Extreme value theory as a risk management tool*

Paul Embrechts, Sidney I. Resnick and Gennady Samorodnitsky

Abstract

The financial industry, including banking and insurance, is undergoing major changes. The (re)insurance industry is increasingly exposed to catastrophic losses for which the requested cover is only just available. Due to an increasing complexity of financial instruments, sophisticated risk management tools have to be put into place. The securitization of risk and alternative risk transfer highlight the convergence of finance and insurance at the product level. Extreme value theory plays an important methodological role within the above.

1 Introduction

Consider the time series in Table 1 of loss–ratios (yearly data) for earthquake insurance in California from 1971 till 1993. The data are taken from Jaffe and Russell (1996).

1971	17.4	1979	2.2	1987	22.8
1972	0	1980	9.2	1988	11.5
1973	0.6	1981	0.9	1989	129.8
1974	3.4	1982	0	1990	47.0
1975	0	1983	2.9	1991	17.2
1976	0	1984	5.0	1992	12.8
1977	0.7	1985	1.3	1993	3.2
1978	1.5	1986	9.3		

Table 1: California earthquake data

*A first version of this paper was presented by the first author as an invited paper at the XXVIIIth International ASTIN Colloquium in Cairns and published in the conference proceedings under the title "Extremes and Insurance". Sidney Resnick and Gennady Samorodnitsky were partially supported by NSF Grant DMS-97-04982 at Cornell University. Paul Embrechts gratefully acknowledges the hospitality of the School of Operations Research and Industrial Engineering, Cornell University during Fall 1996.

On the basis of these data, who would have guessed the 1994 value of 2272.7? Indeed, on the 17th of January of that year the 6.6 Richter scale Northridge earthquake hit California causing an insured damage of USD 10.4 billion and a total damage of USD 30 billion, making 1994 the year with the third highest loss burden (natural catastrophes and major losses) in the history of insurance. The front-runners in this sad hit parade are 1992 (the year of hurricane Andrew) and 1990 (the year of the winter storms Daria and Vivian). For details on these, see Sigma (1995, 1997).

The reinsurance industry experienced a rise in both intensity as well as magnitude of losses due to natural and man–made catastrophes. For the United States alone, Canter, Cole and Sandor (1996) estimate an approximate USD 245 billion of capital in the insurance and reinsurance industry to service a country that has USD 25 to 30 trillion worth of property. No surprise that the finance industry has seized upon this by offering (often in joint ventures with the (re)insurance world) properly securitized products in the realm of catastrophe insurance. At an increasing pace, new products are being born. Some of them only have a short life, and others are reborn under a different shape. Some do not survive. Examples include:

— CAT futures and PCS options (CBOT): in these cases, securitization is achieved through the construction of derivatives written on a newly constructed industry wide loss–ratio index.

— Convertible CAT bonds. The Winterthur convertible hail–bond is an example. This European type convertible has an extra coupon payment contingent on the occurrence of a well–defined cat–event: an excessive number of cars in Winterthur's Swiss portfolio damaged within a hail storm over a specific time period. For details, see Schmock (1997).

Further interesting new products are the so–called multi–line, multi–year, high layer (infrequent event) products, credit lines, the catastrophe risk exchange (CATEX). For a brief review on some of these instruments, see Punter (1997). Excellent overviews stressing more the financial engineering of such products are Doherty (1997) and Tilley (1997). This whole area of alternative risk transfer and securitization has become a major area of applied research in both the banking and insurance industry. Actuaries are actively taking part in some of the new product development, and therefore have to consider the methodological issues underlying these and similar products.

Also recently, similar methods are being introduced in the world of finance through the estimation of Value–at–Risk (VaR) and the so–called shortfall; see Bassi, Embrechts and Kafetzaki (1997) and Embrechts, Samorodnitsky and Resnick (1998). Value At Risk for End–Users (1997) contains a recent summary of some of the more applied issues. More generally, extremes matter eminently within the world of finance. It is no coincidence that Alan Greenspan, Chairman of the Board of Governors of the FED, remarked at a research conference on risk measurement and systemic risk (Washington, D.C., 16 November 1996), that "Work that characterizes the statistical distribution of extreme events would be useful, as well."

For the general observer, extremes in the realm of finance manifest themselves most clearly through stock market crashes or industry losses. In Figure 1, we have plotted the events leading up to and including the 1987 crash for equity data (S&P). Extreme Value Theory (EVT) yields methods for quantifying such events and their consequences in a statistically optimal way. See for instance McNeil (1998) for an interesting discussion on the 1987 crash example. For a general equity book for instance, a risk manager will be interested in estimating the resulting down–side risk which typically can be reformulated in terms of a quantile for a Profit–and–Loss (P & L) function.

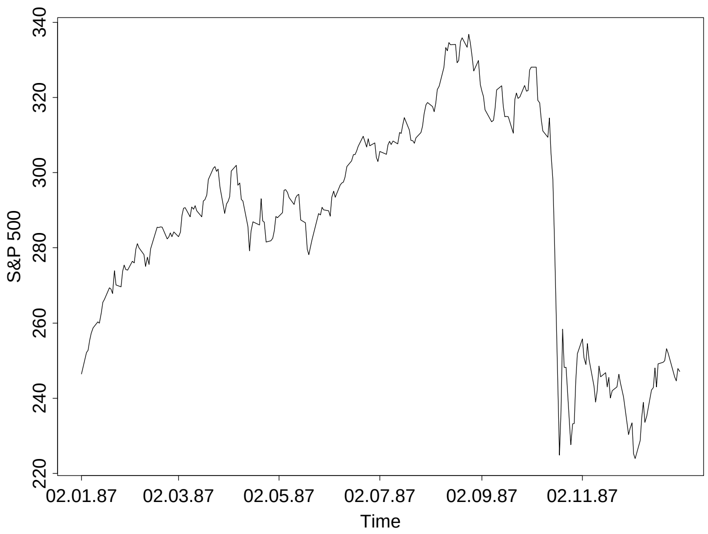

Figure 1: 1987 Crash

Another area of risk–management research where EVT is increasingly playing an important role is in credit risk management. The interested reader may browse J.P. Morgan's web site for information on Credit Metrics. It is no coincidence that big investment banks are looking at actuarial methods for the sizing of reserves to guard against future credit losses. Swiss Bank Corporation for instance introduced ACRA (Actuarial Credit Risk Accounting) for credit risk management; see Figure 2. In their risk measurement framework, they use the following definitions:

expected loss:	the losses which must be assumed to arise on a continuing basis as a consequence of undertaking particular business,
unexpected losses:	the unusual, though predictable, losses which the Bank should be able to absorb in the normal course of its business,
stress loss:	the possible – although improbable – extreme scenarios which the Bank must be able to survive.

EVT offers an important set of techniques for quantifying the boundaries between these different loss–classes. Moreover, EVT offers a scientific language for translating management guidelines on these boundaries into actual numbers. A final example where EVT offers help is in the area of modeling default probabilities and the estimation of so–called diversification factors in the management of bond portfolios. Many more examples could have been added.

