Extremes and Risk Management

M. Ivette Gomes
CEAUL and DEIO (FCUL), University of Lisbon

Abstract. Extreme Value Theory (EVT) plays a relevant role in the area of financial risk modelling. As an example, the Value-at-Risk (VaR) is a parameter of extreme or even rare events: a high quantile of the loss distribution of a portfolio of investments over a fixed period of time. In these notes, after an introduction to the VaR problem, we shall review a few concepts of EVT, together with its application to financial time series, we shall discuss some of the peculiarities of financial data, advancing with the reference to models which capture those peculiarities.

1 Specific features of the data: an introduction

The statistical properties of financial data have revealed to be a challenge both to managers and statisticians. Since there are a large variety of financial data, at a first sight it appears difficult to say something about possible common properties. However, a large variety of financial data, which we shall in the sequel denote either X_t or P_t, $t = 0, 1, 2, \cdots$, where t may be minutes, hours, days, and so on, exhibits similar properties, after the transformation,

$$R_{t+1} = 100 \times (\ln P_{t+1} - \ln P_t)\,, \ t \geq 0.$$

Also, since $(P_{t+1} - P_t)/P_t$ is usually small, we may write,

$$R_{t+1} = 100 \times \ln\left(1 + \frac{P_{t+1} - P_t}{P_t}\right) \approx 100 \times \frac{P_{t+1} - P_t}{P_t} =: S_{t+1}, \ t \geq 0.$$

The resulting R_t is the time series of *Log-Returns* (sometimes called merely *Returns*) and S_t is the time series of *Relative Returns*. As mentioned by several authors (see for instance Embrechts et al. (1997) and Finkenstädt and Rootzén (2004), among others), $\{R_t\}$ can be modelled by a stationary stochastic process if we do not work either at too large or too small time scales (although such an assumption may be also criticized), and there are some "stylized facts" of real life returns, among which we mention:

1. *Heavy-tailed* marginal distributions, i.e., the marginal model F is such that both $F(x)$ as $x \to -\infty$, and $1 - F(x)$ as $x \to +\infty$, approach zero at a slow rate.

2. *Clusters of exceedances of high levels*, i.e., the point process of exceedances of high levels may be no longer a Poisson process but a compound Poisson process.

3. *Long-range dependence* (LRD), i.e., a slow decay of the autocorrelation function.

4. *Aggregational Gaussianity*, i.e., the distribution of log-returns becomes closer and closer to the normal model as the periods of time increase from hourly to monthly, for instance.

In these notes we shall briefly speak about the notion of *Risk* and *Risk Factors*, in section 2. In section 3, we shall briefly review some preliminary *Data Modelling* topics, and in section 4 we shall refer a few concepts in the general field of *Statistics of Extremes*. Since empirical studies of financial data suggest that there are two properties almost always present in the data, *Heavy Tails* and *Long-Range Dependence*, we shall discuss these concepts in Section 5. Finally, in section 6 we review some of the dependent models which try to capture the empirical evidence provided by the analysis of financial data, like the $ARCH$ (*Auto-Regressive Conditionally Heteroscedastic*) and the $GARCH$ (*Generalized Auto-Regressive Conditionally Heteroscedastic*). These models have heavy-tailed margins, exhibit LRD and exceedances of high levels appear in clusters.

2 Introduction to risk and notation

To model *Risk* we often use the language of *Probability Theory*. Risks are represented by random variables (r.v.'s), usually denoted X. These r.v.'s map unforeseen future states of the world into values representing *profits* and *losses*.

Given a portfolio, let X_t denote its value at time t. If we consider the time period $[t, t+1]$, the value X_{t+1} at the end of the time period is unknown to us, and the distribution of $X_{t+1} - X_t$ is known as the *Profit-and-Loss* or $P\&L$ distribution. We are often interested in *Losses*, and we often denote the *Loss* by

$$L_{t+1} = -(X_{t+1} - X_t),$$

so that Losses are positive and Profits negative. But this is just a convention so that we are supposed to pay attention to the right tail, instead of the left tail.

If we think on a portfolio of d stocks, let n_i denote the number of shares in stock i at time t, and let $P_{t,i}$ denote price. Following the standard convention, we shall use the logarithmic prices as *risk factors*,

$$Z_{t,i} = \log P_{t,i}, \quad 1 \leq i \leq d,\ t \geq 0.$$

As said before, the *risk factor changes*, given by

$$X_{t+1,i} = \log P_{t+1,i} - \log P_{t,i}, \quad 1 \leq i \leq d,\ t \geq 0,$$

correspond to the so-called *log-returns* of the stock. We then have a random vector of d risks, generally denoted $\mathbf{X}_t = (X_{t,1}, X_{t,2}, \cdots, X_{t,d})'$.

The *Value* of the portfolio is then given by

$$V_t = \sum_{i=1}^{d} n_i\ P_{t,i} = \sum_{i=1}^{d} n_i\ e^{Z_{t,i}}.$$

Definition 2.1. *Given a r.v. X, which represents a risk or a risk factor, we shall denote F the distribution function (d.f.) of X. We have, for every $x \in \mathbb{R}$*

$$F(x) = \mathbb{P}(X \leq x), \quad \textit{the probability that the event } \{X \leq x\} \textit{ occurs.}$$

Definition 2.2. *The (right) tail function of F, denoted $\overline{F}$, is*

$$\overline{F}(x) = \mathbb{P}(X > x) = 1 - F(x), \quad x \in \mathbb{R}.$$

Definition 2.3. *Whenever there exists the probability density function (p.d.f.) of X, we denote it $f(x)$, and we have*

$$F(x) = \int_{-\infty}^{x} f(t)\ dt, \quad x \in \mathbb{R}.$$

Definition 2.4. *For any $p \in (0,1)$, we shall denote the p-quantile*

$$\chi_p := F^{\leftarrow}(p) = \inf\{x :\ F(x) \geq p\} =: Q(p),$$

where $F^{\leftarrow}$ denotes the generalized inverse function of F, also denoted Q, the so-called quantile function.

In the field of finance it is usually appropriate to interpret this quantile as the *Value-at-Risk* of the risk X, and we write

$$VaR_p(X) = F^{\leftarrow}(p) \equiv Q(p).$$

We are usually interested in large values of p, sometimes larger than $\{1 - 1/n\}$, where n is the available sample size. Consequently, the *Value at Risk* is a parameter of extreme or even rare events: a high quantile of, for instance, the loss distribution of a portfolio of investments over a fixed period of time. The VaR bible is the book by Jorion (2001).

Another alternative risk measure is the expected shortfall:

Definition 2.5. *For any $p \in (0,1)$, the expected shortfall at the level p is the conditional mean value of a risk, given that the risk is larger than VaR_p, i.e.,*

$$ES_p := \mathbb{E}\left(X | X > VaR_p\right).$$

Following closely Frey and McNeil (2000), we present in Figure 1 a possible P&L and the corresponding Loss distribution, which is the mirror of the P&L distribution, i.e., $F_{-L}(x) = 1 - F_{PL}(-x)$, and consequently $f_L(x) = f_{PL}(-x)$.

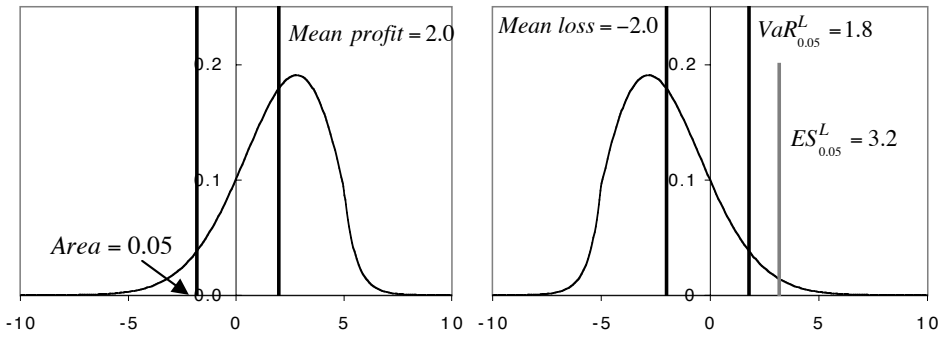

Figure 1: A possible P&L p.d.f. (*left*) and the corresponding Loss p.d.f. (*right*).

Whereas much of the traditional statistics concerns averages, where the normal model is the "king", in the VaR problem we are much more concerned with the extreme, the abnormal and the unexpected. Consequently, *Extreme Value Theory* (EVT) has a very important role to play in this area. We shall pay here some attention to this theory in section 4, but the analysis of financial data overpasses largely the field of *Statistics of Extremes*, and in this field we need to take seriously into account the strong

serial dependencies in financial time series data. Models like the $ARCH$ and the $GARCH$, together with their extremal behaviour are here of main importance.

3 Brief review of Data Modelling

We shall consider that we have got data that may be assumed to be independent, identically distributed (i.i.d.), or at least stationary, from a population X, with unknown d.f. $F(x)$, with an associated quantile function $Q(p) = F^{\leftarrow}(p)$, also unknown. Given the observed sample $(x_1, x_2, \cdots, x_n)$, how to validate a certain model F, as the model underlying the data? We shall give a very brief account of the steps related to a preliminary statistical data analysis:

1. Compute a few *descriptive statistics* (*mean*, *standard deviation*, *asymmetry* and *skewness* coefficients).

2. Plot the *histogram* associated with the data (the statistical counterpart of the p.d.f. $f = F'$), or any equivalent graphical representation.

3. Choose an adequate *model*, among the great diversity of models usual in applications, and perhaps on the basis of theoretical considerations. Note however that we should not be stuck to a theoretical result and should instead *"leave the data speak by themselves"*.

4. Chosen the model, for sure dependent on a set of unknown parameters, *estimate* those parameters.

5. Test *the fit of the model*, adequately, either through goodness-of-fit tests or graphically.

6. If there is no reason to reject the chosen model, advance with the desired *inference* procedures; otherwise, go back to step 3.

Example: Let us consider data associated to the Euro-US$ daily exchange rates from January 4, 1999 till December 15, 2003. This data has been collected by the European System of Central Banks, and was obtained from $http://www.bportugal.pt/rates/cambtx/cambtx_p.htm$.

In Figure 2 we present the Daily Exchange Rates x_t over the above mentioned period and the Log-Returns, $r_t = 100 \times (\ln(x_t) - \ln(x_{t-1}))$, the data to be analyzed. The sample of log-returns has a size $n = 1266$, and let us assume now that we may consider the data to be i.i.d. (or stationary from a model F).

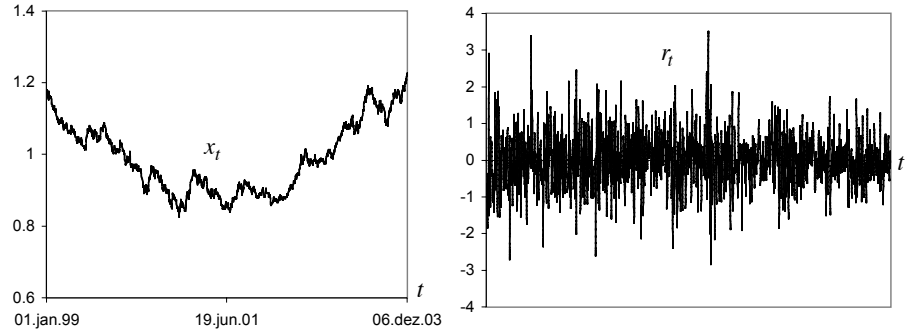

Figure 2: Daily Exchange Rates (*left*) and Daily Log-Returns (*right*) on Euro-US$ Exchange Rate.

Indeed, the graphical representation of the data in Figure 2 (right), leads us to conclude that there is no kind of deterministic trend. We shall here merely try to identify any peculiarities of the model F, underlying the data.

Let us begin with the computation of some descriptive statistics. Using any kind of *package*, we get:

Descriptive statistics

	n	$r_{1:n}$	$r_{n:n}$	$\overline{r}$	s_r	$b_1 = \frac{m_3}{m_2^{3/2}}$	$b_2 = \frac{m_4}{m_2^2} - 3$
Euro/USD	1266	−2.25202	4.20413	0.00008	0.68667	0.32895	1.27468

Brief comments: There is a positive asymmetry (0.329), not very high, but a high kurtosis (1.28), i.e., the underlying model has for sure heavier tails than the classical normal model (a model with null skewness and kurtosis coefficients, in the way they are defined). Note that m_k is the k-th central empirical moment of our sample $(r_1, r_2, \cdots, r_n)$, i.e. $m_k = \left\{\sum_{i=1}^n (r_i - \overline{r})^k\right\}/n$, $\overline{r} = \sum_{i=1}^n r_i/n$, $s_r = n \times m_2/(n-1)$.

The histogram associated to the data would lead us also to detect the existence of a slight positive asymmetry, but a reasonably heavy right tail. And given a random sample $(X_1, X_2, \ldots, X_n)$ (or the observed sample $(x_1, x_2, \ldots, x_n)$), how to estimate F?

Definition 3.1. *The empirical d.f. is either the function,*

$$F_n^*(x) := \frac{\#\{X_i \leq x, 1 \leq i \leq n\}}{n} \quad or \quad F_n^{**}(x) := \frac{\#\{X_i \leq x, 1 \leq i \leq n\}}{n+1}.$$

Both $F_n^*(x)$ and $F_n^{**}(x)$ are consistent estimators of $F(x)$.

With the usual notation, $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$, for the ascending sample of order statistics (o.s.), and with I_A denoting the indicator function of the event A, i.e.,

$$I_A = \begin{cases} 1 & \text{if A occurs} \\ 0 & \text{otherwise} \end{cases} ,$$

we have

$$F_n^*(x) = \frac{1}{n} \sum_{i=1}^{n} I_{[X_i \leq x]} = \begin{cases} 0 & x < X_{1:n} \\ i/n & X_{i:n} \leq x < X_{i+1:n},\ 1 \leq i \leq n-1 \\ 1 & x \geq X_{n:n} \end{cases} .$$

For our data, we get the empirical d.f. presented in Figure 3, where we also picture a zoom of the right tail, comparatively to the normal tail. From such a plot, a quick visual identification of a model is not easy to obtain. But we have methodologies to linearize the d.f. These methodologies enable us to build the so-called probability plots or quantile-quantile (Q-Q) plots associated to a model F, for any particular F.

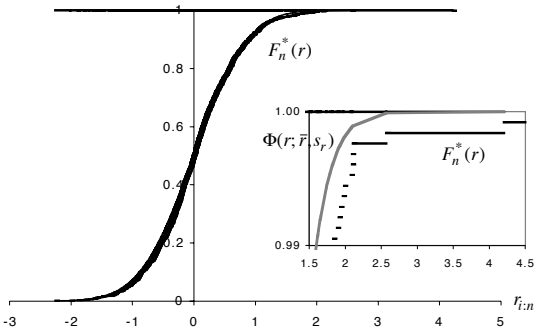

Figure 3: Empirical d.f. of the Daily Log-Returns on Euro-US$ Exchange Rate.

Probability plot. The graphical representation of data in "Probability Paper" became popular in 1914, when Hazen, in a study related with river floods, suggested the linearization of the normal d.f.

Such a technique has been used in the most diversified ways, but its main application has been a *fast visual validation of the fitting of a certain probabilistic model to a set of data,* $(x_1, x_2, \cdots, x_n)$. Such a picture also enables a rough estimation of the unknown parameters. There is a certain ammount of subjectivity in the method, and the statistical properties of the estimates are not easy to obtain. The Probability plot should then be used only as a preliminary exploratory analysis, prior to a more sophisticate fitting of the model and efficient estimation of the unknown parameters.

This technique is easier to apply when the collected observations, $(x_1, x_2, \cdots, x_n)$, may be assumed to be i.i.d. from X, with d.f. of the type

$F((x-\lambda)/\delta)$, with $\lambda \in \mathbb{R}$ and $\delta > 0$ location and scale parameters, respectively. Assuming that the inverse function of F, $F^{\leftarrow}(\cdot)$, exists, and writting

$$p_i = F((x_{i:n} - \lambda)/\delta), \quad 1 \leq i \leq n,$$

we get

$$y_i = F^{\leftarrow}(p_i) = x_{i:n}/\delta - \lambda/\delta, \quad 1 \leq i \leq n,$$

i.e., there exists a linear relationship between $x_{i:n}$ and $y_i = F^{\leftarrow}(p_i)$, where p_i is any adequate estimate of $F((x_{i:n} - \lambda)/\delta)$.

A possible choice for p_i, $1 \leq i \leq n$ (the so-called *plotting positions*), has been given by Weibull (1939):

$$p_i = i/(n+1), \ 1 \leq i \leq n.$$

The justification of such a value for p_i lies on the fact that for absolutely continuous models F, $\mathbb{E}(F((X_{i:n} - \lambda)/\delta)) = i/(n+1)$. We have indeed, $F((X_{i:n} - \lambda)/\delta) = U_{i:n}$, where $U_{i:n}$, $1 \leq i \leq n$ are the ascending o.s.'s associated to a random sample $(U_1, \cdots, U_n)$ from a Uniform(0,1) model, and $U_{i:n}$ is a Beta$(i, n-i+1)$ r.v., $1 \leq i \leq n$. Whenever the graph exhibits a linear relationship between $x_{i:n}$ and $y_i = F^{\leftarrow}(p_i)$ we get an informal validation of the postulated model F. The intersection with any of the axis and the slope of the adjusted line enable us to get rough estimates of λ and δ.

