

Developing Scenarios for Future Extreme Losses Using the POT Model

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

July 6, 1998

Abstract

We use the peaks over thresholds or POT method to derive a natural model for the point process of large losses exceeding a high threshold. This model is used to obtain a joint description of the frequency and the severity with which large losses occur. The issue of trend in both frequency and severity is considered.

We show how the model may be fitted to insurance data on natural catastrophe losses and how it may serve as a basis to examine the future burden of large losses under various scenarios.

Keywords: Extreme value theory; peaks over thresholds method; point processes; homogeneous and inhomogeneous Poisson processes; modelling trends; scenarios.

1 Introduction

In previous work we have looked at the use of extreme value theory (EVT) to model the tails of loss severity distributions (McNeil 1997, McNeil & Saladin 1997). We used the peaks over thresholds (POT) method to obtain the generalized Pareto distribution as an approximation to the distribution of excess amounts over high thresholds.

In this paper we look at ways of extending the POT model to obtain a joint description of both the severity and frequency of losses exceeding high thresholds. We use the point process approach to extreme values to obtain our joint description in the form of a two-dimensional Poisson process. The underlying theory can be found in Leadbetter, Lindgren & Rootzén (1983)

and Embrechts, Klüppelberg & Mikosch (1997) and other examples of its application in an insurance context are Rootzén & Tajvidi (1997) and Smith (1996).

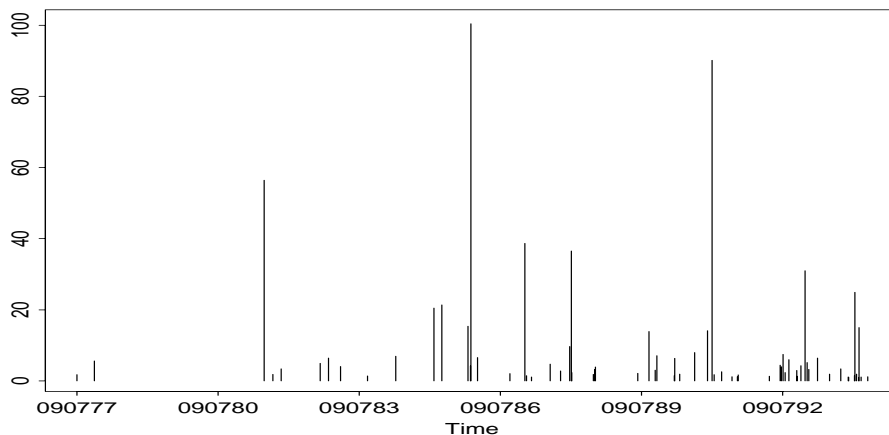


Figure 1: Time Series Plot of all Losses

We apply the method to a small dataset consisting of 67 inflation-adjusted natural catastrophe losses in a Scandinavian country, from the portfolio of a major European insurance company. These losses took place over a period of 17 years and all exceeded a basic inflation-adjusted threshold of one million in the local currency unit. The majority (43) were caused by wind storms; the second largest group (21) are losses due to floods. The process of large losses is shown in Figure 1 and the data are summarized by year in Table 1.

It is very noticeable in these data that there is a trend of increasing loss frequency over the years. This observation motivates us to look at models for the situation where the process of large claims may have a trend. Large losses may become more or less frequent over time and they may also become more or less severe. Both these phenomena can be accommodated in the Poisson process framework by considering processes with time-dependent intensity measures. Such models are also investigated in Smith (1996).

We are particularly interested in finding an effective model for large losses in the past so that we can extrapolate the claims process into the future. We consider the development of stress scenarios based on our fitted models. Under these scenarios loss frequency and loss severity are allowed to develop adversely and we calculate the loss volumes which could typically result.

Year	No. of losses	Median Loss	Total Loss Volume
1977	2	3.71	7.41
1978	0	0	0
1979	0	0	0
1980	0	0	0
1981	3	3.42	61.75
1982	2	5.71	11.41
1983	2	2.79	5.57
1984	1	6.98	6.98
1985	5	20.5	161.96
1986	2	4.33	8.66
1987	6	3.82	58.62
1988	5	3.29	48.06
1989	4	5.11	26.32
1990	5	6.41	32.03
1991	6	1.78	98.88
1992	10	4.17	65.32
1993	7	3.27	22.48
1994*	7	1.50	46.86

Table 1: Yearly loss frequency, median loss and total loss volume in a common inflation-adjusted unit of currency. *Note that 1994 losses are only up to the end of April.

2 Theoretical Foundations of POT Model

2.1 Extremal Types Theorem

Let $X_1, X_2, \dots, X_n, \dots$ be a sequence of independent, identically distributed (iid) insurance losses with unknown severity distribution F . Suppose furthermore that these losses occur at the times $T_1 < T_2 < \dots < T_n < \dots$.