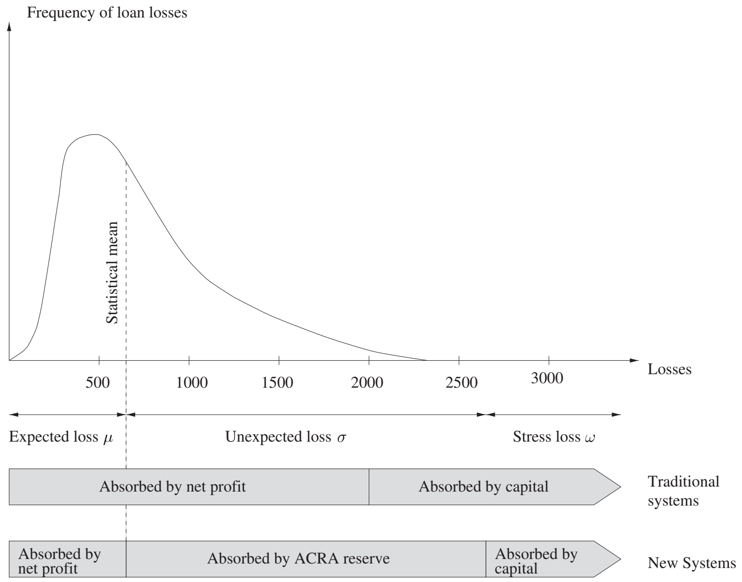

Figure 2: ACRA: Actuarial Credit Risk Accounting

It is our aim in this paper to review some of the basic tools from EVT relevant for industry wide risk management. Some examples towards the end of the paper should give the reader a better idea of the kind of answers EVT provides. Most of the material covered in the paper (and indeed much more) is to be found in Embrechts, Klüppelberg and Mikosch (1997). The latter book also contains an extensive list of further references. For reference of specific results in this book we will occasionally refer to it as EKM $(*)$ where $(*)$ refers to a specific result.

2 The basic theory

As already (implicitly) stated in the introduction, the statistical analysis of extremes is key to many of the risk management problems related to insurance, reinsurance and finance. In order to review some of the basic ideas underlying EVT, we discuss the most important results under the simplifying iid assumption: losses will be assumed to be independent and identically distributed. Most of the results can be extended to much more general models. In Example 4.2 a first indication of such a generalization will be given.

Throughout this paper, losses will always be denoted as positive; consequently we concentrate in our discussion below on one–sided distribution functions (dfs) for positive random variables (rvs).

Given basic loss data

$$X_1, X_2, \dots, X_n \quad \text{iid with df } F\,, \tag{1}$$

we are interested in the random variables

$$X_{n,n} = \min\left(X_1, \dots, X_n\right)\,, \quad X_{1,n} = \max\left(X_1, \dots, X_n\right)\,. \tag{2}$$

Or indeed using the full set of so–called order statistics

$$X_{n,n} \le X_{n-1,n} \le \cdots \le X_{1,n} \,, \tag{3}$$

we may be interested in

$$\sum_{r=1}^{k} h_r \left(X_{r,n} \right) \tag{4}$$

for certain functions h_r, $r = 1, \dots, k$ and $k = k(n)$. An important example corresponds to $h_r \equiv \frac{1}{k}$, $r = 1, \dots, k$, i.e. we average the k largest losses $X_{1,n}, \dots, X_{k,n}$. Another important example would be to take $k = n$, $h_r(x) = (x - u)_+$ where $y_+ = \max(0, y)$, for a given level $u > 0$. In this case we sum all excesses over u of losses larger than u. Typically we would normalize this sum by the number of such exceedances yielding the so–called empirical mean excess function; see Example 4.1. Most of the standard reinsurance treaties are of (or close to) the form (4). The last example given corresponds to an excess–of–loss (XL–) treaty with priority u.

In "classical" probability theory and statistics most of the results relevant for insurance and finance are based on sums

$$S_n = \sum_{r=1}^{n} X_r \,.$$

Indeed the laws of large numbers, the central limit theorem (in its various degrees of complexity), refinements like Berry–Esséen, Edgeworth, saddle–point and normal–power approximations all start from S_n–theory. Therefore, we find in our sum–toolkit such items like

— the normal distributions $N\left(\mu, \sigma^2\right)$,

— the α–stable distributions, $0 < \alpha < 2$,

— Brownian motion,

— α–stable processes, $0 < \alpha < 2$.

We are confident in our sum–toolkit when it comes to modeling/pricing/setting reserves of random phenomena based on averages. Likewise we are confident in statistical techniques based on these tools when applied to estimating distribution tails "not too far" from the mean. Consider however the following easy exercise.

Exercise: It is stated that within a given portfolio, claims follow an exponential df with mean 10 (thousand dollars, say). We have now observed 100 such claims with largest loss 50. Do we still believe in this model? What if the largest loss would have been 100?

Solution:

The basic assumption yields that

$$X_1, \dots, X_{100} \quad \text{are iid with df } P(X_1 \le x) = 1 - e^{-x/10}, \quad x \ge 0\,.$$

Therefore, for $M_n = \max(X_1, \dots, X_n)$,

$$\begin{aligned} P(M_{100} > x) &= 1 - (P(X_1 \le x))^{100} \\ &= 1 - \left(1 - e^{-x/10}\right)^{100}\,. \end{aligned}$$

From this, we immediately obtain

$$\begin{aligned} P(M_{100} \ge 50) &= 0.4914 \\ P(M_{100} \ge 100) &= 0.00453\,. \end{aligned}$$

However, rather than doing the (easy) exact calculations above, consider the following asymptotic argument. First, for all $n \ge 1$ and $x \in \mathbb{R}$,

$$\begin{aligned} P\left(\frac{M_n}{10} - \log n \le x\right) &= P(M_n \le 10(x + \log n)) \\ &= \left(1 - \frac{e^{-x}}{n}\right)^n, \end{aligned}$$

so that

$$\lim_{n\to\infty} P\left(\frac{M_n}{10} - \log n \le x\right) = e^{-e^{-x}} \equiv \Lambda(x)\,.$$

Therefore, use the approximation

$$P(M_n \le x) \approx \Lambda\left(\frac{x}{10} - \log n\right)$$

to obtain

$$\begin{aligned} P(M_{100} \ge 50) &\approx 0.4902\,, \\ P(M_{100} \ge 100) &\approx 0.00453\,, \end{aligned}$$

very much in agreement with the exact calculations above. □

Suppose we were asked the same question, but now we would have much less specific information on $F(x) = P(X_1 \le x)$, could we still proceed? This is exactly the point where classical EVT enters. In the above exercise, we have proved the following.

Proposition 1 *Suppose $X_1, \dots, X_n$ are iid with df $F \sim \mathrm{EXP}(\lambda)$, then for $x \in \mathbb{R}$:*

$$\lim_{n\to\infty} P(\lambda M_n - \log n \le x) = \Lambda(x)\,.$$ □

Here are the key questions:

Q1: What is special about Λ? Can we get other limits, possibly for other dfs F?

Q2: How do we find the norming constants λ and $\log n$ in general? That is, find a_n and b_n so that

$$\lim_{n\to\infty} P\left(\frac{M_n - b_n}{a_n} \le x\right)$$

exists.

Q3: Given a limit coming out of Q1, for which dfs F and norming constants from Q2, do we have convergence to that limit? Can one say something about second order behaviour, i.e. speed of convergence?

