The Peaks over Thresholds Method for Estimating High Quantiles of Loss Distributions

Alexander J. McNeil & Thomas Saladin
Departement Mathematik
ETH Zentrum
CH-8092 Zürich
Tel: +41 1 632 61 62
Fax: +41 1 632 10 85
email: mcneil@math.ethz.ch

April 24, 1997

Abstract

We review the peaks over thresholds or POT method for modelling tails of loss severity distributions and discuss the use of this technique for estimating high quantiles and the possible relevance of this to excess of loss insurance in high layers. We test the method on a variety of simulated heavy-tailed distributions to show what kind of thresholds are required and what sample sizes are necessary to give accurate estimates of quantiles.

Keywords: Loss severity distributions; high excess layers; extreme value theory; excesses over high thresholds; generalized Pareto distribution; quantile estimation; simulation study.

1 Introduction

Historical data on loss severities in insurance are often modelled with a variety of so-called heavy-tailed distributions such as the Pareto, Burr, loggamma, Benktander, Weibull and lognormal. A well-known reference on fitting size-of-loss distributions to data is Hogg & Klugman (1984). Fitting can be

carried out with standard software and it is common practice to fit a number of models to a given dataset and to select the best fitting ones according to goodness-of-fit criteria.

In this paper we are particularly interested in modelling the tails of heavy-tailed loss severity distributions. This is important, for example, in the pricing of excess of loss insurance in high layers. We are less concerned with finding an overall model for all loss severities; we seek a model which describes the largest losses.

Several authors (Beirlant & Teugels 1992, Embrechts & Klüppelberg 1993) have argued that extreme value theory (EVT) motivates a number of sensible approaches to this problem. In particular the peaks over thresholds or POT model has been advocated in this context (Rootzén & Tajvidi 1996, McNeil 1997). In the POT model the excess losses over high thresholds are modelled with the generalized Pareto distribution (GPD). This distribution arises naturally in a key limit theorem in EVT.

The POT model offers a unifying approach to the modelling of the tail of a severity distribution. For many different underlying distributions the GPD can be fitted above a sufficiently high threshold. This presupposes that sufficient data are available above the chosen threshold. However, if this can be assumed, the POT method allows us to use the GPD as the principal modelling tool for large losses.

The aims of this paper are twofold: firstly to review the POT method for modelling tails of loss severity distributions; secondly to test the performance of the technique under the controlled circumstances of a simulation study. We look at the problem of estimating a high quantile of a severity distribution and simulate possible data sets from known heavy-tailed distributions. In this situation the true quantile value is known to us and we can investigate under what circumstances the POT method gives an accurate estimate.

2 Methods

2.1 Theoretical Background

In this section 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).

Suppose we have a sequence of i.i.d. observations $X_1, \ldots, X_n$, from an unknown distribution function F. We are interested in excess losses over a high threshold u. 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 threshold u by

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

for $0 \leq x < x_0 - u$. $F_u(x)$ is thus the probability that a loss exceeds the threshold u by no more than an amount x, given that the threshold is exceeded.

The distribution which comes to the fore in the modelling of excesses is the generalized Pareto distribution (GPD) which is usually expressed as a two parameter distribution with d.f.

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

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 subsumes three other distributions under its parametrization. When $\xi > 0$ we have a reparametrized version of the usual Pareto distribution; if $\xi < 0$ we have a type II Pareto distribution; $\xi = 0$ gives the exponential distribution. The family can be extended by adding a location parameter μ. The GPD $G_{\xi,\mu,\sigma}(x)$ is defined to be $G_{\xi,\sigma}(x - \mu)$.

This distributional choice is motivated by a theorem (Balkema & de Haan 1974, Pickands 1975) which states that, for a certain class of distributions, the GPD is the limiting distribution for the distribution of the excesses, as the threshold tends to the right endpoint. Formally, we can find a positive measurable function $\sigma(u)$ such that

$$\lim_{u \to x_0} \sup_{0 \leq x < x_0 - u} |F_u(x) - G_{\xi,\sigma(u)}(x)| = 0, \tag{2}$$

if and only if F is in the maximum domain of attraction of the extreme value ditribution H_ξ, a condition we write as $F \in \mathrm{MDA}(H_\xi)$. What does this condistion mean?

The generalized extreme value distribution (standard GEV) has the distribution function

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

where x is such that $1 + \xi x > 0$ and ξ is known as the shape parameter. The three extreme value distributions are special cases of the GEV: if $\xi > 0$

we have the Fréchet distribution; if $\xi < 0$ we have the Weibull distribution; $\xi = 0$ gives the Gumbel distribution.

To understand the idea of a maximum domain of attraction we consider maxima of i.i.d. samples and define $M_n = \max(X_1, \ldots, X_n)$. $F \in \mathrm{MDA}(H)$ is equivalent to saying that we can find sequences of real numbers $a_n > 0$ and b_n such that the normalized sequence $(M_n - b_n)/a_n$ converges in distribution to H. That is

$$P\{(M_n - b_n)/a_n \leq x\} = F^n(a_n x + b_n) \to H(x), \text{ as } n \to \infty. \qquad (3)$$

It turns out that the non-degenerate possibilities for this distribution H are limited. It was shown by Fisher & Tippett (1928) that the extreme value distributions (possibly with changes of scale and location) are the only non-degenerate limit distributions for appropriately normalized sample maxima.