Let $M_n = \max(X_1, \dots, X_n)$ be the maximum of the first n losses and suppose that M_n shows regular limiting behaviour in the sense that we can find sequences of real normalizing constants b_n and $a_n > 0$ such that

$$\lim_{n \rightarrow \infty} P \{ (M_n - b_n) / a_n \leq x \} = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = H(x) \quad (1)$$

where $H(x)$ is a non-degenerate distribution function. The extremal types theorem (Fisher & Tippett 1928) tells us that $H(x)$ must be an extreme value distribution. We can find particular choices of the normalizing sequences b_n and a_n so that $H(x)$ has the von Mises parametrization of the generalized

extreme value distribution (GEV)

$$H_\xi(x) = \exp(-V(x)) = \begin{cases} \exp(-(1 + \xi x)^{-1/\xi}) & \text{if } \xi \neq 0, \\ \exp(-e^{-x}) & \text{if } \xi = 0, \end{cases} \quad (2)$$

where $1 + \xi x > 0$ and the case $\xi = 0$ should be viewed as the limit of the distribution function as $\xi \rightarrow 0$. The three cases $\xi < 0$, $\xi = 0$ and $\xi > 0$ correspond to the Weibull, Gumbel and Fréchet extreme value distributions respectively.

Distributions F satisfying condition (1) are said to be in the maximum domain of attraction of an extreme value distribution, written $F \in \text{MDA}(H_\xi)$. Essentially all the common, continuous distributions of statistics and actuarial science are in $\text{MDA}(H_\xi)$ for some value of ξ . The normal, exponential and lognormal distributions are examples of distributions in $\text{MDA}(H_0)$ (the Gumbel case). We call such distributions thin-tailed.

Distributions in $\text{MDA}(H_\xi)$ for $\xi > 0$ (the Fréchet case) we call heavy-tailed. This class includes, amongst others, the Pareto, the Student-t, the Cauchy, the Burr and the log-gamma distributions, all of which are popular models for large losses in insurance. It was shown by Gnedenko (1943) 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 $L(x)$, so that distributions giving rise to this case can be characterized as those whose tails decay essentially like a power function. For a distribution in $\text{MDA}(H_\xi)$ with $\xi > 0$, the parameter $\alpha = 1/\xi$ is known as the tail index of the distribution F .

2.2 Poisson Point Process Approach

It can be shown that condition (1) is also equivalent to saying that there exist normalizing sequences $a_n > 0$ and b_n such that

$$\lim_{n \rightarrow \infty} n(1 - F(a_n x + b_n)) = -\log H_\xi(x). \quad (3)$$

The expression $n(1 - F(a_n x + b_n))$ can be interpreted as the number of normalized observations from the sequence $(X_1 - b_n)/a_n, \dots, (X_n - b_n)/a_n$ which we expect to exceed the threshold x . The actual number of normalized losses exceeding the threshold is a binomial random variable $B(n, 1 - F(a_n x + b_n))$, which, for n large and $1 - F(a_n x + b_n)$ small, can be approximated by a Poisson random variable with mean $n(1 - F(a_n x + b_n))$. This gives us some intuition for the point process formulation.

A point process is constructed in $[0, 1] \times \mathbb{R}$ by considering the points

$$\{(i/n, (X_i - b_n)/a_n), 1 \leq i \leq n\}. \quad (4)$$

Reinterpreting condition (3) we can say that the expected number of points in the area $[0, 1] \times (x, \infty)$ converges to $-\log H_\xi(x)$. Note that, we could also

describe this two-dimensional point process as a marked point process, i/n being the location of the point and $(X_i - b_n)/a_n$ the size of the mark.

It is possible to establish formally that the point process converges weakly, on sets which exclude the lower boundary, to a two-dimensional Poisson process. Consider such a set

$$A = \{(t_1, t_2) \times (x, \infty)\} \subset [0, 1] \times \mathbb{R},$$

with $t_1 < t_2$; the intensity measure Λ of the limiting two-dimensional Poisson process satisfies

$$\begin{aligned} \Lambda(A) &= -(t_2 - t_1) \log H_\xi(x) \\ &= (t_2 - t_1) (1 + \xi x)_+^{-1/\xi}, \end{aligned} \tag{5}$$

where $\Lambda(A)$ is the expected number of points in the set A and the notation $(\cdot)_+$ ensures that $1 + \xi x > 0$ as in (2). Establishing these point process results is mathematically demanding and the interested reader is referred to accounts in Leadbetter et al. (1983), Falk, Hüsler & Reiss (1994) and Embrechts et al. (1997).

2.3 Exploiting this Result for Modelling

Suppose we consider the finite set of losses X_1, \dots, X_N . In general we do not know the appropriate normalizing sequences a_n and b_n , since we do not know the underlying distribution. It is usual practice to consider modelling all losses which exceed a high threshold u with a two-dimensional Poisson process in $[0, 1] \times (u, \infty)$. That is to say, we now put points at $\{(i/N, X_i), 1 \leq i \leq N, X_i > u\}$ and we use the intensity measure

$$\Lambda\{(t_1, t_2) \times (x, \infty)\} = -(t_2 - t_1) \log H_\xi((x - \mu)/\sigma) \tag{6}$$

for $t_1 < t_2$ and $x > u$, where μ and $\sigma > 0$ are location and scaling parameters. These additional parameters take care of the unknown normalizing sequences. The intensity measure can also be expressed as the integral of a constant intensity function along the time axis

$$\Lambda\{(t_1, t_2) \times (x, \infty)\} = (t_2 - t_1)V(x; \xi, \sigma, \mu), \tag{7}$$

where $V(x; \xi, \sigma, \mu) = -\log H_\xi((x - \mu)/\sigma)$ is the constant intensity function, or the integral of an intensity density in the plane

$$\Lambda\{(t_1, t_2) \times (x, \infty)\} = (t_2 - t_1) \int_x^\infty v(w; \xi, \sigma, \mu) dw \tag{8}$$

where $v(x; \xi, \sigma, \mu) = -dV(x; \xi, \sigma, \mu)/dx$ is the intensity density.

