dependent Defaults* Preprint.

[19] Frey R., McNeil A.J., (2001), *VaR and expected shortfall in portfolios of dependent credit risks: conceptual and practical insights*, Preprint.

[20] Frey R., McNeil A.J., Nyfeler M., (2001), *Copulas and credit models*, RISK, October 2001: pages 111-114.

[21] Gencay R., Selcuk F., (2001), Overnight Borrowing, Interest Rates and Extreme Value Theory, Preprint, 26 pages.

[22] Gnedenko B. (1943), *Sur la Distribution Limite du Terme Maximum d'une Série Aléatoire*, Annals of Mathematics, 44, 423-453.

[23] Haan L. de, Resnick S.L, Rootzen H., de Vries C.G., (1989), *Extremal Behaviour of Solutions to a Stochastic Difference Equation with Applications to ARCH Processes*, Stochstic Processes and their Application, 65, 171-185.

[24] Haffernan J.E., (2001), *A Directory of Coefficients of Tail Dependence*, Preprint, 11 pages.

[25] Hosking J.R.M., (1990), *L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics*, Journal of the Royal Statistical Society, Series B, 52, 105-124.

[26] Hosking J.R.M., (1990), *Fortran Routines for the Use with the Method of L-Moments*, Preprint, 33 pages.

[27] Hosking J.R.M., Wallis J.R., (1990), *Regional Frequency Analysis: An Approach Based on L-Moments*, Cambridge University Press, Cambridge, UK.

[28] Kellezi E., Gilli M., (2000), *Extreme Value Theory for Tail-Related Risk Measures*, Preprint, Computational Finance 2000 Conference London, 21 pages.

[29] Lindskog F., McNeil A.J., (2001), *Common Poisson shock models: applictions to insurance and credit risk modelling*, Preprint.

[30] Lindskog F., McNeil A.J., Schmock U., (2001), *Kendall's tau for elliptical distributions*, Preprint.

[31] Longerstaey J., Zangari P., Finger C.C., Howard S., *RiskMetrics - Technical Document*, J.P. Morgan, New York, 4th edition.

[32] McNeil A.J., Saladin T., (1997), *The peaks over thresholds method for estimating high quantiles of loss distributions*, Proceedings of 28th International ASTIN Colloquium.

[33] McNeil A.J., (1998), *On Extremes and Crashes*, RISK, January 1998: page 99.

[34] McNeil A.J., (1998), *Estimating the tails of loss severity distributions using extreme value theory*, ASTIN Bulletin, 27, 117-137.

[35] McNeil A.J., (1998) *Calculating quantile risk measures for financial time series using extreme value theory.* Preprint, ETH E-Collection.

[36] McNeil A.J. (1999), *Extreme value theory for risk managers*, Internal Modelling and CAD II published by RISK Books , 93-113.

[37] 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, 271-300.

[38] McNeil A.J., Saladin T., (2000), *Developing scenarios for future extreme losses using the POT method*, In Extremes and Integrated Risk Management , edt. by P. Embrechts, publ. by RISK books, London.

[39] McNeil A.J., (2000), *Reading the Riskometer*, In Extremes and Integrated Risk Management, edt. by P. Embrechts, publ. by RISK books, London.

[40] Müller, U.A. Dacorogna M.M., Pictet O.V., (1996), Heavy Tails in High-Frequency Financial Data, Preprint, 23 pages.

[41] Parisi F., (2000), *Extreme Value Theory and Standard & Poor's Ratings*, ABS Research, Preprint

[42] Pickands J. (1975), *Statistical Inference Using Extreme Order Statistics*, Annal. Statist. 3, 119-131.

[43] Pictet O., Dacorogna M.M., Müller U.A., (1996), Hill Bootstrap and Jackknife Estimators for Heavy Tails, Preprint, 31 pages.

[44] Resnick S.I, (1997), *Discussion of the Danish Data on Large Fire Insurance Losses*, ASTIN Bulletin 27(2), 139-151.

[45] Resnick S.I, Starica C., (1996), *Tail Index Estimation for Dependent Data*, Preprint TR1174.

[46] Resnick S.I, Starica C., (1997), *Smoothing the Hill Estimator*, Advanced Applied Probability 29, 271-293.

[47] Rockinger M., Jondeau E., (2001), *Conditional Dependency of Financial Series: An Application to Copulas*, Preprint, 32 pages.

[48] Smith R.L, (2000), *Measuring Risk with Extreme Value Theory*, Preprint, 22 pages.

[49] Smith R.L, (2000), *Extreme Value Statistics in Meteorology and the Environment*, Preprint, Chapter 8, pages 701-768.

[50] Starica C., Pictet O., (2001), *The tales the tails of GARCH(1,1) processes tell*, Preprint, 18 pages.

[51] Thas O., Vanrolleghem P., Kops B., Vooren L. van, Ottoy J.P., (1999), *Extreme Value Statistics: Potential Benefits in WaterQuality Management*, Preprint.

[52] Wagner N., Marsh T.A., (2000), *On Adaptive Tail Index Estimation for Financial Return Models*, Preprint, 28 pages.