The solution to Q1 forms part of the famous Fisher–Tippett theorem.

Theorem 2 (EKM(Theorem 3.2.3)) *Suppose $X_1, \ldots, X_n$ are iid with df F and (a_n), (b_n) are constants so that for some non–degenerate limit distribution G,*

$$\lim_{n\to\infty} P\left(\frac{M_n - b_n}{a_n} \le x\right) = G(x)\,. \quad x \in \mathbb{R}\,,$$

Then G is of one of the following types:

— *Type I (Fréchet):*

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

— *Type II (Weibull):*

$$\Psi_\alpha(x) = \begin{cases} \exp\left\{-(-x)^{\alpha}\right\}\,, & x \le 0 \\ 1\,, & x > 0 \end{cases} \quad \alpha > 0\,,$$

— *Type III (Gumbel):*

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

G is of the type H means that for some $a > 0$, $b \in \mathbb{R}$, $G(x) = H((x-b)/a)$, $x \in \mathbb{R}$, and the distributions of one of the above three types are called *extreme value distributions.* Alternatively, any extreme value distribution can be represented as

$$H_{\xi;\mu,\sigma}(x) = \exp\left\{-\left(1 + \xi\frac{x-\mu}{\sigma}\right)_+^{-1/\xi}\right\}\,, \quad x \in \mathbb{R}\,.$$

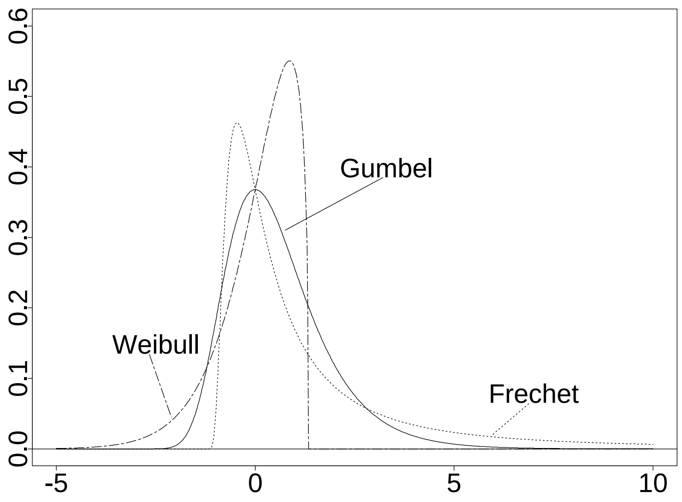

Figure 3: Some examples of extreme value distributions $H_{\xi;0,1}$ for $\xi = 3/4$ (Fréchet), $\xi = 0$ (Gumbel) and $\xi = -3/4$ (Weibull)

Here $\xi \in \mathbb{R}$, $\mu \in \mathbb{R}$ and $\sigma > 0$. The case $\xi > 0$ ($\xi < 0$) corresponds to the Fréchet (Weibull)–type df with $\xi = 1/\alpha$ ($\xi = -1/\alpha$), whereas by continuity $\xi = 0$ corresponds to the Gumbel or double exponential–type df.

In Figure 3, some examples of the extreme value distributions are given. Note that the Fréchet case (the Weibull case) corresponds to a model with finite lower (upper) bound; the Gumbel is two–sided unbounded.

Answering Q2 and Q3 is much more complicated. Below we formulate a complete answer (due to Gnedenko) for the Fréchet case. This case is the most important for applications to (re)insurance and finance. For a general df F, we define the inverse of F as follows:

$$F^{\leftarrow}(t) = \inf\{x \in \mathbb{R} : F(x) \geq t\}, \quad 0 < t < 1.$$

Using this notation, the p–quantile of F is defined as

$$x_p = F^{\leftarrow}(p), \quad 0 < p < 1.$$

Theorem 3 (EKM (Theorem 3.3.7)) *Suppose $X_1, \ldots, X_n$ are iid with df F satisfying*

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(x)} = x^{-\alpha}, \quad x > 0, \quad \alpha > 0. \tag{5}$$

Then for $x > 0$,

$$\lim_{n\to\infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = \Phi_\alpha(x)$$

where $b_n = 0$ and $a_n = F^{\leftarrow}\left(1 - \frac{1}{n}\right)$. The converse of this result also holds true. □

A df F satisfying (5) is called regularly varying with index $-\alpha$, denoted by $\overline{F} = 1 - F \in \mathcal{R}_{-\alpha}$. An important consequence of the condition $\overline{F} \in \mathcal{R}_{-\alpha}$ is that for a rv X with df F,

$$EX^\beta \begin{cases} < \infty & \text{for } \beta < \alpha\,, \\ = \infty & \text{for } \beta > \alpha\,. \end{cases} \tag{6}$$

In insurance applications, one often finds α–values in the range $(1,2)$ whereas in finance (equity daily log–returns say) an interval $(2,5)$ is common. Theorem 3 is also reformulated as: the maximal domain of attraction of Φ_α is $\mathcal{R}_{-\alpha}$, i.e.

$$\mathrm{MDA}\,(\Phi_\alpha) = \mathcal{R}_{-\alpha}\,.$$

Dfs belonging to $\mathcal{R}_{-\alpha}$ are for obvious reasons also called of Pareto type. Though we can calculate the norming constants, the calculation of a_n depends on the tail of F which in practice is unknown. The construction of MDA (Ψ_α) is also fairly easy, the main difference being that for $F \in \mathrm{MDA}\,(\Psi_\alpha)$,

$$x_F \equiv \sup\{x \in \mathbb{R} : F(x) < 1\} < \infty\,.$$

The analysis of MDA(Λ) is more involved. It contains such diverse dfs as the exponential, normal, lognormal, gamma. For details, see Embrechts, Klüppelberg and Mikosch (1997, Section 3.3.3).

3 Tail and quantile estimation

Theorem 3 is the basis of EVT. In order to show how this theory can be put into practice, consider for instance the pricing of an XL–treaty. Typically, the priority (or attachment point) u is determined as a t–year event corresponding to a specific claim event with claim size df F, say. This means that

$$u = u_t = F^{\leftarrow}\left(1 - \frac{1}{t}\right)\,. \tag{7}$$

In our notation used before, $u_t = x_{1-1/t}$. Whenever t is large (typically the case in the catastrophic (i.e. rare) event situation), the following result due to Balkema, de Haan, Gnedenko and Pickands (see Embrechts, Klüppelberg and Mikosch (1997, Theorem 3.4.13(b))) is very useful.