Instead of locating points at the artificial times i/N , if we believe that exceedances of the threshold may occur as a homogeneous Poisson process

in time, it is more natural to use the true times of occurrence of the losses T_i . Suppose our losses occur in a time window $[-T, 0]$ so that time 0 marks the point where the data come to an end and simulation or prediction will begin. In a further step we can create the two-dimensional point process $\{(T_i, X_i), 1 \leq i \leq N, X_i > u\}$.

Under a variety of assumptions it may be a plausible assumption that exceedances of the threshold u occur as a homogeneous Poisson process. For example, if the losses themselves form a Poisson process then the exceedances form a thinned Poisson process which is also Poisson. The assumption of homogeneous Poisson is one we can check and, if it is violated, we can consider non-homogeneous Poisson processes, as will be described later.

So we assume that the point process $\{(T_i, X_i), 1 \leq i \leq N, X_i > u\}$ in $[-T, 0] \times (u, \infty)$ can be modelled by a two-dimensional Poisson process having intensity measure as in (6) but with $t_1, t_2 \in [-T, 0]$, rather than $[0, 1]$. We shall call this model the peaks-over-thresholds or POT model.

2.4 Equivalence to Standard Modelling Procedures

First we note that the GEV may be derived from this model. Consider again the distribution of M_N , the maximum of our N losses. If $M_N \leq x$ (with $x > u$) then there can be no points in the set $[-T, 0] \times (x, \infty)$. The probability of this event is calculated under the POT model using the intensity measure in (6) to be

$$F^N(x) = \exp \left\{ -T(1 + \xi(x - \mu)/\sigma)_+^{-1/\xi} \right\},$$

for $x \geq u$. After rescaling and relocation this is an extreme value distribution as given in (2).

Now consider the excess amounts over the threshold u . The conditional probability that an observation exceeds this threshold by at most an amount x , given that it does indeed exceed the threshold is the so-called excess distribution function

$$F_u(x) = P \{X - u \leq x \mid X > u\} = \frac{F(x + u) - F(u)}{1 - F(u)}, \quad (9)$$

where we use X to denote a generic X_i . It is easier to first consider $1 - F_u(x)$; under the POT model this is equal to the ratio of the intensities for points above the thresholds $u + x$ and u which is given by

$$\frac{V(u + x; \xi, \sigma, \mu)}{V(u; \xi, \sigma, \mu)} = \left(1 + \frac{\xi x}{\sigma + \xi(u - \mu)} \right)^{-1/\xi},$$

where $V(x; \xi, \sigma, \mu)$ is as defined in (7). This means that under the POT model $F_u(x)$ is the df of the generalized Pareto distribution

$$G_{\xi, \phi}(x) = 1 - (1 + \xi x/\phi)^{-1/\xi}, \quad (10)$$

for a positive scaling parameter $\phi = \sigma + \xi(u - \mu)$. The support of this distribution is $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\phi/\xi$ when $\xi < 0$.

This distribution is commonly fitted to data on excess amounts over high thresholds. Hosking & Wallis (1987) show how this is achieved with maximum likelihood and Davison & Smith (1990) provide examples of extreme value analyses based on the generalized Pareto distribution (GPD). Our previous work on large losses in insurance has revolved around this distribution (McNeil 1997, McNeil & Saladin 1997).

In this paper we prefer to fit the full Poisson point process model (6) to data, a technique described in Smith (1989). This approach gives us our joint modelling of frequency and severity of large losses, whereas the GPD approach restricts attention to severity. From the statistical point of view, the POT model has the advantage over the simplified GPD approach that, if the model is a good description of the data, parameter estimates are not dependent on the threshold and should be fairly constant over a range of high thresholds.

The likelihood function for the POT model has the form

$$L(\xi, \sigma, \mu) = \exp\{-TV(u; \xi, \sigma, \mu)\} \cdot \prod_{i=1}^N v(X_i; \xi, \sigma, \mu)$$

where $V(x; \xi, \sigma, \mu)$ and $v(x; \xi, \sigma, \mu)$ are as defined in (7) and (8).

Parameter estimates of ξ , σ and μ are obtained by maximizing this expression. This presents a regular likelihood problem when $\xi > -0.5$. To calculate standard errors for parameter estimates we use the standard method based on the Fisher information matrix.