The use of an *EXCEL* workbook enables us easily to build any probability plot. In figure 4 we picture a Normal and a Gumbel probability plot. Note that the Gumbel model, with d.f. $\Lambda(x) = \exp(-\exp(-x))$, $x \in \mathbb{R}$, is one of the basic models in *Statistics of Extremes*, as we shall see later on. Note that both models are discarded as possible models underlying our data, particularly when we are interested in the right tail. The right tail is heavier than the normal, and even heavier than the Gumbel right tail. The left tail seems to be of an exponential type, similar to the normal left tail.

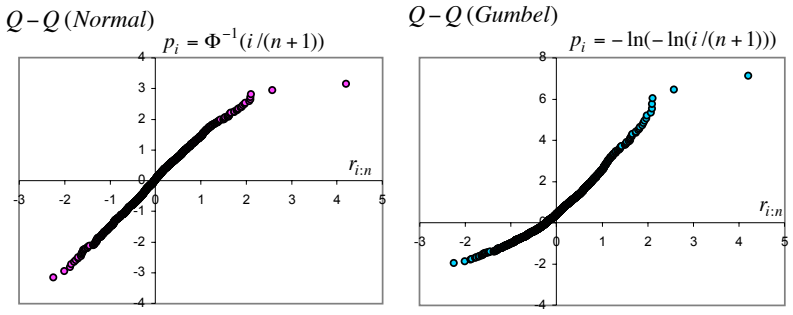

Figure 4: Normal and Gumbel probability plots of the Daily Log-Returns on Euro-US$ Exchange Rate.

And given a sample how to estimate a quantile with probability p, i.e., a value χ_p, which is exceeded with probability $1-p$, i.e., such that

$$F(\chi_p) = p \quad \text{iff} \quad \chi_p = F^{\leftarrow}(p) = \inf\{x : F(x) \geq p\} \equiv Q(p).$$

We should then use the estimate:

$$\chi_p^* = F_n^{**\leftarrow}(p) = x_{i:n} \text{ if } p \in \left[\frac{i}{n+1}, \frac{i+1}{n+1}\right), \text{ i.e. } \chi_p^* = x_{[np]+1:n}.$$

If p is close to 1, i.e., if we want to estimate a high quantile, a level which is overpassed with a "very small" probability, we are never able to extrapolate beyond the sample with the use of either F_n^* or F_n^{**}, i.e,

$$\forall p \geq \frac{n}{n+1}, \ \chi_p^* = x_{n:n}.$$

It is then necessary to find models, which are adequate for the "tails" (right or left) of the unknown d.f. F, so that we may afterwards deal with inference on parameters of extreme or even rare events.

The classical theory of Statistics is dominated by the Normal model, one of the possible limiting models of sums of i.i.d. r.v.'s, linearly normalized. Such a model is no longer relevant in the field of *Statistics of Extremes*, where there appear other models, the so-called *Extreme Value* (EV) models.

4 Statistics of Extremes

The main objectif of *Statistics of Extremes* is to provide inference methodologies, adequate to the tail of F, the model underlying our data, or equivalently, to estimate parameters of extreme or rare events, like for instance *high quantiles*

$$\chi_p = Q(p) := \inf\{x : F(x) \geq p\}, \ p \text{ close to 1 or close to 0}$$

(levels which are overpassed either with a very low or a very high probability). Another important parameter of extreme events is the *return period of a high level* u — the mean waiting time between independent exceedances of the level u or, the mean number of observations needed to get a value exceeding u. In an i.i.d. set-up we get

$$T(u) = \frac{1}{1-F(u)}.$$

Remark 4.1. *If we have for instance access to the log-returns along a certain time period, let us say* $(r_1, r_2, \cdots, r_n)$*, and if we are interested in high log-returns, expressed as a function of* $\max_{1\leq i\leq n} r_i$*, or in small log-returns, related to* $\min_{1\leq i\leq n} r_i$*, we are for sure interested in Statistics of Extremes.*

The exact distributional behaviour of large (or small) values is not usually an easy task. For details on such a theory we refer to Sahran and Greenberg (1962), Galambos (1987), Reiss (1989) and David and Nagaraja (2003). Since we may usually work with a large number of observations, we often use, in *Statistics of Extremes*, asymptotic results for the tails, which provide models, known up to unknown parameters. We shall first review some of the classical results in EVT.

4.1 Limiting behaviour of extremes

4.1.1 Limiting behaviour of the maximum of i.i.d. observations

The d.f. of $X_{n:n} = \max(X_1, X_2, \cdots, X_n)$ is given by

$$P(X_{n:n} \leq x) = F^n(x) \underset{n\to\infty}{\longrightarrow} \begin{cases} 0 & \text{if} \quad 0 \leq F(x) < 1 \\ 1 & \text{if} \quad F(x) = 1 \end{cases} .$$

Consequently, $X_{n:n} \xrightarrow[n\to\infty]{p} x_F := \sup\{x \in \mathbb{R} : F(x) < 1\}$, the right endpoint of F. But:

Definition 4.1. *If there exists sequences of constants* $\{a_n\}$ $(a_n > 0)$ *and* $\{b_n\}$ $(b_n \in \mathbb{R})$*, such that*

$$\frac{X_{n:n} - b_n}{a_n} \xrightarrow[n\to\infty]{d} Z, \text{ non-degenerate,}$$

i.e., such that, at all continuity points of a non-degenerate d.f. H*,*

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n) \underset{n\to\infty}{\longrightarrow} H(x), \tag{4.1}$$

we say that the underlying model F *is in the max-domain of attraction of* H*, and denote such a fact by* $F \in \mathcal{D}_M(H)$*.*

Remark 4.2. *A non-degenerate d.f.* H *is a d.f. which does not put all its mass at a single point.*

Remark 4.3. *Two d.f.'s* H_1 *and* H_2 *are said to be of the same type if there exist* $a > 0$ *and* $b \in \mathbb{R}$ *such that* $H_2(x) = H_1(ax + b)$*.*

We have the following classic result in EVT:

Theorem 4.1. (*Fisher and Tippett, 1928; Gnedenko, 1943*). *If there exist* $\{a_n\}$ $(a_n > 0)$ *and* $\{b_n\}$ $(b_n \in \mathbb{R})$, *such that* (4.1) *holds, with* H *non-degenerate, then* H *is of one of the following three types,*

$$\begin{aligned} \textit{Type I (Gumbel)}: &\quad \Lambda(x) = \exp(-\exp(-x)),\ x \in \mathbb{R} \quad (\gamma = 0) \quad (4.2) \\ \textit{Type II (Fréchet)}: &\quad \Phi_\alpha(x) = \exp(-x^{-\alpha}),\ x \geq 0 \quad (\gamma > 0) \quad (4.3) \\ \textit{Type III (Weibull)}: &\quad \Psi_\alpha(x) = \exp(-(-x)^{\alpha}),\ x \leq 0 \quad (\gamma < 0). \quad (4.4) \end{aligned}$$

Remark 4.4. *Note that essentially all common continuous distributions used in Risk Management, Insurance Mathematics, and even other real life fields, are in* $\mathcal{D}_M(EV_\gamma)$ *for some value* γ, *as we shall see later, for heavy tails.*

The three types in (4.2), (4.3) and (4.4) may be combined in an unified *Extreme Value* (*EV*) d.f. of the type:

$$G_\gamma \equiv EV_\gamma(x) = \begin{cases} e^{-(1+\gamma x)^{-1/\gamma}},\ 1+\gamma x > 0 & \text{if} \quad \gamma \neq 0 \\ e^{-e^{-x}},\ x \in \mathbb{R} & \text{if} \quad \gamma = 0 \end{cases}. \tag{4.5}$$

We get $\Lambda(x) = EV_0(x)$, $\Phi_\alpha(x) = EV_\gamma((x-1)/\gamma)$, $\gamma = 1/\alpha > 0$, and $\Psi_\alpha(x) = EV_\gamma(-(x+1)/\gamma)$, $\gamma = -1/\alpha < 0$. The *EV* d.f. has been introduced in von Mises (1954) and Jenkinson (1955), being sometimes called *von-Mises-Jenkinson* d.f.

Consequently, when the sample size $n \to \infty$,

$$P[X_{n:n} \leq x] = F^n(x) \approx EV_\gamma\left(\frac{x-\lambda}{\delta}\right),\ \lambda \in \mathbb{R},\ \delta > 0.$$

The parameter γ, usually called the *tail index*, is the shape parameter, which rules the the tail behaviour of F. In an informal language, for $\gamma = 0$ we have *exponential tails*, for $\gamma > 0$ we have *heavy tails*, of a negative polynomial type, and for $\gamma < 0$ we have *light tails*.

The limiting result in Theorem 4.1 led Gumbel (1958) to suggest the first approach to Statistics of Extremes, the so-called *Annual Maxima* approach, or *Univariate EV* approach or even *Gumbel's* approach. In this approach we split the N observations in k sub-samples (usually corresponding to k years) of size n (with $N = nk$) and fit one of the extreme models in (4.2), (4.3) or (4.4) (Type I, Type II or Type III) or the EV_γ distribution in (4.5), any of the models with unknown location $\lambda \in \mathbb{R}$ and scale $\delta > 0$, to the sample of the k maximum values of any of the k sub-samples.

All the statistical inference is then related to these models, being necessary to deal with the proper estimation of λ, δ, and γ (or α), to be addressed later on, in sections 4.3, 4.4 and 4.5. This is then the basis for the estimation of other extreme event's parameters.

Note that any result related to maxima has a counterpart for minima. Such a counterpart comes from the identity

$$\min_{1\le i\le n} X_i = -\max_{1\le i\le n}(-X_i).$$

That's the reason why we sometimes symmetrize the data, and work with right tails — remember what has been said about losses, at the very beginning of these notes.

As the sample size $n \to \infty$,

$$P[X_{1:n} \le x] = 1 - (1 - F(x))^n \approx EV^*_{\gamma^*}\left(\frac{x-\lambda^*}{\delta^*}\right),$$

with $EV^*_{\gamma^*}(x) = 1 - EV_{\gamma^*}(-x)$, EV_{γ^*} given in (4.5), i.e.,

$$G^*_{\gamma^*}(x) \equiv EV^*_{\gamma^*}(x) = \begin{cases} 1 - e^{-(1-\gamma^* x)^{-1/\gamma^*}},\ 1-\gamma^* x > 0 & \text{if} \quad \gamma^* \neq 0 \\ 1 - e^{-e^x},\ x \in \mathbb{R} & \text{if} \quad \gamma^* = 0 \end{cases}.$$

4.1.2 Limiting distribution of the k largest observations in an i.i.d. set-up

In most fields of application, we have often no natural seasonality in the data. The sub-sample method looks than a bit artificial and subjective. It has been argued that seems more sensible to take the k largest observations from our original sample of size $N = n \times k$. Indeed, if we have for instance yearly data, some of these years may contain some of those k top observations (for sure relevant to infer on the tail of F), whereas other years may not contain any of those k top observations. In a certain sense we may say that this new approach provides some additional information, lost by the classical Gumbel method.

Such an approach depends on the joint distributional behaviour of the top o.s., which we refer briefly. First of all, and for the distribution of the k-th extreme:

Theorem 4.2. (*Smirnov, 1952*). *Under the conditions of Theorem 4.1, and for fixed $k \ge 1$, if* (4.1) *holds, then $(X_{n-k+1:n} - b_n)/a_n$ converges weakly, as*

$n \to \infty$, towards a non-degenerate r.v., with d.f.

$$G_\gamma^{(k)}(x) \equiv EV_\gamma^{(k)}(x) = EV_\gamma(x) \sum_{j=0}^{k-1} \frac{[-\ln EV_\gamma(x)]^j}{j!}, \tag{4.6}$$

with $EV_\gamma(x) \equiv EV_\gamma^{(1)}(x)$ given in (4.5).

More generally we may say:

Theorem 4.3. (*Lamperti, 1964; Dwass, 1964*). *Under the conditions of Theorem 4.1, again for fixed k, if* (4.1) *holds, then*

$$\left(\frac{X_{n:n} - b_n}{a_n}, \cdots, \frac{X_{n-k+1:n} - b_n}{a_n} \right) \xrightarrow[n\to\infty]{d} (Z_1, \cdots, Z_k)$$

where, with G_γ an univariate EV_γ d.f., and g_γ the associated p.d.f., $(Z_1, \cdots, Z_k)$ has a p.d.f. of the type

$$\begin{aligned} h_\gamma(z_1, \cdots, z_k) &= \frac{\partial^k}{\partial z_1 \cdots \partial z_k} H_\gamma(z_1, \cdots, z_k) \\ &= g_\gamma(z_k) \prod_{j=1}^{k-1} \left(\frac{g_\gamma(z_j)}{G_\gamma(z_j)} \right), \quad \text{if } z_1 > z_2 > \cdots z_k, \end{aligned} \tag{4.7}$$

the so called multivariate EV_γ distribution.

Whenever $N \to \infty$, with fixed k

$$P[X_{N:N} \le x_1, \cdots, X_{N-k+1:N} \le x_k] \approx H_\gamma \left(\frac{x_1 - \widetilde{\lambda}}{\widetilde{\delta}}, \cdots, \frac{x_k - \widetilde{\lambda}}{\widetilde{\delta}} \right),$$

where H_γ is the multivariate d.f. associated to the multivariate p.d.f. h_γ in (4.7).

We have then a second approach to *Statistics of Extremes*. In this approach we base our inference on this multivariate structure. We then work with the so-called *Multivariate EV_γ model*, the model in (4.7). In this model it is easier to increase the sample size k, and consequently we do not need to have a large initial sample size, $N = n \times k$, say, with n large, so that we may use the asymptotic results.

We may also consider the so-called *Multidimensional EV* approach. We then colect a multivariate random sample

$$(\underline{X}_1, \underline{X}_2, \cdots, \underline{X}_m), \text{ where } \underline{X}_r = (X_{1r} > \cdots > X_{i_r r}),\ 1 \le r \le m,$$

are multivariate extreme vectors (Gomes, 1981; Smith, 1986).

4.1.3 Limiting behaviour of the excesses over a high threshold

Another methodology, equivalent to the *Multivariate* EV_γ approach is the following: consider only the observations above a certain high level or *threshold*, and fit an appropriate statistical model, either to the *excesses* or to the *peaks* above that level (maximum value of a sequence of consecutive observations above a level).

The following classical result holds true:

Theorem 4.4. (*Balkema and de Haan, 1974; Pickands, 1975*). *A model F is in the max-domain of attraction of EV_γ, i.e., $F \in \mathcal{D}_{_M}(EV_\gamma)$, with EV_γ in* (4.5) *if and only if*

$$\lim_{u \to x_{_F}} \sup_{0 \le x < x_{_F} - u} \left| F_u(x) - GP_{\gamma,\beta(u)}(x) \right| = 0,$$

where, for $0 \le x < x_{_F} - u$, with $x_{_F} := \sup\{x : F(x) < 1\} \le \infty$, the right endpoint of F,

$$F_u(x) = P(X - u \le x | X > u) = \frac{F(x+u) - F(u)}{1 - F(u)},$$

and, for $\beta > 0$, $GP_{\gamma,\beta}(x) = GP_\gamma(x/\beta)$, where

$$GP_\gamma(x) = \begin{cases} 1 - (1 + \gamma\ x)^{-1/\gamma}\ ,\ x \ge 0,\ 1 + \gamma x > 0 & \text{if} \quad \gamma \ne 0 \\ 1 - e^{-x},\ x \ge 0 & \text{if} \quad \gamma = 0 \end{cases} \tag{4.8}$$

is the so-called univariate Generalizade Pareto (GP) model.

Remark 4.5. *Note the close parallel between limit results for sample maxima and limit results for exceedances over high thresholds, a topic which has been extensively exploited in modern Statistics of Extremes.*

The adequate model for the excesses over a high level is thus a Generalized Pareto model. When $u \to x_0^F$, the upper limit of the support of F, we get

$$P\left[X - u \le x | X > u\right] = 1 - \frac{1 - F(x+u)}{1 - F(u)} \approx GP_\gamma(x/\beta), \quad \beta = \beta_u,$$

and with GP_γ given in (4.8).

We have thus another approach to *Statistics of Extremes*. We consider a high level, and work with the excesses (the differences between the observations over that level and the chosen level), which are well modelled by a Generalized Pareto model. This approach is often called the *Paretian Excesses* approach or the *Peaks Over Thresholds* (*POT*) approach.

Remark 4.6. *Note that to implement the POT method we must choose a suitable threshold u, and such a choice is equivalent to the choice of the number k of top o.s. to take in the Largest Observations methodology. There are a few data-analytic tools, to help us in such a choice, but simulations suggest that inference is often robust to the choice of the threshold, provided the underling model F is "reasonably well behaved".*

More recently, both the *POT* method and the Method of the Largest Observations have been used in a semi-parametric way. We no longer need to identify a parametric model, dependent on a location parameter λ, a dispersion δ and a shape γ. We merely assume that F is in the domain of attraction for maxima of EV_γ, with γ the primary parameter to be estimated, on the basis of some top observations, and according to the adequate methodology, discussed later on, in sections 4.6. and 5.3.

