# Estimating the Tails of Loss Severity Distributions using Extreme Value Theory

Alexander J. McNeil
Departement Mathematik
ETH Zentrum
CH-8092 Zürich

December 9, 1996

**Abstract**

Good estimates for the tails of loss severity distributions are essential for pricing or positioning high-excess loss layers in reinsurance. We describe parametric curve-fitting methods for modelling extreme historical losses. These methods revolve around the generalized Pareto distribution and are supported by extreme value theory. We summarize relevant theoretical results and provide an extensive example of their application to Danish data on large fire insurance losses.

**Keywords:** Loss severity distributions; high excess layers; extreme value theory; excesses over high thresholds; generalized Pareto distribution.

# 1 Introduction

Estimating loss severity distributions from historical data is an important actuarial activity in insurance. A standard reference on the subject is Hogg & Klugman (1984).

In this paper we are specifically interested in estimating the tails of loss severity distributions. This is of particular relevance in re-insurance if we are required to choose or price a high-excess layer. In this situation it is essential to find a good statistical model for the largest observed historical losses; it is less important that the model explains smaller losses. In fact, as we shall see, a model chosen for its overall fit to all historical losses may not provide a particularly good fit to the large losses. Such a model may not be suitable for pricing a high-excess layer.

Our modelling is based on extreme value theory (EVT), a theory which until comparitively recently has found more application in hydrology and climatology (de Haan 1990, Smith 1989) than in insurance. As its name suggests, this theory is concerned with the modelling of extreme events and in the last few years various authors (Beirlant & Teugels 1992, Embrechts & Klüppelberg 1993) have noted that the theory is as relevant to the modelling of extreme insurance losses as it is to the modelling of high river levels or temperatures.

For our purposes, the key result in EVT is the Pickands-Balkema-de Haan theorem (Balkema & de Haan 1974, Pickands 1975) which essentially says that, for a wide class of distributions, losses which exceed high enough thresholds follow the generalized Pareto distribution (GPD). In this paper we are concerned with fitting the GPD to data on exceedances of high thresholds. This modelling approach was developed in Davison (1984), Davison & Smith (1990) and other papers by these authors.

To illustrate the methods, we analyse Danish data on major fire insurance losses. We provide an extended worked example where we try to point out the pitfalls and limitations of the methods as well as their considerable strengths.

# 2 Modelling Loss Severities

## 2.1 The Context

Suppose insurance losses are denoted by the random variables $X_1, X_2, \ldots$. We assume that we are dealing with losses of the same general type and that these loss amounts are adjusted (for inflation and social inflation) so as to be comparable.We also assume that they are independent of one another.

That is, we consider $X_1, X_2, \ldots$ to be independent, identically distributed random variables. We denote their common distribution function by $F_X(x) = P\{X \le x\}$ where $x > 0$.

Now, suppose we are interested in a high-excess loss layer with lower and upper attachment points $r$ and $R$ respectively, where $r$ is large and $R > r$. This means that the payout $Y_i$ on a loss $X_i$ is given by

$$Y_i = \begin{cases} 0 & \text{if } 0 < X_i < r, \\ X_i - r & \text{if } r \le X_i < R, \\ R - r & \text{if } R \le X_i < \infty. \end{cases}$$

The process of losses becoming payouts is sketched in Figure 1. Of six losses, two pierce the layer and generate a non-zero payout. One of these losses overshoots the layer entirely and generates a capped payout.

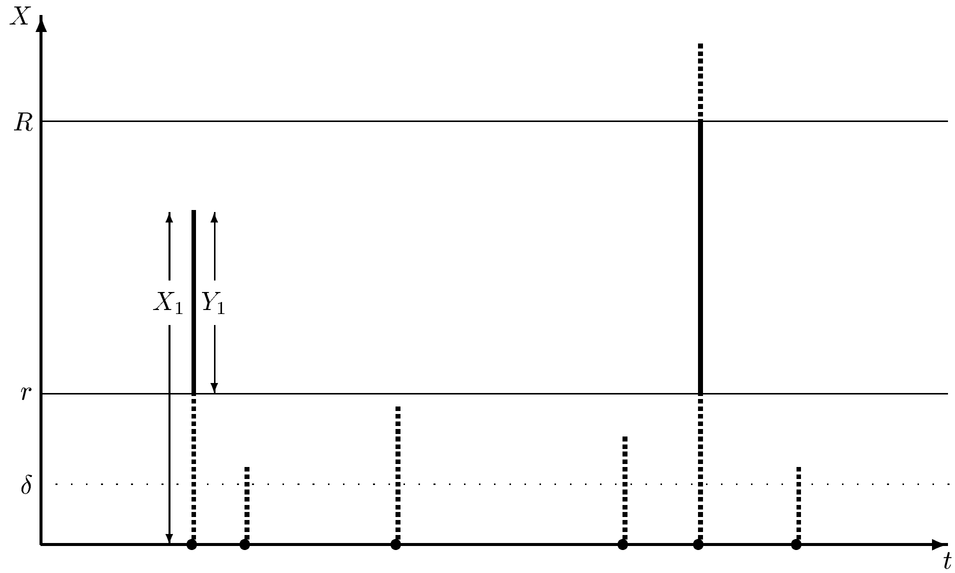


Figure 1: Possible realizations of losses in future time period

Two related actuarial problems concerning this layer are:

1. The pricing problem. Given $r$ and $R$ what should this insurance layer cost a customer?
2. The optimal attachment point problem. If we want payouts greater than a specified amount to occur with at most a specified frequency, how should we choose $r$?

To answer these questions we need to fix a period of insurance and know something about the frequency of losses incurred by a customer in such a time period. Denote the unknown number of losses in a period of insurance by $N$ so that the losses are $X_1, \ldots, X_N$. Thus the aggregate payout would be $Z = \sum_{i=1}^{N} Y_i$.

The pricing problem usually reduces to finding moments of $Z$. A common pricing formula is $Price = E[Z] + k.var[Z]$, where the price is the expected payout plus a risk loading which is $k$ times the variance of the payout, for some $k$. The expected payout $E[Z]$ is also known as the pure premium and it can be shown to be $E[Y_i]E[N]$. It is clear that if we wish to price the cover provided by the layer $(r, R)$ we must be able to calculate, amongst other things, $E[Y_i]$, the pure premium for a single loss. We will calculate $E[Y_i]$ as a price indication in later analyses in this paper.

One possible way of formulating the attachment point problem might be: choose $r$ such that $P\{Z > 0\} < p$ for some stipulated $p$. That is to say, $r$ is determined so that in the period of insurance a non-zero aggregate payout occurs with probability less than $p$ where, for a high-excess layer, $p$ will be set to very small. The attachment point problem essentially boils down to the estimation of a high quantile of the loss severity distribution $F_X(x)$.

In both of these problems we need a good estimate of the loss severity distribution for $x$ large, that is to say, in the tail area. We must also have a good estimate of the loss frequency distribution of $N$, but this will not be a topic of this paper.

## 2.2 Data Analysis

Typically we will have historical data on losses which exceed a certain amount known as a displacement. It is practically impossible to collect data on all losses and data on small losses are of less importance anyway. Insurance is generally provided against significant losses and insured parties deal with small losses themselves and may not report them.

