A survival kit on quantile estimation

Abstract

Questions concerning risk management in finance and premium calculation in non-life insurance often involve quantile estimation. We give an introduction to the basic extreme value theory which yields a methodological basis for the analysis of such questions.

1. Introduction

The following two dates are important with respect to risk management:

- February 1, 1953,
- January 28, 1986.

Upon asking "why?" to audiences of bankers and/or (re)insurers, we hardly ever got the right answer. First of all, during the night of February 1, 1953, at various locations the sea dykes in the Netherlands collapsed during a severe storm, causing major flooding in large parts of coastal Holland and killing over 1800 people. An extremal event (a record surge) caused a protective system (the sea dykes) to break down. In the wake of this disaster, the Dutch government established the Delta-committee. Under van Dantzig, statistical problems discussed included the task of answering the following question: "What would the dyke heights have to be (at various locations) so as to prevent future catastrophic flooding with a sufficiently high probability?". Though the above formulation is a gross simplification of the questions discussed, it pays to look a bit more in detail at the issues underlying a possible answer. To (over)simplify matters even further, suppose we are given data $X_1, \ldots, X_n$ interpreted as iid observations on maximum annual wave heights, say, with common distribution function F. The iid assumption is clearly only a first approximation especially now where, due to changing climatological factors, more general models will have to be looked at more carefully. The Dutch government required the Delta-committee to estimate the necessary dyke heights which would imply a so-called t-year event for t sufficiently large. In engineering language, a t-year event related to a random variable (rv) X with distribution function (df) F corresponds to a value

$$x_t = F^{\leftarrow}(1 - t^{-1})$$

where $F^{\leftarrow}$ denotes the generalised inverse of F, i.e.

$$F^{\leftarrow}(t) = \inf\{x \in \mathbb{R} : F(x) \geq t\}.$$

Its obvious interpretation is that, with probability $1 - t^{-1}$, the level x_t will not be superceeded. Formulated differently, x_t is the $(1 - t^{-1})$ upper quantile of the df F. In the case of the dyke problem, what were the available historical data? The basic data consisted of observations on maximum annual heights. Moreover, the previous record surge occured in 1570 with a recorded value of (NAP+4)m, where NAP stands for normal Amsterdam level. The 1953 flooding corresponded to a (NAP+3.85)m surge. The statistical analysis in the Delta-report led to an estimated $(1 - 10^{-4})$-quantile of (NAP+5.14)m for the yearly maximum, based on a total of 166 historical observations. We most strongly recommend the paper by de Haan [Ha] for further reading on this topic. Clearly, the task of estimating such a 10 000-year event leads to

$$\boxed{\text{estimating well beyond the range of the data.}}$$

The only way in which such a problem can be solved is by making extra assumptions on the underlying model, i.e. on the df F of the annual maxima. Below some such models will be introduced.

The previous example already stressed the importance of estimating quantiles well beyond the range of the data. The second date, January 28, 1986 corresponds to the tragic explosion of the Space Shuttle Challenger. Our account concerning the search for possible causes is taken from the splendid book by Feynman [Fe] and the statistical analysis in Dalal, Fowlkes and Hoadley [DFH]. In his contribution to the Presidential Committee investigating the Challenger disaster, Richard Feynman found out that a potential cause may have been the insufficient functioning of so-called O-rings used for sealing the various segments of the external rocket boosters. The latter was mainly due to the exceptionally low temperature ($31^o F$) the night before launching. The data-set plotted in Figure 1 plays an important role in the discussion in [DFH].

Each plotted point represents a shuttle flight that experienced thermal distress on the O-rings; the x-axis shows the ground temperature at launch and the y-axis shows the number of O-rings that experienced some thermal distress. From an analysis of the effect of (low) temperature on O-ring performance the night before the accident, using the data in Figure 1a, one decided that "The temperature data are not conclusive on predicting primary O-ring blowby". After the accident a presidential commission [PC], p.145, noted that a mistake in the analysis of the thermal-distress data (Figure 1a) was that the flights with zero incidents were left off the plot because it was felt that these flights did not

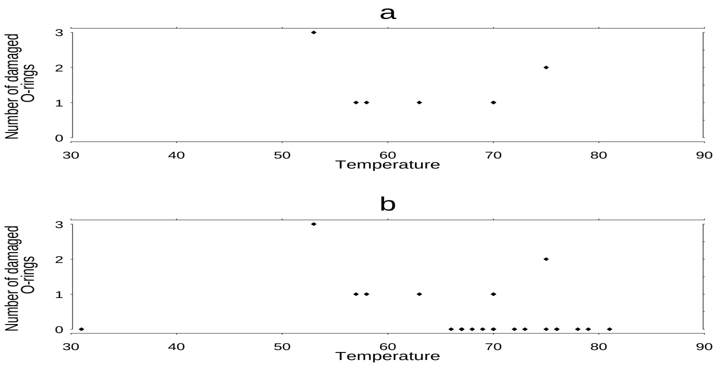

Figure 1: *Temperature versus number of O-rings having some thermal distress. Panel b includes flights with no incidents.*

contribute any information about the temperature effect. Including these data (see Figure 1b), a significant dependence between O-ring damages and temperature was concluded. In [DFH] a detailed statistical analysis, based on a logistic regression model, is performed for the data in Figures 1a and b. The outcome of which showed a very significant increase in the failure probability in the case of Figure 1b as compared to Figure 1a. Richard Feynman summarised his findings as follows:

> It appears that there are enormous differences of opinion as to the probability of the failure with loss of vehicle and human life. The estimates range from roughly 1 in 100 to 1 in 100 000. The higher figures come from management. As far as I can tell "engineering judgement" means that they are just going to make up numbers!...It was clear that the numbers were chosen so that when you add up everything you get 1 in 100 000.

The resulting message from the above for risk management should be clear!

Always look very critically at the available data.