2.5 Extending POT to Model Trend

Suppose we replace the parameters μ , σ and ξ by $\mu(t, \Theta)$, $\sigma = \sigma(t, \Theta)$ and $\xi = \xi(t, \Theta)$ so that the Poisson intensity parameters are now functions of the time t and a set of model parameters Θ . Contrasting with (7) and (8) we now have a time-dependent intensity measure of the form

$$\begin{aligned} \Lambda\{(t_1, t_2) \times (x, \infty)\} &= \int_{t_1}^{t_2} V(x; \xi(t, \Theta), \sigma(t, \Theta), \mu(t, \Theta)) dt \\ &= \int_{t_1}^{t_2} \int_x^\infty v(w; \xi(t, \Theta), \sigma(t, \Theta), \mu(t, \Theta)) dw dt \end{aligned} \quad (11)$$

where both

$$V(x; \xi(t, \Theta), \sigma(t, \Theta), \mu(t, \Theta)) = -\log H_{\xi(t, \Theta)} \left(\frac{x - \mu(t, \Theta)}{\sigma(t, \Theta)} \right)$$

and

$$v(x; \xi(t, \Theta), \sigma(t, \Theta), \mu(t, \Theta)) = -dV(x; \xi(t, \Theta), \sigma(t, \Theta), \mu(t, \Theta))/dx$$

are now time-dependent. If we think in terms of our two-dimensional point process model, introducing time-dependency in this way has an effect on the intensity of occurrence of the points in both the temporal and size dimensions. If we think of the occurrence of the points in time, we are moving from a (one-dimensional) homogeneous Poisson process to a non-homogeneous Poisson process. At the same time we are introducing time-dependence into our modelling of the sizes of the losses.

The likelihood implied by the intensity measure in (12) now has the form

$$L(\Theta) = \exp \left\{ - \int_{-T}^0 V(u; \xi(s, \Theta), \sigma(s, \Theta), \mu(s, \Theta)) ds \right\} \cdot \prod_{i=1}^N v(X_i; \xi(T_i, \Theta), \sigma(T_i, \Theta), \mu(T_i, \Theta)). \quad (12)$$

3 A POT Analysis of the Data

To use the POT model approach we are first required to choose a threshold value u . Choosing the threshold is perhaps the greatest practical problem which besets the method and we have written in more detail on this question in McNeil (1997). One particular tool to help us here 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 N th 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 to be $e(u) = E[X - u \mid X > u]$. 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 if the excesses have a GPD with shape and scale parameters ξ and ϕ then as long as $\xi < 1$

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

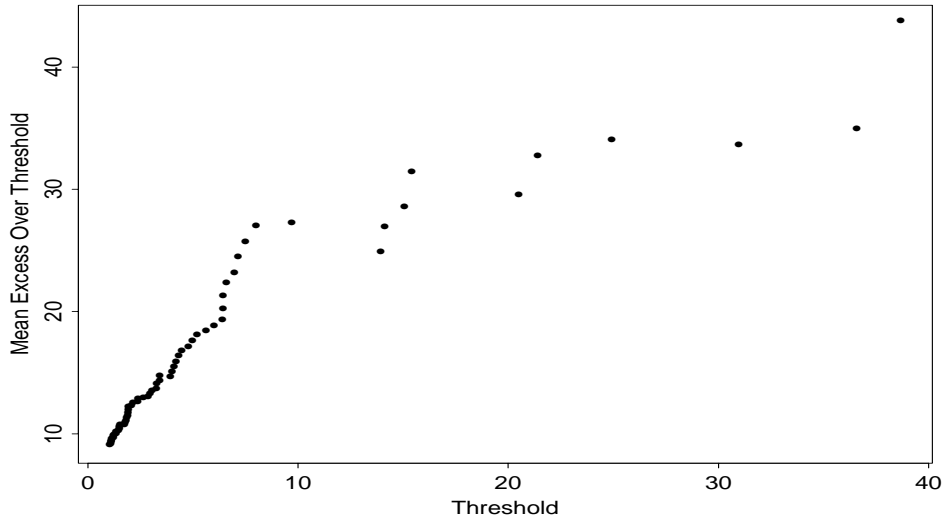


Figure 2: Mean Excess Plot of all Losses

where $\phi + u\xi > 0$ determines the range of u .

For our 67 natural catastrophe losses the sample mean excess plot (Figure 2) shows a pattern which is not obviously inconsistent with the GPD assumption, even using a threshold as low as one which gives 67 excesses. The plot shows a straight line to begin which then bends somewhat; the final 11 points seem to show a different gradient but it should be remembered that these are averages of only small numbers of excesses which provide inaccurate estimates of the mean excess function. A sample mean excess plot of 67 simulated losses from the GPD distribution could easily give a similar result (Embrechts et al. 1997). If we accept the GPD hypothesis, this in turn provides some support for the POT model above a threshold of one, since the GPD distribution for excesses flows necessarily from the POT model. Of course, we also need to check other aspects of the fit, including whether or not the process of exceedances of this level is plausibly a homogeneous Poisson process.

The huge and rather mysterious gap between the 23rd November 1977 and the first of July 1981 when no natural catastrophe losses were recorded argues against the Poisson assumption for all losses. However if the first two losses are removed from the dataset and the analysis is restricted to the 65 losses occurring after the first of July 1981 the assumption becomes more plausible although, as we shall see, still open to serious objections. Henceforth our full dataset will consist of these 65 losses.