Our argument is summarized as follows. If, for a given distribution F, appropriately normalized sample maxima converge in distribution to a non-degenerate limit (3), then this is equivalent to saying that F is in the maximum domain of attraction of an extreme value distribution H_ξ for some value of ξ. In this case, by the theorem of Pickands, Balkema and de Haan (2), it follows that the distribution function of the excesses above a high threshold converges to generalized Pareto with shape parameter ξ, as the threshold tends to the right endpoint of F.

The class of distributions F for which the condition $F \in \mathrm{MDA}(H_\xi)$ holds is large; essentially all commonly encountered continuous distributions show the kind of regular behaviour for sample maxima described by (3). 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 ($F \in \mathrm{MDA}(H_\xi)$ where $\xi > 0$). This is of particular interest because insurance loss data are generally heavy-tailed.

Gnedenko (1943) showed that for $\xi > 0$, $F \in \mathrm{MDA}(H_\xi)$ 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.

Distributions in this class we will define formally to be the heavy-tailed distributions. For these distributions the theorem of Pickands, Balkema and De Haan, tells us that above sufficiently high thresholds the distribution of the excesses may be approximated by a GPD with positive shape parameter, which is to say a reparametrized version of the Pareto distribution.

Distributions in MDA(H_0), the maximum domain of attraction of the Gumbel, include the normal, exponential, gamma and lognormal distributions. These distributions we call medium-tailed distributions and the distribution of excesses over sufficiently high thresholds may in this case be approximated by the exponential distribution. Note that, 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_ξ) for $\xi < 0$.

Distributions in the domain of attraction of the Weibull (H_ξ for $\xi < 0$) are short tailed distributions such as the uniform and beta distributions. This class is of lesser interest in insurance applications.

It is customary to say that a distribution in MDA(H_ξ) has tail index or extremal value index $1/\xi$. For positive ξ, the smaller the tail index the larger the weight of the tail.

2.2 Fitting the GPD

These theoretical considerations suggest that when we have data from unknown underlying severity distributions we may be able to successfully approximate the distribution of excess amounts over sufficiently high thresholds by a generalized Pareto distribution, $G_{\xi,\sigma}(x)$ for some values of ξ and σ. This is a modelling approach which can be found in several papers in the statistical literature (Smith 1989, Davison & Smith 1990).

This approach presupposes that we can find a suitable high threshold above which the approximation suggested by the theorem of Pickands, Balkema and de Haan is good and above which we still have sufficient data to give accurate estimates of unknown parameters. Such practical problems are addressed in McNeil (1997).

One tool for choosing suitable thresholds is the sample mean excess plot

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

where $X_{1:n}$ and $X_{n:n}$ are the first and nth order statistics of the data sample 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 exceedance occurs. The interpretation of the plot is explained in Beirlant, Teugels & Vynckier (1996), Embrechts et al. (1997) and Hogg & Klugman (1984). 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 excesses over this threshold follow a generalized Pareto distribution with positive shape parameter. This is clear since for the GPD

$$e(u) = (\sigma + \xi u)/(1 - \xi),$$

where $\sigma + u\xi > 0$.