The basic aim of the present paper is to show that relevant methodology from extreme value theory offers a modelling tool which helps answer some of the technical questions in modern risk management. Our paper should be viewed as a key opening the door to this essential toolkit. The reader is referred to the other papers in this volume for further details on some of the techniques mentioned. A general overview on the for insurance and finance relevant extreme value theory, together with many examples and references is to be found in Embrechts, Klüppelberg and Mikosch [EKM].

2. On Value-at-Risk

One of the catchwords now frequently heard at all levels of financial institutions is "VaR" standing for Value-at-Risk. Though not uniquely describing one well-defined mathematical concept, in general terms VaR comes about as follows. Suppose a rv X with df F and density f describes the profit and losses of a certain financial instrument. The latter can be a particular asset, a portfolio or even the global financial exposure of a bank. Typically, the so-called Profit-Loss (P&L-) distribution may look like the example given in Figure 2.

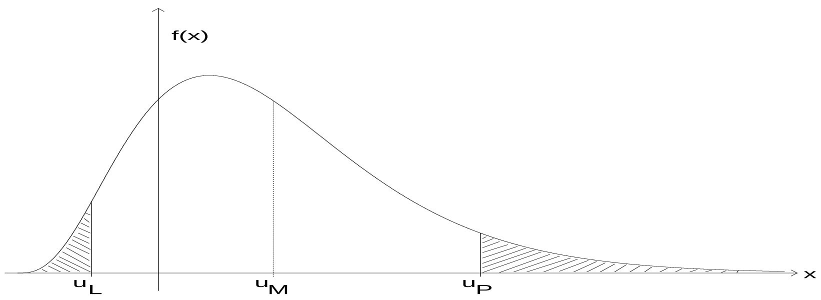

Figure 2: *Possible P&L-density with α_L-quantile u_L, $(1-\alpha_P)$-quantile u_P, for $0 < \alpha_P, \alpha_L < 1$, and mean u_M.*

A typical VaR calculation (for a portfolio say) has the following ingredients: "Calculate a possible extreme loss resulting from holding the portfolio for a fixed period (10 days say) using as a measure of underlying riskiness the volatility over the last 100 days". The notion *extreme loss* is translated as the left 1 or 5%-quantile u_L for the P&L-distribution. Hence, in its easiest form, VaR estimates u_L for sufficiently small values of α_L. As in the dyke example, this quantile estimation will typically be outside (or at the edge of) the range of available data. Moreover, management may also be interested in setting targets at the profit side of the P&L-curve, i.e. it wants estimates for u_P. One step further would involve questions concerning the estimation of the size of under- (or over-) shoot of the levels u_L, respectively u_P, given that these events have occurred. Consequently, we want to estimate the following functions:

- $P(X - u_L \leq -x \mid X < u_L) = \theta(x \mid u_L), \quad x \geq 0,$

- $P(X - u_P > \ x \mid X > u_P) = \theta(x \mid u_P), \quad x \geq 0.$

In the finance context, $\theta(x \mid u_L)$ is often referred to as the *shortfall distribution*.

One of the basic messages of the present paper is that modern extreme value theory offers the necessary tools and techniques for answering the following questions:

- What are potential parametric models for the P&L-distribution function?
- Given such a model how can we estimate the quantiles u_L (VaR) and u_P?
- What are the resulting models for the conditional probabilities $\theta(\cdot \mid u_L)$ and $\theta(\cdot \mid u_P)$? How can these be estimated efficiently?
- Can these procedures be extended to more complex situations including non-stationarity (e.g. trends, seasonality), the use of exogenous economic variables, a multivariate set-up,...?

The good news is that all of the above questions can be answered. A detailed discussion however would be beyond the scope of this paper. Below we have listed some of the terminology which every risk management group should be aware of.

A survival kit on quantile estimation
• Quantiles • QQ-plots • The mean excess function • The generalised extreme value distribution • The generalised Pareto distribution • The point process of exceedances • The POT-method: peaks over thresholds • The Hill and Pickands estimators, together with their ramifications and extentions

Though this list is not complete, it contains some of the relevant items from extreme value theory. Omissions for instance are methods for continuous-time stochastic processes and, more importantly, multivariate techniques. The reader is referred to Embrechts et al. [EKM] for an extensive list of references and indications for further reading. In the following paragraphs, we explain some of the items above more in detail.

3. Basic definitions

Suppose F is the distribution function of a real-valued random variable X. If F has a density we denote it by f. The (right) tail of F is denoted by $\bar{F} = 1 - F$. In the sequel we only concentrate on $\bar{F}(x)$ for $x \geq 0$; similar results may be obtained for left tails, i.e. $F(-x)$ for $x \geq 0$ or for the two-sided tail $T(x) = 1 - F(x) + F(-x),\ x \geq 0$. Given a threshold probability $p \in (0, 1)$, the p-quantile x_p of F is defined as (see Figure 3):

$$x_p = F^{\leftarrow}(p) = \inf\{x \in I\!R : F(x) \geq p\}.$$

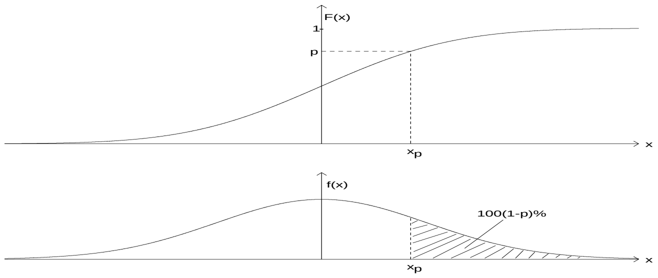

Figure 3: *Graphical representation of the p-quantile; once by means of the distribution function and once by means of the density.*

A key question now concerns the statistical inference of x_p, typically for p close to 1. Suppose the data $X_1, \ldots, X_n$ are iid with df F. An obvious estimator uses the so-called empirical distribution function

$$F_n(x) = \frac{\#\{i : 1 \leq i \leq n \text{ and } X_i \leq x\}}{n}, \quad x \in I\!R.$$