Thus the data should be thought of as being realizations of random variables truncated at a displacement $\delta$, where $\delta \ll r$. This displacement is shown in Figure 1; we only observe realizations of the losses which exceed $\delta$.

The distribution function (d.f.) of the truncated losses can be defined as in Hogg & Klugman (1984) by

$$F_{X^\delta}(x) = P\{X \leq x \mid X > \delta\} = \begin{cases} 0 & \text{if } x \leq \delta, \\ \frac{F_X(x) - F_X(\delta)}{1 - F_X(\delta)} & \text{if } x > \delta, \end{cases}$$

and it is, in fact, this d.f. that we shall attempt to estimate.

With adjusted historical loss data, which we assume to be realizations of independent, identically distributed, truncated random variables, we attempt to find an estimate $\widehat{F_{X^\delta}(x)}$ of the truncated severity distribution $F_{X^\delta}(x)$. One way of doing this is by fitting parametric models to data and obtaining parameter estimates which optimize some fitting criterion - such as maximum likelihood. But problems arise when we have data as in Figure 2 and we are interested in a very high-excess layer.

Figure 2 shows the empirical distribution function of the Danish fire loss data evaluated at each of the data points. The empirical d.f. for a sample of size $n$ is defined by $F_n(x) = n^{-1} \sum_{i=1}^{n} 1_{\{X_i \leq x\}}$; i.e. the number of observations less than or equal to $x$ divided by $n$. The empirical d.f. forms an approximation to the true d.f. which may be quite good in the body of the distribution; however, it is not an estimate which can be successfully extrapolated beyond the data.

The Danish data comprise 2157 losses over one million Danish Krone. We work in units of one million and consider the displacement $\delta$ to be one. The x-axis is shown on a log scale to indicate the great range of the data.

Suppose we are required to price a high-excess layer running from 50 to 200. In this interval we have only six observed losses. If we fit some overall parametric severity distribution to the whole dataset it may not be a particularly good fit in this tail area where the data are sparse.

There are basically two options open to an insurance company. Either it may choose not to insure such a layer, because of too little experience of possible losses. Or, if it wishes to insure the layer, it must obtain a good estimate of the severity distribution in the tail.

Even with a good tail estimate we cannot be sure that the future does not hold some unexpected catastrophic loss. The extreme value methods which we explain in the next section do not predict the future with certainty, but they do offer good models for explaining the extreme events we have seen in

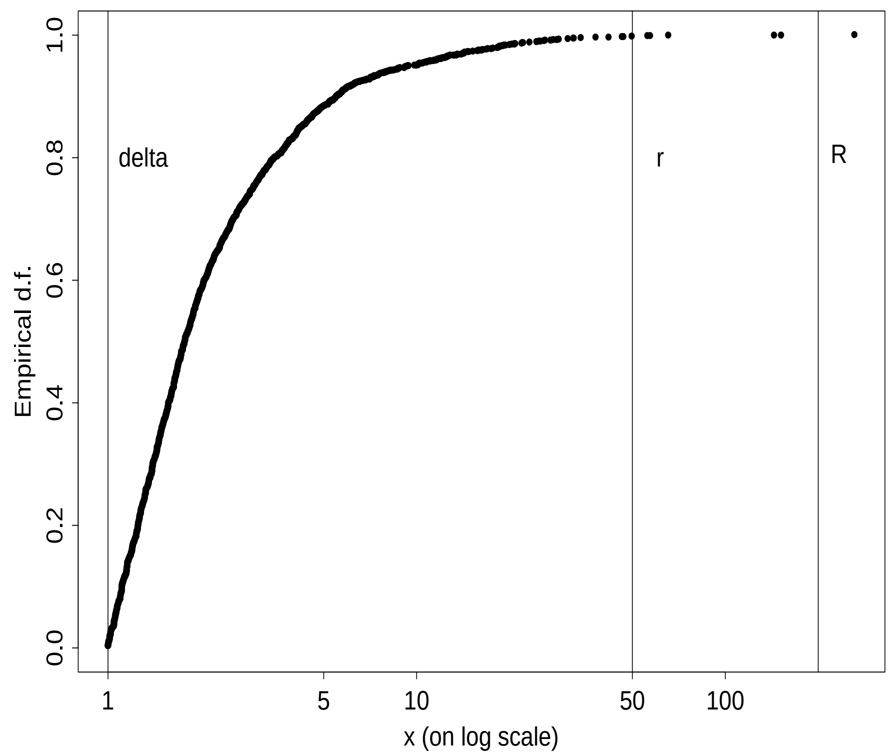


Figure 2: High-excess layer in relation to available data

the past. These models are not arbitrary but based on rigorous mathematical theory concerning the behaviour of extrema.

# 3 Extreme Value Theory

In the following we summarize the results from EVT which underlie our modelling. General texts on the subject of extreme values include Falk, Hüsler & Reiss (1994), Embrechts, Klüppelberg & Mikosch (1997) and Reiss & Thomas (1996).

## 3.1 The generalized extreme value distribution

Just as the normal distribution proves to be the important limiting distribution for sample sums or averages, as is made explicit in the central limit theorem, another family of distributions proves important in the study of the limiting behaviour of sample extrema. This is the family of extreme value distributions.

This family can be subsumed under a single parametrization known as the generalized extreme value distribution (GEV). We define the d.f. of the standard

GEV by

$$H_\xi(x) = \begin{cases} exp(-(1+\xi x)^{-1/\xi}) & \text{if } \xi \neq 0, \\ exp(-e^{-x}) & \text{if } \xi = 0, \end{cases}$$

where $x$ is such that $1 + \xi x > 0$ and $\xi$ is known as the shape parameter. Three well known distributions are special cases: if $\xi > 0$ we have the Fréchet distribution with shape parameter $\alpha = 1/\xi$; if $\xi < 0$ we have the Weibull distribution with shape $\alpha = -1/\xi$; if $\xi = 0$ we have the Gumbel distribution.

If we introduce location and scale parameters $\mu$ and $\sigma > 0$ respectively we can extend the family of distributions. We define the GEV $H_{\xi,\mu,\sigma}(x)$ to be $H_\xi((x-\mu)/\sigma)$ and we say that $H_{\xi,\mu,\sigma}$ is of the type $H_\xi$.

### 3.2 The Fisher-Tippett Theorem

The Fisher-Tippett theorem is the fundamental result in EVT and can be considered to have the same status in EVT as the central limit theorem has in the study of sums. The theorem describes the limiting behaviour of appropriately normalized sample maxima.

Suppose we have a sequence of i.i.d. random variables $X_1, X_2, \ldots$ from an unknown distribution $F$ - perhaps a loss severity distribution. We denote the maximum of the first $n$ observations by $M_n = \max(X_1, \ldots, X_n)$. Suppose further that we can find sequences of real numbers $a_n > 0$ and $b_n$ such that $(M_n - b_n)/a_n$, the sequence of normalized maxima, converges in distribution. That is

$$P\{(M_n - b_n)/a_n \leq x\} = F^n(a_n x + b_n) \to H(x), \text{ as } n \to \infty, \qquad (1)$$

for some non-degenerate df $H(x)$. If this condition holds we say that $F$ is in the maximum domain of attraction of $H$ and we write $F \in \text{MDA}(H)$.