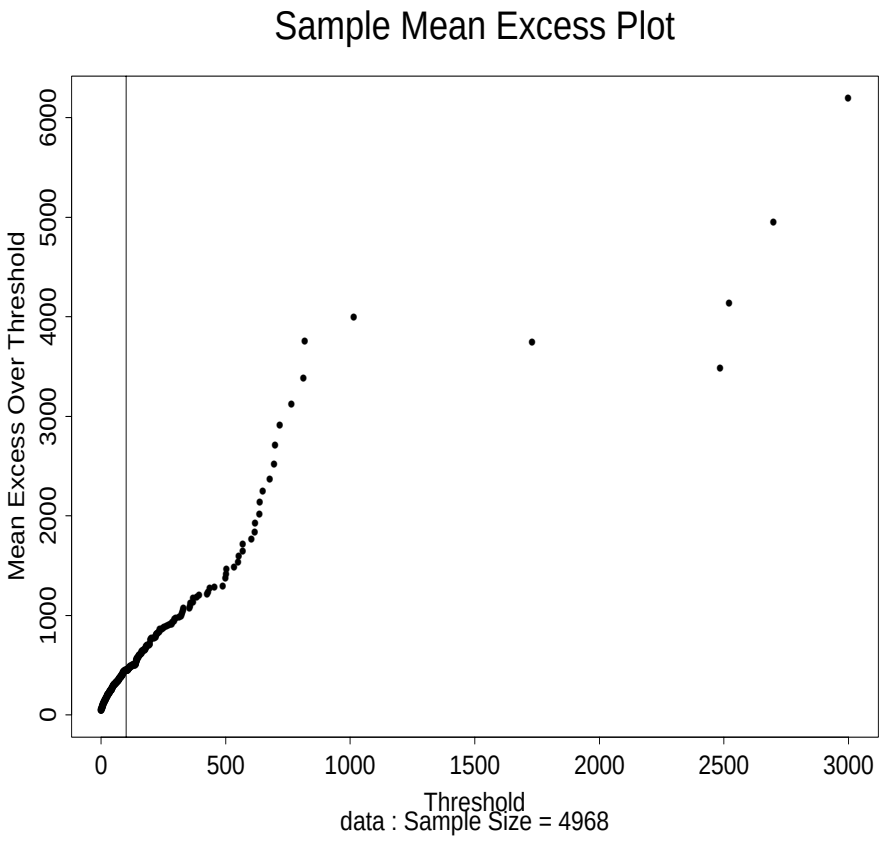

Figure 1: Sample mean excess plot for 4968 property losses. (Note that the three largest losses have been omitted since these tend to distort plot.)

An example of the sample mean excess plot is shown in figure 1 for a dataset comprising 4968 property losses normalized so that the smallest loss is 1. (The money unit is not important to us.) In fact, apart from the extreme right hand points which are averages of only a small number of large

excesses, the whole plot is approximately linear, suggesting that the GPD might be fitted from a reasonably low threshold. We have chosen a very high threshold at 100 (marked by a vertical line), since there is some evidence that the plot kinks slightly around this threshold. This threshold is exceeded by 149 losses and represents a threshold at about the 97th percentile of the data.

As well as being able to approximate the distribution of the excesses above thresholds, it is also possible to find closely related approximations to the distribution of the ground-up exceedances of a threshold (the excesses plus u) and to the tail of the original unknown underlying severity distribution. The distribution of the exceedances we define as $F_u(x-u)$ for $x \geq u$ and this may be approximated by $G_{\xi,\sigma}(x-u) = G_{\xi,u,\sigma}(x)$.

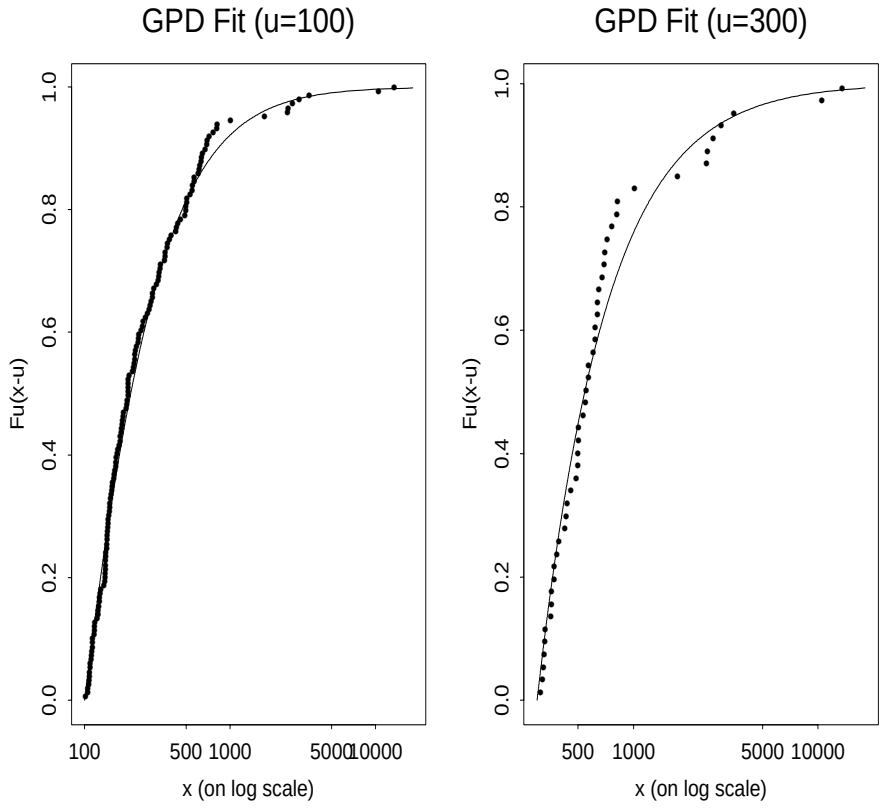

Figure 2: In left plot GPD is fitted to 149 exceedances of the threshold 100. The parameter estimates are $\xi = 0.747$ and $\sigma = 118.4$. In right plot GPD is fitted to 49 exceedances of the threshold 300. The parameter estimates are $\xi = 0.814$ and $\sigma = 261.4$.

Examples of this approximation can be seen in figure 2 for the 4968

property losses with high thresholds set at 100 and 300. These thresholds give 149 and 49 exceedances respectively. Parameters are estimated by maximum likelihood and the resulting GPD curves are superimposed on the empirical estimate of the distribution function of the ground-up exceedances. Note that values of ξ are around 0.8 which translates to a tail index around 1.25 and indicates the kind of heavy tail which is typical of insurance data.