From the Glivenko-Cantelli theorem we know that with probability 1,

$$\sup_{x \in I\!R} |F_n(x) - F(x)| \to 0, \qquad n \to \infty.$$

Consequently, for given $p \in (0, 1)$, we may estimate $x_p = F^{\leftarrow}(p)$ by

$$\hat{x}_{p,n} = F_n^{\leftarrow}(p).$$

The properties of this estimator can most easily be expressed in terms of the ordered data. Denote by $X_{k,n}$ the kth largest observation. For the sake of simplicity, we assume that F is continuous so that with probability 1, there are no ties in $X_1, \ldots, X_n$. The elements of the ordered sample

$$X_{1,n} = \max(X_1, \ldots, X_n) \geq X_{2,n} \geq \cdots \geq X_{n,n} = \min(X_1, \ldots, X_n)$$

are called the *order statistics*. One readily verifies that, for $k = 1, \ldots, n$,

$$\hat{x}_{p,n} = F_n^{\leftarrow}(p) = X_{k,n}, \quad 1 - \frac{k}{n} < p \leq 1 - \frac{k-1}{n}. \tag{3.1}$$

Using the Central Limit Theorem (CLT), it is not difficult to show that for a df F with continuous density f with $f(x_p) \neq 0$ and $k = k(n)$ so that $n - k = np + o(n^{\frac{1}{2}})$,

$$\hat{x}_{p,n} \sim \mathrm{AN}\left(x_p, \frac{p(1-p)}{nf^2(x_p)}\right). \tag{3.2}$$

Using a slight, though obvious abuse of notation, $\mathrm{AN}(\mu, \sigma^2)$ stands for asymptotically normal with mean μ and variance σ^2. Also, using the iid assumption on $X_1, \ldots, X_n$, one immediately obtains, for $i < j$:

$$P\left(X_{j,n} \leq x_p < X_{i,n}\right) = \sum_{r=i}^{j-1} \binom{n}{r} p^{n-r}(1-p)^r. \tag{3.3}$$

Either from (3.2) or (3.3), one can obtain confidence intervals (approximate in the case of (3.2)) for the required quantiles. However notice the main problems.

- In the case of (3.2), asymptotics for $n \to \infty$ are used. This allows for the estimation of p close to 1 if sufficient data are available. A crucial choice to be made is that of $k = k(n)$.

- In the case of (3.3), exact confidence intervals can be obtained for n not too large. One very much relies on the fact that p is such that x_p lies *within* the range of the data.

Though the above approaches can be refined, one clearly faces basic problems for p-values such that $1 - p \ll n^{-1}$. As stated in the Section 1, in order to estimate quantiles for p close to 1 we must assume extra (parametric) conditions on F. Standard methods from statistical data fitting can (and should) be used:

- likelihood techniques,

- goodness-of-fit procedures,

- Bayesian methods.

Two graphical techniques deserve special mention. First of all the so-called *QQ-* (or *quantile-*) *plot* as a graphical goodness-of-fit procedure. Suppose $F(.;\theta)$ is a parametric model to be fitted to $X_1, \ldots, X_n$, resulting in an estimate $\hat{\theta}$ and hence a fitted model $\hat{F} = F(.;\hat{\theta})$. The plot of the points

$$\left\{\left(X_{k,n}, \hat{F}^{\leftarrow}\left(p_{k,n}\right)\right) : k = 1, \cdots, n\right\}$$

for an appropriate plotting sequence $(p_{k,n})$ is the QQ-plot of the data $X_1, \ldots, X_n$ when fitting the parametric family $F(.;\theta)$. One often takes $p_{k,n} = (n - k +$

1)/(n + 1), or some variant of this. A good fit results in a straight line of the plot, whereas model deviations (like heavier tails, skewness, outliers, ...) can easily be diagnosed. An example is shown in Figure 4.

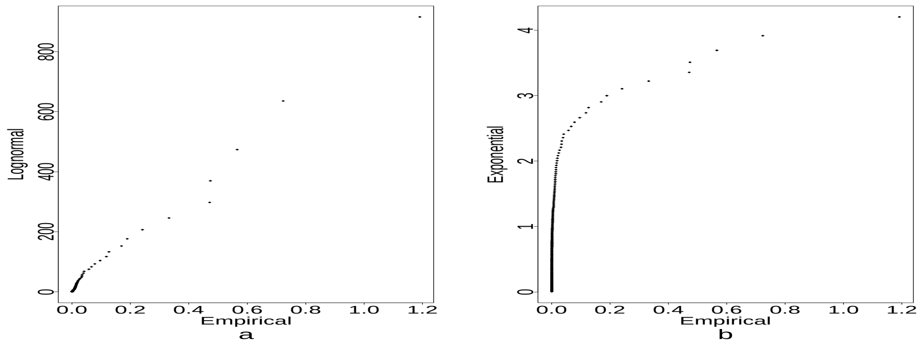

Figure 4: *QQ-plot for lognormal versus lognormal* (a) *and versus the exponential distribution* (b). *The sample size is 200. One sees in the right picture that curvature appears, the empirical quantiles grow faster than the theoretical ones: the simulated lognormal data is heavy-tailed in comparison to the exponential model.*

A second graphical technique, used mainly in biostatistics, insurance and reliability, is the so-called *mean excess function* (*mef*)

$$e(u) = E\left(X - u | X > u\right), \quad u \geq 0, \tag{3.4}$$

and its empirical counterpart

$$e_n(u) = \frac{\sum_{i=1}^{n} (X_i - u)^+}{\sum_{i=1}^{n} I_{\{X_i > u\}}}. \tag{3.5}$$

Here $y^+ = \max(y, 0)$ and I_A denotes the indicator function of A, i.e. $I_A(x) = 0$ or 1 depending on whether $x \notin A$ or $x \in A$. By plotting $\{(u, e_n(u)) : u \geq 0\}$, as in Figure 5, one easily distinguishes between short- and long-tailed distributions. Figure 6 shows an application of these graphical techniques to a real data-set. For applications in insurance, see for instance Hogg and Klugman [HK] and references therein.