Theorem 4 *Suppose $X_1, \ldots, X_n$ are iid with df F. Equivalent are:*

i) $F \in \mathrm{MDA}\,(H_\xi)$, $\xi \in \mathbb{R}$,

ii) for some function $\beta : \mathbb{R}^+ \to \mathbb{R}^+$,

$$\lim_{u \uparrow x_F} \sup_{0 < x < x_F - u} \left| F_u(x) - G_{\xi,\beta(u)}(x) \right| = 0\,, \tag{8}$$

where $F_u(x) = P(X - u \le x \mid X > u)$, *and the* generalized Pareto *df is given by*

$$G_{\xi,\beta}(x) = 1 - \left(1 + \xi \frac{x}{\beta}\right)_+^{-1/\xi}, \tag{9}$$

for $\beta > 0$. □

It is exactly the so–called *excess* df F_u which both risk managers as well as reinsurers should be interested in. Theorem 4 states that for large u, F_u has a generalized Pareto df (9). Now in order to estimate the tail $\overline{F}(u+x)$ for a fixed large value of u and all $x \ge 0$ consider the trivial identity

$$\overline{F}(u+x) = \overline{F}(u)\,\overline{F}_u(x)\,, \quad u, x \ge 0\,. \tag{10}$$

In order to estimate $\overline{F}(u+x)$, one first estimates $\overline{F}(u)$ by the empirical estimator

$$\widehat{\left(\overline{F}(u)\right)} = \frac{N_u}{n}$$

where $N_u = \#\{1 \le i \le n : X_i > u\}$. In order to have a "good" estimator for $\overline{F}(u)$, we need u not too large: the level u has to be well within the data. Given such a u–value, we approximate $\overline{F}_u(x)$ via (8) by

$$\widehat{\left(\overline{F}_u(x)\right)} = \overline{G}_{\hat{\xi},\hat{\beta}(u)}(x)$$

for some estimators $\hat{\xi}$ and $\hat{\beta}(u)$ depending on u. For this to work well, we need u large (indeed in Theorem 4ii $u \uparrow x_F$, the latter being $+\infty$ in the Fréchet case). A "good" estimator is obtained via a trade–off between these two conflicting requirements on u.

The whole statistical theory developed in order to work out the above program runs under the name *Peaks over Thresholds Method* and is discussed in detail in Embrechts, Klüppelberg and Mikosch (1997, Section 6.5), McNeil and Saladin (1997) and references therein. Software (S–plus) implementation can be found on

http://www.math.ethz.ch/~mcneil/software.

This maximum likelihood based approach also allows for modeling of the excess intensity N_u, as well as the modeling of time (or other co–variable) dependence in the relevant model parameters. As such, a highly versatile modeling methodology for extremal events is available. Related approaches with application to insurance are to be found in Beirlant, Teugels and Vynckier (1996), Reiss and Thomas (1997) and the references therein. Interesting case studies using up–to–date EVT methodology are McNeil (1997), Resnick (1997) and Rootzén and Tajvidi (1997). The various steps needed to perform a quantile estimation within the above EVT context are nicely reviewed in McNeil and Saladin (1997), where also a simulation study is to be found. In the next section, we illustrate the methodology on real and simulated data relevant for insurance and finance.

4 Examples

4.1 Industrial fire insurance data

In order to highlight the methodology briefly discussed in the previous sections, we first apply it to 8043 industrial fire insurance claims. We show how a tail–fit and the resulting quantile estimates can be obtained. Clearly, a full analysis (as for instance to be found in Rootzén and Tajvidi (1997) for windstorm data) would require a lot more work. We have grouped the figures towards the end of the example.

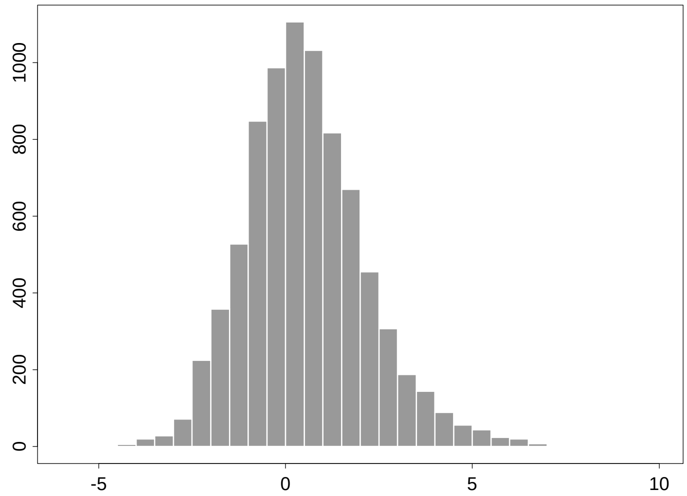

Figure 4: Log–histogram of the fire insurance data

Figure 4 contains the log–histogram of the data. The right–skewness stresses the long–tailed behaviour of the underlying data. A useful plot in order to specify the long–tailed nature of data is the so–called mean–excess plot as given in Figure 5. In it, the mean–excess function $e(u) = E(X - u \mid X > u)$ is estimated by its empirical counterpart

$$e_n(u) = \frac{1}{\#\{1 \le i \le n : X_i > u\}} \sum_{i=1}^{n} (X_i - u)^+ \ .$$

The Pareto df can be characterized by linearity (positive slope) of $e(u)$. In general, long–tailed dfs exhibit an upwards sloping behaviour, exponential–type dfs have roughly a constant mean excess plot, whereas short–tailed data yield a plot decreasing to 0. In our case, the upward trend clearly stresses the long–tailed behaviour. The increase in variability towards the upper end of the plot is characteristic of the technique, since towards the largest observation $X_{1,n}$, only few data points go into the calculation of $e_n(u)$. The main aim of our EVT–analysis is to find a fit of the underlying df $F(x)$ (or of its tail $\overline{F}(x)$) by a generalized Pareto df, especially for the larger values of x. The empirical df $\overline{F}_n$ is given in Figure 6 on a doubly logarithmic scale. This scale is used to highlight the tail region. Here an exact Pareto df corresponds to a linear plot.

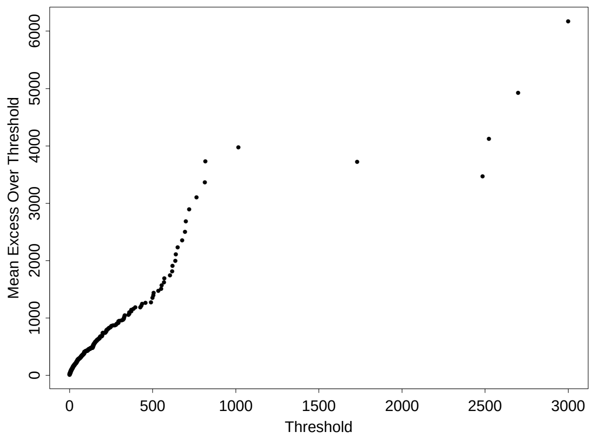

Figure 5: Mean–excess plot of the fire insurance data

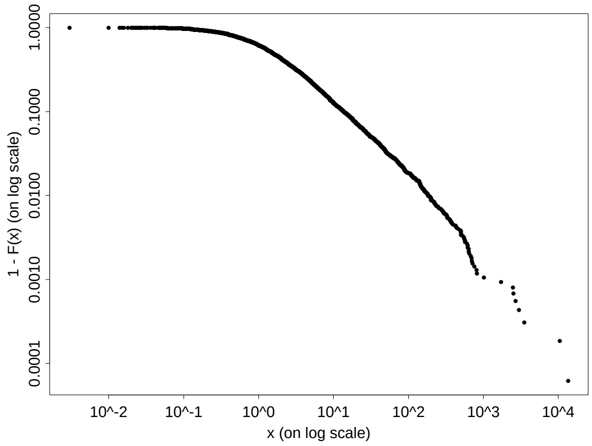

Figure 6: Empirical estimator of $\overline{F}$ on doubly logarithmic scale

Using the theory presented in Theorems 2 and 4, a maximum likelihood based approach yields estimates for the parameters of the extreme value df $H_{\xi;\mu,\sigma}$ and the generalized Pareto df $G_{\xi;\beta}$. In order to start this procedure, a threshold value u has to be chosen as estimates depend on the excesses over this threshold. The estimates of the key shape parameter ξ as a function of u (alternatively, as a function of the number of order statistics used) is given in Figure 7. Approximate 95% confidence intervals are given. The picture shows a rather stable behaviour for values of u below 300 say. An estimate in the range $(0.7, 0.9)$ results, which corresponds to an α–value in the range $(1.1, 1.4)$. It should be remarked that the "optimal" value of the threshold u to be used is difficult (if not impossible) to obtain. See Embrechts, Klüppelberg and Mikosch (1997, p. 351) and Beirlant, Teugels and Vynckier (1996) for some discussion. We also would like to stress that in order to produce Figure 7, a multitude of models (one for each u chosen) has to be estimated.