It was shown by Fisher & Tippett (1928) that

$$F \in \text{MDA}(H) \Longrightarrow H \text{ is of the type } H_\xi \text{ for some } \xi.$$

Thus, if we know that suitably normalized maxima converge in distribution, then the limit distribution must be an extreme value distribution for some value of the parameters $\xi$, $\mu$, and $\sigma$.

The class of distributions $F$ for which the condition (1) holds is large. A variety of equivalent conditions may be derived (see Falk et al. (1994)). One such result is a condition for $F$ to be in the domain of attraction of the heavy-tailed Fréchet distribution ($H_\xi$ where $\xi > 0$). This is of interest to us because insurance loss data are generally heavy-tailed.

Gnedenko (1943) showed that for $\xi > 0$, $F \in \text{MDA}(H_\xi)$ if and only if $1 - F(x) = x^{-1/\xi}L(x)$, for some slowly varying function $L(x)$. This result essentially says that if the tail of the d.f. $F(x)$ decays like a power function, then the distribution is in the domain of attraction of the Fréchet. The class of distributions where the tail decays like a power function is quite large and includes the Pareto, Burr, loggamma, Cauchy and t-distributions as well as various mixture models. These are all so-called heavy tailed distributions.

Distributions in the maximum domain of attraction of the Gumbel ($\text{MDA}(H_0)$ include the normal, exponential, gamma and lognormal distributions. The lognormal distribution has a moderately heavy tail and has historically been a popular model for loss severity distributions; however it is not as heavy-tailed as the distributions in MDA($H_\xi$) for $\xi < 0$ . Distributions in the domain of attraction of the Weibull ($H_\xi$ for $\xi < 0$) are short tailed distributions such as the uniform and beta distributions. This class is not of interest in insurance applications.

The Fisher-Tippett theorem suggests the fitting of the GEV to data on sample maxima, when such data can be collected. There is much literature on this topic (see Embrechts et al. (1997)), particularly in hydrology where the so-called annual maxima method has a long history. A well-known reference is Gumbel (1958).

### 3.3 The generalized Pareto distribution

Further results in EVT describe the behaviour of large obervations which exceed high thresholds, and these are the results which lend themselves most readily to the modelling of insurance losses. They address the question: given an observation is extreme, how extreme might it be? The distribution which comes to the fore in these results is the generalized Pareto distribution (GPD).

The GPD is usually expressed as a two parameter distribution with d.f.

$$G_{\xi,\sigma}(x) = \begin{cases} 1 - (1 + \xi x/\sigma)^{-1/\xi} & \text{if } \xi \neq 0, \\ 1 - exp(-x/\sigma) & \text{if } \xi = 0, \end{cases} \tag{2}$$

where $\sigma > 0$, and the support is $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\sigma/\xi$ when $\xi < 0$. The GPD again subsumes other distributions under its parametrization. When $\xi > 0$ we have a reparametrized version of the usual Pareto distribution with shape $\alpha = 1/\xi$; if $\xi < 0$ we have a type II Pareto distribution; $\xi = 0$ gives the exponential distribution.

Again we can extend the family by adding a location parameter $\mu$. The GPD $G_{\xi,\mu,\sigma}(x)$ is defined to be $G_{\xi,\sigma}(x - \mu)$.

### 3.4 The Pickands-Balkema-de Haan Theorem

Consider a certain high threshold $u$ which might, for instance, be the lower attachment point of a high-excess loss layer. At any event $u$ will be greater than any possible displacement $\delta$ associated with the data. We are interested in excesses above this threshold, that is, the amount by which observations overshoot this level.

Let $x_0$ be the finite or infinite right endpoint of the distribution $F$. That is to say, $x_0 = \sup\{x \in \mathbb{R} : F(x) < 1\} \leq \infty$. We define the distribution function of the excesses over the high threshold $u$ by

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

for $0 \leq x < x_0 - u$.

The theorem (Balkema & de Haan 1974, Pickands 1975) shows that under MDA conditions (1) the generalized Pareto distribution (2) is the limiting distribution for the distribution of the excesses, as the threshold tends to the right endpoint. That is, we can find a positive measurable function $\sigma(u)$ such that

$$\lim_{u \to x_0} \sup_{0 \leq x < x_0 - u} |F_u(x) - G_{\xi,\sigma(u)}(x)| = 0,$$

if and only if $F \in \mathrm{MDA}(H_\xi)$.

This theorem suggests that, for sufficiently high thresholds $u$, the distribution function of the excesses may be approximated by $G_{\xi,\sigma}(x)$ for some values of $\xi$ and $\sigma$. Equivalently, for $x - u \geq 0$, the distribution function of the ground-up exceedances $F_u(x-u)$ (the excesses plus $u$) may be approximated by $G_{\xi,\sigma}(x-u) = G_{\xi,u,\sigma}(x)$.

The statistical relevance of the result is that we may attempt to fit the GPD to data which exceed high thresholds. The theorem gives us theoretical grounds to expect that if we choose a high enough threshold, the data above will show generalized Pareto behaviour. This has been the approach developed in Davison (1984) and Davison & Smith (1990).

## 3.5 Tail fitting

If we can fit the GPD to the conditional distribution of the excesses above a high threshold, we may also fit it to the tail of the original distribution above the high threshold (Reiss & Thomas 1996). For $x \geq u$, i.e. points in the tail of the distribution,

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

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

This means that for $x \geq u$ we can use the tail estimate

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

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

## 3.6 Statistical Aspects

The theory makes explicit which models we should attempt to fit to historical data. However, as a first step before model fitting is undertaken, a number of exploratory graphical methods provide useful preliminary information about the data and in particular their tail. We explain these methods in the next section in the context of an analysis of the Danish data.

The generalized Pareto distribution can be fitted to data on excesses of high thresholds by a variety of methods including the maximum likelihood method (ML) and the method of probability weighted moments (PWM). We choose to use the ML-method. For a comparison of the relative merits of the methods we refer the reader to Hosking & Wallis (1987) and Rootzén & Tajvidi (1996).

For $\xi > -0.5$ (all heavy tailed applications) it can be shown that maximum likelihood regularity conditions are fulfilled and that maximum likelihood estimates $(\hat{\xi}_n, \hat{\sigma}_n)$ based on a sample of $n$ excesses are asymptotically normally distributed (Hosking & Wallis 1987).

Specifically we have

$$n^{1/2}\begin{pmatrix}\hat{\xi}_n\\ \hat{\sigma}_n\end{pmatrix}\xrightarrow{\text{d}}N\left[\begin{pmatrix}\xi\\ \sigma\end{pmatrix},\begin{pmatrix}(1+\xi)^2 & \sigma(1+\xi)\\ \sigma(1+\xi) & 2\sigma^2(1+\xi)\end{pmatrix}\right].$$

This result enables us to calculate approximate standard errors for our maximum likelihood estimates.

# 4 Analysis of Danish Fire Loss Data

The Danish data consist of 2157 losses over one million Danish Krone (DKK) from the years 1980 to 1990 inclusive. The loss figure is a total loss figure for the event concerned and includes damage to buildings, damage to furniture and personal property as well as loss of profits. For these analyses the data have been suitably adjusted to reflect 1985 values.