We would like to stress that both the QQ-plot and the mef belong to the realm of exploratory data analysis. The QQ-plot gives a quick assessment of the plausibility of certain distributional model assumptions. More importantly, it quickly shows how the data deviates from a model proposed. For instance concavity hints at a heavier-tailed model as is clearly shown in Figure 4 (b).

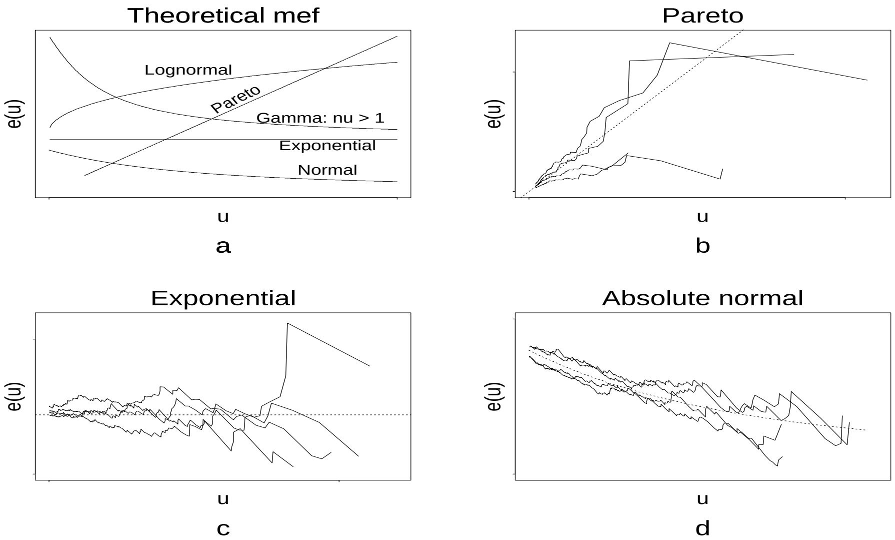

Figure 5: *Theoretical mefs (Figure a) and some examples of empirical mefs for simulated data (n=200). Normal in Figure a stands for the absolute value of a normally distributed rv. In Figure b the empirical mef of 4 simulated* $Pareto(1.5)$ *distributed data-sets are plotted, Figure c corresponds to* $Exp(1)$*, and d to the absolute value of a normally distributed rv. The dotted line shows the theoretical mef of the underlying distribution.*

Though mefs could serve a similar purpose, we tend to use them only to distinguish between heavy-tailed and short-tailed distributions. The former typically exhibit an increasing pattern, wheras the latter are either constant or decreasing; see Figure 5 (a). In the Figures 5 (b), (c) and (d) we have simulated 4 realisations of the mef for a given model. Especially in the heavy-tailed Pareto case, there is great instability towards the higher u-values. The latter is due to the sparseness of the data in that range. It clearly shows that mefs have to be treated with care.

4. Extremes versus sums

We find it pedagogically usefull to compare and contrast the Central Limit Theorem with the main results from Extreme Value Theory. One reason for this is that the mathematical technology used in both cases is fairly similar. A second reason is that a comparison of both theories often leads to important new insight in extremal event problems at hand. For a detailed discussion on this, together with many examples, see Embrechts, Klüppelberg and Mikosch

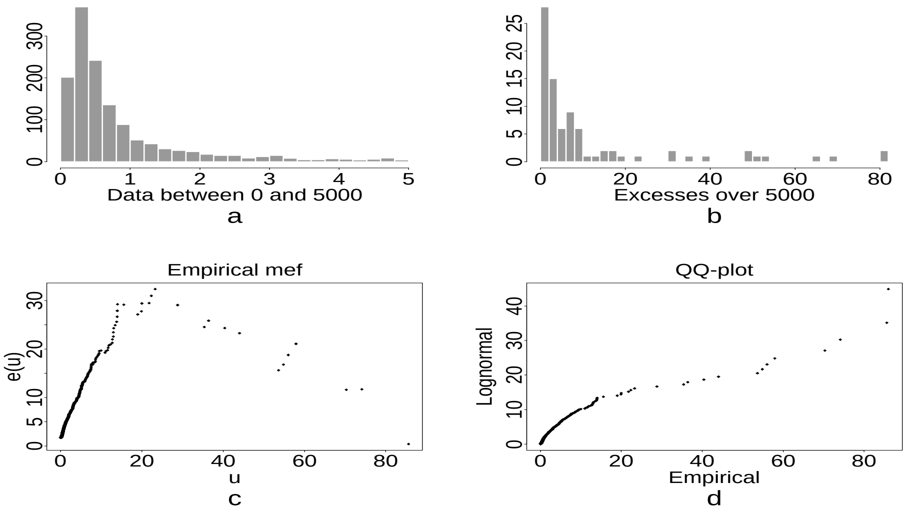

Figure 6: *Graphical representation of fire insurance data (in units of 1 000 Sfr). The sample size is 1 438. First two histogramms are plotted: one with the "smaller" data (a) and one with the higher data (b). A QQ-plot against the lognormal (see (d)) shows that the data are more heavy-tailed. Indeed (c) hints at a Pareto like model.*

[EKM].

The general formulation of the CLT goes as follows. Suppose $X_1, \ldots, X_n$ are iid random variables with df F. Define the partial sum $S_n = X_1 + \cdots + X_n$. The CLT now solves the following problems.

(CLT 1) Given F, find constants $a_n > 0$ and $b_n \in \mathbb{R}$ so that

$$\frac{S_n - b_n}{a_n} \xrightarrow{d} Y, \quad n \to \infty, \tag{4.6}$$

where Y is a non-degenerate rv with df G say.

(CLT 2) Characterise the df G of Y in (4.6) above .

(CLT 3) Given a possible df G in (4.6), find all dfs F which satisfy (4.6) (the domain of attraction problem), characterise the sequences (a_n) and (b_n).