For points in the tail of the distribution ($x \geq u$) we note that

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

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

Thus for $x \geq u$ we can use the tail estimate

$$\widehat{F(x)} = (1 - F_n(u))G_{\xi,u,\sigma}(x) + F_n(u) \tag{4}$$

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

An example of this estimate is seen in figure 3. The GPD is again fitted to 149 exceedances of a threshold of 100. The estimate of the shape parameter is as in figure 2; the estimates of the shape and location parameters have been adjusted so that the curve fits the empirical distribution function of all losses in the tail area. $1 - \widehat{F(x)}$ is plotted against x with both axes on the log scale.

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). These methods have been compared for GPD distributed data in simulation studies by Hosking & Wallis (1987) and Rootzén & Tajvidi (1996).

Hosking & Wallis (1987) found that for GPD data with shape parameter in the range $0 \leq \xi \leq 0.4$ and particularly for small sample sizes, the PWM method has advantages over the ML method since PWM estimates show less dispersion around the true value (less mean squared error). However, as the sample size increases the difference becomes less pronounced.

On the other hand Rootzén & Tajvidi (1996) showed that for heavy tailed data with $\xi \geq 0.5$ the PWM method gives seriously biased parameter estimates whereas ML estimates are consistent. Since we are considering the modelling of heavy tailed insurance data we favour the ML method. Maximum likelihood has the further attraction to the statistician that models can

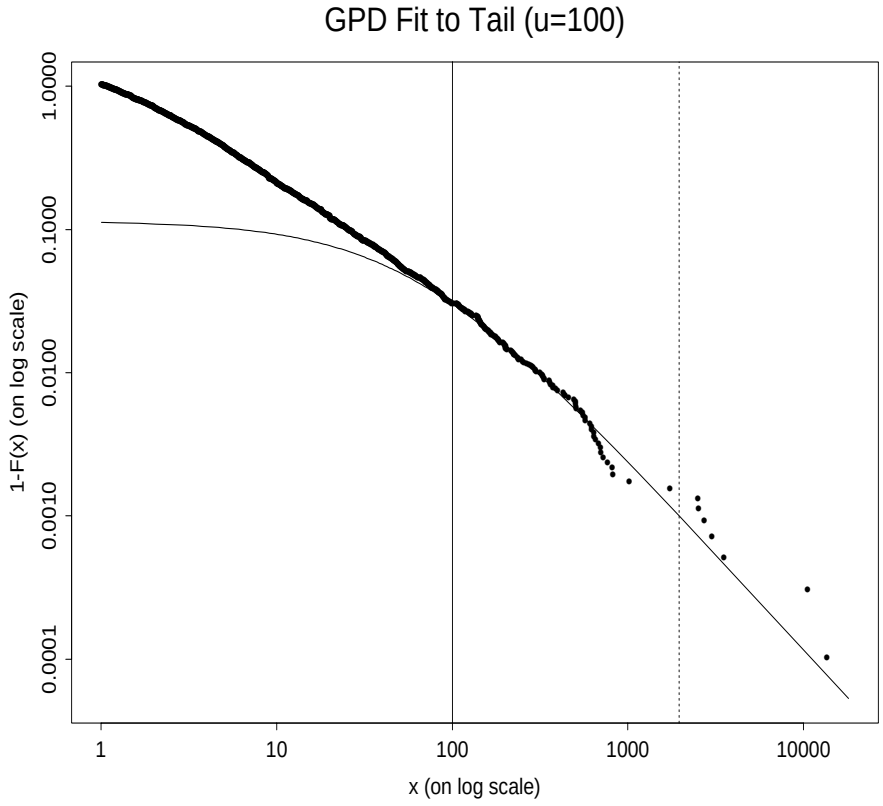

Figure 3: Fitting the GPD to tail of severity distribution above threshold 100. The parameter estimates are $\xi = 0.747$, $\mu = -46.94$ and $\sigma = 8.65$.

be easily extended to encompass regression relationships between data and other explanatory variables.

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

Specifically we have

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

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

2.3 Quantile Estimation

We will assume that the goal of our modelling is to estimate a high quantile in the tail of a loss severity distribution such as $x_{0.999}$. This represents a one in a thousand loss event. In the context of high excess layer insurance $x_{0.999}$ is the lower layer boundary we would choose if we wished the layer to be pierced at most once every thousand losses.

For a distribution F and a general p, $x_p = F^{\leftarrow}(p)$ where $F^{\leftarrow}$ denotes the gereralized inverse of F;

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

For a known distribution this is something we can calculate. When the distribution is unknown this is something we wish to estimate.

The POT estimator of x_p is obtained by inverting the tail estimation formula given in (4) and substituting unknown parameters of the GPD by estimates $\hat{\xi}$ und $\hat{\sigma}$ to get

$$\widehat{x_p} = \widehat{F^{\leftarrow}(p)} = G^{-1}_{\hat{\xi},u,\hat{\sigma}}\left(\frac{p - F_n(u)}{1 - F_n(u)}\right) = u + \frac{\hat{\sigma}}{\hat{\xi}}\left(\left(\frac{1-p}{1-F_n(u)}\right)^{-\hat{\xi}} - 1\right).$$

If we write N_u for the number of exceedances of the threshold u and n for the total number of realizations we have from the distribution F, our quantile estimator is

$$\widehat{x_p} = u + \frac{\hat{\sigma}}{\hat{\xi}}\left(\left(\frac{n}{N_u}(1-p)\right)^{-\hat{\xi}} - 1\right). \tag{5}$$

An example of the use of this estimator is seen in figure 3 where $\widehat{x_{0.999}}$ has been added to the picture as a dotted line.