4.1.4 Generalizations to original schemes not necessarily i.i.d.

Whenever the original scheme is no longer identically distributed, but there is independence, for instance T-periodicity, the limiting results presented before may still hold true. Also, under adequate weak dependence conditions, similar limiting results hold true, and we get the same (*max-stable*) limit for the suitably linearly normalized maximum. For details on extremes of dependent sequences see, for instance, Leadbetter et al. (1983). It appears now an additional parameter, relevant for any inferential procedure — the so-called *extremal index*, easy to relate with the exceedances over a high level

$$u_n : \quad F(u_n) = 1 - \tau/n + o(1/n), \text{ as } n \to \infty, \quad \tau \text{ fixed.} \tag{4.9}$$

What is then the asymptotic behaviour of the exceedances of high levels? Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. r.v.'s. Then, any limiting point process of exceedances of high levels is a *Poisson Process*.

Let us denote $\{Y_n\}_{n\geq 1}$ a stationary sequence, under "general" local and asymptotic dependence conditions, not to be specified here. Then, any limiting process of exceedances of high levels, after a suitable time normalization, is a *Compound Poisson Process*, i.e., there exists a cluster of exceedances of high levels: the positions of the clusters are still well modelled by a Poisson process, but there are multiplicities associated to those points, the sizes of the clusters. For these classes of stationary sequences, there exists and is well defined the so-called *extremal index* θ, $0 \leq \theta \leq 1$, which plays an important role in the obtention of the sizes of the clusters of exceedances: for several interesting practical situations (but not in general) the *extremal index* is the reciprocal of the limiting cluster's mean size. We have obviously

$\theta = 1$ for i.i.d. sequences, i.e., the exceedances of high levels appear isolated, and $\theta > 0$ for "almost all" interesting cases. Let us look at the sample paths of the following sequences:

(a) $\{X_n\}_{n\geq 1}$ is a sequence of i.i.d. r.v.'s from an underlying model $F(x) = (1 - \exp(-x))^2$, $x \geq 0$,

(b) $\{Y_n\}_{n\geq 1}$ is a 2-dependent sequence, $Y_n = \max(Z_{n-1}, Z_n)$, $n \geq 1$, where $\{Z_n\}_{n\geq 1}$ are i.i.d. r.v.'s from $H(z) = 1 - \exp(-z)$, $z \geq 0$. We consequently have the same underlying model, $F(y) = (1-\exp(-y))^2$, $y \geq 0$.

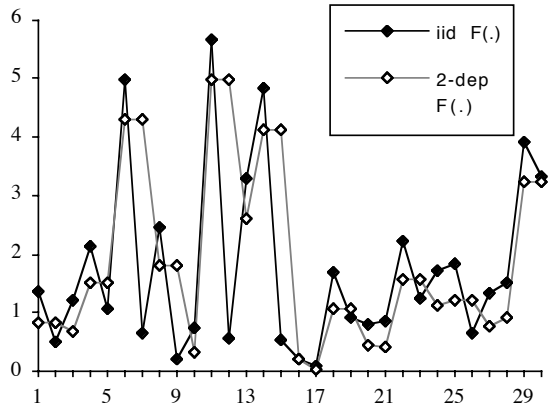

Figure 5: Sample paths of an i.i.d. process and a 2-dependent process

It is obvious from Figure 5 that we have, for the $\mathbf{Y}$-sequence, a size equal to 2 for the clusters of exceedances of high levels. This implies an extremal index equal to 1/2 for this sequence. It is also evident the *shrinkage* of the largest observations, whenever we consider the 2-dependent sequence, instead of the i.i.d. one, although we have the same model F, underlying both sequences.

What is the influence of the *extremal index* in the limiting behaviour of the upper o.s.'s? Let us consider a high level, such that (4.9) holds. For any Borel set $B \in \mathcal{B}(0,1)$, let

$$N_n(B) = \sum_{i=1}^{n} \epsilon_{i/n}(B) I_{[X_i > u_n]} = \# \left\{ \frac{i}{n}, 1 \leq i \leq n : \frac{i}{n} \in B \text{ and } X_i > u_n \right\},$$

be the *exceedances point process*, and

$$\widetilde{N}_n(B) = \sum_{i=1}^{n} \epsilon_{i/n}(B) I_{[u_n \leq X_i < u_{n+1}]}, \ B \in \mathcal{B},$$

the *upper-crossings point processs*. For details on point processes in EVT, see, for instance Resnick (1987) and Reiss (1993). Then, in an i.i.d. set-up

$$N_n \to N, \quad \text{a Poisson Process}(\tau)$$

and

$$\widetilde{N}_n \to \widetilde{N}, \quad \text{a Poisson Process}(\tau).$$

Under a weak dependence condition, like the one we will find, later on, in the $ARCH$ and $GARCH$ processes, for instance,

$$\widetilde{N}_n \to \widetilde{N}, \quad \text{a Poisson Process}(\theta\tau), \quad \text{with } \theta \text{ the extremal index,}$$

and

$$N_n \to N = \sum_{i=1}^{\widetilde{N}} Z_i,$$

where $\{Z_i\}$ are i.i.d. discrete r.v.'s, representing the clusters' sizes.

Let $\{X_n\}_{n\geq 1}$ be an i.i.d. sequence from a model F and $\{Y_n\}_{n\geq 1}$ a stationary dependent sequence, for which general dependence restrictions hold true, so that we may assure the existence of an extremal index θ. Let also F be the marginal d.f. of $Y_n,\ n \geq 1$. Let us denote $M_n^X = M_{n,X}^{(1)} = \max\limits_{1\leq i\leq n} X_i$ and $M_n^Y = M_{n,Y}^{(1)} = \max\limits_{1\leq i\leq n} Y_i$.

For the same normalized level u_n in (4.9), let $S_n^X = \sum_{j=1}^n I_{[X_j>u_n]}$ and $S_n^Y = \sum_{j=1}^n I_{[Y_j>u_n]}$. Then S_n^X is a binomial r.v., and converges towards a Poisson r.v. with mean value τ, as $n \to \infty$, whereas S_n^Y converges towards a compound Poisson r.v. In particular:

$$P[M_n^X \leq u_n] = P[S_n^X = 0] \to \exp(-\tau),$$

where τ is the intensity of the limiting Poisson positions, and

$$P[M_n^Y \leq u_n] = P[S_n^Y = 0] \to \exp(-\theta\tau), \quad 0 < \theta < 1,$$

i.e., the intensity is now $\theta\tau < \tau$. Consequently,

Theorem 4.5. (*Leadbetter et al., 1983*). *If there exist sequences* $\{a_n\}_{n\geq 1}$ $(a_n > 0)$, $\{b_n\}_{n\geq 1}$ $(b_n \in \mathbb{R})$, *and a non-degenerate r.v.* Z, *with the* $EV_\gamma(x)$ *d.f. in* (4.5), *such that*

$$\frac{M_n^X - b_n}{a_n} \to Z, \ \textit{as } n \to \infty,$$

then

$$\frac{M_n^Y - b_n}{a_n} \to Z^*, \ \textit{as } n \to \infty,$$

where Z^* *has a d.f.* $EV_\gamma^\theta(x)$.

This means that for dependent structures, we have a *shrinkage* of maximum values. But the type of limiting law remains the same, since the EV_γ model is *stable for maxima* or *max-stable*, and consequently,

$$EV_\gamma^\theta(x) = EV_\gamma\left(\frac{x-\lambda_\theta}{\delta_\theta}\right), \quad \lambda_\theta = \frac{\theta^{-\gamma}-1}{\gamma}, \quad \delta_\theta = \theta^{-\gamma}.$$

However, the same does not happen with the other upper o.s.'s. Let us denote $M_{n,X}^{(k)}$ the k-th largest o.s. among $(X_1, \cdots, X_n)$ and $M_{n,Y}^{(k)}$ the k-th largest o.s. among $(Y_1, \cdots, Y_n)$,

$$P[M_{n,X}^{(k)} \leq u_n] = P[S_n^X < k] \to e^{-\tau} \sum_{j=0}^{k-1} \frac{\tau^j}{j!},$$

whereas

$$P[M_{n,Y}^{(k)} \leq u_n] = P[S_n^Y < k] \to e^{-\theta\tau} \sum_{j=0}^{k-1} \frac{(\theta\tau)^j}{j!} \sum_{i=j}^{k-1} \pi^{*j}(i),$$

where $\pi^{*j}(\cdot)$ is the j-th convolution of a discrete p.d.f. $\pi(.)$, the limiting p.d.f. of clusters' sizes $\left(\pi^{*0}(i) = 1, \ \forall i \geq 1\right)$, i.e.

$$\pi(j) = \lim_{n\to\infty} \pi_n(j), \text{ with } \pi_n(j) = P\left[\sum_{i=1}^{r_n} I_{[Y_i>u_n]} = j | \sum_{i=1}^{r_n} I_{[Y_i>u_n]} > 0\right]$$

(the distribution of the number of events in a cluster of exceedances, conditional to the existence of at least one exceedance).

This result gives rise to a limiting law for the k-th largest o.s., in an i.i.d. scheme, given by (4.6), i.e., by

$$EV_{\gamma,X}^{(k)}(x) = EV_\gamma(x) \sum_{j=0}^{k-1} \frac{(-\ln EV_\gamma(x))^j}{j!},$$

whereas in this kind of weak dependence schemes,

$$EV_{\gamma,Y}^{(k)}(x) = EV_\gamma^\theta(x) \sum_{j=0}^{k-1} \frac{(-\theta \ln EV_\gamma(x))^j}{j!} \ p_{j,k}, \quad p_{j,k} = \sum_{i=j}^{k-1} \pi^{*j}(i).$$

Note that the limiting d.f. of $M_{n,Y}^{(k)}$ may be written as a mixture of the limiting d.f.'s of the i largest o.s.'s, $1 \leq i \leq k$, in an i.i.d. structure where the maximum is attracted towards $EV_\gamma^\theta(x) = EV_\gamma\left(\frac{x-\lambda_\theta}{\delta_\theta}\right)$, i.e., may be written

as

$$\sum_{j=1}^{k} \alpha_j \, EV_{\gamma,X}^{(j)} \left(\frac{x-\lambda_\theta}{\delta_\theta} \right), \; \alpha_j \geq 0, \; \sum_{j=1}^{k} \alpha_j = 1, \; \lambda_\theta = \frac{\theta^{-\gamma}-1}{\gamma}, \; \delta_\theta = \theta^{-\gamma},$$

$$\alpha_1 = 1 - p_{1,k} = 1 - (\pi_1 + \cdots + \pi_{k-1}),$$

$$\alpha_j = p_{j-1,k} - p_{j,k}, \quad 2 \leq j \leq k-1, \quad \alpha_k = p_{k-1,k}.$$

The difference is even more drastic when we consider the joint distribution of $\left(M_n^{(1)}, \cdots, M_n^{(k)} \right)$ in both cases, i.e., the existence of clusters of exceedances of high levels does not change the type of limiting distribution of the maximum, but changes the limiting distribution of the other top o.s.'s, due to the fact that for instance the second maximum may occur in the same cluster as the maximum.

In *Statistics of Extremes*, as it is easy to deduce from the facts pointed out before, only the classical Gumbel approach is not drastically affected by the existence of an extremal index smaller than 1, at least in an initial phase. All the other approaches presented are disturbed. It is thus necessary to have good estimates of the extremal index (may be even of the probabilities $\pi_i, \; i \geq 1$), to go on with inference on the tail, a topic to be dealt with in section 4.2.

4.1.5 Bivariate Extremes

For bivariate extremes, or even multivariate extremes, the situation is identical, although a bit more complicated. Let us assume that we have two risks related, for instance, to the log-returns of the exchange rates of Dolar agains Euro and Sweddish crown against Euro, exactly during the same period of time. It seems sensible to assume that those observations are positively correlated. From the bivariate sample of daily log-returns, we may take for each year the maximum observed value for each sequence, and work with that sample of pairs of maximum values. What kind of model should we next use for that bivariate sample?

Once again we rely on the asymptotic theory, which provides models, now dependent not only on unknown location, scale and shape parameters, but also on a *dependence function*, to be estimated on the basis of our sample of bivariate maxima.

Let us consider, for instance, the particular case of *Gumbel* margins.

Given the sample of independent random pairs

$$(X_1, Y_1), \cdots, (X_n, Y_n),$$

with d.f. $F(x, y)$, the limiting distribution of the pair

$$\left(\max_{1 \le i \le n} X_i, \max_{1 \le i \le n} Y_i \right),$$

suitably normalized, is,

$$\Lambda(x, y) = [\Lambda(x)\Lambda(y)]^{k(y-x)}$$

if both sequences of maximum values are attracted towards the Gumbel law $\Lambda(x) = \exp(-\exp(-x)),\ x \in \mathbb{R}$. The dependence function $k(\cdot)$ needs to obey to certain rules, and this enables us to be able to get different models, among which we refer, for θ in the range $[0, 1]$:

Logistic Model (Gumbel, 1962):	$k(w\|\theta) = \frac{(1+e^{-w/(1-\theta)})^{1-\theta}}{1+e^{-w}}$;
Mixed Model (Gumbel, 1966):	$k(w\|\theta) = 1 - \theta \frac{e^w}{(1+e^w)^2}$;
Bi-extremal Model (Tiago de Oliveira, 1968):	$k(w\|\theta) = 1 - \frac{\min(\theta, e^w)}{1+e^w}$;
Gumbel Model (Tiago de Oliveira, 1971):	$k(w\|\theta) = 1 - \theta \frac{min(1, e^w)}{1+e^w}$.

For further details on multivariate models see Embrechts et al. (1997) and Tiago de Oliveira (1997). For asymmetric extensions of the Logistic model, useful in the applications, see Tawn (1988).

4.2 Approaches to Statistics of Extremes and estimation of the extremal index

4.2.1 Approaches to Statistics of Extremes

In an abridged way we may thus speak of the following approaches:

1. *Parametric approaches*:

   I. Univariate EV model (for the k maximum values of sub-samples of size n, $N = nk$) — Gumbel's classical approach.

   II. Multivariate EV model (for the k largest o.s.'s associated to the whole lot of N observations.)

   III. Multi-dimensional EV model (multivariate EV model for the i_j largest observations, $j = 1, 2, .., m$, in sub-samples of size m', $m \times m' = N$).
   If $m = k$ and $i_j = 1$ for $1 \le j \le k$ we get model I;
   If $m = 1$ and $i_1 = k$ we get model II.

IV. Paretian model (for the excesses $X_j - u$, of a high level u, adequately chosen) [*POT* approach (*Peaks Over Thresholds*)].

2. *Semi Parametric approaches*:

V. We work then with the k largest o.s.'s associated to the global N observations, as in II, but we no longer consider a parametric model; we instead consider that $F \in D(G_\gamma)$.

Approach I will be discussed in section 4.3. Section 4.4 is devoted to approach III, which generalizes II (as well as I). In sections 4.5 and 4.6 we deal with approaches IV and V, respectively.

4.2.2 Extremal index estimation

Most frequently, to estimate θ, we use the following identities:

$$\begin{aligned} \theta &= \frac{1}{\text{limiting mean cluster size}} \\ &= \lim_{n\to\infty} P[X_2 \le u_n | X_1 > u_n]. \end{aligned}$$

We shall speak about two alternative estimators.

I Let $u = X_{n-\tau:n}, \quad [\sqrt{n}/2] \le \tau \le [4\sqrt{n}]$.
Compute for any value of u

$$\begin{aligned} \widehat{\theta}^{(1)}_{n,u} &:= \frac{\sum_{j=1}^{n-1} I_{[X_j<u\le X_{j+1}]}}{\sum_{j=1}^{n} I_{[X_j>u]}} \\ &= \frac{\text{total number of upper-crossings}}{\text{total number of excedances}}. \end{aligned} \tag{4.10}$$

Select as estimate of θ, for instance, the mean or the median of the estimates in the stability region.

II Given k_n blocks of size r_n, arbitrarily chosen, compute

$$\begin{aligned} \widehat{\theta}^{(2)}_{n,u_i,1\le i\le k_n} &= \frac{k_n}{\sum_{i=1}^{k_n} \sum_{j=(i-1)r_n+1}^{ir_n} I_{[X_j>u_i]}} \\ &= \frac{\text{total number of blocks}}{\text{total number of exceedances}}. \end{aligned}$$

The choice of $u_i, 1 \le i \le k_n$ is not usually a problem: such a choice is based on the fact that the mean number of crossings in any of the

k_n blocks of size r_n ($k_n r_n = n$) should be equal to one. Then, the second local maximum in each block is an eligible candidate! We may for instance take k_n = smallest integer dividing n, larger thane $\sqrt[3]{n}$.

For an illustration of the behaviour of these estimators see, for instance Gomes (1993). For the log-return data in Figure 2 (right), we now picture, in Figure 6, the sample path of the estimates in (4.10), as a function of u.

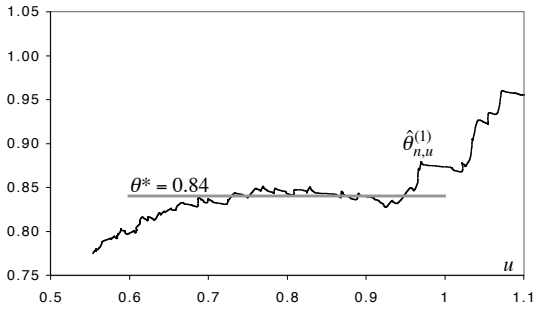

Figure 6: Sample path of the extremal index estimates in (4.10), as a function of the threshold u.