For each given u, a tail–fit for $\overline{F}_u$ and $\overline{F}$ (as in (10)) can be obtained. For the former, in the case of $u = 100$ an estimate $\hat{\xi} = 0.747$ results. A graphical represen-

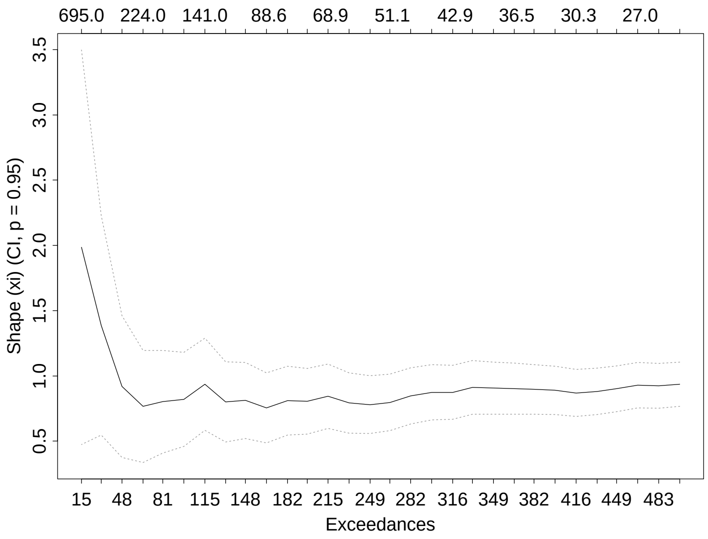

Figure 7: Maximum likelihood estimate of ξ as a function of the threshold u (top), alternatively, as a function of the number of exceedances

tation of $\hat{F}_{100}$ is given in Figure 8. Using the parameter estimates corresponding to $u = 100$ in (10) the tail–fit of $\overline{F}$ on doubly logarithmic scale is given in Figure 9.

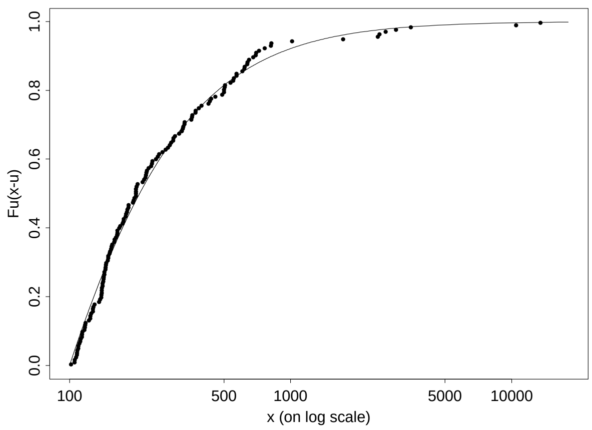

Figure 8: Maximum likelihood fit of the mean excess tail df $\overline{F}_u$ based on exceedances above $u = 100$

Though we have extended the generalized Pareto fit to the left of $u = 100$, clearly only the range above this u–value is relevant. The fitting method is only designed for the tail. Below u (where typically data are abundant) one could use a smooth version of the empirical df. From the latter plot, quantile estimates can be deduced.

Figure 10 contains as an example the estimate for the 99.9% quantile $x_{0.999}$ together with the profile likelihood. The latter can be used to find confidence intervals for $x_{0.999}$. The 95% and 99% intervals are given.

Figure 11 contains the same picture but the (symmetric) confidence intervals are

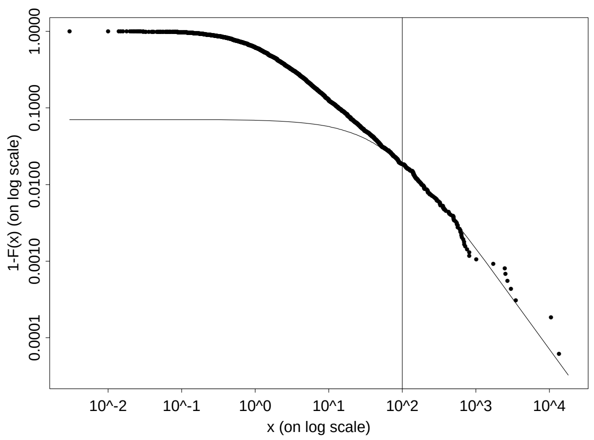

Figure 9: Tail–fit for $\overline{F}$ based on a threshold value of $u = 100$, doubly logarithmic scale

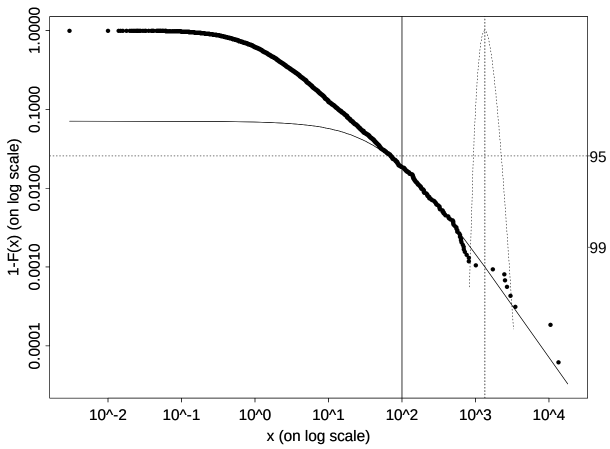

Figure 10: Tail–fit with an estimate for $x_{0.999}$ and the corresponding profile likelihood

calculated using the Wald statistic. Finally, the 99.9% quantile estimates across a whole range of models (depending on the threshold value, or number of exceedances used) are given in Figure 12. Though the estimate of $x_{0.999}$ settles between 1400 and 1500, the 95% Wald intervals are rather wide, ranging from 500 to about 2200.

The above analysis yields a summary about the high quantiles of the fire insurance data based on the information on extremes available in the data. The analysis can be used as a tool in the final pricing of risks corresponding to high layers (i.e. catastrophic, rare events). All the methods used are based on extremes and are fairly standard.