All the above points can be solved and yield results typically belonging to the *sum world*. We all know the special case for when $\sigma^2 < \infty$, indeed then $a_n = \sqrt{n}\sigma,\ b_n = n\mu,\ G = \mathrm{N}(0,1)$, the standard normal, and all dfs with finite second moment are attracted to $\mathrm{N}(0,1)$. An essential ingredient in the analysis of the above problems are the so-called functions of *regular variation.* First of all, a positive, Lebesgue measurable function $L: \mathbb{R}^+ \to \mathbb{R}^+ \setminus \{0\}$ is *slowly varying* if

$$\text{for all } \lambda > 0, \quad \lim_{x\to\infty} \frac{L(\lambda x)}{L(x)} = 1. \tag{4.7}$$

Notation: $L \in \mathcal{R}_0$. A positive, Lebesgue measurable function h is *regularly varying with index* $\alpha \in \mathbb{R}$ if $h(x) = x^\alpha L(x)$. Notation: $h \in \mathcal{R}_\alpha$. For a detailed analysis of $\mathcal{R}_\alpha$, together with a discussion of many applications to probability theory, including the CLT, see Bingham, Goldie and Teugels [BGT].

The full solution of (4.6) leads to the class of *stable distributions*, elements of which are mainly characterised by a parameter $\alpha \in (0, 2]$. The case $\alpha = 2$ corresponds to the normal distribution, the case $\alpha = 1$ leads to the Cauchy distribution with density $f(x) = (\pi(1 + x^2))^{-1}$, $x \in \mathbb{R}$. Whenever $\alpha < 2$, $\sigma^2 = \infty$ and $E|X| = \infty$ for $\alpha < 1$, so that stable distributions with $\alpha < 2$ are sometimes used as models for heavy-tailed log-returns, a fact used by Mandelbrot [Ma] to promote these models in the context of finance.

Parallel to the above sum-world, similar questions can now be asked upon replacing S_n by $M_n = X_{1,n} = \max(X_1, \ldots, X_n)$. The following theorem (going back to Fisher and Tippett [FT]) summarises the for us most important results.

Theorem 4.1 *Suppose $X_1, \ldots, X_n$ are iid with df F. If there exist constants $a_n > 0$ and $b_n \in \mathbb{R}$ so that*

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} Y, \quad n \to \infty, \tag{4.8}$$

where Y is non-degenerate with df G, then G is of one of the following types :

1. *(Gumbel)*

$$\Lambda(x) = \exp\left\{-e^{-x}\right\}, \quad x \in \mathbb{R},$$

2. *(Fréchet)*

$$\Phi_\alpha(x) = \begin{cases} 0 & \text{if } x \le 0 \\ \exp\left\{-x^{-\alpha}\right\} & \text{if } x > 0 \end{cases} \quad \text{for } \alpha > 0,$$

3. *(Weibull)*

$$\Psi_\alpha(x) = \begin{cases} \exp\left\{-(-x)^\alpha\right\} & \text{if } x < 0 \\ 1 & \text{if } x \ge 0 \end{cases} \quad \text{for } \alpha > 0.$$

□

The limit distribution in (4.8) is determined up-to type: two dfs F_1 and F_2 are of the same type if there exist constants $a_1 \in \mathbb{R}, a_2 > 0$ so that $F_2(x) = F_1((x - a_1)/a_2)$, i.e. F_1 and F_2 belong to the same location and scale family.

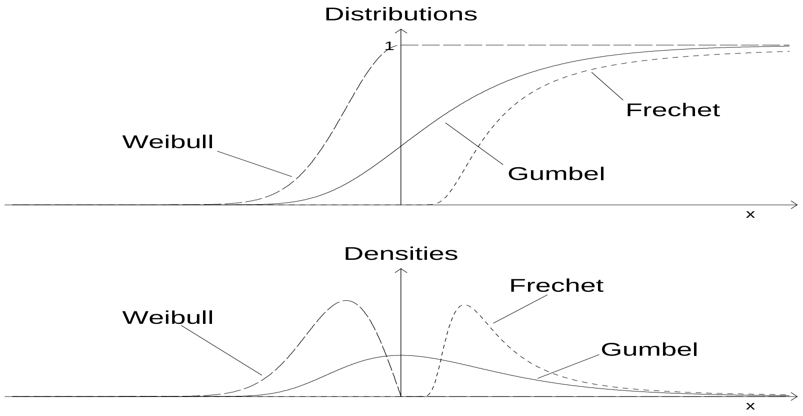

Figure 7: *Distribution functions and densities of the Gumbel, Fréchet (with $\alpha = 2$) and Weibull (with $\alpha = 2$) law.*

Dfs of the type of Λ, Φ_α and Ψ_α are called the (*generalised*) *extreme value distributions.* A useful parametrisation of these three classes is $H_{\xi,\mu,\beta}$ where

$$H_{\xi,\mu,\beta}(x) = \exp\left\{-\left(1+\xi\frac{x-\mu}{\beta}\right)^{-\frac{1}{\xi}}\right\}, \quad 1+\xi\frac{x-\mu}{\beta} \geq 0. \tag{4.9}$$

The adjective "generalised" has no special meaning here; $H_{\xi,\mu,\beta}$ is merely a useful reparametrisation of the Gumbel-, Fréchet- and Weibull-types. This terminology has become widely accepted. The case $\xi = 0$ in (4.9) has to be interpreted as $\xi \to 0$, resulting in the Gumbel type. The df $H_{\xi,\mu,\beta}$ is referred to as the *generalised extreme value* (GEV) *distribution with shape parameter* $\xi \in I\!R$, *location parameter* $\mu \in I\!R$ and *scale parameter* $\beta > 0$. The standard case $\mu = 0,\ \beta = 1$ will be denoted by $H_\xi = H_{\xi,0,1}$. The case $\xi < 0$ corresponds to the Weibull, $\xi = 0$ to the Gumbel and $\xi > 0$ to the Fréchet distribution.