2.4 Simulation Procedure

In the kind of data analyses shown in figures 2 and 3 the POT method for tail estimation appears to work well. To understand better under what circumstances the POT method works we can use simulated datasets. We can simulate data from known heavy-tailed distributions and use quantile estimation as the performance measure. For known distributions high quantiles can be calculated and the true values can be compared with POT estimates.

The principal problem is determining suitable high thresholds. In a simulation study it is impractical to perform interactive data analysis with each of several hundred simulated datasets and to choose thresholds on the basis

of exploratory techniques like the sample mean excess plot. To get around this problem we do not simulate datasets of a fixed size. Rather we fix the threshold to be a reasonably high quantile of the distribution and we simulate fixed numbers of excesses above this threshold; we then attempt to estimate an even higher quantile.

Specifically, our procedure is as follows:

1. Choose a heavy-tailed distribution F and fix the probabilities $0 << q < p < 1$ and the number of data points above the threshold $N_u > 0$.

2. Calculate $u = x_q$; the high threshold is taken to be the qth quantile of F.

3. Calculate x_p the true value of the quantile to be estimated.

4. Sample N_u independent points from F above the threshold u by the rejection method and record the total number of sampled points n that this requires.

5. Estimate ξ and σ by fitting the GPD to the N_u excesses over the threshold u using maximum likelihood.

6. Evaluate $\widehat{x_p}$ the POT quantile estimate.

7. Repeat 500 times the above steps 4 to 6 and estimate bias and root mean square error of the quantile estimate.

The bias of the estimator is defined to be

$$\text{Bias}(\widehat{x_p}) = E\left[\widehat{x_p} - x_p\right],$$

the expected difference between the estimator and the true quantile value; it is estimated in our study by

$$\frac{1}{500}\sum_{i=1}^{500} \widehat{x_{pi}} - x_p$$

where $\widehat{x_{pi}}$ denotes the ith POT quantile estimate from 500 replications. In order to be able to compare estimates for different distributions we then express this bias as a percentage of the true value x_p; we denote the percentage bias by %Bias.

The mean square error of the quantile estimator is defined to be

$$\text{MSE}(\widehat{x_p}) = \left[(\widehat{x_p} - x_p)^2\right],$$

and it can be shown that

$$\mathrm{MSE}(\widehat{x_p}) = \mathrm{var}\,[\widehat{x_p}] + \mathrm{Bias}(\widehat{x_p})^2.$$

Unbiased estimators are often compared using their variances but when there is a possibility of bias it is better to compare mean squared errors which give the dispersion of estimates around the true value. In our study we estimate the root mean square error $\mathrm{RMSE}(\widehat{x_p})$ by

$$\sqrt{\frac{1}{500}\sum_{i=1}^{500}(\widehat{x_{pi}} - x_p)^2}.$$

The root mean square error is on the same scale as the original data and in order to compare quantile estimates across distributions we again calculate the RMSE as a percentage of the true value x_p and we denote the result by %RMSE.

For any given distribution there are three parameters of the method which can be altered: q, p and N_u. That is, we can alter the threshold, the target quantile and the size of the sample of exceedances. In describing our results we will discuss how the accuracy of quantile estimates depends on these parameters.

2.5 Particular Distributions

Of interest to us in the simulation study are the distributions described below.

2.5.1 Pareto

The Pareto is a very commonly used heavy-tailed loss distribution with distribution function

$$F_{\alpha,a}(x) = 1 - (a/x)^{\alpha}$$

for positive parameters a and α and $x \geq a$. By the theorem of Gnedenko it is clearly a member of $\mathrm{MDA}(H_{1/\alpha})$.

2.5.2 Student's t

The Student's t distribution $F_\alpha(x)$ has probability density function

$$f_\alpha(x) \propto (1 + \alpha^{-1}x^2)^{-(\alpha+1)/2}$$

where $\alpha > 0$ is known as the degrees of freedom. The case when $\alpha = 1$ is more usually referred to as the Cauchy distribution.

$F_\alpha(x)$ is in MDA($H_{1/\alpha}$). This is clear by using L'Hopital's rule and observing that

$$\lim_{x\to\infty} \frac{1-F_\alpha(x)}{\alpha^{(\alpha-1)/2}x^{-\alpha}} = \lim_{x\to\infty} \frac{f_\alpha(x)}{\alpha^{(\alpha+1)/2}x^{-\alpha-1}} = \lim_{x\to\infty} \frac{\alpha^{-(\alpha+1)/2}x^{\alpha+1}}{(1+\alpha^{-1}x^2)^{(\alpha+1)/2}}$$

$$= \lim_{x\to\infty} \frac{\alpha^{-(\alpha+1)/2}x^{\alpha+1}}{\alpha^{-(\alpha+1)/2}x^{\alpha+1}(1+\alpha x^{-2})^{(\alpha+1)/2}} = 1.$$

Thus for x large

$$1-F_\alpha(x) \sim cx^{-\alpha}$$

for some constant c so that the assertion follows by an adaptation of Gnedenko's theorem.