The first model we fitted was the standard POT model using a threshold value of one. This led to parameter estimates for (ξ, σ, μ) of $(0.92, 0.049,$

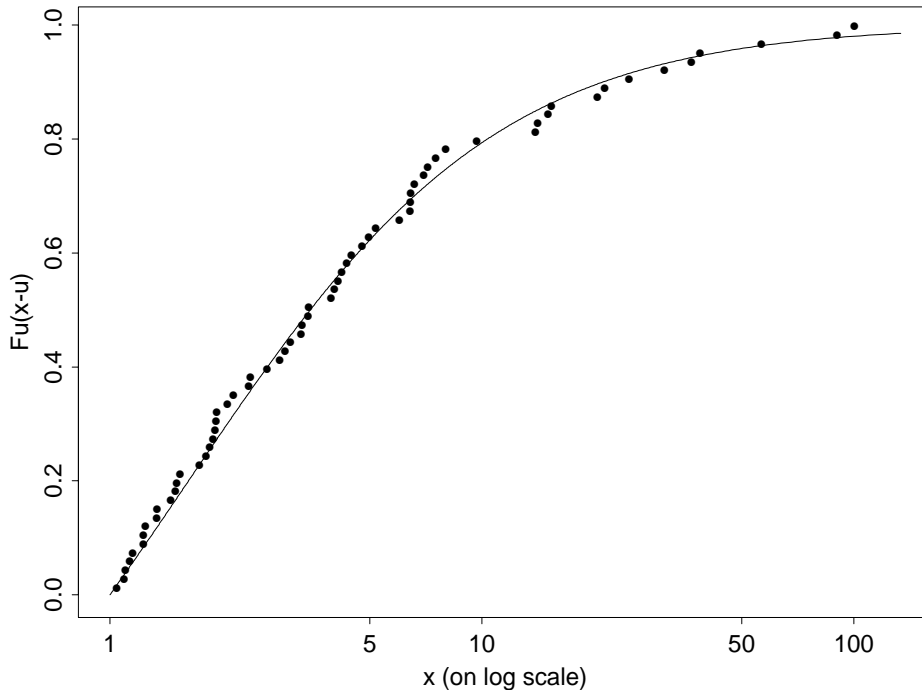


Figure 3: Implied GPD fit from the POT model using a threshold level of 1. The POT parameter estimates and standard errors are $\xi = 0.92$ (0.25), $\sigma = 0.049$ (0.060) and $\mu = -1.69$ (1.18). The implied GPD scaling parameter was $\phi = 2.53$ (0.64).

-1.69) and standard errors of (0.25, 0.060, 1.18). The large ξ value tells us we are clearly in the domain of large losses from a very heavy-tailed severity distribution. These parameter estimates imply a GPD distribution for the excess losses over one with estimated shape parameter $\hat{\xi} = 0.92$ (0.25) and estimated scaling parameter $\hat{\phi} = 2.53$ (0.64).

In Figure 3 we show the implied GPD fit to the excess amounts over the threshold. The losses appear to be well modelled but this can be checked more carefully using residuals proposed by Davison (1984). On the assumption that the excess amounts $(X_i - u)$ are iid from the GPD with parameters given by (ξ, σ, μ) crude residuals in the sense of Cox & Snell (1968) can be defined to be

$$W_i = \frac{1}{\xi} \log \left(1 - \xi \frac{X_i - u}{\sigma + \xi(u - \mu)} \right). \quad (13)$$

These should be iid unit exponentially distributed and this hypothesis can be checked using graphical diagnostics. In the upper right panel of Figure 4 we

have plotted these residuals against order of occurrence and superimposed a smooth curve fit. There seems to be a trend whereby the residuals become smaller than unit exponential towards the end of the series. This would indicate that the losses themselves are becoming smaller on average. On the other hand, the QQplot in the lower right panel shows no evidence against the exponential distributional assumption, although this diagnostic does not examine the ordering of the residuals. It would appear we can model the excesses as GPD but that the assumption of identical distribution may not be correct.