## 4.1 Exploratory data analysis

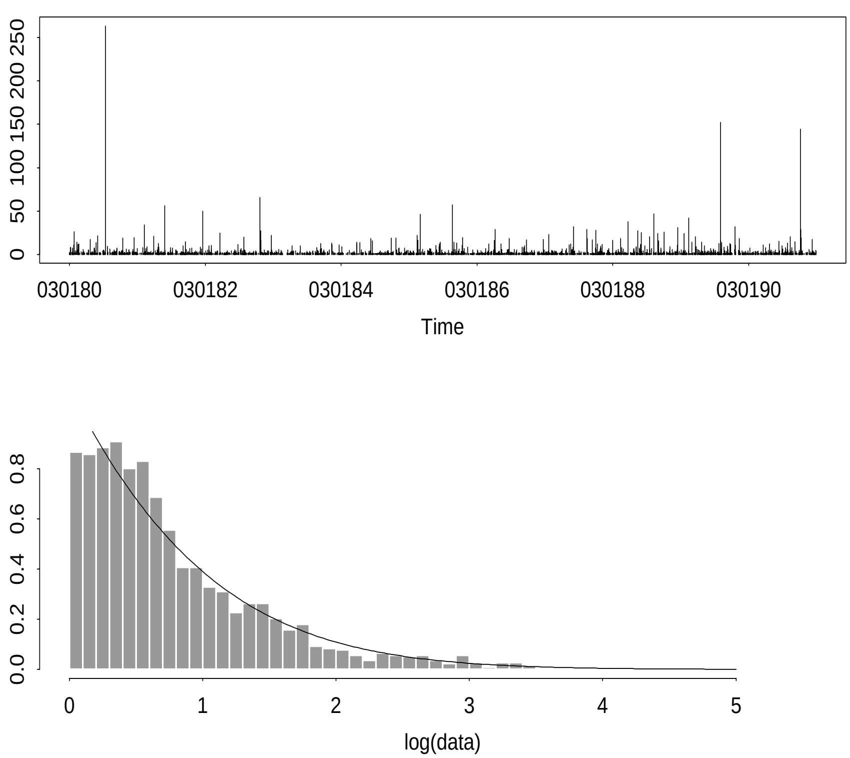


Figure 3: Time series and log data plots for the Danish data. Sample size is 2157.

The time series plot (Figure 3, top) allows us to identify the most extreme losses and their approximate times of occurrence. We can also see whether there is evidence of clustering of large losses, which might cast doubt on our assumption of i.i.d data. This does not seem to be the case with the Danish data; an autocorrelation plot (not shown) also reveals no serial correlation structure.

The histogram on the log scale (Figure 3, bottom) shows the wide range of the data. It also allows us to see whether the data may perhaps have a lognormal right tail, which would be indicated by a familiar bell-shape in the log plot.

We have fitted a truncated lognormal distribution to the dataset using the maximum likelihood method and superimposed the resulting probability density function on the histogram. The scale of the y-axis is such that the total area under the curve and the total area of the histogram are both one. The truncated lognormal appears to provide a reasonable fit but it is difficult to tell from this picture whether it is a good fit to the largest losses in the high-excess area in which we are interested.

The QQ-plot against the exponential distribution (Figure 4) is a very useful guide to heavy-tailedness. It examines visually the hypothesis that the losses come from an exponential distribution, i.e. from a distribution with a medium-sized tail. The quantiles of the empirical distribution function on the x-axis are plotted against the quantiles of the exponential distribution function on the y-axis. The plot is

$$\left\{(X_{k:n}, G_{0,1}^{-1}(\frac{n-k+1}{n+1})),\ k=1,\ldots,n\right\},$$

where $X_{k:n}$ denotes the $k$th order statistic, and $G_{0,1}^{-1}$ is the inverse of the d.f. of the exponential distribution (a special case of the GPD). The points should lie approximately along a straight line if the data are an i.i.d. sample from an exponential distribution.

A concave departure from the ideal shape (as in our example) indicates a heavier tailed distribution whereas convexity indicates a shorter tailed distribution. We would expect insurance losses to show heavy-tailed behaviour.

The usual caveats about the QQ-plot should be mentioned. Even data generated from an exponential distribution may sometimes show departures from typical exponential behaviour. In general, the more data we have, the clearer the message of the QQ-plot. With over 2000 data points in this analysis it seems safe to conclude that the tail of the data is heavier than exponential.

A further useful graphical tool is the plot of the sample mean excess function (see again Figure 4) which is the plot

$$\{(u, e_n(u)),\ X_{n:n} < u < X_{1:n}\},$$

where $X_{1:n}$ and $X_{n:n}$ are the first and $n$th order statistics and $e_n(u)$ is the sample mean excess function defined by

$$e_n(u) = \frac{\sum_{i=1}^n (X_i - u)^+}{\sum_{i=1}^n 1_{\{X_i > u\}}};$$

i.e. the sum of the excesses over the threshold $u$ divided by the number of data points which exceed the threshold $u$.

The sample mean excess function $e_n(u)$ is an empirical estimate of the mean excess function which is defined as $e(u) = E[X - u \mid X > u]$. The mean excess function describes the expected overshoot of a threshold given that exceeedance occurs.

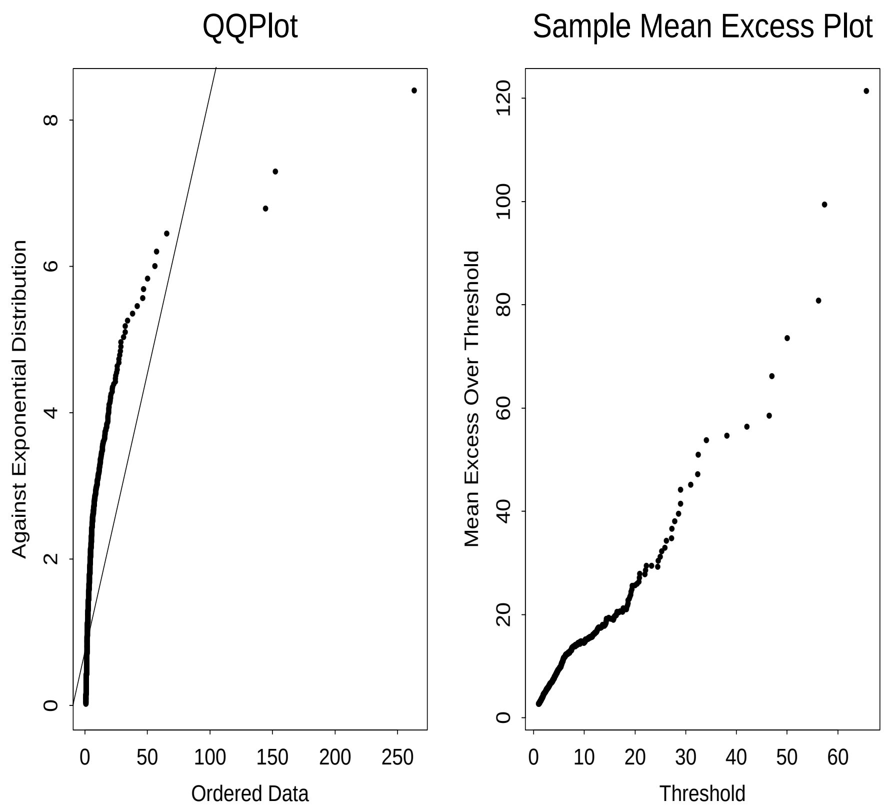


Figure 4: QQ-plot and sample mean excess function

In plotting the sample mean excess function we choose to end the plot at the fourth order statistic and thus omit a possible three further points; these points, being the averages of at most three observations, may be erratic.