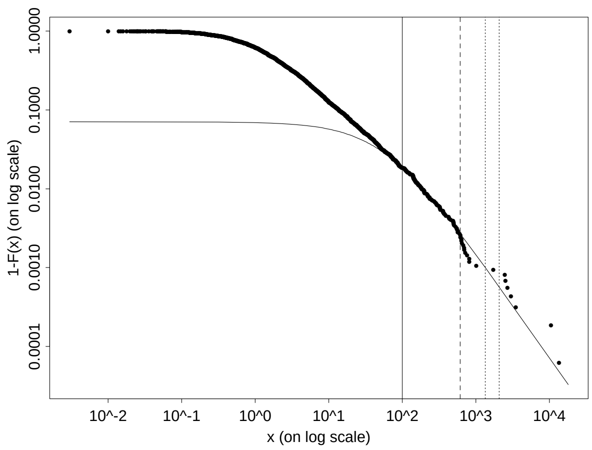

Figure 11: Estimate of $x_{0.999}$ with 95% Wald–statistic confidence interval

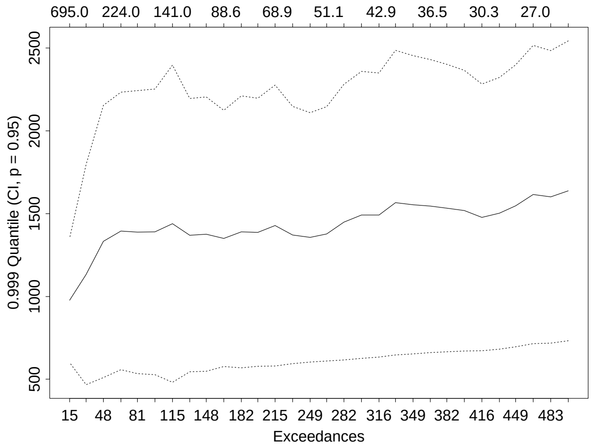

Figure 12: Estimates of the quantile $x_{0.999}$ as a function of the threshold u. The vertical line indicates the model corresponding to $u = 100$

4.2 An ARCH–example

To illustrate further some of the available techniques, we simulated an ARCH(1) time series of length 99,000. The time series, called *testarch*, has the form

$$\xi_n = X_n \left(\beta + \lambda \xi_{n-1}^2\right)^{1/2}, \quad n \geq 1, \tag{11}$$

where $\{X_n\}$ are iid $N(0,1)$ random variables. In our simulation, we took

$$\beta = 1, \quad \lambda = 0.5.$$

From known results of Kesten (1973) (see also Embrechts, Klüppelberg and Mikosch (1997, Theorem 8.4.12), Goldie (1991), Vervaat (1979))

$$P(\xi_1 > x) \sim c x^{-2\kappa}, \quad x \to \infty, \tag{12}$$

and we get from Table 3.2 of de Haan et al. (1989) that

$$\kappa = 2.365$$

(see also Hooghiemstra and Meester (1995)).

There are several reasons why we choose to simulate an ARCH process.

— Despite the fact that the ARCH process is dependent, much of the classical extreme value analysis applies with suitable modifications.

— The ARCH process has heavy tails which matches what is observed in data sets emerging from finance.

— Although it is often plausible to model large insurance claims as iid, data from the finance industry such as exchange rate data are demonstrably not iid. Some of these examples have the property that the data look remarkably uncorrelated but squares or absolute values of the data appear to have high correlations. It is this property that the ARCH process and its cousins were designed to model. See for instance Taylor (1986) for more details.

To experiment with this ARCH data we took the first 10,000 observations to form a data set *shortarch* which will be used for estimation. Then based on the estimation, some model based predictions can be made and compared with actual data in $testarch \backslash shortarch$.

Figure 13 shows a time series plot of *shortarch*. The plot exhibits the characteristic heavy tail appearance.

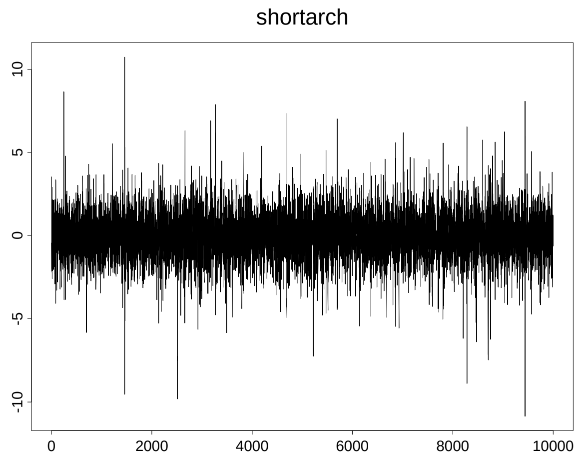

Figure 13: Time series plot of shortarch

The *Hill estimator* is a popular way of detecting heavy tails and estimating the Pareto index $1/\xi$ for $\xi > 0$ in (9). See for instance Embrechts, Klüppelberg and Mikosch (1997) for an introduction. Figure 14 contains four views of the Hill estimator applied to *shortarch*.

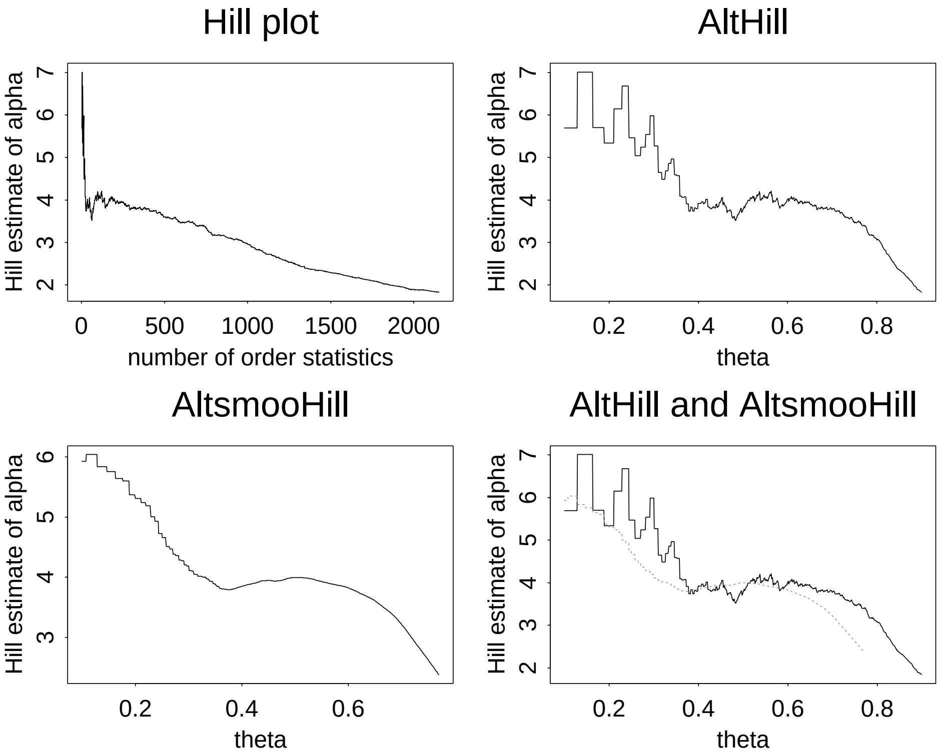

Figure 14: Hill plots of shortarch

The Hill estimator is known to be a consistent estimator of 2κ for the ARCH process (Resnick and Stărică (1996)). In our case, the Hill estimator is trying to estimate $2\kappa \approx 2\times 2.365 \approx 4.7$. A review of Figure 14 yields an estimate of about 4. If $H_{k,n}$ represents the Hill estimator when the sample is n and k upper order statistics are used, i.e.