2.5.3 Loggamma

The loggamma distribution $F_{\alpha,\beta}$ has probability density function

$$f_{\alpha,\beta}(x) \propto (\log x)^{\beta-1}x^{-\alpha+1}$$

for $x > 1$ where $\alpha > 0$ and $\beta > 0$. X is loggamma distributed for $X > 1$ if and only if $Y = \log X$ is gamma distributed for $Y > 0$. It can be shown, again for instance by use of L'Hopital, that for x large

$$1-F_{\alpha,\beta}(x) \sim cx^{-\alpha}(\log x)^{\beta-1}.$$

The loggamma as parametrized above is in MDA($H_{1/\alpha}$) so that the parameter α dictates the weight of the tail. However β also plays an interesting role in this model; if $\beta = 1$ we simply have a Pareto distribution so that this case is not of interest being already covered above. For other values of β we have essentially a Pareto distribution contaminated by a slowly varying function $(\log x)^{\beta-1}$; the larger the value of β the greater the contamination.

2.5.4 Lognormal

If $\log X$ has the standard normal distribution then X has a standard lognormal distribution. Both distributions are in MDA(H_0) (for proofs see Embrechts et al. (1997)) and technically have medium-sized tails, although the tail of the lognormal is clearly alot longer than that of the normal. Because this distribution is a very popular size-of-loss distribution we also simulate data from it.

2.5.5 Other distributions

Study of the loggamma distribution and the result of Gnedenko gives a clue how we may construct other heavy-tailed distributions which are challenging to the POT method. We may choose particular slowly varying functions $L(x)$ and then simulate data with distribution function $F(x) = 1 - x^{-1/\xi}L(x)$ for positive values of ξ.

We will not go that far in this paper but will limit ourselves to considering the distributions in the table 1 below plus the lognormal. The distributions in the table are all in the maximum domain of attraction of the Fréchet with ξ parameter 1 or 0.5. This translates to extremal indexes of 1 or 2 which correspond to the kind of distributions and tail weights frequently observed in insurance. Note that the insurance data we analysed earlier had tail index around 1.25.

Note also that distributions in MDA($H_{0.5}$) have infinite second and higher moments whilst distributions in MDA(H_1) have no finite moments (not even a finite mean).

For the loggamma we will consider two different values of β: $\beta = 2$ for a light contamination of the Pareto and $\beta = 10$ for a much heavier contamination.

MDA($H_{0.5}$)	MDA(H_1)
Pareto ($\alpha = 2, a = 1$)	Pareto ($\alpha = 1, a = 1$)
Student's t ($\alpha = 2$)	Student's t ($\alpha = 1$) = Cauchy
Loggamma ($\alpha = 2, \beta = 2$)	Loggamma ($\alpha = 1, \beta = 2$)
Loggamma ($\alpha = 2, \beta = 10$)	Loggamma ($\alpha = 1, \beta = 10$)

Table 1: Summary of distributions used in simulation study

3 Simulation Results

We choose to estimate two target quantiles, $x_p = x_{0.99}$ and $x_p = x_{0.999}$; that is, we try to estimate the magnitude of a one in a hundred and a one in a thousand loss.

We investigate datasets of $N_u = 25$, 50, 100 and 200 exceedances. 25 exceedances of a threshold we consider to be the most meagre resources that one would attempt to work with; estimates derived from less data would be too unreliable. 200 exceedances would be an ideal situation and 50 and 100 realistic situations.

			%Bias		%RMSE	
q	N_u	$E[n]$	p=0.99	p=0.999	p=0.99	p=0.999
			Standard Lognormal Distribution			
0.7	25	83	*0.90*	32.56	40.48	161.08
	50	167	*-1.73*	*8.58*	*24.56*	62.02
	100	333	*-0.57*	*7.90*	*16.58*	*43.09*
	200	667	*-0.96*	*5.04*	*11.76*	*26.80*
0.9	25	250	*0.62*	*3.09*	*18.84*	*54.28*
	50	500	*-1.15*	*1.85*	*12.95*	*39.26*
	100	1000	*0.30*	*3.79*	*10.05*	*28.46*
	200	2000	*-0.71*	*0.94*	*6.61*	*16.51*

Table 2: Simulation results for lognormal in MDA(H_0)

Thresholds are set at $x_q = x_{0.7}$ and $x_q = x_{0.9}$. In the former case we would expect to have to generate approximately $n = 83$, 167, 333 and 667 data points from the distribution before we obtained 25,50,100 and 200 exceedances respectively. In the latter case we would expect to have to generate 250,500,1000 and 2000 data points respectively. Thus we get a feel for the total amount of data we require in order that we have a reasonable number of exceedances of only moderately high thresholds.

The results of the simulation study for distributions in MDA($H_{0.5}$) are shown in table 3 and those for MDA(H_1) are in table 4. Results for the lognormal distribution are in table 2.