4.2.3 Back to the appraoches to Statistics of Extremes

As pointed out before, the classical Gumbel approach is not affected by the existence of a extremal index smaller than one. *And what can we do in any of the non-classical approaches*? Choose representatives of the clusters of exceedances, and work with them and the structures defined for i.i.d. set-ups.

How to choose those representatives?
The estimate of θ may provide us with indications on:

- The number of top observations to consider (approximately $k\ \theta$) in order to get a sample of size k from the *multivariate EV* model; or equivalently,
- The high level which enables us to get k independent Paretian excesses.

4.3 Gumbel's classical approach

As we have mentioned before, the extreme univariate models are based on the asymptotic probabilistic result in Theorem 4.1, related with the limiting behaviour of the sequence of maximum values $M_n = \max_{1\leq i\leq n} X_i$,

similar to the Central Limit Theorem for sums $S_n = \sum_{i=1}^{n} X_i$.

The possible limiting laws for M_n, suitably normalized, are of the type (4.5), the EV_γ model, which generalizes the three classical types in (4.2), (4.3) and (4.4). As we know, the usefulness of a model in practice depends partially on the existence of reliable procedures of estimation of the unknown parameters. Since the estimation of the unknown parameters in an EV model is not a very easy job, it is still usual, whenever we are dealing with a sample of maximum values, to try the fitting of one of the three models, *Gumbel* (the easiest one), *Fréchet* or *Weibull*. Beforehand we should perform a *statistical choice test*, to select one of those three models. Such a test may be for instance a *probability plot*.

4.3.1 Gumbel's model

In its general form, the Gumbel d.f. is given by

$$\Lambda(x; \lambda, \delta) = \exp\left(-\exp\left(-\frac{x-\lambda}{\delta}\right)\right), \ x \in \mathbb{R}.$$

The main population characteristics of this model are:

$$\mathbb{E}(X) = \lambda - \Gamma'(1)\ \delta,$$

$$\mathbb{V}ar(X) = (\pi\delta)^2/6,$$

$$\mathbb{M}ode = \lambda,$$

$$\mathbb{M}edian = \lambda - \delta \ln\ln 2$$

where Γ is the complete *Gamma function*, $\Gamma(x) = \int_0^\infty u^{x-1}e^{-u}du, \ x > 0$, Γ' its derivative, $\Gamma'(x) = \int_0^\infty (\ln u)u^{x-1}e^{-u}du, \ x > 0$.

Remark 4.7. *The Gumbel d.f. plays, as a limiting d.f. of maximum values, suitably normalized, the same role that the Normal d.f. plays, as a limiting d.f. of suitably normalized sums, and the Log-Normal plays, as limiting distribution of suitably normalized products.*

The Gumbel d.f. is without doubt the most frequently used in practice. The main reason for this lies on the fact that statistical inference procedures are easier for Λ than for either Φ_α or Ψ_α.

The *Maximum Likelihood* (ML) estimates of the unknown parameters (λ, δ) are easy to obtain numerically. We get

$$\widehat{\delta} = \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{\sum_{i=1}^{n} x_i \exp\left(-x_i/\widehat{\delta}\right)}{\sum_{i=1}^{n} \exp\left(-x_i/\widehat{\delta}\right)}.$$

The fixed point method enables us to obtain in a fast way the ML-estimate of δ, and

$$\widehat{\lambda} = -\widehat{\delta}\ \ln\left\{\frac{1}{n}\sum_{i=1}^{n}\exp\left(-x_i/\widehat{\delta}\right)\right\}.$$

4.3.2 The Fréchet Model

The Fréchet d.f. is given by

$$\Phi_\alpha(x;\lambda,\delta) = \exp\left(-\left(\frac{x-\lambda}{\delta}\right)^{-\alpha}\right),\ x \geq \lambda \quad (\alpha > 0),$$

and its main characteristics are

$$\mathbb{E}(X) = \lambda + \delta\ \Gamma(1-1/\alpha) \qquad \text{if} \quad \alpha > 1;$$

$$\mathbb{V}ar(X) = \delta^2\left\{\Gamma(1-2/\alpha) - \Gamma^2(1-1/\alpha)\right\} \quad \text{if} \quad \alpha > 2;$$

$$\mathbb{M}ode = \lambda + \delta\ (1+1/\alpha)^{-1/\alpha};$$

$$\mathbb{M}edian = \lambda + \delta\ (\ln 2)^{-1/\alpha}.$$

For a Fréchet model, the ML-estimate of δ may be explicitly obtained as a function of the ML-estimates of λ and α, through the equation,

$$\widehat{\delta} = \left\{n/\sum_{i=1}^{n}(x_i - \widehat{\lambda})^{-\widehat{\alpha}}\right\}^{1/\widehat{\alpha}}.$$

The ML-estimates of λ and α are obtained trhrough the numerical solution of the following system of two equations:

$$\begin{cases} \frac{\widehat{\alpha}+1}{n}\sum_{i=1}^{n}(x_i-\widehat{\lambda})^{-1} - \frac{\widehat{\alpha}\sum_{i=1}^{n}(x_i-\widehat{\lambda})^{-\widehat{\alpha}-1}}{\sum_{i=1}^{n}(x_i-\widehat{\lambda})^{-\widehat{\alpha}}} = 0 \\ \\ \frac{1}{\widehat{\alpha}} + \frac{\sum_{i=1}^{n}(x_i-\widehat{\lambda})^{-\widehat{\alpha}}\ln(x_i-\widehat{\lambda})}{\sum_{i=1}^{n}(x_i-\widehat{\lambda})^{-\widehat{\alpha}}} - \frac{1}{n}\sum_{i=1}^{n}\ln(x_i-\widehat{\lambda}) = 0 \end{cases}.$$

To obtain the maximum likelihood estimates of λ and α, it is possible to use, for instance, the Newton-Raphson's algorithm.

4.3.3 The Weibull Model

It is a model with d.f.

$$\Psi_\alpha(x;\lambda,\delta) = \exp\left(-\left(-\frac{x-\lambda}{\delta}\right)^\alpha\right),\ x \le \lambda \quad (\alpha > 0).$$

Its main characteristics are:

$$\mathbb{E}(X) = \lambda + \delta\ \Gamma(1+1/\alpha),$$

$$\mathbb{V}ar(X) = \delta^2\left\{\gamma(1+2/\alpha) - \gamma^2(1+1/\alpha)\right\},$$

$$\mathbb{M}ode = \lambda + \delta\ (1-1/\alpha)^{1/\alpha},$$

$$\mathbb{M}edian = \lambda + \delta\ (\ln 2)^{1/\alpha}.$$

Things are now not too much different from before. For a Weibull model, the ML estimate of δ may also be explicitly given as a function of the ML estimates of λ and α, through the equation,

$$\widehat{\delta} = \left\{ n / \sum_{i=1}^{n} (\widehat{\lambda} - x_i)^{\widehat{\alpha}} \right\}^{-1/\widehat{\alpha}}.$$

The ML estimates of λ and α are the solution of the system of non-linear equations:

$$\begin{cases} \frac{\widehat{\alpha}-1}{n} \sum_{i=1}^{n} (\widehat{\lambda} - x_i)^{-1} - \frac{\widehat{\alpha} \sum_{i=1}^{n} (\widehat{\lambda}-x_i)^{\widehat{\alpha}-1}}{\sum_{i=1}^{n} (\widehat{\lambda}-x_i)^{\widehat{\alpha}}} = 0 \\ \\ -\frac{1}{\widehat{\alpha}} + \frac{\sum_{i=1}^{n} (\widehat{\lambda}-x_i)^{\widehat{\alpha}} \ln(\widehat{\lambda}-x_i)}{\sum_{i=1}^{n} (\widehat{\lambda}-x_i)^{\widehat{\alpha}}} - \frac{1}{n} \sum_{i=1}^{n} \ln(\widehat{\lambda} - x_i) = 0 \end{cases}.$$

Here, to get the solution of this system of equations, it is sometimes necessary to use algorithms more sophisticate than Newton-Raphson's algorithm. More than that: standard properties of ML estimators apply only if $\gamma > -0.5$.

4.3.4 The EV_γ model

As mentioned before, the EV_γ model has an associated d.f. given by:

$$G_\gamma(x;\lambda,\delta) = \exp\left(-\left(1+\gamma\ \frac{x-\lambda}{\delta}\right)^{-1/\gamma}\right),\ 1+\frac{\gamma\ (x-\lambda)}{\delta} > 0,\ (\gamma \in \mathbb{R}),$$

and its main characteristics are

$$\mathbb{E}(X) = \lambda - \frac{\delta}{\gamma}\left[1 - \Gamma(1-\gamma)\right] \qquad \text{if} \quad \gamma < 1;$$

$$\mathbb{V}ar(X) = \left(\frac{\delta}{\gamma}\right)^2 \left[\Gamma(1-2\gamma) - \Gamma^2(1-\gamma)\right] \qquad \text{if} \quad \gamma < 1/2;$$

$$\mathbb{M}ode = \lambda + \frac{\delta}{\gamma}\left[(1+\gamma)^{-\gamma} - 1\right] \qquad \text{if} \quad \gamma \neq 0;$$

$$\beta(X) = -sign(\gamma)\frac{\Gamma(1-3\gamma) - 3\ \Gamma(1-2\gamma)\ \Gamma(1-\gamma) + 2\ \Gamma^3(1-\gamma)}{\left[\Gamma(1-2\gamma) - \Gamma^2(1-\gamma)\right]^{3/2}} \qquad \text{if} \quad \gamma < 1/3.$$

The EV_γ model has a positive asymmetry for $\gamma > -0.28$. In Figure 7 we picture the standard $EV_\gamma(x)$ for $\gamma = -0.5,\ 0$ and 0.5.

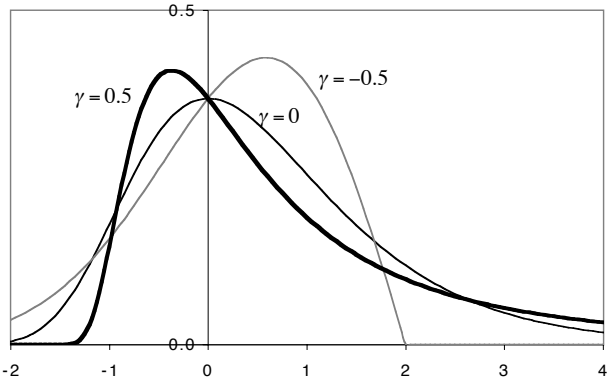

Figure 7: Extreme Value p.d.f.'s for $\gamma = -0.5,\ 0$ and 0.5.

The ML method is again one of the possible methods, but we need to solve a non-linear system of highly unstable ML equations, often computationally expensive. The standard asymptotic results of ML-estimators hold true if $\gamma > -1/2$ (Smith, 1985). An alternative method, often used in applications, is the *Weighted Moments* Method (Hosking, 1985):

If we consider, for $\gamma < 1$, the weighted moments

$$\beta_r = \mathbb{E}\left\{X(F(X))^r\right\} = \frac{1}{r+1}\left(\lambda - \frac{\delta}{\gamma}\left(1 - \frac{\Gamma(1-\gamma)}{(r+1)^{-\gamma}}\right)\right),$$

and

$$b_r = \frac{1}{n}\sum_{j=1}^{n} \frac{(j-1)(j-2)\cdots(j-r)}{(n-1)(n-2)\cdots(n-r)}\ x_{j:n},$$

unbiased estimators of β_r, we get the following estimators of λ, δ and γ, in the EV model:

$$\gamma^* = -7.8590c - 2.9554c^2, \quad c = \frac{2b_1 - b_0}{3b_2 - b_0} - \frac{\ln 2}{\ln 3},$$
$$\delta^* = \frac{\gamma^*(2b_1 - b_0)}{\Gamma(1-\gamma^*)(2^{\gamma^*} - 1)},$$
$$\lambda^* = b_0 + \delta^* \left\{1 - \Gamma(1-\gamma^*)\right\} / \gamma^*.$$

4.3.5 Statistical choice among extremal models.

Given a sample of maximum values, or in a certain way related to possible maximum values of underlying non-observed data, how to choose among *Gumbel*, *Fréchet* and *Weibull* models? We may for instance use the probability paper's technique or any statistical choice test statistic.

1. **Probability paper.** If F is the Gumbel d.f.

$$\Lambda(x; \lambda, \delta) = \exp(-\exp\left(-\frac{x-\lambda}{\delta}\right) \Longrightarrow -\ln(-\ln p_i) = -\frac{\lambda}{\delta} + \frac{x_{i:n}}{\delta}.$$

Consequently, as exhibited in Figure 8, if there exists a linear relationship between $x_{i:n}$ and $-\ln(-\ln p_i)$ we have an informal validation of Gumbel's model. More than that: if the points $(x_{i:n}, -\ln(-\ln p_i))$, $1 \leq i \leq n$, are concave, we get an informal validation of the Fréchet model, and if the points $(x_{i:n}, -\ln(-\ln p_i))$, $1 \leq i \leq n$, are convex, we get an informal validation of the Weibull model.

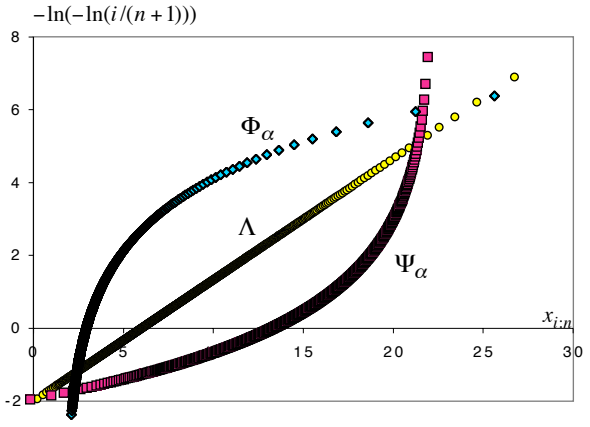

Figure 8: Plot of extreme data in a Gumbel probability plot.

2. **Statistical choice tests.** We may also perform a statistical choice test of extremal models. The easiness of inference for Gumbel models, leads us to test the hypothesis:

$$H_0 : \gamma = 0 \text{ versus } H_1 : \gamma \neq 0 \quad (\text{or } \gamma > 0) \ (\text{or } \gamma < 0)$$

in an EV_γ model, to solve the "trilemma" in Statistics of Extremes (Tiago de Oliveira, 1981). If such an hypothesis H_0 is not rejected, at an adequate significance level α, let us say $\alpha = 0.05$, we go on with the fitting of a Gumbel model. Otherwise, the observed value of the test statistic will provide the most adequate choice among the other two types of extremal models, either Fréchet or Weibull.

Given the sample $(X_1, X_2, \cdots, X_n)$, a possible statistical choice test statistic, studied in Gomes (1982, 1987) and later in Fraga Alves and Gomes (1996) is:

$$W_n = \frac{X_{n:n} - X_{[n/2]+1:n}}{X_{[n/2]+1:n} - X_{1:n}}.$$

Under the validity of H_0,

$$W_n^* = \ln\ln n \left\{ W_n - \frac{\ln n + \ln\ln 2}{\ln\ln n - \ln\ln 2} \right\} \xrightarrow{w} Z, \text{ as } n \to \infty,$$

where Z has a standard Gumbel d.f., $\Lambda(x) = \exp(-\exp(-x)),\ x \in \mathbb{R}$. For small values of n, critical points for W_n^* are provided in Table 1.

Table 1: Critical points of W_n^*

α	0.025	0.05	0.10	0.90	0.95	0.975
$n = 10$	-1.06	-0.98	-0.87	1.42	2.12	2.83
$n = 50$	-1.30	-1.15	-0.94	1.80	2.47	3.11
$n = 100$	-1.35	-1.16	-0.95	1.77	2.34	3.04
$n \to \infty$	-1.31	-1.10	-0.83	2.25	2.97	3.68

If we are led to the rejection of H_0, with a high value of W_n^*, we should assume a Fréchet model, of the type of the one in (4.3); otherwise, a Weibull model of the type of the one in (4.4) should be considered.