Similar to the CLT, one may try to find a solution to the *maximum domain of attraction problem*, i.e. given H_ξ, find conditions on F so that (4.8) holds for appropriate sequences (a_n) and (b_n). In this case we say that F belongs to the *maximum domain of attraction of* H_ξ *with norming constants* a_n, b_n. The latter is denoted by $F \in MDA(H_\xi)$. For extremes the solution turns out to be much more dependent on the fine behaviour of $1-F$. Two examples may illustrate this in the case of the Gumbel distribution.

1. Suppose $F =$ Exp(1), then as $n \to \infty$,

$$M_n - \ln n \xrightarrow{d} \Lambda.$$

2. Suppose $F = \mathrm{N}(0,1)$, then as $n \to \infty$,

$$(2 \ln n)^{\frac{1}{2}} \left(M_n - (2 \ln n)^{\frac{1}{2}} + \frac{1}{2} \frac{\ln \ln n + \ln 4\pi}{(2 \ln n)^{\frac{1}{2}}} \right) \xrightarrow{d} \Lambda .$$

For applications in insurance and finance, the Gumbel and the Fréchet family turn out to be the most important models for extremal events. The domain of attraction condition for the Fréchet takes on a particularly easy form. The following theorem is due to Gnedenko.

Theorem 4.2 *Suppose $X_1, \ldots, X_n$ are iid with df F. Then there exist sequences (a_n) and (b_n) so that (4.8) holds for Y Fréchet distributed with parameter $\alpha > 0$ (i.e. $F \in MDA(H_{1/\alpha}) = MDA(\Phi_\alpha)$) if and only if*

$$1 - F(x) = x^{-\alpha} L(x), \quad x > 0, \tag{4.10}$$

with L slowly varying as defined in (4.7). □

We have seen that from a methodological point of view, the class of distribution functions with regularly varying tail ((4.10)) enter very naturally as standard models in the area of extreme value theory. From a statistical point of view, the specification (4.10) is a semi-parametric one: parametric in α and non-parametric in L. It should be clear from the defining property (4.7) that a conclusive test for slow variation of $x^\alpha \bar{F}(x)$ in any practical situation is impossible. However, based on the general asymptotic theory, (4.10) is not just a valiant assumption.

A useful property which trivially holds for a positive rv with df satisfying (4.10) is that for $\beta > \alpha, EX^\beta = \infty$ whereas for $\beta < \alpha, EX^\beta < \infty$. It now turns out that parameter estimates for α in (4.10) or alternatively for the range of divergent versus convergent moments rather consistently points at values for $\alpha \in (1,2)$ for claim size distributions in non-life insurance, whereas in finance, often log-return distributions yield $\alpha \in (3,4)$. The former implies the divergence of second moments, whereas the latter corresponds to infinite fourth moment. References on this topic are for instance Hogg and Klugman [HK] for loss data in insurance, and Taylor [Tay] for data in finance. Embrechts, Klüppelberg and Mikosch [EKM] contain numerous examples and references for further reading.

The above suggests which parametric models to use in practice when estimating quantiles beyond the range of the data: a possible answer is the GEV distribution H_ξ, $\xi > 0$. This observation forms the corner-stone to the use of extreme value theory in finance and insurance applications.

Suppose now that $X_1, \ldots, X_n$ are iid with df $F \in MDA(H_\xi)$, $\xi \in \mathbb{R}$. Recall the definition of p-quantile, $F(x_p) = p$. Then

$$\begin{aligned} P\left(a_n^{-1}(M_n - b_n) \le x\right) &= P\left(M_n \le a_n x + b_n\right) \\ &= F^n\left(a_n x + b_n\right) \longrightarrow H_\xi(x), \quad n \to \infty, \end{aligned}$$

whence, by using Taylor expansion,

$$n\bar{F}(u) \approx \left(1+\xi\frac{u-b_n}{a_n}\right)^{-\frac{1}{\xi}}.$$

where $u = a_n x + b_n$. A resulting tail estimator for $\bar{F}(u)$ could take on the form

$$\widehat{\bar{F}(u)} = \frac{1}{n}\left(1+\hat{\xi}\frac{u-\hat{b}_n}{\hat{a}_n}\right)^{-\frac{1}{\hat{\xi}}},$$

yielding as an estimator for the quantile x_p

$$\hat{x}_p = \hat{b}_n + \frac{\hat{a}_n}{\hat{\xi}}\left((n(1-p))^{-\hat{\xi}}-1\right).$$

Here $\hat{a}_n$, $\hat{b}_n$ and $\hat{\xi}$ are suitable estimators of a_n, b_n and ξ. As a candidate estimator for the crucial shape parameter ξ one often encounters the so-called *Hill estimator*

$$\hat{\xi}^{(H)}_{n,k} = \frac{1}{k}\sum_{j=1}^{k}\ln X_{j,n} - \ln X_{k+1,n},$$

where $k = k(n)$ has to be chosen in an appropriate way. Another classical estimator for ξ is the *Pickands estimator*

$$\hat{\xi}^{(P)}_{n,k} = \frac{1}{\ln 2}\ln\frac{X_{k,n}-X_{2k,n}}{X_{2k,n}-X_{4k,n}},$$

where $k = k(n)$, as in the Hill case. The choice of k turns out to be the Achilles heel of these estimators. In order to achieve weak consistency of these estimators, one needs $k \to \infty$ and $k/n \to 0$. More complicated growth conditions on k are required in order to obtain strong consistency and asymptotic normality. We prefer to plot the estimator accross a wide range of k-values and look (hope?) for stability. For a detailed discussion of the properties of $\hat{\xi}^{(H)}$, $\hat{\xi}^{(P)}$ and related estimators we refer to other papers in this volume and to Embrechts el al. [EKM] and the references therein. An illustration of these estimators is to be found in Figure 8.