To help compare tables we have introduced a very arbitrary definition of a good estimate. We consider that a good estimate of $x_{0.99}$ will have proportional bias less than 5% and proportional RMSE less than 30%. For a good estimate of $x_{0.999}$ bias should be less than 10% and RMSE less than 60%. Where one of these criteria is fulfilled the relevant number is italicized in the table so that it can be more quickly seen which combinations of threshold and sample size lead to more accurate quantile estimates.

3.1 Domains of attraction compared

It is clearly seen (by the relative proportion of italicized entries) that estimation of quantiles for distributions in MDA($H_{0.5}$) is much easier than for distributions in MDA(H_1). The heavier the tail the higher we have to set the threshold and the larger our sample sizes should be to obtain the desired accuracy. For the lognormal distribution (see table 2) the GPD quantile method gives very satisfactory results, even with a threshold set at the 70%

			%Bias		%RMSE	
q	N_u	$E[n]$	p=0.99	p=0.999	p=0.99	p=0.999
			Pareto Distribution ($\alpha = 2, a = 1$)			
0.7	25	83	*4.38*	58.28	56.72	281.12
	50	167	*3.39*	28.42	37.69	145.04
	100	333	*1.07*	10.87	*23.36*	63.93
	200	667	*1.55*	*7.18*	*16.45*	*42.09*
0.9	25	250	*2.50*	23.60	*29.17*	132.18
	50	500	*2.58*	19.16	*20.88*	91.93
	100	1000	*0.83*	*4.74*	*13.58*	*44.51*
	200	2000	*0.12*	*1.85*	*9.32*	*26.77*
			Student's t ($\alpha = 2$)			
0.7	25	83	*3.86*	*8.48*	46.25	119.40
	50	167	*2.77*	*-0.99*	34.88	93.54
	100	333	*-0.17*	-12.46	*23.91*	*54.82*
	200	667	*0.19*	-15.06	*16.63*	*33.83*
0.9	25	250	*1.12*	*4.61*	*25.17*	81.24
	50	500	*0.95*	*3.22*	*20.72*	65.16
	100	1000	*0.09*	*-3.55*	*14.37*	*41.42*
	200	2000	*0.62*	*-2.77*	*9.84*	*26.50*
			Loggamma ($\alpha = 2, \beta = 2$)			
0.7	25	83	*4.56*	74.87	67.42	374.99
	50	167	*4.85*	42.54	51.56	442.87
	100	333	*2.70*	17.00	*27.89*	79.14
	200	667	*0.84*	*6.79*	*18.88*	*45.80*
0.9	25	250	*3.23*	27.38	32.80	134.60
	50	500	*1.97*	13.48	*22.30*	73.68
	100	1000	*0.74*	*5.45*	*15.58*	*45.23*
	200	2000	*1.08*	*5.60*	*10.61*	*19.95*
			Loggamma ($\alpha = 2, \beta = 10$)			
0.7	25	83	51.44	675.96	233.10	5345.21
	50	167	16.61	130.81	77.28	355.60
	100	333	7.35	73.14	47.03	187.29
	200	667	*4.86*	52.49	32.43	111.33
0.9	25	250	20.95	110.12	228.63	594.55
	50	500	10.66	51.32	289.09	477.33
	100	1000	*0.50*	28.72	*25.25*	94.15
	200	2000	*-0.70*	17.01	*16.49*	*53.98*

Table 3: Simulation results for distributions in MDA($H_{0.5}$)

			%Bias		%RMSE	
q	N_u	$E[n]$	p=0.99	p=0.999	p=0.99	p=0.999
			Pareto Distribution ($\alpha = 1, a = 1$)			
0.7	25	83	38.27	351.65	179.64	1849.76
	50	167	13.76	92.87	80.48	496.42
	100	333	8.92	41.36	57.52	193.08
	200	667	*4.21*	17.44	33.21	80.93
0.9	25	250	10.52	207.49	83.28	1611.14
	50	500	7.69	51.70	50.25	254.02
	100	1000	*2.10*	15.16	*27.54*	88.32
	200	2000	*0.50*	*8.27*	*19.32*	*55.89*
			Student's t ($\alpha = 1$)			
0.7	25	83	26.13	158.98	149.31	864.71
	50	167	7.31	37.46	76.32	241.56
	100	333	*2.76*	10.79	53.65	161.05
	200	667	*-2.35*	-11.74	30.38	*58.78*
0.9	25	250	14.75	131.52	69.82	495.95
	50	500	*3.98*	31.86	40.16	163.80
	100	1000	*3.16*	14.53	*26.92*	79.44
	200	2000	*0.74*	*6.52*	*20.84*	60.69
			Loggamma ($\alpha = 1, \beta = 2$)			
0.7	25	83	74.04	768.32	383.92	5761.78
	50	167	24.17	189.04	110.95	1030.45
	100	333	11.85	72.18	67.34	271.01
	200	667	5.14	40.26	41.23	133.50
0.9	25	250	15.72	313.97	125.13	2629.67
	50	500	6.60	82.57	56.27	408.71
	100	1000	*4.01*	31.06	33.82	120.87
	200	2000	*-0.02*	14.13	*21.44*	64.58
			Loggamma ($\alpha = 1, \beta = 10$)			
0.7	25	83	371.63	23637.73	1421.20	210663.20
	50	167	186.41	4484.23	650.76	28013.25
	100	333	69.83	732.98	161.78	1711.59
	200	667	49.46	478.71	105.28	894.76
0.9	25	250	50.12	934.70	236.02	5206.99
	50	500	27.28	433.77	122.76	1691.70
	100	1000	10.42	185.10	62.87	548.44
	200	2000	6.22	116.79	43.38	321.13

Table 4: Simulation results for distributions in MDA(H_1)

quantile provided there are at least 50 excesses to work with.