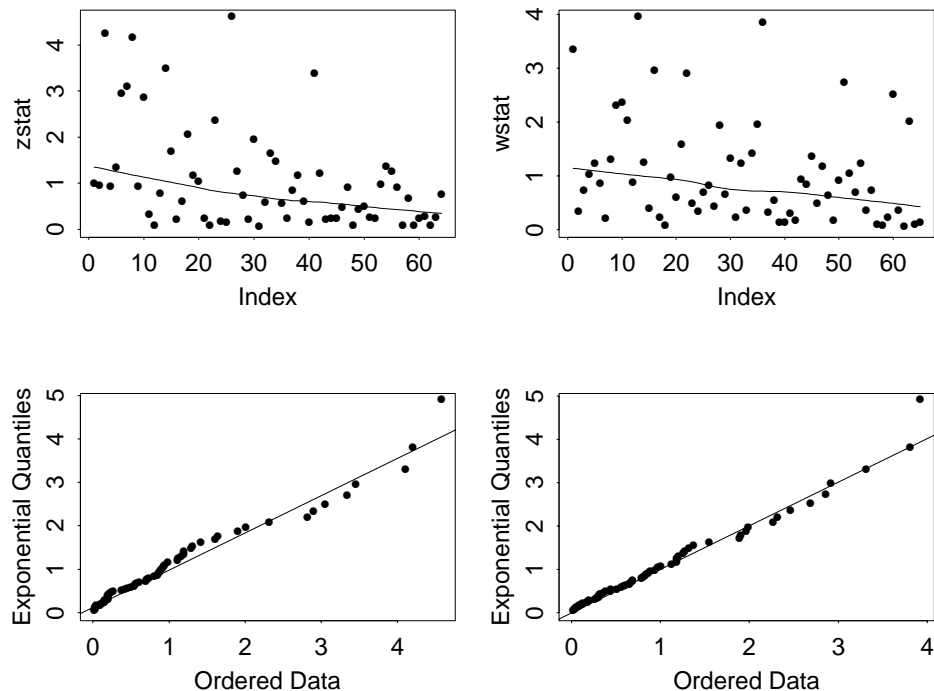


Figure 4: Four diagnostic residual plots for the adequacy of the basic POT model; these plots are described in some detail in the text.

The next aspect of the model to examine is the Poisson assumption for the occurrence times. If the threshold exceedances really occur as a homogeneous Poisson process with a constant intensity which we shall write as

$$\lambda = \left(1 + \xi \frac{u - \mu}{\sigma}\right)^{-1/\xi}$$

then the scaled inter-exceedance times given by

$$Z_k = \lambda \times (T_k - T_{k-1}), \quad k = 1, \dots, N + 1,$$

where $T_0 = -T$ and $T_{N+1} = 0$ should also be unit exponentially distributed. The left-hand plots of Figure 4 are diagnostics for these residuals. The upper left plot is a scatterplot of the Z -values in order of occurrence. A smooth fit has been superimposed to reveal a trend of decreasing Z values; in other words, our initial observation that losses are becoming more frequent is supported by this plot.

The lower left plot is a QQplot of these gaps against the exponential distribution; although some departure from linearity is observed, this is not obviously inconsistent with the exponential assumption. Once again, although the assumption of exponential inter-exceedance times may be tenable, the assumption of identical distribution seems more problematic.

In summary we can say that there is evidence against the POT model we have fitted, both from the point of view of modelling the temporal occurrence of losses and modelling the size of losses. To explain our data we need to incorporate time-dependence into the intensity measure of the two-dimensional Poisson model we are fitting; in other words, we need to consider trend.

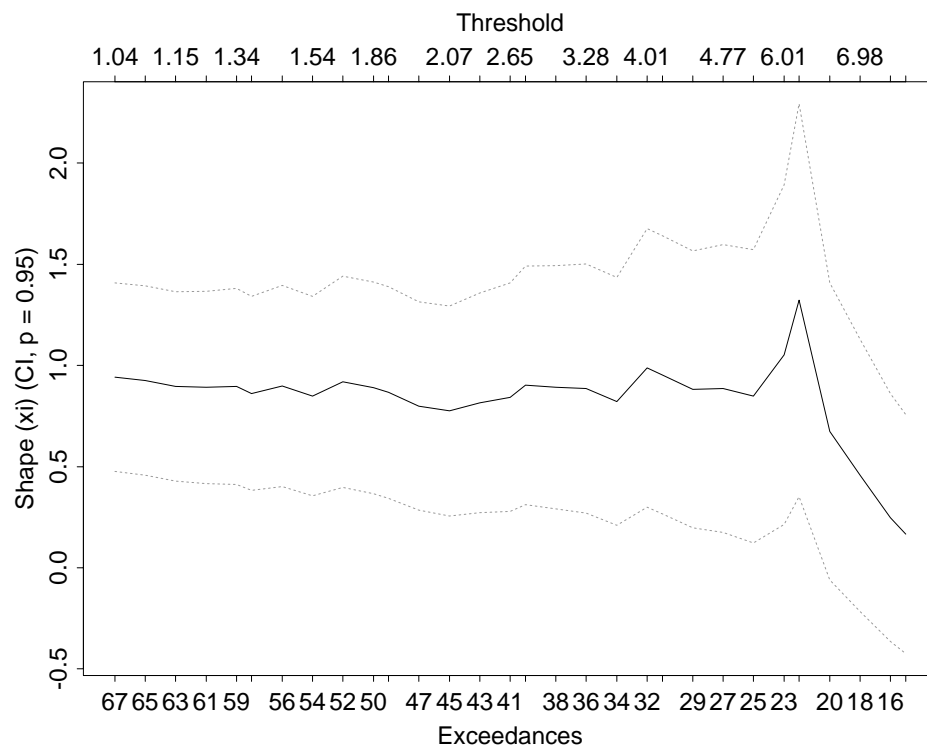


Figure 5: Shape Plot

Further exploratory analyses gave further confirmation of the findings in the residual plots. A linear regression model of the logarithm of the inter-

exceedance times against their order of occurrence was highly significant (p-value of 0.0015). A linear regression model of the log of the loss sizes against time was also significant (p-value of 0.033).

We have also analysed the data with a variety of thresholds greater than one. The trend of increasingly frequent threshold exceedances persists at higher levels; the trend of decreasing average loss sizes becomes slightly weaker. However, when we raise the threshold we quickly have very few data to work with and analysis becomes unsatisfactory. We prefer to continue with our threshold at 1 and to model the trend features we have discovered explicitly, rather than to try and remove trend by increasing the threshold.