The basic message of the above is that extreme value theory, through results like Theorems 4.1 and 4.2, yields a possible methodological basis on which to base quantile estimation.

In Section 2 on Value-at-Risk we discussed the importance of estimating the conditional dfs $\theta(\cdot|u_L)$ and $\theta(\cdot|u_P)$, also referred to as *shortfall distributions* in the finance literature. The key result needed to treat this issue is given in Theorem 4.3 known under the name of *Gnedenko-Pickands-Balkema-de Haan Theorem.* In order to state it in its most general form, we first define the right-endpoint x_F of a df F by

$$x_F = \sup\{x \in I\!R : F(x) < 1\}.$$

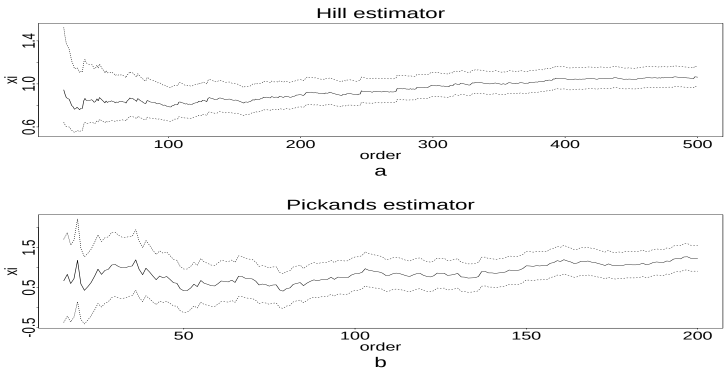

Figure 8: *The Hill and Pickands estimator for the fire data in Figure 6. Both plots suggest a value for ξ in the range (0.5,1.5) corresponding to an estimate for α in the range (0.6,2).*

Clearly $x_\Lambda = x_{\Phi_\alpha} = \infty$, $x_{\Psi_\alpha} = 0$. One can also show that for $F \in MDA(\Phi_\alpha)$, $x_F = \infty$, whereas for $F \in MDA(\Psi_\alpha)$, $x_F < \infty$. The former result is one of the reasons why in insurance and finance, the Fréchet distribution and its MDA provide the relevant models: they allow for unbounded upper values. Skeptics may eliminate the Fréchet distribution precisely because of the fact that $x_F = \infty$, on the basis of the somewhat tired argument that everything in the universe is bounded. Besides throwing away the normal distribution with the same argument, the main point however is that one can often clearly distinguish between data for which the larger values seem to cluster toward some well-defined upper limit, whereas in genuinely heavy-tailed data there does not seem to be such a clustering effect. Especially in applications in (re)insurance it may be much more prudent to model without an a priori fixed, finite x_F. Therefore all standard models used in non-life insurance have $x_F = \infty$. In what follows, we assume without loss of generality that $x_F \geq 0$. The following notion of *excess distribution* $F_u(x)$ is also important:

$$F_u(x) = P\left(X - u \leq x \mid X > u\right), \quad x \geq 0.$$

In the VaR example of Section 2,

$$\theta(x|u_P) = \bar{F}_{u_P}(x), \quad x \geq 0.$$

In the sequel we will use a new class of distributions. The *generalised Pareto distribution* with parameters $\xi \in I\!R$, $\beta > 0$ is defined as follows:

$$G_{\xi,\beta}(x) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-\frac{1}{\xi}}, & \xi \neq 0, \\ 1 - \exp\left\{-\frac{x}{\beta}\right\}, & \xi = 0. \end{cases}$$

where $x \geq 0$ for $\xi \geq 0$ and $0 \leq x \leq -\beta/\xi$ for $\xi < 0$. Figure 9 plots the densities of the GPD for some values of the parameter ξ.

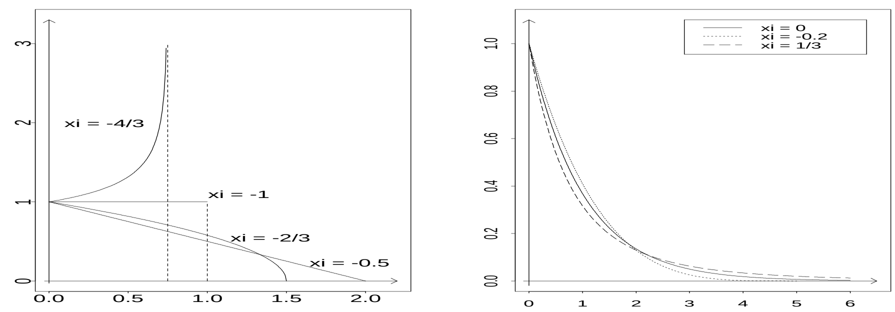

Figure 9: *Densities of some generalised Pareto distributions. The parameter ξ is given in the pictures, β is always 1.*

We are now ready to state the main result concerning the estimation of the overshoot probabilities $\theta(\cdot|\cdot)$ in Section 2.

Theorem 4.3 *Suppose F is a df with excess distributions F_u, $u \geq 0$. Then, for $\xi \in \mathbb{R}, F \in MDA(H_\xi)$ if and only if there exists a positive measurable function $\beta(u)$ so that*

$$\lim_{u \uparrow x_F} \sup_{0 \leq x \leq x_F - u} |\bar{F}_u(x) - \bar{G}_{\xi,\beta(u)}(x)| = 0. \tag{4.11}$$

□

Remark that in most cases interesting to us, $x_F = \infty$ so that (4.11) yields an approximation to $\bar{F}_u$ by a generalised Pareto distribution; *the so-called threshold u has to be taken large* ($u \uparrow \infty$). From this approximation we immediately obtain an upper-tail estimator. Indeed, for $x, u \geq 0$,

$$\bar{F}(u+x) = \bar{F}_u(x)\bar{F}(u) \approx \bar{G}_{\xi,\beta(u)}(x)\bar{F}(u).$$

Based on $X_1, \ldots, X_n$ iid with df F, we obtain as a natural estimator (writing $\beta = \beta(u)$):

$$\widehat{\bar{F}(u+x)} = \bar{G}_{\hat{\xi},\hat{\beta}}(x)\frac{N_u}{n}, \tag{4.12}$$

where $N_u = \#\{i : 1 \leq i \leq n,\ X_i > u\}$. The estimators $\hat{\xi}$ and $\hat{\beta}$ can for instance be obtained through maximum likelihood estimation.