Within a maximum domain of attraction quantile estimates for the ordinary Pareto distribution and Student's t seem to present problems of comparable difficulty to the GPD method. As expected, for the loggamma with $\beta = 2$, estimation is slightly more difficult and when $\beta = 10$ estimation is very much more difficult.

3.2 Distributions in MDA($H_{0.5}$)

For these distributions (see table 3) a threshold at $q_{0.7}$ is sometimes sufficient to obtain the target accuracy. In this case, for the Pareto distribution and for the loggamma with $\beta = 2$, we require around 100 excesses to estimate the 99th percentile to the target accuracy and around 200 excesses to estimate the 99.9th percentile to the target accuracy. For the Student's t we require 100 excesses for $q_{0.99}$ but estimation of $q_{0.999}$ appears to require a higher threshold since a small negative bias arises for large samples, so that some underestimation of the quantile occurs. For the loggamma with $\beta = 10$ we cannot obtain the desired accuracy with a threshold at $q_{0.7}$.

A threshold at $q_{0.9}$ is sufficient to obtain the target accuracy in all cases except for the 99.9th percentile of the loggamma with $\beta = 10$. For the Pareto distribution and the Student's t we can estimate the 99th percentile to the target accuracy with as few as 25 excesses and we can estimate the 99.9th percentile with 100 excesses. For the loggamma with $\beta = 2$ we require 50 excesses for $q_{0.99}$ and 100 for $q_{0.999}$. For the loggamma with $\beta = 10$ we require 100 excesses for $q_{0.99}$ but the threshold must be raised (or possibly more data must be available) to estimate $q_{0.999}$.

3.3 Distributions in MDA(H_1)

For these distributions (see table 4) a threshold at $q_{0.7}$ appears never to be quite high enough to estimate the target quantiles to the desired accuracy.

With a threshold at $q_{0.9}$ we can estimate $q_{0.99}$ for the Pareto with 100 excesses and $q_{0.999}$ with 200 excesses. We can estimate $q_{0.99}$ for the Student's t with 100 excesses and for the loggamma with $\beta = 2$ we require 200 excesses, but for both these distributions the target accuracy was not reached for $q_{0.999}$. In the case of the loggamma with $\beta = 10$ we did not reach the target accuracy for either quantile.

Particularly for the loggamma it seems that higher thresholds are required before we can obtain the desired accuracy with the POT quantile estimation method. This means we need to generate very large datasets

from the loggamma before we have enough data over a high enough threshold to implement the POT method. What this essentially tells us is that, in the case of loggamma with high values of β, the convergence in the theorem of Pickands, Balkema and de Haan is very slow.

Note that the biases in table 4 are mostly positive so that the tendency is mostly to overestimate the quantile. Data from the loggamma tend to point to heavier tails than is actually the case. This means that the high quantile estimates we derive are actually exceeded less often than we calculate. Our error is at least in a conservative direction as far as insurance would be concerned.

4 Conclusion

Estimation of high quantiles (such as $q_{0.99}$ and $q_{0.999}$) of an unknown distribution from datasets of several hundred or several thousand observations is an inherently difficult problem. The POT method is a theoretically well supported technique for fitting a parametric distribution to the tail of an unknown underlying distribution and reading off quantile estimates from the fitted curve. The POT method is to our knowledge the best parametric method available for this estimation problem.

The Achilles heel of the method is the choice of a suitable threshold above which tail behaviour can be said to begin. The simulation experiment has indicated for a number of common loss severity distributions where we should position thresholds and how many excesses we require to obtain comparable accuracy of quantile estimates. When confronted by a dataset we would tend to use exploratory techniques to decide on an appropriate threshold but these techniques seldom point unequivocably to a single choice of treshold. The simulation study provides us with supplementary information on the kind of errors which may arise for known distributions.

The most serious challenge to the method was estimating high quantiles of the loggamma distribution with tail index 1 and β parameter 10. It may however be argued that this is an unrealistically difficult case and a β parameter this high would not occur in practice in insurance.

Acknowledgements

The first author is Swiss Re Research Fellow at ETH Zurich and gratefully acknowledges Swiss Re for their financial support and many fruitful discussions.

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

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 Verlag, Berlin. 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.

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.

McNeil, A. (1997), 'Estimating the tails of loss severity distributions using extreme value theory', *ASTIN Bulleitin* . To appear.

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.