$$H_{k,n} = \left(\frac{1}{k} \sum_{j=1}^{k} \log X_{j,n} - \log X_{k,n} \right)^{-1},$$

the usual methodology is to make a *Hill plot* $\left\{ \left(k, H_{k,n}^{-1}\right), 1 \le k \le n \right\}$. The upper left graph is a Hill plot with some values for small and large k deleted to make the picture scale attractively. The upper right plot is the Hill plot in *alt scale* (see Resnick and Stărică, 1997) where we plot $\left\{ \left(\theta, H^{-1}_{\left[n^\theta\right],n} \right), 0 \le \theta \le 1 \right\}$. The lower left plot applies a smoother (Resnick and Stărică, 1997) and plots in *alt scale*.

A supplementary tool for estimating the Pareto index is the QQ plot (see Embrechts, Klüppelberg and Mikosch (1997, Section 6.2.1)) and this has the added advantage that it allows simultaneous estimation of the constant c appearing in (12). The method is sensitive to the choice of the number of upper order statistics and some trial and error is usually necessary. In Figure 15 we give the QQ plot based on the upper 400 order statistics. This gives estimates of $2\kappa = 3.851904$ and $c = 1.15389$. (Applying this technique to the full *testarch* data produced estimates of $2\kappa = 3.861008$ and $c = 1.319316$, when the number of upper order statistics was

300.)

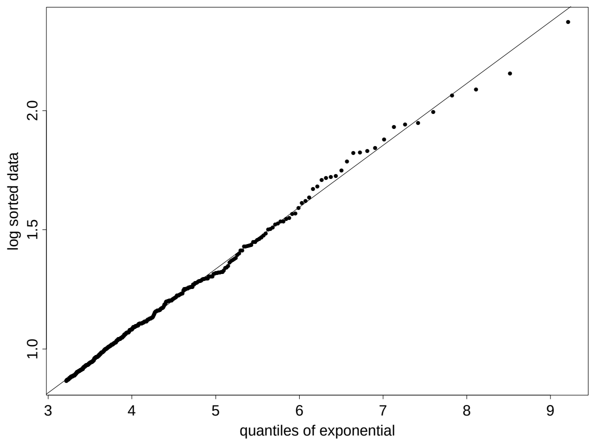

Figure 15: QQ plot of shortarch

Based on these estimates, we experiment with some predictions and compare these predictions with what is observed from the part of the data set *testarch*, called *playarch* obtained by removing the 10,000 *shortarch* observations. Thus the length of *playarch* is $99,000 - 10,000 = 89,000$. In the following table we give estimated marginal probabilities that the ARCH variable exceeds x for $x = 5, 10, 15, 20$. Note that we are predicting values that are beyond the range of the data and have not been observed. The second row gives the estimate (12) based on the fitted values for c and 2κ. In the third row we compute the empirical frequency that elements of *playarch* exceed x. The last row gives the corresponding probabilities $1 - \Phi(x, \mu, \sigma^2)$ based on a normal distribution whose mean and variance are the sample mean and variance computed from *shortarch*. One can see from the table the penalty paid for ignoring extreme value analysis and relying on more conventional normal distribution based analysis.

x	5	10	15	20
$P(X > x)$	0.002309062	0.0001589119	0.0000332	0.0000109
$\widehat{P}(X > x)$	0.002640449	0.0001797753	0.0000449	0.0000112
$1 - \Phi(x, \mu, \sigma^2)$	0.000186	5.29×10^{-13}	0	0

The extreme value theory for the ARCH process is somewhat complicated by the ARCH dependence structure not present for an iid sequence. A quantity called the *extremal index* must be accounted for; see Embrechts, Klüppelberg and Mikosch (1997, Section 8.1). From (11) and de Haan et al. (1989), Table 3.2 we have

$$P(\max\{\xi_i, \dots, \xi_n\} \le y) \approx \exp\left\{-\frac{1}{2}\, c\theta' n y^{-2\kappa}\right\}, \tag{13}$$

where the extremal index $\theta' = 0.835$ accounts for the effective reduction in sample size due to dependence. From this formula, estimates of upper quantiles can be worked out. The upper 100p% quantile x_p would be

$$x_p \approx \left(\left(- \log(1-p) \frac{2}{c\theta' n} \right) \right)^{-1/2\kappa} . \tag{14}$$

We give a few representative values:

y	15	20	25	30
$P\left(\max\left\{\xi_i, \ldots, \xi_n\right\} \leq y\right)$	0.28229	0.65862	0.83794	0.91613

p	0.05	0.01	0.005	0.0005
x_p	34.47153	52.63015	63.04769	114.692

4.3 Value–at–Risk: a word of warning

We already pointed out the similarity in estimating attachment points or retentions in reinsurance and VaR calculations in finance. Both are statistically based methods, where the basic underlying risk measure corresponds to a quantile estimate $\widehat{x}_p$ of an unknown df. Through the work of Artzner et al. (1996, 1998) we know that a quantile based risk measure for general (non–normal) data fails to be coherent, i.e. such a measure is not subadditive creating inconsistencies in the construction of risk capital based upon them. This situation typically occurs in portfolios containing non–linear derivatives. Further critical statements concerning VaR are to be found in Danielsson, Hartmann and de Vries (1992), Garman (1997), Longin (1997a,b) and Cárdenas et al. (1997).

A much better, and indeed (almost) coherent risk measure is the conditional VaR (or so–called mean excess)

$$E\left(X \mid X > \widehat{x}_p\right) . \tag{15}$$

For the precise formulation of the "almost" bove, see Artzner et al. (1996). We only want to point out that the latter measure is well–known in insurance, but only gradually is being recognized as fundamental to risk management. It is one of the (many) examples where an exchange of ideas between actuaries and finance experts may lead to improved risk measurement. Note that in the equation above, $E\left(X \mid X > \widehat{x}_p\right) = e\left(\widehat{x}_p\right) + \widehat{x}_p$. One could use the mean excess plot $\left\{\left(u, e_n(u)\right), u \geq 0\right\}$ in order to visualize the tail behavior of the underlying data and hence get some insight on the coherent risk measure (15) above.