The interpretation of the mean excess plot is explained in Beirlant, Teugels & Vynckier (1996), Embrechts et al. (1997) and Hogg & Klugman (1984). If the points show an upward trend, then this is a sign of heavy tailed behaviour. Exponentially distributed data would give an approximately horizontal line and data from a short tailed distribution would show a downward trend.

In particular, if the empirical plot seems to follow a reasonably straight line with positive gradient above a certain value of $u$, then this is an indication that the data follow a generalized Pareto distribution with positive shape parameter in the tail area above $u$.

This is precisely the kind of behaviour we observe in the Danish data (Figure 4). There is evidence of a straightening out of the plot above a threshold of 10, and perhaps again above a threshold of 20. In fact the whole plot is sufficiently straight to suggest that the GPD might provide a reasonable fit to the entire dataset.

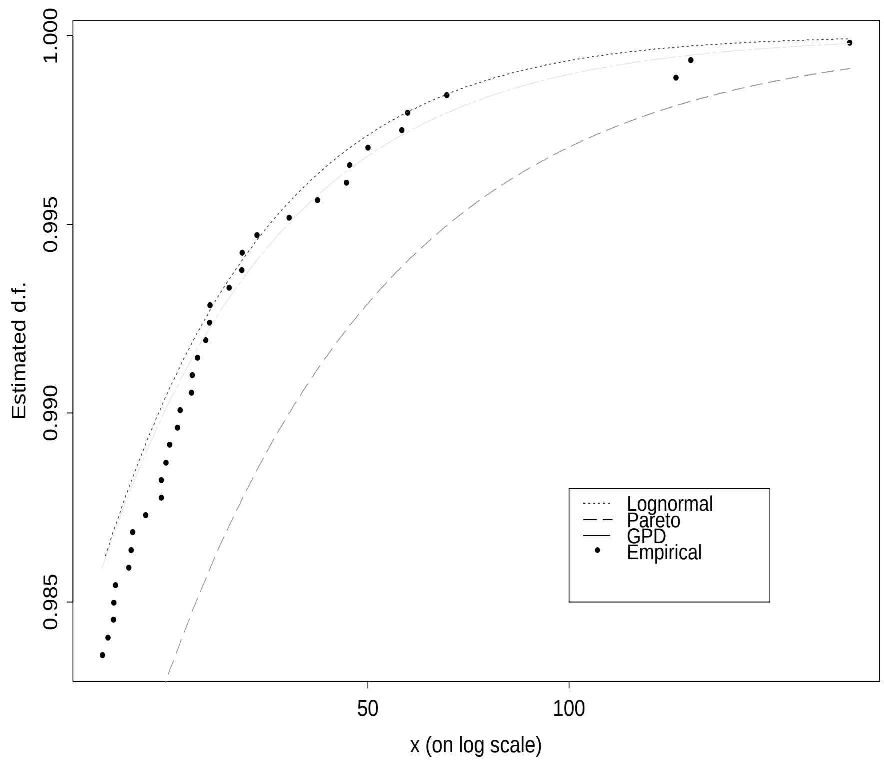


Figure 5: Performance of overall fits in the tail area

## 4.2 Overall fits

In this section we look at standard choices of curve fitted to the whole dataset. We use two frequently used severity models - the truncated lognormal and the ordinary Pareto - as well as the GPD.

By ordinary Pareto we mean the distribution with d.f. $F(x) = 1 - (a/x)^\alpha$ for unknown positive parameters $a$ and $\alpha$ and with support $x > a$. On making the change of variable $y = x - a$ (so that we look at excesses above $a$) and the reparametrization $\alpha = \xi^{-1}$ and $a = \sigma/\xi$ we obtain the GPD for the case $\xi > 0$. The distributions are thus closely related but maximum likelihood estimation leads to different fits; the parametric form of the GPD is in general more flexible.

As discussed earlier, the lognormal distribution is not strictly speaking a heavy tailed distribution. However it is moderately heavy tailed and in many applications it is quite a good loss severity model.

In Figure 5 we see the fit of these models in the tail area above a threshold of 20. The lognormal is a reasonable fit, although its tail is just a little too thin to capture the behaviour of the very highest observed losses. The Pareto, on the other hand, seems to overestimate the probabilities of large losses. This, at first sight, may seem a desirable, conservative modelling feature. But it may be the case, that this d.f. is so conservative, that if we use it to answer our attachment point and premium calculation problems, we will arrive at answers that are unrealistically high.

The GPD is somewhere between the lognormal and Pareto in the tail area and actually seems to be quite a good explanatory model for the highest losses. The data are of course truncated at 1M DKK, and it seems that, even above this low threshold, the GPD is not a bad fit to the data. By raising the threshold we can, however, find models which are even better fits to the larger losses.

Estimates of high quantiles and layer prices based on these three fitted curves are given in table 1.

### 4.3 Fitting to data on exceedances of high thresholds

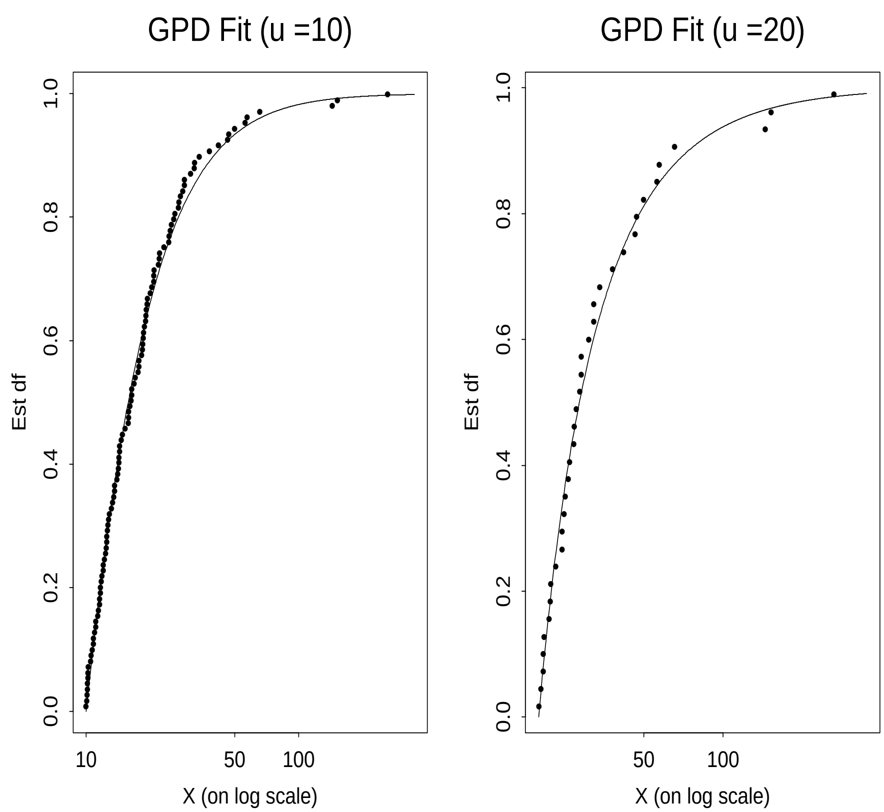


Figure 6: In left plot GPD is fitted to 109 exceedances of the threshold 10. The parameter estimates are $\xi = 0.497$, $\mu = 10$ and $\sigma = 6.98$. In right plot GPD is fitted to 36 exceedances of the threshold 20. The parameter estimates are $\xi = 0.684$, $\mu = 20$ and $\sigma = 9.63$.