4.3.6 High quantiles and return periods' estimation.

Let us assume we have access to $N = n \times k$ original observations. Let us consider k maxima (of sub-samples of size N/k). Under Gumbel's approach,

we have the approximation,

$$F^{N/k}(x) \approx G_\gamma \left(\frac{x - \lambda_{N/k}}{\delta_{N/k}} \right),$$

which leads us easily to the estimation of high quantiles,

$$\chi_p : F(\chi_p) = p, \quad p \text{ large.}$$

We often consider

$$\widehat{\chi}_p = \begin{cases} \widehat{\lambda}_{N/k} - \widehat{\delta}_{N/k} \left(\frac{1-(N(1-p)/k)^{-\widehat{\gamma}}}{\widehat{\gamma}} \right) & \text{if} \quad \gamma \neq 0 \\ \widehat{\lambda}_{N/k} - \widehat{\delta}_{N/k} \ln(N(1-p)/k) & \text{if} \quad \gamma = 0 \end{cases}.$$

To estimate a return period, we may use the approximation:

$$T(u) = \frac{1}{1-F(u)} \approx \begin{cases} \frac{N}{k} \left(1 + \gamma \frac{u - \lambda_{N/k}}{\delta_{N/k}} \right)^{1/\gamma} & \text{if} \quad \gamma \neq 0 \\ \frac{N}{k} \, e^{\frac{u-\lambda_{N/k}}{\delta_{N/k}}} & \text{if} \quad \gamma = 0 \end{cases}.$$

The existence of an extremal index θ would imply the need of the modifications mentioned before. We then have the approximation:

$$F^{N\theta/k}(x) \approx G_\gamma \left(\frac{x - \lambda^*_{N/k}}{\delta^*_{N/k}} \right).$$

We then consider:

$$\widehat{\chi}_p = \begin{cases} \widehat{\lambda}^*_{N/k} - \widehat{\delta}^*_{N/k} \frac{1-(N(1-p)\theta/k)^{-\widehat{\gamma}}}{\widehat{\gamma}} & \text{if} \quad \gamma \neq 0 \\ \widehat{\lambda}^*_{N/k} - \widehat{\delta}^*_{N/k} \ln(N(1-p)\theta/k) & \text{if} \quad \gamma = 0 \end{cases},$$

and may use now the approximation

$$\begin{aligned} T(u) &= \frac{1}{\theta} \, \frac{1}{1-F(u)} - \frac{1}{\theta} + 1 \\ &\approx \begin{cases} \frac{N}{k} \left(1 + \gamma \frac{u - \lambda^*_{N/k}}{\delta^*_{N/k}} \right)^{1/\gamma} - \frac{1}{\theta} + 1 & \text{if} \quad \gamma \neq 0 \\ \frac{N}{k} \, e^{\left(u - \lambda^*_{N/k}\right)/\delta^*_{N/k}} - \frac{1}{\theta} + 1 & \text{if} \quad \gamma = 0 \end{cases} \end{aligned}$$

to estimate the return period.

4.4 Multivariate and multi-dimensional EV model

In a period of m years, assume that, for instance, we have collected the r largest yearly observations. We have thus collected a sample,

$$(\underline{X}_1, \underline{X}_2, \cdots, \underline{X}_m), \text{ with } \underline{X}_j = (X_{1j} > X_{2j} > \cdots > X_{rj}),\ 1 \le j \le m.$$

For details on such an approach see Gomes (1981) and Smith (1986).

For $\gamma = 0$, we then consider that those observations come from a model with p.d.f.

$$f(x_1, \cdots, x_r) = \delta^{-r} \exp\left[-\exp\left(-\frac{x_r - \lambda}{\delta}\right) - \sum_{j=1}^{r} \frac{x_j - \lambda}{\delta} \right],$$
$$\text{if } x_1 \ge x_2 \ge \cdots \ge x_r.$$

For $\gamma \neq 0$, the multivariate p.d.f. is

$$\delta^{-r} \exp\left(-\left[1 + \gamma\left(\frac{x_r - \lambda}{\delta}\right)\right]^{-1/\gamma} - (1/\gamma + 1) \sum_{j=1}^{r} \ln\left[1 + \gamma\left(\frac{x_j - \lambda}{\delta}\right)\right] \right)$$
$$\text{if } x_1 \ge x_2 \ge \cdots \ge x_r;\ 1 + \gamma(x_j - \lambda)/\delta > 0,\ 1 \le j \le r.$$

Under the validity of a Gumbel model, we get the following estimates for λ and δ:

$$\widehat{\lambda}_r = -\widehat{\delta}_r \ln\left\{ \frac{1}{m\ r} \sum_{j=1}^{m} \exp\left(-x_{rj}/\widehat{\delta}\right) \right\},$$

and

$$\widehat{\delta}_r = \frac{1}{mr} \sum_{j=1}^{m} \sum_{i=1}^{r} x_{ij} - \frac{\sum_{j=1}^{m} x_{rj} \exp\left(-x_{rj}/\widehat{\delta}_r\right)}{\sum_{j=1}^{m} \exp\left(-x_{rj}/\widehat{\delta}_r\right)}.$$

The estimators become a bit more complicated when we consider as an underlying model, the Fréchet, the Weibull or, more generally, a unified EV model as the limiting distribution of maximum values. Anyway, they generalize easily the ML equations obtained for $r = 1$.

If we consider $m = 1$ (i.e. if we assume to be working only with the $r =: k$ largest observations, among the global set of data, the estimators are easier to obtain. Also for the Gumbel model, we get, with $X_{11} > X_{21} > \cdots > X_{k1}$,

$$\widehat{\delta} = \frac{1}{k} \sum_{i=1}^{k} (X_{i1} - X_{k1})\,; \qquad \widehat{\lambda} = \widehat{\delta}\ \ln k + X_{k1}.$$

The estimation of quantiles and return periods is usually based on the maximum values, using a technique similar to the one in 4.3.6., with the adequate changes.

4.5 The POT methodology — the Generalized Pareto model

The *Generalized Pareto* model has a d.f.

$$P_{\gamma,\beta}(x) = \begin{cases} 1 - \left(1 + \gamma \, \frac{x}{\beta}\right)^{-1/\gamma}, \; 1 + \gamma \, \frac{x}{\beta} > 0, \; x \geq 0 & \text{if} \quad \gamma \neq 0 \\ 1 - \exp\left(-x/\beta\right), \; x \geq 0 & \text{if} \quad \gamma = 0 \end{cases}$$

and may exhibit different shapes, as the ones in Figure 9.

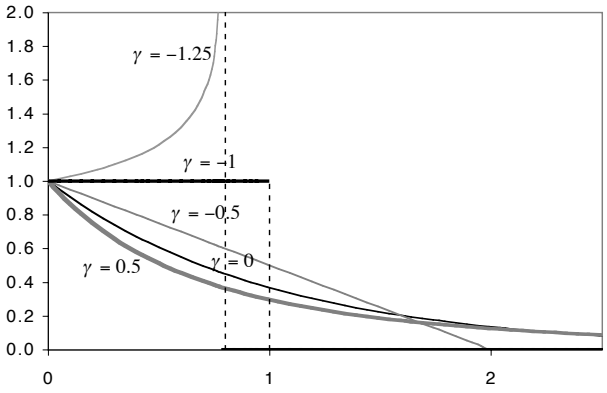

Figure 9: Generalized Pareto p.d.f.'s for $\gamma = -1.25, \; -1, \; -0.5, \; 0, \;$ and 0.5.

The Pareto p.d.f. is an exponential p.d.f. whenever $\gamma = 0$ and a uniform p.d.f. whenever $\gamma = -1$. We get the following model characteristics,

$$\mathbb{E}(X) = \tfrac{\beta}{1-\gamma} \qquad \text{if} \quad \gamma < 1;$$

$$\mathbb{V}ar(X) = \tfrac{\beta^2}{(1-\gamma)^2(1-2\gamma)} \qquad \text{if} \quad \gamma < 1/2;$$

$$\beta(X) = \tfrac{2(1+\gamma)\sqrt{1-2\gamma}}{1-3\gamma} \qquad \text{if} \quad \gamma < 1/3;$$

$$\mathbb{E}\left(X - u | X > u > 0\right) = \tfrac{\beta + \gamma \, u}{1-\gamma} \quad \text{if} \quad \gamma < 1.$$

As before, we may test an exponential model ($\gamma = 0$) for the excesses over a high level. Such a test may again be a graphical one, performed

through the use of a probability plot, devised for the exponential model (i.e., the plotting of $x_{i:n}$ versus $-\ln(i/(n+1))$, for $1 \leq i \leq n$).

Moment estimators:

$$\begin{array}{lll} \beta^* = \frac{\overline{x}}{2}\ \left(\left(\frac{\overline{x}}{s}\right)^2 + 1\right); \quad \gamma^* = \frac{1}{2}\left(1 - \left(\frac{\overline{x}}{s}\right)^2\right) & \text{if} & \gamma < 1/2\ (\gamma \neq 0) \\ \beta^* = \overline{x} & \text{if} & \gamma = 0 \end{array} .$$

For $\gamma = 0$, $\beta^* = \overline{x}$ is also the ML estimate of β.

When $\gamma \geq 1/4$ these estimators have a very bad performance, being then necessary to use alternative estimators, among which we refer the Weighted Moments estimators (Hosking et al., 1987). For $\gamma < 1$, they are given by

$$\beta^{**} = \frac{2a_0a_1}{a_0 - 2a_1}; \qquad \gamma^{**} = 2 - \frac{a_0}{a_0 - 2a_1},$$

with

$$a_r = \frac{1}{n}\sum_{j=1}^{n} \frac{(n-j)(n-j-1)...(n-j-r+1)}{(n-1)(n-2)...(n-r)} x_{j:n},$$

now the empirical counterparts of

$$\alpha_r = \mathbb{E}\left[X\left(1 - F(X)\right)^r\right] = \frac{\beta}{(r+1)(r+1-\gamma)}, \quad r \geq 0.$$

The ML estimation has problems similar to the ones we get for the ML estimation in an EV model, but may be done numerically, through any optimization method, used to maximize the log-likelihood, or through the numerical solution of a system of non-linear equations. For a detailed discussion on the subject see, for instance, Davison and Smith (1990).

For the estimation of high quantiles or return periods we use the approximation

$$\frac{F(x+u) - F(u)}{1 - F(u)} = 1 - \frac{1 - F(x+u)}{1 - F(u)} \approx 1 - \left(1 + \gamma\ \frac{x}{\beta}\right)^{-1/\gamma},$$

or equivalently,

$$1 - F(x) \approx (1 - F(u)) \left(1 + \gamma\ \frac{x-u}{\beta}\right)^{-1/\gamma},$$

for large values of x. The estimate of $\{1 - F(u)\}$ is often chosen equal to k/N, where k is the number of excesses of the level u, out of N observations.

4.6 Semi-parametric estimation of tail index and extreme quantiles

4.6.1 Heavy tails — The Hill estimator

We shall here consider that a heavy-tailed model is a model $F \in \mathcal{D}_{_M}(EV_\gamma),\ \gamma > 0$. Let us consider the function

$$U(t) = F^{\leftarrow}(1 - 1/t), \quad t > 1.$$

It is possible to show (de Haan, 1970) that

$$F \in \mathcal{D}_{_M}(EV_\gamma),\ (\gamma > 0) \iff \lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\gamma,\ \forall\ x \geq 0. \tag{4.11}$$

The most common semi-parametric estimator of the tail index γ is the Hill estimator (Hill, 1975),

$$\widehat{\gamma}_n^H(k) := \frac{1}{k} \sum_{i=1}^{k} \{\ln X_{n-i+1:n} - \ln X_{n-k:n}\}\,. \tag{4.12}$$

To get consistency, k needs to be intermediate, i.e.

$$k = k_n \to \infty \text{ and } \frac{k}{n} \to 0,\ \text{as } n \to \infty. \tag{4.13}$$

Remark 4.8. *Note that this is the ML estimator that we obtain, whenever we use, for an underlying standard Fréchet model, the log-likelihood in the extremal process, i.e., the joint limiting distribution of the top k extremes in* (4.7) (*Weissman, 1978*).

We may state the following:

Theorem 4.6. (*de Haan and Peng, 1998*). *Let $F \in \mathcal{D}_{_M}(EV_\gamma)$, $\gamma > 0$. Then, if there exists a function $A(t)$ such that*

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \frac{x^\rho - 1}{\rho},\ x > 0, \tag{4.14}$$

where $\rho \leq 0$ rules the rate of convergence in the first order condition in (4.11), *and if*

$$\lim_{n\to\infty} \sqrt{k}\ A(n/k) = \lambda,\ \textit{finite},$$

then

$$\sqrt{k}\left(\widehat{\gamma}_n^H(k) - \gamma\right) \xrightarrow{\ d\ } N\left(\frac{\lambda}{1-\rho},\ \gamma^2\right).$$

The possible non-null asymptotic bias, equal to $\lambda/(1-\rho)$, whenever $\sqrt{k}\ A(n/k) \to \lambda$, as $n \to \infty$, provides for some models in $\mathcal{D}_M(EV_\gamma)$, a very high bias for large k. And whereas the bias increases with k the variance may be very high for small values of k. Indeed, it is sometimes difficult to detect a stability region in the sample path of Hill's estimates, as a function of k. This has led to a large set of alternative estimators, which enable to remove the dominant component of bias, giving rise to a wider stability region for the sample paths. For instance, Beirlant et al. (1999) and Feuerverger and Hall (1999) provide implicit "maximum likelihood" estimators, based on the fact that the scaled log-spacings are approximately independent exponential r.v.'s, i.e., for $1 \le i \le k$, and with $\{E_i\}$ unit exponential i.i.d. r.v.'s.,

$$\begin{aligned} U_i &= i\,\{\ln X_{n-i+1:n} - \ln X_{n-i:n}\} \qquad (4.15)\\ &\approx \gamma\ e^{\beta(n/i)^\rho}\ E_i \qquad \text{(Feuerverger and Hall, 1999)}\\ &\approx \gamma\ (1+\beta(n/i)^\rho)\ E_i \qquad \text{(Beirlant et al., 1999)} \end{aligned}$$

The three parameters (γ, β, ρ) are then estimated through maximum likelihood. The parameters β and ρ appear in the function $A(t)$ in (4.14), whenever such a function may be chosen as $A(t) = \gamma\ \beta\ t^\rho$.

Gomes et al. (2000) suggest the use of estimators based on the Generalized Jackknife methodology, with the mispecification of the second order parameter ρ in $\rho = -1$, and Gomes and Martins (2002) provide explicit estimators based on the external estimation of ρ. One of those estimators is similar in spirit to the estimators provided in Beirlant et al. (1999) and Feuerverger and Hall (1999), but is explicit and given by:

$$\widehat{\gamma}_n^{ML(\widehat{\rho})}(k) := \frac{1}{k}\sum_{i=1}^{k} \exp\left(\widehat{\beta}\ (i/n)^{-\widehat{\rho}}\right)\ U_i \qquad (4.16)$$

where

$$\widehat{\beta} \equiv \widehat{\beta}(k) := \frac{1}{n^{\widehat{\rho}}}\ \frac{\left(\sum_{i=1}^k i^{-\widehat{\rho}}\right)\left(\sum_{i=1}^k U_i\right) - k\left(\sum_{i=1}^k i^{-\widehat{\rho}} U_i\right)}{\left(\sum_{i=1}^k i^{-\widehat{\rho}}\right)\left(\sum_{i=1}^k i^{-\widehat{\rho}} U_i\right) - k\left(\sum_{i=1}^k i^{-2\widehat{\rho}} U_i\right)},$$

with U_i denoting the scaled log-spacings in (4.15). The estimate $\widehat{\rho}$ is one of the estimators in Fraga Alves et al. (2003), and will be addressed in the subsection 4.6.2.

Another class of "asymptotically unbiased" estimators of a positive tail index γ to be considered here, also based on an external estimation of the second order parameter ρ, and based on the results in Caeiro and Gomes

(2002), is given by

$$\widehat{\gamma}_n^{(\widehat{\alpha})}(k) := \frac{\Gamma(\widehat{\alpha})}{M_n^{(\widehat{\alpha}-1)}(k)} \left(\frac{M_n^{(2\widehat{\alpha})}(k)}{\Gamma(2\widehat{\alpha}+1)} \right)^{1/2}, \quad \text{with} \quad \widehat{\alpha} = -\frac{\ln\left[1-\widehat{\rho}-\sqrt{(1-\widehat{\rho})^2-1}\right]}{\ln(1-\widehat{\rho})}, \quad (4.17)$$

where generally $\widehat{\rho}$ may be any consistent estimator of ρ, and where

$$M_n^{(\alpha)}(k) := \frac{1}{k} \sum_{i=1}^{k} \{\ln X_{n-i+1:n} - \ln X_{n-k:n}\}^{\alpha}, \quad \alpha > 0. \quad (4.18)$$

4.6.2 The estimation of the second order parameter ρ

The class of ρ-estimators provided in Fraga Alves et al. (2003) is parameterised in a tuning parameter τ and depends on the statistics

$$T_n^{(\tau)}(k) := \begin{cases} \dfrac{\left(M_n^{(1)}(k)\right)^{\tau} - \left(M_n^{(2)}(k)/2\right)^{\tau/2}}{\left(M_n^{(2)}(k)/2\right)^{\tau/2} - \left(M_n^{(3)}(k)/6\right)^{\tau/3}} & \text{if} \quad \tau > 0 \\ \dfrac{\ln\left(M_n^{(1)}(k)\right) - \frac{1}{2}\ln\left(M_n^{(2)}(k)/2\right)}{\frac{1}{2}\ln\left(M_n^{(2)}(k)/2\right) - \frac{1}{3}\ln\left(M_n^{(3)}(k)/6\right)} & \text{if} \quad \tau = 0, \end{cases}$$

which converge towards $3(1-\rho)/(3-\rho)$, independently of τ, whenever the second order condition (4.14) holds and k is such that $k = o(n)$ and $\sqrt{k}\, A(n/k) \to \infty$, as $n \to \infty$. The estimators are thus given by

$$\widehat{\rho}_n^{(\tau)}(k) := \min\left(0,\ \frac{3(T_n^{(\tau)}(k)-1)}{T_n^{(\tau)}(k)-3}\right). \quad (4.19)$$

Remark 4.9. *The theoretical and simulated results in Fraga Alves et al. (2003), together with the use of these estimators in the Generalized Jackknife statistics of Gomes et al. (2000) (Gomes and Martins, 2002) led us to advise in practice the consideration of the level*

$$k_1 = \min(n-1, [2n/\ln\ln n]) \quad (4.20)$$

(not chosen in any optimal way), and of the tuning parameters $\tau = 0$ for the region $\rho \in [-1, 0)$ and $\tau = 1$ for the region $\rho \in (-\infty, -1)$. Anyway, we advise practitioners not to choose blindly the value of τ in (4.19)*. It is sensible to draw a few sample paths of $\widehat{\rho}_n^{(\tau)}(k)$ in* (4.19)*, as functions of k, and for a few values of τ, like $\tau = 0, 0.5, 1$, electing the value of τ which provides higher stability for large k, by means of any stability criterion. The value of k_1 in* (4.20)*, although not chosen in any optimal way, may be considered for any chosen value of the tuning parameter τ.*

4.6.3 The semi-parametric estimation of high quantiles for heavy tails

We want to estimate

$$\chi_p \equiv VaR_p : \ F(\chi_p) = p, \quad p \text{ large},$$

i.e.,

$$\chi_p := U\left(\frac{1}{1-p}\right), \quad p = p_n \to 1, \quad n(1-p_n) \to K \geq 0, \text{ as } n \to \infty.$$

We shall assume to be working in Hall's class of models (Hall and Welsh, 1985), where $\rho < 0$ and $U(t) \sim C\ t^{\gamma}$. More specifically, we shall assume that

$$U(t) = Ct^{\gamma}\left(1 + Dt^{\rho} + o(t^{\rho})\right), \quad \rho < 0, \quad \text{as } t \to \infty.$$

We are going to base inference on the largest k top order statistics (o.s.), and as usual in semi-parametric estimation of parameters of rare events, we shall assume that k is an intermediate sequence of integers, i.e., (4.13) holds.