For further details see Embrechts et al. [EKM] and references therein. We would like to stress that $\hat{\xi}$ and $\hat{\beta}$ are functions of u! From the estimate for $\bar{F}$ above one can immediately obtain the necessary quantile estimator. For instance, an estimator for the p-quantile $x_p = F^{\leftarrow}(p)$ becomes

$$\hat{x}_p = u + \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{n}{N_u}(1-p) \right)^{-\hat{\xi}} - 1 \right). \tag{4.13}$$

In practice, its construction involves the following steps:

1. The data are $X_1, \ldots, X_n$, and for given threshold u, denote N_u the number of exceedances of the level u.

2. Determine u large enough so that the approximation (4.11) is valid. One graphical tool is based on the mef, i.e. look for u-value above which the mef "looks linear". In the case of the fire insurance data (Figure 6(c)) relatively small u-values could be considered. A first choice of $u = 10$ may be good.

3. Use maximum likelihood theory to fit the GPD to the excesses over u (= 10, say). From this, a tail fit (4.12) and an estimator for the quantile x_p ((4.13)) are obtained. In Figure 10 two of such plots are shown.

4. Repeat step 3. for various u-values. This implies refitting a model for each new u. A summary plot of the type

$$\{(u, \hat{x}_p) : u \geq 0\}$$

gives a good idea on the (non-)stability of an estimation procedure as a function of the excesses used.

The statistical theory underlying the above methodology runs under the acronym POT: Peaks Over Threshold method. A nice introduction is given in Davison and Smith [DS]. The software used in the above analysis was written in S-plus by Alexander McNeil and is available via `http://www.math.ethz.ch/~mcneil/`.

5. Conclusion

Going back to the "Survival kit on quantile estimation" given in Section 2, we have touched upon all the items listed. Much more can, and indeed should be said. This volume, and similar ones will hopefully contribute to making extreme value methodology more widely known in fields like insurance and finance. Natural or man-made disasters, crashes on the stockmarket or other extremal events form part of society. The methodology given above may be useful in contributing towards a sound scientific discussion concerning their causes and effects.

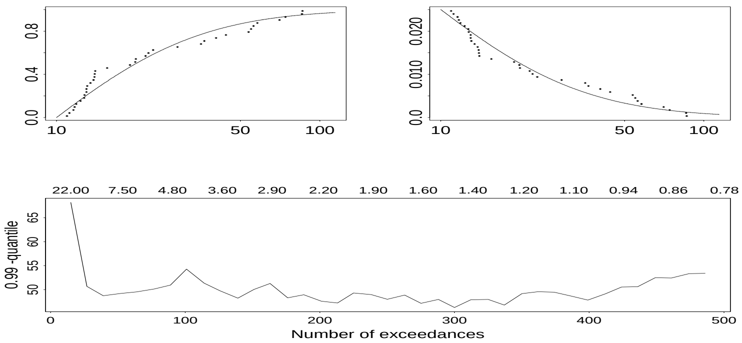

Figure 10: *Analysis of the fire data: the top left figure contains the fit for the excess distribution* $F_u(x-u)$ *where* $u = 10$*; the top right figure gives the fit for* $\bar{F}(u+x)$*, again for* $u = 10$*. The estimated shape parameter equals 0.43. Both figures have logarithmic horizontal scale. Finally, the bottom figure yields a quantile plot* $\{(u, \hat{x}_p) : u \geq 0\}$ *for* $p = 0.99$*. Note that in the bottom x-axes the number of exceedances is given, on the other hand on the top x-axes the corresponding threshold is reported.*

Acknowledgment

We would like to thank an anonymous referee for the very careful reading of a first version of the paper. His/her comments lead to numerous improvements.

References

[BGT] N. H. Bingham, C. M. Goldie, J. L. Teugels, Regular Variation. *Cambridge University Press*, Cambridge (1987).

[DFH] S. R. Dalal, E. B. Fowlkes and B. Hoadley, Risk analysis of the space shuttle: pre-Challenger prediction of failure. *Journal of the American Statistical Association* **408** (1989), 945-957.

[DS] A. Davison and R. L. Smith, Models for exeedances over high thresholds (with discussion). *J. Roy. Statist. Soc.* **Ser. B 52** (1990), 393-442.

[EKM] P. Embrechts, C. Klüppelberg, and T. Mikosch, Modelling Extremal Events for Insurance and Finance. *Springer, Berlin* (1997) (to appear).

[Fe] R. Feynman, What do you care what other people think? *Bantam books* (1989).

[FT] R. A. Fisher and L. H. C. Tippett, Limiting forms of the frequency distribution of the largest or smallest number of the sample. *Proc. Cambridge Philos. Soc.* **24**, (1928), 180-190.

[Ha] L. de Haan, Fighting the arch-enemy with mathematics. *Statistica Neerlandica* **44** (1990), 45-68.

[HK] R. V. Hogg and S. A. Klugman, Loss Distributions. *Wiley, New York* (1984).

[Ma] B. Mandelbrot, The variation of certain speculative prices. *Journal of Business, Univ. Chicago* **36**, (1963), 394-419.

[PC] *Report of the Presidential Commission on the Space Challenger Accident*, **Vols. 1&2**, Washington DC (1986).

[Tay] S. J. Taylor, Modelling Financial Time Series. *Wiley, Chichester* (1986).

Franco Bassi, Paul Embrechts[1], *Maria Kafetzaki*

Department of Mathematics, ETHZ, CH-8092 Zürich, Switzerland

[1]This paper is based on a talk given by the second author at Olsen and Associates, Zurich.