The sample mean excess function for the Danish data suggests we may have success fitting the GPD to those data points which exceed high thresholds of 10 or 20; in Figure 6 we do precisely this. We use the three parameter form of the GPD with the location parameter set to the threshold value and we obtain fits to these data which seem reasonable to the naked eye.

The estimates we obtain are estimates of the conditional distribution of the losses, given that they exceed the threshold. Quantile estimates derived from these curve are conditional quantile estimates which indicate the scale of losses which could be experienced if the threshold were to be exceeded.

As described in section 3.5, we can transform scale and location parameters to obtain a GPD model which fits the severity distribution itself in the tail area above the threshold. Since our data are truncated at the displacement of one million we actually obtain a fit for the tail of the truncated severity distribution $F_{X^\delta}(x)$. This is shown for a threshold of 10 in Figure 7. Quantile estimates derived from this curve are quantile estimates conditional on exceedance of the displacement of one million.

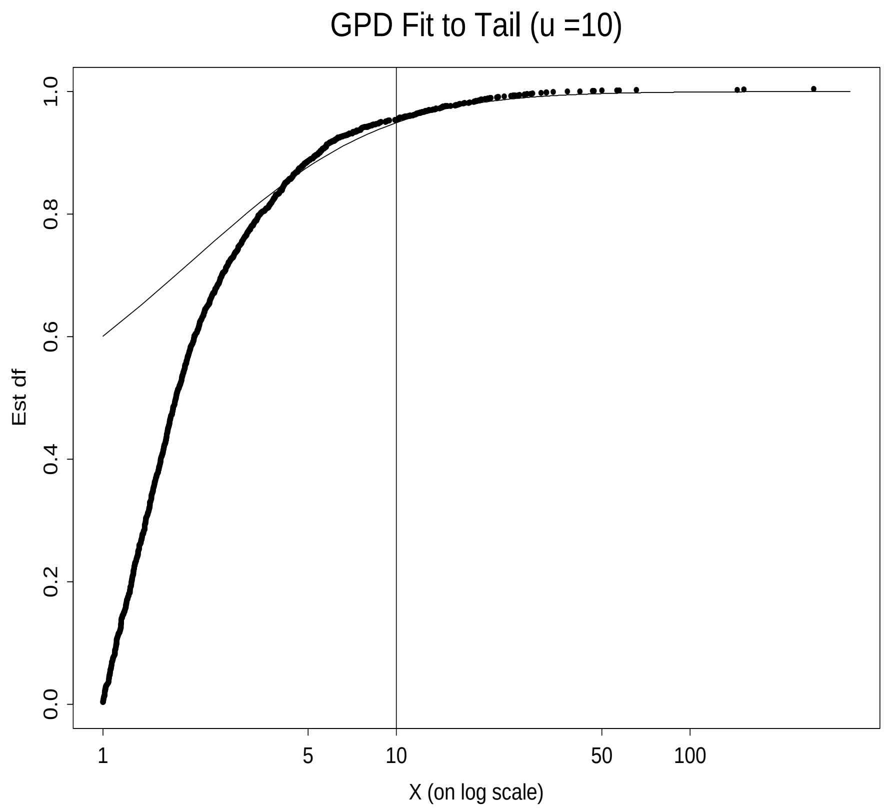


Figure 7: Fitting the GPD to tail of severity distribution above threshold 10. The parameter estimates are $\xi = 0.497$, $\mu = -0.845$ and $\sigma = 1.59$.

So far we have considered two arbitrary thresholds. In the next sections we consider the question of optimizing the choice of threshold by investigating the different estimates we get for model parameters, high quantiles and prices of high-excess layers.

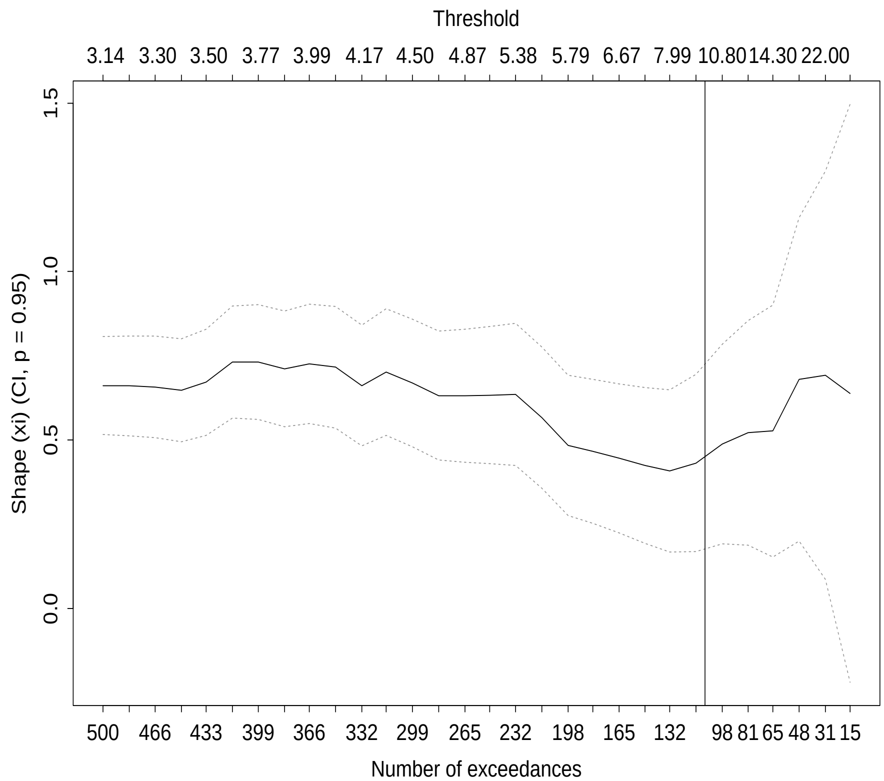


Figure 8: Estimates of shape by increasing threshold on the upper x-axis and decreasing number of exceedances on the lower x-axis; in total 30 models are fitted.

## 4.4 Shape and quantile estimates

As far as pricing of layers or estimation of high quantiles using a GPD model is concerned, the crucial parameter is $\xi$, the tail index. Roughly speaking, the higher the value of $\xi$ the heavier the tail and the higher the prices and quantile estimates we derive. For a three-parameter GPD model $G_{\xi,\mu,\sigma}$ the $p$th quantile can be calculated to be $\mu + \sigma/\xi((1-p)^{-\xi} - 1)$.