Notice that since $\chi_p = U(1/(1-p))$ and $U(t) \sim C\ t^{\gamma}$, an obvious estimator of χ_p is $\widehat{\chi}_p = \widehat{C}(1-p)^{-\widehat{\gamma}}$, with $\widehat{\gamma}$ any consistent estimator of γ. Also, with $Y_{i:n},\ 1 \leq i \leq n$, denoting the set of ascending o.s. associated to a standard Pareto i.i.d. sample from the model $F_{_Y}(y) = 1 - y^{-1},\ y \geq 1$,

$$X_{n-k:n} = U(Y_{n-k:n}) \sim C\ Y_{n-k:n}^{\gamma} \sim C\left(\frac{n}{k}\right)^{\gamma}.$$

Consequently, an obvious estimator of C is $\widehat{C} = \left(\frac{k}{n}\right)^{\widehat{\gamma}} X_{n-k:n}$, and,

$$\widehat{\chi}_p(k) = X_{n-k:n}\left(\frac{k}{n(1-p)}\right)^{\widehat{\gamma}} \tag{4.21}$$

is the obvious quantile estimator, often considered in practice, whenever dealing with heavy tails. Such an estimator depends strongly on an adequate estimation of the tail index γ, and we propose the use of the "asymptotically unbiased" estimators discussed before, either the one in (4.16) or the one in (4.17).

4.6.4 A general tail — Pickands' and Moment's estimators

Let us assume that $F \in \mathcal{D}_{_M}(EV_{\gamma}),\ \gamma \in \mathbb{R}$. Pickands (1975) proposed the estimator

$$\widehat{\gamma}_n^P(k) := \frac{1}{\ln 2}\ \ln\left\{\frac{X_{n-k+1:n} - X_{n-2k+1:n}}{X_{n-2k+1:n} - X_{n-4k+1:n}}\right\} \tag{4.22}$$

as a possible estimator of γ, showing that such an estimator is weakly consistent, again if k is intermediate, i.e., (4.13) holds. Dekkers and de Haan (1989) show that $\widehat{\gamma}_n^P$ in (4.22) is also strongly consistent if we further assume that $k \ln \ln n \to \infty$.

The Moment estimator of $\gamma \in \mathbb{R}$ was proposed by Dekkers and de Haan (1989), and is based on the moment statistics in (4.18). Such an estimator is given by

$$\widehat{\gamma}_n^M(k) := M_n^{(1)}(k) + 1 - \frac{1}{2}\left(\frac{M_n^{(2)}(k)}{M_n^{(2)}(k) - \left(M_n^{(1)}(k)\right)^2}\right) \tag{4.23}$$

Remark 4.10. *Note that* $M_n^{(1)}(k)$ *is the Hill estimator in* (4.12) $\left(M_n^{(1)}(k) \equiv \widehat{\gamma}_n^H(k)\right)$.

Remark 4.11. *It is possible to derive results similar to the the one presented in Theorem 4.6, both for Pickands' estimator in* (4.22) *and for the Moment estimator in* (4.23). *See, for instance, David and Nagaraja* (2003).

4.6.5 Quantile estimation for a general tail

For a general tail, and following a way similar to the one used for heavy tails, we may consider the quantile estimator (de Haan and Rootzén, 1993; Ferreira et al., 2003):

$$\widehat{\chi}_p(k) := X_{n-k:n} + \widehat{a}(k)\ \frac{\left(\frac{k}{n(1-p)}\right)^{\widehat{\gamma}(k)} - 1}{\widehat{\gamma}(k)}, \tag{4.24}$$

with

$$\widehat{a}(k) = X_{n-k:n} M_n^{(1)}(k)(1 - \min(\widehat{\gamma}(k), 0)),$$

$M_n^{(1)}$ given in (4.18).

Remark 4.12. *Note that for a heavy-tailed model we may choose* $\widehat{a}(k) = X_{n-k:n} \times \widehat{\gamma}(k)$. *Hence* (4.24) *gives* (4.21).

5 Heavy tails and long-range dependence

As mentioned before, empirical studies of financial log-returns suggest that there are two properties often present in the data: *heavy tails* and *long-range dependence*. We shall here briefly review these concepts. Let F be the d.f. of a positive r.v. X (perhaps made positive through a shift), and let us denote $\overline{F}$ the tail function of F.

5.1 Regularly varying tails

A common definition of heavy-tailed models is the one we have considered before, i.e., models $F \in \mathcal{D}_M(EV_\gamma),\ \gamma > 0$, i.e.,

Definition 5.1. *The d.f. F (or the r.v. X) has a heavy tail, as $x \to \infty$, if and only if*

$$\overline{F}(x) = P(X > x) = x^{-\alpha} L(x) \quad (\alpha > 0), \tag{5.1}$$

where $L(x)$ is a slowly varying function, i.e.

$$\frac{L(tx)}{L(t)} \underset{t\to\infty}{\longrightarrow} 1, \qquad \textit{for all } x > 0.$$

Remarks:

1. The function $\overline{F}$ in (5.1) is said to be *regularly varying* with *index of regular variation* equal to $-\alpha$. We shall denote such a property by $\overline{F} \in RV_{-\alpha}$.

2. Note that $F \in \mathcal{D}_M(EV_\gamma)$, $\gamma > 0$, if and only if $\overline{F} = 1 - F \in RV_{-1/\gamma}$, or equivalently, if and only if $U \in RV_\gamma$, with U the quantile function at $1 - 1/t$, i.e., the function in (4.11) and (4.14) (Gnedenko, 1943; de Haan, 1970).

3. A slowly varying function is a regularly varying function with index of regular variation equal to 0.

4. If $\alpha < 2$ the variance of X is infinite. Indeed, some authors consider that a model is heavy-tailed only if $\alpha \in (0, 2)$, i.e., $\gamma > 1/2$.

5. If $\alpha < 1$ the mean value is infinite.

6. A particular important case of the model in (5.1) is provided by the the slowly varying functions $L(x) = C(1 + o(1))$, i.e. $L(x) \to C$, $0 < C < \infty$, as $x \to \infty$. Notice however that a slowly varying function may converge to zero, like $\{1/\ln x\}$, or diverge to infinity, like $\{\ln x\}$.

Examples of heavy-tailed models

1. The clasical heavy-tailed model is the Pareto model, with a d.f.

$$F(x) = 1 - \left(1 + \frac{x}{\delta}\right)^{-\alpha}, \ x \geq 0 \quad (\alpha, \ \delta > 0).$$

   This is a re-parametrization of the *Generalized Pareto* model in (4.8), with $\gamma = 1/\alpha$ and $\beta = \delta/\alpha$. The second order parameter in (4.14) is $\rho = -\gamma = -1/\alpha$.

2. Another important heavy-tailed model is the Fréchet model, with a d.f.

$$F(x) = \exp\left\{-\left(\frac{x}{\delta}\right)^{-\alpha}\right\} \ x \geq 0 \quad (\alpha, \ \delta > 0).$$

   For this model we have a tail index $\gamma = 1/\alpha$ and a second order parameter $\rho = -1$.

3. The Student-t model with ν degrees of freedom, with a standard p.d.f. given by

$$t_\nu(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\sqrt{\nu\pi}\right)} \ \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}, \quad x \in \mathbb{R} \quad (\nu > 0),$$

   is also heavy-tailed. For this model both the right and the left tail are heavy, $\gamma = 1/\nu$ and $\rho = -2/\nu$.

4. Some other models, also common in the literature related to heavy tails, are the so called *stable models*, which appear as limits related to sum's schemes.

Definition 5.2. *A stable d.f., $S_\alpha(\sigma, \beta, \mu)$, is characterized by four parameters: the index of stability $\alpha \in (0, 2]$, a scale parameter $\sigma > 0$, a skewness parameter $\beta \in [-1, 1]$ and a location parameter $\mu \in \mathbb{R}$. The characteristic function $\mathbb{E}\left(e^{itx}\right)$ of a stable r.v. is given by*

$$\begin{aligned} &\exp\left\{-\sigma^\alpha |t|^\alpha (1 - i\beta \ \text{sign}(t) \tan(\pi\alpha/2)) + i\mu t\right\} \quad \textit{if} \quad \alpha \neq 1 \\ &\exp\left\{-\sigma|t|(1 + 2i\beta\pi^{-1}\text{sign}(t) \ln|t|) + i\mu t\right\} \quad \textit{if} \quad \alpha = 1. \end{aligned}$$

Although any stable distribution has a density, in general an explicit expression of the density in terms of elementary functions is unknown. Exceptions (excluding the degenerate case) are the Lévy distribution $S_{1/2}(\sigma, 1, \mu)$, the Cauchy d.f. $S_1(\sigma, 0, \mu)$ and the Gaussian d.f. $S_2(\sigma, 0, \mu) \equiv N(\mu, 2\sigma^2)$.

The parameter β is a skewness parameter, and if $\beta = 0$ the distribution is symmetric around μ. The mean of of a stable r.v. is μ if $\alpha \in (1, 2]$, and for $\alpha \leq 1$, $\mathbb{E}|X| = \infty$. For $\alpha \in (0, 2)$ the variance is infinite. Then (Samorodnitsky and Taqqu, 1994, Property 1.2.15), as $x \to \infty$, S_α has tails given by

$$P(X > x) \sim C_\alpha \sigma^\alpha \frac{1+\beta}{2} x^{-\alpha}, \quad \text{and} \quad P(X \leq x) \sim C_\alpha \sigma^\alpha \frac{1-\beta}{2} x^{-\alpha},$$

where

$$C_\alpha = \begin{cases} \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)} & \text{if} \quad \alpha \neq 1 \\ \frac{2}{\pi} & \text{if} \quad \alpha = 1 \end{cases}.$$

Remark 5.1. *The adoption of stable modelling in finance seems to be promising, as may be seen in the book by Rachev and Mittnik (2000).*

A full treatment of regular variation may be found in Bingham et al. (1987).

5.2 Sub-exponential tails

Some authors consider a class of heavy-tailed distributions, much larger than the class of regularly varying functions with a negative index of regular variation, the class of *sub-exponential* distribution functions:

Definition 5.3. *The d.f. F (or the r.v. X) is sub-exponential if for some $n \geq 2$, and equivalently for all $n \geq 2$,*

$$\lim_{x \to \infty} \frac{P(X_1 + \cdots + X_n > x)}{P(\max(X_1, \cdots, X_n) > x)} = 1. \tag{5.2}$$

Remark 5.2. *This is a much larger class of models, containing the class of regularly varying tails. Further examples of sub-exponential d.f.'s are the lognormal and the Weibull d.f.'s, $F(x) = 1 - \exp(-x^\beta), x \geq 0,\ \beta \in (0, 1)$. These distributions are in $\mathcal{D}_M(EV_0)$, but exhibit a penultimate tail behaviour similar to that of distributions in $\mathcal{D}_M(EV_\gamma)$, with $\gamma > 0$. For details on penultimate approximations see, for instance, Gomes* (1984)*, Gomes and Pestana* (1986) *and Gomes and de Haan* (1999)*.*

Remark 5.3. *Equation* (5.2) *has a possible physical interpretation: the sum of i.i.d. subexponential r.v.'s is likely to be large if and only if their maximum is. This fact accounts for large values in subexponential samples.*

For relations among different classes of heavy-tailed distributions see Embrechts et al. (1997), section 1.4.

5.3 How to detect heavy tails?

We may use different graphical tools, usually based on the notion of regularly varying tails. We shall refer two of those methods:

1. The *log-log cumulative distribution* plot (LLCD plot) is frequently used. The idea is to plot $1 - F_n^*$ on log-log scales, where F_n^* is, as mentioned before, the empirical cumulative d.f.,

$$F_n^*(x) = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \leq x\}}.$$

If the theoretical d.f. has a heavy right tail, for large n and moderately large values of x, the LLCD plot should consist of points randomly scattered around a straight line with slope equal to $-\alpha$. An estimate of α may be obtained through least squares regression. Analogously we may plot $\{\ln x_{i:n}\}$ versus $\{-ln(1 - i/(n + 1))\}$. The points should then be approximately over a line with slope $\alpha = 1/\gamma$.

For the log-returns of Euro against US Dollar, we get the LLCD plot in Figure 10. The fitting of a least squares line to the points $\ln r_{i:n} > 0$ would provide an α-estimate $\widetilde{\alpha} = 3.54$, and consequently a tail index estimate $\widetilde{\gamma} = 1/3.54 = 0.28$.

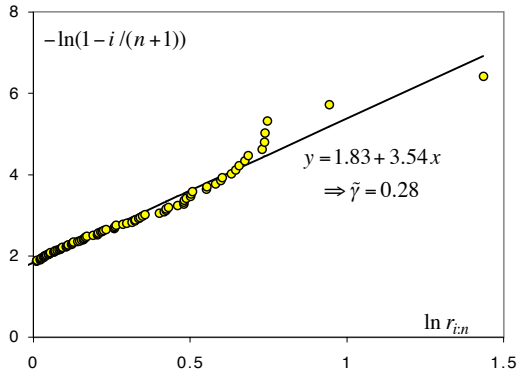

Figure 10: LLCD plot associated to the Daily Log-Returns on the Euro-US$.

2. As mentioned before, the most frequently used method for estimating the tail parameter α is the Hill estimator of $\alpha = 1/\gamma$:

$$\widehat{\alpha}_n(k) := \left(\frac{1}{k} \sum_{i=1}^{k} \ln \frac{X_{n-i+1:n}}{X_{n-k:n}} \right)^{-1}.$$

In case of a heavy tail, the plot of $\{\ln k\}$ versus $\widehat{\alpha}_n(k)$ for small k will stabilize around a value $\alpha > 0$. For other alternatives to Hill's estimator of $\gamma = 1/\alpha$, which exhibit smaller bias for large k, enabling thus their plot as functions of k, see for instance Gomes and Martins (2002).

The log-returns clearly exhibit graphically an indication of an underlying heavy-tailed model. In Figure 11, and working merely with the $n^+ = 611$ positive log-returns, we now present the sample path of the ρ_τ estimates in (4.19) (*left*), as function of k, for $\tau = 0$ and $\tau = 1$, together with the sample paths of the classical tail index estimate (H) and of the two "asymptotically unbiased" tail index estimates discussed before, $\widehat{\gamma}_n^{ML(\widehat{\rho}_0)}$ and $\widehat{\gamma}_n^{\alpha_0(\widehat{\rho}_0)}$ in (4.16) and (4.17), respectively, associated to the value $\widehat{\rho} = \widehat{\rho}_0(611) = -0.69$ (*right*).

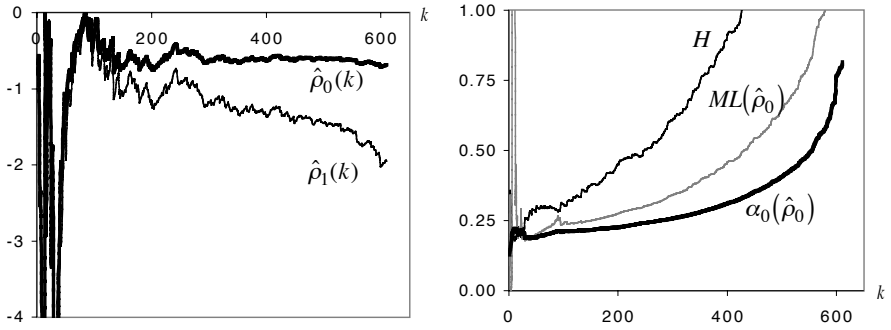

Figure 11: Estimates of the second order parameter ρ (*left*) and of the tail index γ (*right*) for the Daily Log-Returns on the Euro-US$.