As a final example we have plotted in Figure 16 the daily log–returns of BMW over the period January 23, 1973—July 12, 1996, together with the mean excess plot of the absolute values of the negative returns (hence corresponding to downside risk). In the latter plot, the heavy–tailed nature of the returns is very clear. Also clear is the change in curvature of the plot around 0.03. This phenomenon is regularly observed in all kinds of data. One can look at it as a pictorial view of Theorem 4: indeed, Smith (1990) indicates how to base on this observation a graphical tool to determine an initial threshold value for an extreme value analysis. See also

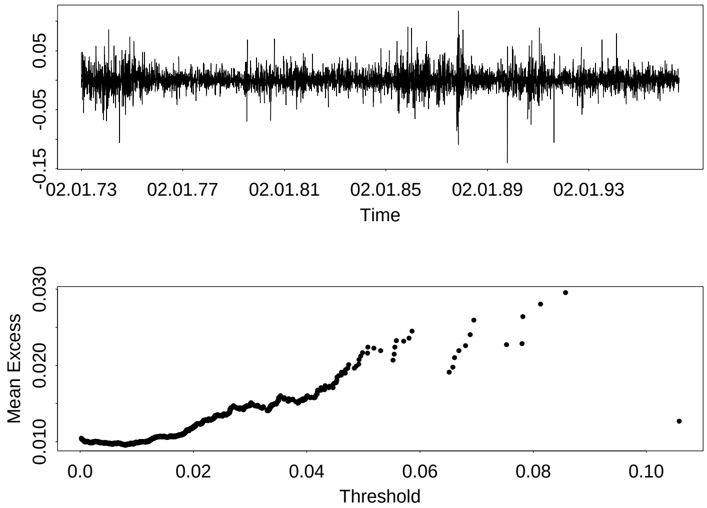

Figure 16: Time series and mean excess plots of BMW return data

Embrechts, Klüppelberg and Mikosch (1997, p. 356) for a discussion. We would like to warn the reader however that due to the intricate dependencies in finance data, one should be careful in using these plots beyond the mere descriptive level.

More work is needed to combine the ideas presented in this paper with detailed statistical information on financial time series before risk measures such as conditional VaR (15) can be formulated precisely and estimated reliably. Once more, the interplay between statisticians, finance experts and actuaries should prove to be fruitful towards achieving this goal.

References

Artzner, P., Delbaen, F., Eber, J.-M. and D. Heath (1996) A characterization of measures of risk. Université de Strasbourg preprint.

Artzner, P., Delbaen, F., Eber, J.-M. and D. Heath (1998) Thinking coherently. *RISK*, **10**, 68–71.

Bassi, F., Embrechts, P. and M. Kafetzaki (1997) Risk Management and quantile estimation. In: Adler, R., Feldman, R. and Taqqu, M.S. (Eds.) *A Practical Guide to Heavy Tails: Statistical Techniques for Analysing Heavy–Tailed Distributions.* Birkhäuser, Boston.

Beirlant, J., Teugels, J.L.T. and P. Vynckier (1996) Practical analysis of extreme values. Leuven University Press, Leuven.

Canter, M.S., Cole, J.B. and R.L. Sandor (1996) Insurance derivatives: a new asset class for the capital markets and a new hedging tool for the insurance industry. *Journal of Derivatives*, 89–104.

Cárdenas, J., Fruchard, E., Koehler, E., Michel, C. and I. Thomazeau (1997) VaR: One step beyond. *RISK* **10**, 72–74.

Danielsson, J., Hartmann, P. and C.G. de Vries (1998) The cost of conservatism. *RISK* **11**, 101–103.

Doherty, N.A. (1997) Financial innovation in the management of catastrophe risk. Joint Day Proceedings, XXVIIIth International ASTIN Colloquium and 7th International AFIR Colloquium, Cairns (Australia), 1–26.

Embrechts, P., Klüppelberg, C. and T. Mikosch (1997) *Modelling extremal events for insurance and finance.* Springer–Verlag, Berlin.

Embrechts, P., Samorodnitsky, G. and S.I. Resnick (1998) Living on the edge. *RISK* **11**, 96–100.

Garman, M. (1997) Taking VaR to pieces. *RISK* **10**.

Goldie, C.M. (1991) Implicit renewal theory and tails of solutions of random equations. *Annals of Applied Probability* **1**, 126–166.

Haan, L. de, Resnick, S.I., Rootzén, H. and C.G. de Vries (1989) Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes. *Stoch. Proc. Appl.* **32**, 213–224.

Hooghiemstra, G. and L. Meester (1995) Computing the extremal index of special Markov chains and queues. *Stochastic Process. Appl.* **65**, 171–185.

Jaffe, D.M. and T. Russell (1996) Catastrophe insurance, capital markets and uninsurable risk. Financial Institutions Center, The Wharton School, 96–12.

Kesten, H. (1973) Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131**, 207–248.

Longin, F.M. (1997a) From value at risk to stress testing: the extreme value approach. ESSEC and CEPR Paris preprint.

Longin, F.M. (1997b) Beyond the VaR. ESSEC and CEPR Paris preprint.

McNeil, A.J. (1997) Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin* **27**(2), 117–137.

McNeil, A.J. (1998) On extremes and crashes. *RISK* **11**, 99.

McNeil, A.J. and T. Saladin (1997) The peaks over threshold method for estimating high quantiles of loss distributions. Proceedings of the XXVIIIth International ASTIN Colloquium, Cairns, 23–43.

Punter, A. (1997) A critical evaluation of progress to date. In: *Practical applications of financial market tools to corporate risk management.* International Risk Management, p. 7–12.

Reiss, R.–D. and M. Thomas (1997) *Statistical Analysis of Extremal Values*, Birkhäuser, Basel.

Resnick, S.I. (1997) Discussion of the Danish data on large fire insurance losses. *ASTIN Bulletin* **27**(2), 139–151.

Resnick, S.I. and C. Stărică (1996) Tail index estimation for dependent data. Available as TR1174.ps.Z at http://www.orie.cornell.edu/trlist/trlist.html. To appear in *Ann. Applied Probability.*

Resnick, S.I. and C. Stărică (1997) Smoothing the Hill estimator. *Adv. Applied Probability* **29**, 271–293.

Rootzén, H. and N. Tajvidi (1997) Extreme value statistics and windstorm losses: a case study. *Scand. Actuarial J.* **1**, 70–94.

Schmock, U. (1997) Estimating the value of the Wincat coupons of the Winterthur Insurance convertible bond. Joint Day Proceedings, XXVIIIth International ASTIN Colloquium and 7th International AFIR Colloquium, Cairns (Australia), 231–259.

Sigma (1995) Natural catastrophes and major losses in 1994: Third highest loss burden in the history of insurance. Swiss Reinsurance Company, 3/95.

Sigma (1997) Natural catastrophes and major losses in 1996: high losses from man–made disasters, but not extremely costly losses from natural catastrophes. Swiss Reinsurance Company, 3/97.

Smith, R.L. (1990) Extreme value theory. In: Ledermann, W. (Chief Ed.) *Handbook of Applicable Mathematics, Supplement*, pp. 437–472. Wiley, Chichester.

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

Tilley, J.A. (1997) The securitization of catastrophic property risks. Joint Day Proceedings, XXVIIIth International ASTIN Colloquium and 7th International AFIR Colloquium, Cairns (Australia), 27–53.

Value at Risk for End–Users (1997) *Risk Magazine*, Supplement, March 1997.

Paul Embrechts
Department of Mathematics
ETHZ
CH – 8092 Zurich, Switzerland
e-mail: embrechts@math.ethz.ch

Sidney I. Resnick and Gennady Samorodnitsky
School of Operations Research
and Industrial Engineering
Cornell University
Rhodes Hall/ETC Building
Ithaca, New York 14853, USA
e-mail: sid@orie.cornell.edu
e-mail: gennady@orie.cornell.edu