We note that estimation of the parameter ξ , the parameter determining the weight of the tail of the severity distribution, seems relatively robust to changes in the threshold. In Figure 5 we see that estimates of ξ are fairly stable around 0.95 for a variety of different thresholds. In the next section we will investigate the stability of ξ and the other model parameters μ and σ over time.

4 Modelling Trend

Let the full set of model parameters be $\Theta = \{\alpha_1, \beta_1, \alpha_2, \beta_2, \alpha_3, \beta_3\}$. In the remaining sections of the paper we will drop explicit mention of Θ in our notation. We consider the following basic trends models.

1. $\mu(t) = \alpha_1 + \beta_1 t$ linear growth in μ
2. $\sigma(t) = \exp(\alpha_2 + \beta_2 t)$ exponential growth in σ
3. $\xi(t) = \alpha_3 + \beta_3 t$ linear growth in ξ

These are the simplest trend models we can entertain. We model the location and shape parameters as linear functions of time and, since $\sigma > 0$, we model the log of the scaling parameter as a linear function of time.

We also consider all possible combinations of these models where, for example, Model 12 denotes a combination of Model 1 and Model 2, and Model 123 denotes a combination of all three trends. Model 0 will denote the basic model of the previous section; we now notate this model using the regression parameters Θ with β_1, β_2 and β_3 set to be zero. Parameter estimates for all eight possible models are given in Table 2 and Table 3. The log likelihoods for the fitted models are given and the deviance statistic for a comparison with the basic Model 0 using the asymptotic chi-squared test.

Models 1,2 and 3 all offered a significant improvement in fit with respect to the basic model. Model 1 provided the best fit among the single trend models, followed by Model 2. It seems that a trend for the location parameter is more effective than a trend for the scaling or shape parameters. Of the models with two trends, Models 12 was the best followed by Model 13. However, neither these models nor Model 123 provided a significant improvement on Model 1. We conclude that Model 1 is the best parsimonious

	Model 0	Model 1	Model 2	Model 3
$\widehat{\alpha}_3$	0.922 (0.247)	0.983 (0.239)	0.948 (0.245)	1.154 (0.294)
$\widehat{\beta}_3 \times 1000$				0.077 (0.031)
$\widehat{\alpha}_2$	-3.01 (1.22)	-3.22 (1.14)	-2.44 (1.04)	-3.34 (1.24)
$\widehat{\beta}_2 \times 1000$			0.349 (0.128)	
$\widehat{\alpha}_1$	-1.69 (1.18)	0.122 (0.629)	-1.58 (1.10)	-1.41 (1.04)
$\widehat{\beta}_1 \times 1000$		1.10 (0.361)		
$-\ln L$	528.34	516.70	521.07	523.87
Deviance		23.3	14.5	8.9

Table 2: Trend models compared

description of the data. In the remainder of this paper we will work with Models 1, 2 and 12 only.

	Model 12	Model 13	Model 23	Model 123
$\widehat{\alpha}_3$	0.905 (0.240)	0.899 (0.272)	0.561 (0.215)	1.134 (0.351)
$\widehat{\beta}_3 \times 1000$		-0.019 (0.034)	-0.200 (0.096)	0.109 (0.102)
$\widehat{\alpha}_2$	-3.08 (1.11)	-2.99 (1.18)	-1.21 (0.86)	-3.96 (1.46)
$\widehat{\beta}_2 \times 1000$	-0.129 (0.144)		1.012 (0.400)	-0.570 (0.397)
$\widehat{\alpha}_1$	0.227 (0.577)	0.113 (0.637)	-1.82 (1.20)	0.569 (0.424)
$\widehat{\beta}_1 \times 1000$	1.44 (0.62)	1.25 (0.49)		1.72 (0.68)
$-\ln L$	516.4	516.6	518.3	515.9
Deviance	24.0	23.6	20.1	24.9

Table 3: Trend models compared

We further examined the fit of Models 1 and 2 using ideas developed in Smith & Shively (1995). They propose residual statistics which are really simple generalizations of the residuals from the previous section. Suppose the intensity of the exceedance process over a threshold u is as implied by the likelihood in (13). We rewrite this intensity as a simple function of t

and drop the Θ notation

$$\lambda(t) = \left(1 + \xi(t) \frac{u - \mu(t)}{\sigma(t)} \right)_+^{-\frac{1}{\xi(t)}}.$$

Adopting terminology used by (Smith 1996), we define the Z-Statistics to be

$$Z_k = \int_{T_{k-1}}^{T_k} \lambda(s) ds, \quad k = 1, \dots, N + 1,$$

where T_k is the time of the k th exceedance and T_0 and T_{N+1} are as before. Assuming the model to be correct, the Z_i are again iid from the exponential distribution with mean one. They can thus provide a guide to our success in modelling the temporal behaviour of the exceedances of the threshold u .