Remark 5.4. *Note that the sample path of the ρ-estimates associated to $\tau = 0$ and $\tau = 1$ leads us to chosse, on the basis of any stability criterion for large k, the estimate associated to $\tau = 0$. From the experience we have with this type of estimates, this means that the underlying ρ-value is larger or equal to -1, and the consideration of $\tau = 0$ is then advisable.*

Remark 5.5. *Regarding the tail index estimation, note that whereas the Hill estimator is unbiased for the estimation of the tail index γ if the underlying model is a Pareto model, with d.f. $F(x) = 1 - x^{-1/\gamma}$, $x \geq 0$ ($\gamma > 0$), it exhibits a relevant bias when we have only Pareto-like tails, as happens here, and may be seen from Figure 11. The other estimators, which are "asymptotically unbiased" reveal a smaller bias, and enable us to take a decision upon the estimate of γ to be used, with the help of any stability criterion or any heuristic procedure, like the "largest run". Here, if we consider the tail index estimates with* 2

decimal figures, the largest run is achieved by the sample path of the estimator in (4.17)*. Such a run has a size equal to* 43 $(86 \leq k \leq 128)$*, and is associated to* $\widehat{\gamma}_n^{(\alpha_0(\widehat{\rho}_0))} = 0.21$*. Should we be happy with one decimal figure, would we also get the largest run for the same estimator: a run of size* 278 $(4 \leq k \leq 281)$ *associated to* $\widehat{\gamma}_n^{(\alpha_0(\widehat{\rho}_0))} = 0.2$*. For the ML-estimate, and with one decimal figure, we would also get a tail index estimate equal to* 0.2*, but with a run of size* 130*. With this same criterion, the Hill estimator would provide an estimate equal to* 0.3*, with a run of size* 102*.*

5.4 Long-range dependence

Whenever we are working with a weakly stationary (finite variance) stochastic process $\{X_t\}$, $t = 0, 1, 2, \cdots$, dependence between observations at times t and $t+k$ is usually measured by the autocorrelation function (ACF), $\rho(k)$, at lag k, given by

$$\rho(k) = \frac{\mathbb{C}ov(X_t, X_{t+k})}{\mathbb{V}ar(X_t)} = \frac{\mathbb{C}ov(X_0, X_k)}{\mathbb{V}ar(X_0)} =: \frac{\gamma(k)}{\gamma(0)}, \quad k = 0, 1, 2, \cdots . \tag{5.3}$$

The plotting of $\rho(k)$ versus k gives us an idea of the second order structure of the process. We obviously expect $|\rho(k)|$ to be decreasing in k. The notion of LRD (*long-range dependence*) depends on the size of $\rho(k)$ for large k. A possible definition of LRD (Beran, 1994) is,

Definition 5.4. *The stationary process* $\{X_t\}$ *exhibits LRD if, as* $k \to \infty$*, the ACF in* (5.3) *is such that*

$$\rho(k) \sim C\ k^{-\beta}, \tag{5.4}$$

where C *is a positive constant and* $\beta \in (0, 1)$*. Equivalently, we may say that* $\sum_k \rho(k)$ *is not summable.*

This property contrasts to *short-range dependent* processes like the *autoregressive* (AR) or more generally the *autoregressive moving average* ($ARMA$) processes, where $\rho(k)$ decays at an exponential rate. Popular processes, like the fractional ARIMA (Brockwell and Davis, 1991, section 13.2) and the *fractional Gaussian noise*, introduced later on, satisfy (5.4) with positive values $\rho(k)$.

5.4.1 Detection of long-range dependence

Several heuristic graphical tools may be used to detect LRD in a time series. A statistical evaluation of these exploratory methods may be found in Taqqu

and Teverovsky (1995, 1996). *LRD* seems to be present in most of the data in the field of finance.

We shall here describe essentially two methods:

1. Suppose we have observed a stationary time series $(X_t,\ t = 1, 2, \cdots, n)$. One way to detect *LRD* makes use of the sample autocorrelations

$$\widehat{\rho}(k) = \frac{\widehat{\gamma}(k)}{\widehat{\gamma}(0)}, \quad \widehat{\gamma}(k) = \frac{1}{n}\sum_{t=1}^{n-k}\left(X_t - \overline{X}_n\right)\left(X_{t+k} - \overline{X}_n\right),\ 0 \le k \le n-1.$$

   If *LRD* is present, a plot $\{\ln|\widehat{\rho}(k)|\}$ agains $\{\ln k\}$ (*the log-log correlogram*) will provide points randomly scattered around a straight line with slope $-\beta$ for appropriate values of k and large n. A disadvantage of this method is that the estimate $\widehat{\rho}(k)$ is unreliable for large k with respect to n (see Brockwell and Davis, Section 7.2)

2. The oldest and perhaps better known *LRD* technique detection is the R/S method. This method has its origins in the field of hydrology and was first used by Hurst when he discovered *LRD*-like features in the yearly minimal water levels of the Nile river (Hurst, 1951). For a stationary process $(X_t,\ t = 1, 2, \cdots, n)$, let us consider partial sums $Y_k = \sum_{t=1}^{k} X_t$ and the sample variance $S_k^2 = \sum_{t=1}^{k} X_t^2/k - (Y_k/k)^2$. The R/S statistic is

$$(R/S)(k) := \frac{1}{S_k}\left\{\max_{0\le t\le k}\left(Y_t - \frac{t\ Y_k}{k}\right) - \min_{0\le t\le k}\left(Y_t - \frac{t\ Y_k}{k}\right)\right\}.$$

   It can be shown that if X_t is Gaussian, stationary, ergodic and (5.4) holds, then, with $H = 1 - \beta/2$,

$$k^{-H}\ (R/S)\,(k) \xrightarrow{d} Z,\ \text{as } k \to \infty, \tag{5.5}$$

   where Z is a non-degenerate r.v. (see Mandelbrot, 1975, Theorems 5 and 11, as well as Taqqu, 1975). The parameter H is the so-called *Hurst parameter* and is frequently used as a measure of strength of the *LRD* present in the data: H close to 1 corresponds to a strong presence of *LRD*. For various short memory processes, (5.5) holds with $H = 1/2$. For fractional Gaussian noise and fractional ARIMA (with gaussian innovations),

$$\mathbb{E}(R/S)(k) \sim c\ k^H,\ \text{as } k \to \infty \quad (c > 0). \tag{5.6}$$

   For an observed time series $(X_t,\ t = 1, 2, \cdots, n)$ one can try to verify (5.6) as follows:

   - partition the series in $[n/m]$ blocks of size m;

- then, for each k, compute $(R/S)_{m_i}(k)$, starting at points $m_i = im + 1$, $i = 0, 1, \cdots$, such that $m_i + k \leq n$;
- for values of $k \leq m$ we get $[n/m]$ different estimates of $(R/S)(k)$. For values of k approaching n, we get fewer values, as few as 1 as $k \geq n - m$;
- next plot $\ln(R/S)_{m_i}(k)$ against $\{\ln k\}$ and get, for each k several values on the plot;
- for large k, the points shoud lie around a straight line with slope H.

Another popular method to detect LRD (see Brockwell and Davis, 1975, section 10.3) is based on the *periodogram*

$$I(\lambda) = \frac{1}{2\pi\ n} \left| \sum_{t=1}^{n} X_t\ e^{-it\lambda} \right|, \quad \lambda \in [-\pi, \pi]. \tag{5.7}$$

It is common use to evaluate the periodogram at the *Fourier frequencies* $\lambda_j = 2\pi j/n,\ j = -[(n-1)/2], \cdots, [n/2]$. If LRD is present, a plot of $\ln I(\lambda_j)$ against $\ln(\lambda_j)$ would result in an approximately straight line with slope $\{\beta - 1\}$ for small frequencies λ_j.

5.4.2 Long-range dependence or non-stationarity?

It is clear that the concept of LRD, as given in (5.4), applies only to stationary processes, and we may say that no general test for the stationarity of an observed time series is indeed available in the literature. More than that: the graphical methods used to detect LRD in a time series are not fully reliable, and it has been observed that non-stationarities, like shifts in the mean or a slowly decaying trend can also be the cause of slowly decaying autocorrelations. For more details on this subject see Stegman (2002). It seems sensible to assume that "a financial time series is stationary only over short periods": at different time scales different factors may induce non-stationarities in the measurement series.

5.4.3 $ARIMA$ models and long-range dependence

The class of auto-regressive moving averages ($ARMA$) processes is the class of models most frequently applied to time series that exhibit no apparent deviations from stationarity and have rapidly decreasing autocorrelations. $ARMA$ processes are defined as follows: for non-negative integers p and q, let the polynomials ϕ and θ be given by

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \quad \text{and} \quad \theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q.$$

Let us define the *backward shift operator* B, by

$$B^j X_t = X_{t-j}, \quad j = 0, 1, 2, \cdots .$$

Definition 5.5. *The process* $\{X_t, t = \pm 1, \pm 2, \cdots\}$ *is an* $ARMA(p, q)$ *process if* $\{X_t\}$ *is stationary and if for every* t

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \tag{5.8}$$

where $\{Z_t, t = \pm 1, \pm 2, \cdots\}$ *is a sequence of white noise, i.e., uncorrelated and identically distributed r.v.'s with mean 0 and variance* σ_z^2. *Compactly we may write* (5.8) *as*

$$\phi(B) X_t = \theta(B) Z_t,$$

where ϕ *is the autoregressive part and* θ *the moving average part of the process.*

Notice that the $ARMA$ process has mean equal to 0. We say that $\{X_t\}$ is an $ARMA(p, q)$ with mean μ if $\{X_t - \mu\}$ satisfies (5.8).

Definition 5.6. *An* $ARMA(p, q)$ *process is called casual if the polynomials* ϕ *and* θ *have no common zeros and* ϕ *has no zeros inside or on the unit circle in the complex plane. A causal* $ARMA$ *process can be written as an infinite moving average*

$$X_t = \left(\frac{\theta}{\phi}\right)(B) Z_t,$$

which is the unique stationary solution of (5.8).

An $ARMA$ process has *short memory* in the sense that the autocorrelation function $\rho(k)$ satisfies

$$|\rho(k)| \leq b a^k, \text{ for some constants } b \in (0, \infty) \text{ and } a \in (0, 1). \tag{5.9}$$

The $ARMA$ class is very convenient for modelling stationary short memory time series. However, in practice, stationarity of a time series is not always observed. Often, the series is transformed to "make it look more stationary". A frequently applied method is differencing, i.e., instead of the original series one looks at the series

$$(1 - B) X_t = X_t - X_{t-1},$$

which will usually remove a global linear trend. This is exactly what has been done at the very beginning of these notes to the risk factors $\ln P_t$, $t \geq 0$. Differencing twice will remove second degree polynomials, and usually we stop here, because, on a finite interval, many functions can be well approximated by polynomials of a reasonably low degree.

Definition 5.7. *A process that, after differencing finitely many times, reduces to an ARMA process, is called an autoregressive integrated moving average (ARIMA) process. An ARIMA process that is ARMA(p, q) after d times differencing is denoted ARIMA(p, d, q), i.e., if* $\{X_t\}$ *is an* $ARIMA(p, d, q)$ *then* $(1 - B)^d X_t$ *is an* $ARMA(p, q)$.

Note that if $d \geq 1$ the $ARIMA$ process is not stationary, and may appear as a model for series exhibiting a "LRD" type property.

5.4.4 The Fractional Brownian motion

The fractional Brownian motion is a stochastic process exhibiting LRD in the sense of (5.4):

Definition 5.8. *A process* $\{\sigma_0 B_{_H}(t),\ t \geq 0\}$ *with* $\sigma_0 > 0$ *is called a fractional Brownian motion if*

1. $B_{_H}(t) \sim N\left(\sigma^2 t^{2H}\right)$.
2. $\mathbb{C}ov\left(B_{_H}(s), B_{_H}(t)\right) = \frac{1}{2}\left(s^{2H} + t^{2H} - |t-s|^{2H}\right)$ *for some* $H \in (0, 1)$.
3. $B_{_H}$ *has continuous sample paths a.s.*

Since the fractional Brownian motion is a Gaussian process, its finite-dimensional distributions are completely determined by *1.* and *2.* Using *2.* it can be shown that for $s < t$,

$$\mathbb{V}ar(B_{_H}(t) - B_{_H}(s)) = \mathbb{V}ar(B_{_H}(t-s)).$$

Hence, the fractional Brownian motion has stationary increments. Notice that for $H = 1/2$, $B_{_H}$ is an ordinary Brownian motion. Fractional Brownian motions have continuous sample paths which become smoother as H increases.

The increment process $Y_t = B_{_H}(t) - B_{_H}(t-1),\ t = 1, 2, \cdots$, of a fractional Brownian motion is called a *fractional Gaussian noise*. It is a mean-zero stationary Gaussian process with ACF

$$\rho_{_Y}(k) = \frac{1}{2}\left(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\right), \quad k = 0, 1, 2, \cdots.$$

For $H = 1/2$, Y is a sequence of i.i.d. $N(0, \sigma^2)$ variables. If $H \neq 1/2$, Y is a dependent sequence. It has been shown (Samorodnitsky and Taqqu, 1994, Proposition 7.2.10) that for $H \neq 1/2$, as $k \to \infty$,

$$\rho_{_Y}(k) \sim H(2H-1)k^{2H-2}.$$

Hence, for $H \in (1/2,\ 1)$ the fractional Gaussian noise exhibits LRD. If $H \in (0, 1/2)$ the autocorrelations $\rho_{_Y}(k)$ are absolutely summable, i.e., the process has short memory.

6 ARCH and GARCH processes

In the classical time series analysis, *linear processes* $\{X_t\}$, like the one in (5.8), are the basic models, all based on an i.i.d. *noise* sequence $\{Z_t\}$. Heavy tails of the distribution of X_t may be explained by a power law behaviour for the distribution of the noise Z_t. But, due to the other peculiarites of return data, one is often forced to model financial data by *non-linear* processes.

6.1 Multiplicative models and the ARCH process

Most models for return data are multiplicative, i.e., we may write

$$X_t = \mu + \sigma_t \, Z_t, \quad t \in \mathbb{Z}, \tag{6.1}$$

where $\{Z_t\}$ is an i.i.d. sequence and $\{\sigma_t\}$ is the so-called *volatility process*, being Z_t and σ_t independent for any fixed t. It is often assumed that Z_t is symmetric around zero, with unit variance. It is also convenient to assume $\mu = 0$, and this will be done in what follows.

Remark 6.1. *Note that, if Z_t is standard normal, conditionally on the past values $X_{t-1}, \sigma_{t-1}, X_{t-2}, \sigma_{t-2}, \cdots$, today's return X_t has a $N(0, \sigma_t^2)$ distribution. This allows us to give "conditional confidence bands for X_t, and it is then easy to obtain the conditional VaR of the sequence X_t".*

Engle (1982) suggested the following model:

Definition 6.1. *The multiplicative model in* (6.1)*, with $\mu = 0$, and with a volatility σ_t given by*

$$\sigma_t^2 = \phi_0 + \sum_{i=1}^{p} \phi_i X_{t-i}^2, \quad t \in \mathbb{Z}, \tag{6.2}$$

where $\phi_i > 0$ if $i \geq 1$, is an autoregressive conditionally heteroscedastic model of order p, also called an $ARCH(p)$ model.

This model may be related to the $ARMA$ models in (5.8). Let us write

$$\epsilon_t := X_t^2 - \sigma_t^2 = \sigma_t^2 (Z_t^2 - 1).$$

Then

$$\phi(B) X_t^2 = \phi_0 + \epsilon_t, \quad t \in \mathbb{Z}. \tag{6.3}$$

If $\{Z_t\}$ is an i.i.d. sequence with mean zero and unit variance, and $\{X_t\}$ is second order stationary, then $\{\epsilon_t\}$ is a white noise sequence. There is thus

a formal analogy with the $AR(p)$ process in (5.8) with $\theta(B) \equiv 1$. But there is also a big difference: whereas the right hand-side of (5.8) depends on an i.i.d. sequence Z_t, the right hand-side of (6.3) depends on the sequence ϵ_t, a dependent sequence.

Since the $ARCH(p)$ provides usually a bad fit to log-returns, unless p is quite large, several people have dealt with the improvement of the $ARCH(p)$ process.

A useful model in finance is the $AR(1)$ with an $ARCH(1)$ error, studied in Borkovec (2001) and Borkovec and Klüppelberg (2001).

Definition 6.2. *A process $\{X_t\}$ is an AR(1) with an $ARCH(1)$ error process if*

$$X_t = \alpha\ X_{t-1} + \sqrt{\left(\phi_0 + \phi_1 X_{t-1}^2\right)}\ Z_t, \quad t \geq 1, \tag{6.4}$$

where $\alpha > 0$, $\phi_0,\ \phi_1 > 0$, $\{Z_t\}$ are i.i.d. symmetric r.v.'s with unit variance, being X_0 independent of Z_t.

For further details on this process see Klüppelberg (2004).

The $GARCH(p,q)$ process generalizes the $ARCH$ process. It has been introduced independently by Bollerslev (1986) and Taylor (1986):

Definition 6.3. *The multiplicative model in (6.1), with $\mu = 0$, is a $GARCH(p,q)$ process if the volatility σ_t is given by*

$$\sigma_t^2 = \phi_0 + \sum_{i=1}^{p} \phi_i X_{t-i}^2 + \sum_{j=1}^{q} \theta_j \sigma_{t-j}^2, \quad t \in \mathbb{Z}, \tag{6.5}$$

with $\phi_i,\ \theta_i > 0,\ i \geq 1$.