In Figure 8 we fit GPD models with different thresholds to obtain maximum likelihood estimates of $\xi$, as well as asymptotic confidence intervals for the parameter estimates. On the lower x-axis the number of data points exceeding the threshold is plotted; on the upper x-axis the threshold itself. The shape estimate is plotted on the y-axis. A vertical line marks the location of our first model with a threshold at 10.

In using this picture to choose an optimal threshold we are confronted with a bias-variance tradeoff. Since our modelling approach is based on a limit theorem which applies above high thresholds, if we choose too low a threshold we may get biased estimates because the theorem does not apply. On the other hand, if we set too high a threshold we will have few data points and our estimates will be prone to high standard errors.

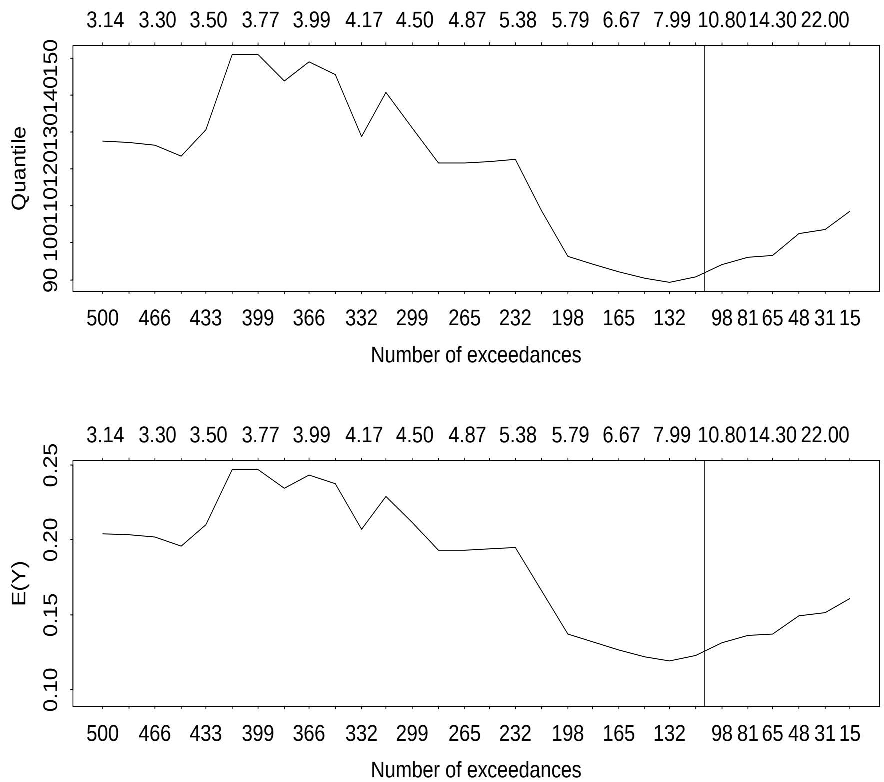


Figure 9: .999 quantile estimates (upper picture) and price indications for a (50,200) layer (lower picture) for increasing thresholds and decreasing numbers of exceedances.

Figure 9 (upper panel) is a similar plot showing how quantile estimates depend on the choice of threshold. We have chosen to plot estimates of the .999th quantile. Roughly speaking, if the model is a good one, one in every thousand losses which exceed 1M DKK might be expected to exceed this quantile; such losses are rare but threatening to the insurer.

We have tabulated quantile estimates for some selected thresholds in table 1 and given the corresponding estimates of the shape parameter. Using the model with a threshold at 10 the .999th quantile is estimated to be 95. But if we push the threshold back to 4 the quantile estimate goes up to 147. There is clearly a considerable difference these two estimates and if we attempt to estimate higher quantiles such as the .9999th this difference becomes more pronounced. This underlines the fact that estimates of high quantiles are extremely model dependent.

| Model | Threshold | Excesses | Shape | s.e. | .999th | .9999th | $P$ |
|---|---|---|---|---|---|---|---|
| GPD | 3 | 532 | 0.67 | 0.07 | 129 | 603 | 0.21 |
| GPD | 4 | 362 | 0.72 | 0.09 | 147 | 770 | 0.24 |
| GPD | 5 | 254 | 0.63 | 0.10 | 122 | 524 | 0.19 |
| GPD | 10 | 109 | 0.50 | 0.14 | 95 | 306 | 0.13 |
| GPD | 20 | 36 | 0.68 | 0.28 | 103 | 477 | 0.15 |
| Models fitted to whole dataset | | | | | | | |
| GPD | all | data | 0.60 | 0.04 | 101 | 410 | 0.15 |
| Pareto | all | data | | | 235 | 1453 | 0.10 |
| Lognormal | all | data | | | 82 | 239 | 0.41 |
| Scenario models | | | | | | | |
| GPD | 10 | 108 | 0.39 | 0.13 | 77 | 201 | 0.09 |
| GPD | 10 | 106 | 0.17 | 0.11 | 54 | 96 | 0.02 |
| GPD | 10 | 110 | 0.60 | 0.15 | 118 | 469 | 0.19 |

Table 1: Comparison of shape and quantile estimates for various models

## 4.5 Calculating price indications

To give an indication of the prices we get from our model we calculate $P = E[Y_i \mid X_i > \delta]$ for a layer running from 50 to 200 million (as in Figure 2). It is easily seen that, for a general layer $(r, R)$, $P$ is given by

$$P = \int_r^R (x - r) f_{X^\delta}(x) dx + (R - r)(1 - F_{X^\delta}(R)), \tag{3}$$

where $f_{X^\delta}(x) = dF_{X^\delta}(x)/dx$ denotes the density function for the losses truncated at $\delta$. Picking a high threshold $u$ $(< r)$ and fitting a GPD model to the excesses, we can estimate $F_{X^\delta}(x)$ for $x > u$ using the tail estimation procedure. We have the estimate

$$\widehat{F_{X^\delta}(x)} = (1 - F_n(u)) G_{\hat{\xi}, u, \hat{\sigma}}(x) + F_n(u),$$

where $\hat{\xi}$ and $\hat{\sigma}$ are maximum-likelihood parameter estimates and $F_n(u)$ is an estimate of $P\left\{X^\delta \leq u\right\}$ based on the empirical distribution function of the data. We can estimate the density function of the $\delta$-truncated losses using the derivative of the above expression and the integral in (3) has an easy closed form.

In Figure 9 (lower picture) we show the dependence of $P$ on the choice of threshold. The plot seems to show very similar behaviour to that of the .999th percentile estimate, with low thresholds leading to higher prices.

The question of which threshold is ultimately best depends on the use to which the results are to be put. If we are trying to answer the optimal attachment point problem or to price a high layer we may want to err on the side of conservatism and arrive at answers which are too high rather than too low. In the case of the Danish data we might set a threshold lower than 10, perhaps at four. The GPD model may not fit the data quite so well above this lower threshold as it does above the high threshold of 10, but it might be safer to use the low threshold to make calculations.

On the other hand there may be business reasons for trying to keep the attachment point or premium low. There may be competition to sell high excess policies and this may mean that basing calculations only on the highest observed losses is favoured, since this will lead to more attractive products (as well as a better fitting model).