Remark 6.2. *The condition for the GARCH equation to define a covariance stationary process with finite variance is that*

$$\sum_{i=1}^{p} \phi_i + \sum_{j=1}^{q} \theta_j < 1.$$

Remark 6.3. *The ARCH and GARCH are technically white noise processes. Indeed, the fact that Z_t and σ_t are independent, and that $\{Z_t\}$ is a strict white noise, enables us to write:*

$$\gamma(h) = \mathbb{C}ov(X_t, X_{t+h}) = \mathbb{E}\left(\sigma_{t+h} Z_{t+h} \sigma_t Z_t\right) = \mathbb{E}(Z_{t+h})\mathbb{E}(\sigma_{t+h}\sigma_t Z_t) = 0.$$

Let us assume that the noise $\{Z_t\}$ is i.i.d. with null mean and unit variance, and let us think on the $GARCH(1,1)$ process. If we denote

$$A_t := \phi_1 Z_{t-1}^2 + \theta_1, \qquad B_t := \phi_0 \quad \text{and} \quad Y_t := \sigma_t^2,$$

we get

$$\sigma_t^2 = \phi_0 + \phi_1 X_{t-1}^2 + \theta_1 \sigma_{t-1}^2 = \phi_0 + \left(\phi_1 Z_{t-1}^2 + \theta_1\right) \sigma_{t-1}^2,$$

i.e.,

$$Y_t = A_t Y_{t-1} + B_t, \quad t \in \mathbb{Z}. \tag{6.6}$$

Note that A_t and Y_{t-1} are independent, and $\{A_t, B_t\}$ is an i.i.d. bivariate sequence. The process in (6.6) is often called a *random coefficient autoregressive* process. The stationary solution may then be obtained through the iteration of (6.6). We may state the following:

Theorem 6.1. (*Nelson, 1990; Bougerol and Picard, 1992*). *There exists an a.s. unique strictly stationary solution (dependent only on past and present values of the Z's) of* (6.6)*, with* $B_t \equiv \phi_0$*, if and only if* $\phi_0 > 0$ *and* $\mathbb{E} \ln A_1 < 0$.

A general $GARCH(p,q)$ process can be embedded in a similar stochastic recurrence equation, where $\mathbf{A}_t$'s are i.i.d. ramdom matrices of size $(p+q-1) \times (p+q-1)$ and the $\mathbf{B}_t$'s are i.i.d. $(p+q-1)$-dimensional random vectors. See Bougerol and Picard (1992) for a generalization of Theorem 6.1.

To understand better the motivation behind the $ARCH$ process, see Engle (1995). From an extremal point of view, the $ARCH$, and more generally the $GARCH$ processes have regularly varying marginal and finite-dimensional distributions, even for a large class of light-tailed noises. This is due to a fundamental result of Kesten (1973), with the following univariate version:

Theorem 6.2. *Let* $\{A_t, B_t\}$ *be an i.i.d. bivariate sequence of positive r.v.'s, and consider the process* $\{Y_t\}$ *in* (6.6)*. Let us assume that the distribution of the random walk*

$$\ln \prod_{i=1}^{n} A_i = \sum_{i=1}^{n} \ln A_i, \quad n \geq 1, \tag{6.7}$$

is not concentrated on a lattice. If $\mathbb{E}\ \ln A_1 < 0$*, and if there exists a constant* $\alpha > 0$ *such that* $\mathbb{E}\ A_n^\alpha = 1$*, then the right tail of* Y_t *is regularly varying with exponent* $-\alpha$*, i.e.,*

$$P(Y_t > x) \sim C_\alpha x^{-\alpha}, \quad \text{as } x \to \infty. \tag{6.8}$$

Remark 6.4. *Clearly a similar feature is true for the autoregressive conditional heteroscedastic (ARCH) process introduced by Engle (1982). The first order ARCH process is defined by*

$$X_t = \left(\phi_0 + \phi_1\ X_{t-1}^2\right)^{1/2} Z_t, \quad t \geq 1, \tag{6.9}$$

where Z_t *are independent standard normal random variables,* $\beta > 0$*,* $0 < \lambda < 1$*. Consequently* $\left\{Y_t := X_t^2\right\}$ *satisfies* (6.6)*, with* $A_t = \phi_1 Z_t^2$*,* $B_t = \phi_0 Z_t^2$*.*

Under general conditions of the type of the ones in Theorem 6.2, it is possible to prove that these processes possess an extremal index $\theta < 1$. There are explicit expressions for such an extremal index, but these expressions are difficult to evaluate numerically. For conclusions on the modelling of financial time series data, or more specifically log-return data, we advise the reading of Mikosch (2004), who refers that "there are indeed a lot of questions without answers, which will keep the academic community and practioners busy for the next few years". A major challenge is the multivariate case. Fortunately a lot has been done recently: see, for instance, Dias (2003). Indeed, *Extreme Value Theory*, *Statistics of Extremes* and *Time Series Analysis* are very useful tools in the analysis of financial data. However, Statistics of Extremes is based on asymptotic results, and so we need large sample sizes; and those large samples make the possibility of changes in the data more likely, and the statistical analysis less reliable. See Dias and Embrechts (2002) for a detection of change points, with the fitting of possibly different models to different time regions.

References

[1] Balkema, A. and L. de Haan (1974). Residual life time at great age. *Ann. Probab.* **2**, 792-804.

[2] Basrak, B., Davis, R. and T. Mikosch (2002). Regular variation of GARCH processes. *Stoch. Processes Appl.* **99**, 95-115.

[3] Beirlant, J., Dierckx, G., Goegebeur, Y. and G. Matthys (1999). Tail index estimation and an exponential regression model. *Extremes* **2**:2, 177-200.

[4] Beran, J. (1994). *Statistics for Long-Memory Processes*. Chapman and Hall.

[5] Binghman, N. H., Goldie, C. M. and J. Teugels (1987). *Regular Variation*. Cambridge Univ. Press.

[6] Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *J. Econometrics* **31**, 307-327.

[7] Borkovec, M. (2001). Asymptotic behaviour of the sample autocovariance and autocorrelation function of the AR(1) process with ARCH(1) errors. *Bernoulli* **7**, 847-872.

[8] Borkovec, M. and C. Klüppelberg (2001). The tail of the stationary distribution of an autoregressive process with ARCH(1) errors. *Ann. Appl. Probab.* **11**, 1220-1241.

[9] Bougerol, P. and N. Picard (1992). Stationarity of GARCH processes and of some nonnegative time series. *J. Econometrics* **52**, 115-127.

[10] Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods*, 2nd edition. Springer-Verlag.

[11] Caeiro, F. and M. I. Gomes (2002). A class of "asymptotically unbiased" semiparametric estimators of the tail index. *Test* **11**:2, 345-364.

[12] David, H. N. and H. N. Nagaraja (2003). *Order Statistics*, 3rd ed. Wiley.

[13] Davison, A. C. and R. Smith (1990). Models for exceedances over high thresholds. *J. Royal Statist. Soc.* **B52**, 393- 442.

[14] Dekkers, A. and L. de Haan (1989). On the estimation of the extreme value index and large quantile estimation. *Ann. Statist.* **17**, 1795-1832.

[15] Dias, A. (2003). *Copula Inference for Finance and Insurance*. Ph. D. Thesis, Swiss Federal Institute of Technology, Zurich, Switzerland.

[16] Dias, A. and Embrechets, P. (2002). Change point analysis for dependence structures in finance and insurance. In L. Carvalho et al. (eds.), *Novos Rumos em Estatstica*, 69-86.

[17] Dwass, M. (1964). Extremal processes. *Ann. Math. Statist.* **35**, 1718-1725.

[18] Embrechts, P., Kluppelberg, C. and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag.

[19] Engle, R. F. (1982). Autoregressive conditional heteroscedastic models with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987-1007.

[20] Engle, R. F. (ed.) (1995). *ARCH Selected Readings*. Oxford University Press, UK.

[21] Ferreira, A. Peng, L. and L. de Haan (2003). On optimising the estimation of high quantiles of a probability distribution. *Statistics* **37**(5), 401-434.

[22] Feuerverger, A. and P. Hall (1999). Estimating a tail exponent by modelling departure from a Pareto distribution. *Ann. Statist.* **27**, 760-781.

[23] Finkenstädt, B. and H. Rootzén (eds.) (2004). *Extreme Values in Finance, Telecommunications, and the Environment*. Chapman & Hall/CRC.

[24] Fisher, A. and L. H. C. Tippett (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Camb. Phil. Soc.* **24**, 180-190.

[25] Fraga Alves, M. I. and M. I. Gomes (1996). Statistical choice of extreme value domains of attraction — a comparative analysis. *Communications in Statistics* **25**, 789-811.

[26] Fraga Alves, M. I., Gomes, M. I. and L. de Haan (2003). A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica* **60**:2, 193-214.

[27] Frey, R. and A. J. McNeil (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *J. Emp. Finance* **7**, 271-300.

[28] Galambos, J. (1987). *The Assymptotic Theory of Extreme Order Statistics*. Krieger.

[29] Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* **44**, 423-453.

[30] Gomes, M. I. (1981). An i-dimensional limiting distribution function of largest values and its relevance to the statistical theory of extremes. In C. Taillie et al. (eds.), *Statistical Distributions in Scientific Work*, Vol. **6**, D. Reidel, 389-410.

[31] Gomes, M. I. (1982). A note on statistical choice of extremal models. In *Actas IV Matemticas Hispano-Lusas*, Vol. **II**, 653-656.

[32] Gomes, M. I. (1984). Penultimate limiting forms in extreme value theory. *Ann. Inst. Statist. Math.* **36A**, 71-85.

[33] Gomes, M. I. (1987). Extreme Value Theory — statistical choice. In P. Revész et al. (eds.), *Goodness-of-Fit*, 195-209. North-Holland, Amsterdam.

[34] Gomes, M. I. (1993). On the estimation of parameters of rare events in environmental time series. In *Statistics for the Environment*, V. Barnett and K. F. Turkman (eds.), Wiley, 225-241.

[35] Gomes, M. I. and L. de Haan (1999). Approximation by penultimate extreme value distributions. *Extremes* **2**:1, 71-85.

[36] Gomes, M. I. and M. J. Martins (2002). "Asymptotically unbiased" estimators of the tail index based on external estimation of the second order parameter. *Extremes* **5**:1, 5-31.

[37] Gomes, M. I., Martins, M. J. and M. Neves (2000). Alternatives to a semi-parametric estimator of parameters of rare events — the Jackknife methodology. *Extremes* **3**:3, 207-229.

[38] Gomes, M. I. and D. D. Pestana (1987). Nonstandard domains of attraction and rates of convergence. In M.L. Puri *et al.* (eds.), *New Perspectives in Theoretical and Applied Statistics*, Wiley, New York, 467-477.

[39] Gumbel, E. J. (1958). *Statistics of Extremes*. Columbia University Press, New York.

[40] Gumbel, E. J. (1962). Statistical theory of extreme values (main results). In Sarhan, A. R. and B. G. Greenberg eds. *Contributions to Order Statistics*, Wiley, New York, 56-93.

[41] Gumbel, E. J. (1966). Two systems of bivariate extremal distributions. *Bull. Intern. Statist. Inst.* **41**.

[42] Haan, L. de (1970). *On Regular Variation and Its Application to Weak Convergence of Sample Extremes*. Mathematical Centre Tracts **32**, Mathematisch Centrum, Amsterdam.

[43] Haan, L. de and L. Peng (1998). Comparison of tail index estimators. *Statistica Neerlandica* **52**, 60-70.

[44] Haan, L. de and H. Rootzén (1993). On the estimation of high quantiles. *J. Statist. Planning and Inference* **35**, 1-13.

[45] Hall, P. and A. H. Welsh (1985). Adaptive estimates of parameters of regular variation. *Ann. Statist.* **13**, 331-341.

[46] Hazen, A. (1914). Storage to be provided in impounding reservoirs for municipal water supply. *Trans. Amer. Soc. Civil Engineers* **77**, 1529-1669.

[47] Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**, 1163-1174.

[48] Hosking, J. R. M. and J. R. Wallis (1987). Parameter and quantile estimation for the Generalized Pareto distribution. *Technometrics* **28**, 334-349.

[49] Hosking, J. R. M., Wallis, J. R. and E. F. Wood (1985). Estimation of the GEV distribution by the method of the probability-wheigthed moments. *Technometrics* **27**, 251-261.

[50] Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng.* **116**, 770-799.

[51] Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological events. *Quarterly J. Royal Meteor. Soc.* **81**, 158-172.

[52] Jorion, P. (2001). *Value at Risk: the New Benchmark for Measuring Financial Risk*. McGraw-Hill, New York.

[53] Kesten, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Mathematica* **131**, 207-248.

[54] Klüppelberg, C. (2004). Risk management with Extreme Value Theory. In Finkenstädt and Rootzén (eds.) *Extreme Values in Finance, Telecommunications, and the Environment*, 101-168, Chapman & Hall/CRC.

[55] Lamperti, J. (1964). On extreme order statistics. *Ann. Math. Statist.* **35**, 1726-1737.

[56] Leadbetter, M. R. Lindgren, G. and H. Rootzén (1983). *Extremes and Related Properties of Random Sequences and Series*. Springer-Verlag.

[57] Mandelbrot, B. B. (1975). Limit theorems on the self-normalized range for weakly and strongly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie and verwandte Gebiete* **31**, 271-285.

[58] Mikosch, T. (2004). *Modeling dependence and tails of financial time series*. In Finkenstädt and Rootzén (eds.) *Extreme Values in Finance, Telecommunications, and the Environment*, 185-286, Chapman & Hall/CRC.

[59] Nelson, D. B. (1990). Stationary and persistence in the GARCH(1,1) model. *Econometric Theory* **6**, 318-334.

[60] Pickands, J. (1975). Statistical inference using extreme-order statistics. *Ann. Statist.* **3**, 119-131.

[61] Rachev, S. and S. Mittnik (2000). *Stable Paretian Models in Finance*. Wiley.

[62] Reiss, R.-D. (1989). *Asymptotic Distribution of Order Statistics*. Springer-Verlag.

[63] Reiss, R.-D. (1993). *A Course on Point Processes*. Springer-Verlag.

[64] Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag.

[65] Sahran A. E., and B. G. Greenberg (1962). *Contributions to Order Statistics*. Wiley.

[66] Samorodnitsky, G. and M. S. Taqqu (1994). *Stable Non-Gaussian Random Processes. Stochastic Models with Infinite Variance*. Wiley, New York.

[67] Smirnov, N. V. (1952). Limit distributions for the terms of a variational series. *Amer. Math. Soc. Transl.*, Ser. **1**, N. 67.

[68] Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72**, 67-90.

[69] Smith, R. L. (1986). Extreme value theory based on the largest annual events. *J. Hidrology* **86**, 27-43.

[70] Stegeman, A. (2002). *The Nature of the Beast — Analyzing and Modeling Computer Network Traffic*. Rijksuniversiteit Groningen.

[71] Taqqu, M. S. (1975). Weak convergence of fractional Brownian motion to the Rosenblatt process. *Zeitschrift für Wahrscheinlichkeitstheorie and verwandte Gebiete* **31**, 287-302.

[72] Taqqu, M. S. and V. Teverovsky (1995). Estimators for long-range dependence: an empirical study. *Fractals* **3**, 785-798.

[73] Taqqu, M. S. and V. Teverovsky (1996). On estimating the intensity of long-range dependence in finite and infinite variance time-series. In Adler et al. 8eds.). *A Practical Guide to heavy tails: Statistical Techniques and Applications*. Birkhauser, 177-217.

[74] Tawn, J. (1988) Bivariate extreme value theory — models and estimation. *Biometrika* **75**, 397-415.

[75] Tawn, J. (1992). Estimating probabilities of extreme sea levels. *Appl. Statist.* **41**, 77-93.

[76] Taylor, S. J. (1986). *Modeling Financial Time Series*. Springer, New York.

[77] Tiago de Oliveira, J. (1968). Extremal processes; definition and properties. *Publ. Inst. Statist. Univ. Paris* **XII**, 25-36.

[78] Tiago de Oliveira, J. (1971). A new model of bivariate extremes; statistical decision. In *Studi di Probabilit, Statistica e Ricerca Operativa in onore di Giuseppe Pompilj*, 1-13.

[79] Tiago de Oliveira, J. (1981). Statistical choice of univariate extreme models. In C. Taillie et al. (eds.), *Statistical Distributions in Scientific Work*, Vol. **6**, D. Reidel, 367-387.

[80] Tiago de Oliveira, J. (ed.) (1984). *Statistical Extremes and Applications*. D. Reidel, Dordrecht.

[81] Tiago de Oliveira, J. (1997). *Statistical Analysis of Extremes*. Ed. Pendor, vora, Portugal.

[82] von Mises, R. (1954). La distribution de la plus grande de n valeurs. In *Selected Papers*, **Vol II**, 271-294. American Mathematical Society.

[83] Weibull, W. (1939). The phenomenon of systems in solids. *Ing. Vetenskaps. Akad. Handl.* **153**.

[84] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Assoc.* **73**, 812-815.