In other insurance datasets the effect of varying the threshold may be different. Inference about quantiles might be quite robust to changes in threshold or elevation of the threshold might result in higher quantile estimates. Every dataset is unique and the data analyst must consider what the data mean at every step. The process cannot and should not be fully automated.

### 4.6 Sensitivity of Results to the Data

We have seen that inference about the tail of the severity distribution may be sensitive to the choice of threshold. It is also sensitive to the largest losses we have in our dataset. We show this by considering several scenarios in Table 1.

If we return to our first model with a threshold at 10, our estimate of the .999th quantile is 95. Dropping the largest loss from the dataset (so that we now have only 108 exceedances) reduces this estimate to 77. The shape parameter is reduced from 0.50 to 0.39. Removing the next two largest losses causes a further large reduction in the estimates. The price indications $P$ are similarly affected.

On the other hand if we introduce a new largest loss of 350 to the dataset (the previous largest being 263) the shape estimate goes up to 0.60 and the estimate of the .999th quantile to 118. Such changes are caused by tampering with the most extreme values in the dataset. Adding or deleting losses of lower magnitude has much less effect.

## 5 Discussion

We hope to have shown that fitting the generalized Pareto distribution to insurance losses which exceed high thresholds is a useful method for estimating the tails of loss severity distributions. In our experience with several insurance datasets we have found consistently that the generalized Pareto distribution is a good approximation in the tail.

This is not altogether surprising. As we have explained, the method has solid foundations in the mathematical theory of the behaviour of extremes; it is not simply a question of ad hoc curve fitting. It may well be that, by trial and error, some other distribution can be found which fits the available data even better in the tail. But such a distribution would be an arbitrary choice, and we would have less confidence in extrapolating it beyond the data.

It is our belief that any practitioner who routinely fits curves to loss severity data should know about extreme value methods. There are, however, a number of caveats to our endorsement of these methods. We should be aware of various layers of uncertainty which are present in any data analysis, but which are perhaps magnified in an extreme value analysis.

On one level, there is parameter uncertainty. Even when we have abundant, good- quality data to work with and a good model, our parameter estimates are still subject to a standard error. We obtain a range of parameter estimates which are compatible with our assumptions. As we have already noted, inference is sensitive to small changes in the parameters, particularly the shape parameter.

Model uncertainty is also present - we may have good data but a poor model. Using extreme value methods we are at least working with a good class of models, but they are applicable over high thresholds and we must decide where to set the threshold. If we set the threshold too high we have few data and we introduce more parameter uncertainty. If we set the threshold too low we lose our theoretical justification for the model. In the analysis presented in this paper inference was very sensitive to the threshold choice (although this is not always the case).

But probably more serious than parameter and model uncertainty is data uncertainty. In a sense, it is never possible to have enough data in an extreme value analysis. Whilst a sample of 1000 data points may be ample to make inference about the mean of a distribution using the central limit theorem, our inference about the tail of the distribution is less certain, since only a few points enter the tail region. As we have seen, inference is very sensitive to the largest observed losses and the introduction of new extreme losses to the dataset may have a substantial impact. For this reason, there is still be a role for stress scenarios in loss severity analyses, whereby historical loss data are enriched by hypothetical losses to investigate the consequences of unobserved, adverse events.

Another aspect of data uncertainty is that of dependent data. In this paper we have made the familiar assumption of independent, identically distributed data. In practice we may be confronted with clustering, trends, seasonal effects and other kinds of dependencies. When we consider fire losses in Denmark it may seem a plausible first assumption that individual losses are independent of one another; however, it is also possible to imagine that circumstances conducive or inhibitive to fire outbreaks generate dependencies in observed losses. Destructive fires may be greatly more common in the summer months; buildings of a particular vintage and building standard may succumb easily to fires and cause high losses. Even after adjustment for inflation there may be a general trend of increasing or decreasing losses over time, due to an increasing number of increasingly large and expensive buildings, or due to increasingly good safety measures.

These issues lead to a number of interesting statistical questions in what is very much an active research area. Papers by Davison (1984) and Davison & Smith (1990) discuss clustering and seasonality problems in environmental data and make suggestions concerning the modelling of trends using regression models built into the extreme value modelling framework. The modelling of trends is also discussed in Rootzén & Tajvidi (1996).

We have developed software to fit the generalized Pareto distribution to exceedances of high thresholds and to produce the kinds of graphical output presented in this paper. It is written in Splus and is available over the World Wide Web at `http://www/math.ethz.ch/~mcneil`.

# 6 Acknowledgements

Much of the work in this paper came about through a collaborative project with Swiss Re Zurich. I gratefully acknowledge Swiss Re for their financial support as well as for many fruitful discussions [1].

[1]The author is supported by Swiss Re as a research fellow at ETH Zurich

I thank Mette Rytgaard of Copenhagen Re for making the Danish fire loss data available and Richard Smith for advice on software and algorithms for fitting the GPD to data.

Special thanks are due to Paul Embrechts for introducing me to the theory of extremes and its many interesting applications.

# References


Balkema, A. & de Haan, L. (1974), 'Residual life time at great age', *Annals of Probability* **2**, 792–804.

Beirlant, J. & Teugels, J. (1992), 'Modelling large claims in non-life insurance', *Insurance: Mathematics and Economics* **11**, 17–29.

Beirlant, J., Teugels, J. & Vynckier, P. (1996), *Practical analysis of extreme values*, Leuven University Press, Leuven.

Davison, A. (1984), Modelling excesses over high thresholds, with an application, *in* J. de Oliveira, ed., 'Statistical Extremes and Applications', D. Reidel, pp. 461–482.

Davison, A. & Smith, R. (1990), 'Models for exceedances over high thresholds (with discussion)', *Journal of the Royal Statistical Society, Series B* **52**, 393–442.

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

Embrechts, P. & Klüppelberg, C. (1993), 'Some aspects of insurance mathematics', *Theory of Probability and its Applications* **38**, 262–295.

Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997), *Modelling extremal events for insurance and finance*, Springer. To appear.

Falk, M., Hüsler, J. & Reiss, R. (1994), *Laws of Small numbers: extremes and rare events*, Birkhäuser, Basel.

Fisher, R. & Tippett, L. (1928), 'Limiting forms of the frequency distribution of the largest or smallest member of a sample', *Proceedings of the Cambridge Philosophical Society* **24**, 180–190.

Gnedenko, B. (1943), 'Sur la distribution limite du terme maximum d'une série aléatoire', *Annals of Mathematics* **44**, 423–453.

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

Hogg, R. & Klugman, S. (1984), *Loss Distributions*, Wiley, New York.

Hosking, J. & Wallis, J. (1987), 'Parameter and quantile estimation for the generalized Pareto distribution', *Technometrics* **29**, 339–349.

Pickands, J. (1975), 'Statistical inference using extreme order statistics', *The Annals of Statistics* **3**, 119–131.

Reiss, R. & Thomas, M. (1996), 'statistical analysis of extreme values'. Documentation for XTREMES software package.

Rootzén, H. & Tajvidi, N. (1996), 'Extreme value statistics and wind storm losses: a case study'. To appear in Scandinavian Actuarial Journal.

Smith, R. (1989), 'Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone', *Statistical Science* **4**, 367–393.