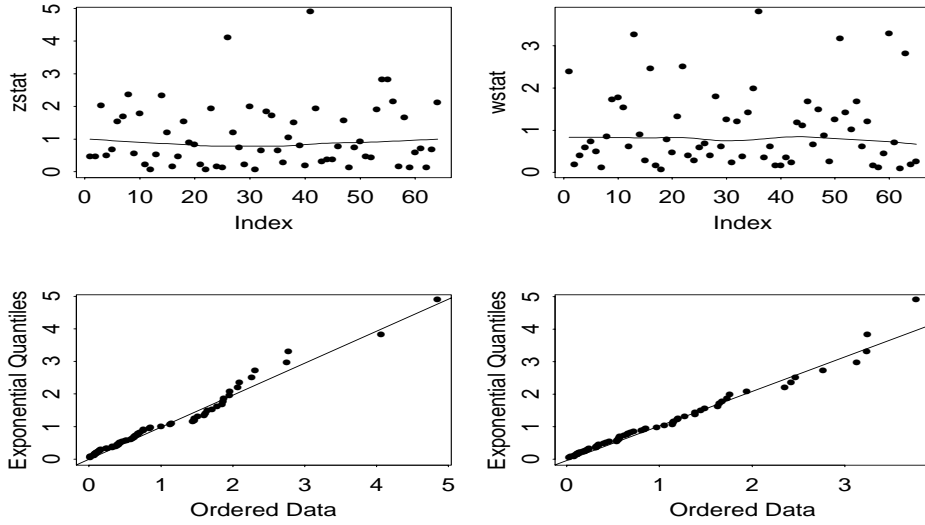


Figure 6: Diagnostic plots for the Z- and W-statistics applied to Model 1.

The W -Statistics are defined to be

$$W_k = \frac{1}{\xi(T_k)} \ln \left(1 + \xi(T_k) \frac{X_k - u}{\sigma(T_k) + \xi(T_k)(u - \mu(T_k))} \right).$$

If the implied GPD model for the excesses is correct then the W_k should also be iid exponentially distributed.

We can again use diagnostic plots to examine the Z - and W -Statistics. Scatterplots of Z_k and W_k against T_k can be used to search for systematic deviation from unit exponential, which might indicate the persistence of trends which are not being adequately modelled. Q-QPlots of Z_k and W_k

values against the exponential distribution can be used to check the distributional assumption. We show these plots for Models 1 and 2 in Figures 6 and 7 respectively. Model 1 gives very satisfactory results; the unit exponential hypothesis is plausible for both residuals. Model 2 seems to model the temporal behaviour well, but the W–statistics suggest that a trend in size of losses is not being adequately modelled. This cannot be avoided in Model 2, as will be seen.

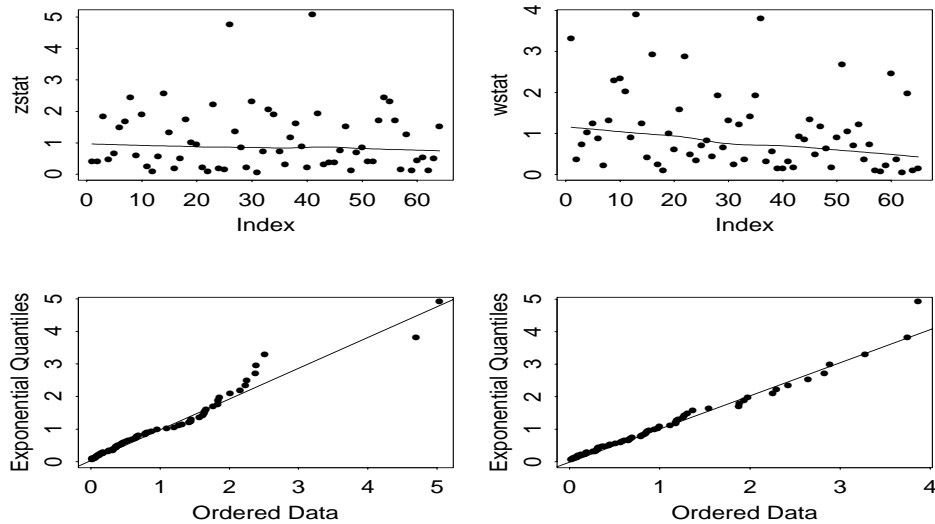


Figure 7: Diagnostic plots for the Z– and W–statistics applied to Model 2.

All three of our selected models (1,2 and 12) allow us to capture the trend of increasing intensity of large losses which seems to be revealed by the data. However it is interesting to consider what the exact consequences of our models are. For all three models the estimated intensity increases, although the nature of the increase is different (see Figure 8). Models 1 and 12 give a very dramatic increase in intensity, whereas Model 2 shows a gentler increase.

We also consider how the average loss size varies with time for the three fitted models. Because all ξ estimates are close to 1 and the mean of the implied GPD distribution is close to being undefined we take the median loss as our guide to average loss size. The median loss at time t is then one plus the median of a GPD distribution with shape parameter ξ and scaling parameter $\phi(t) = \sigma(t) + \xi(u - \mu(t))$. The three models behave very differently as a function of time. Models 1 and 12 model a decrease in average loss whereas Model 2 yields a very slight increase. Thus Models 1 and 12 capture the trend of decreasing loss sizes whereas Model 2 does not.

Note that Model 1, if it is to model an increasing intensity of losses

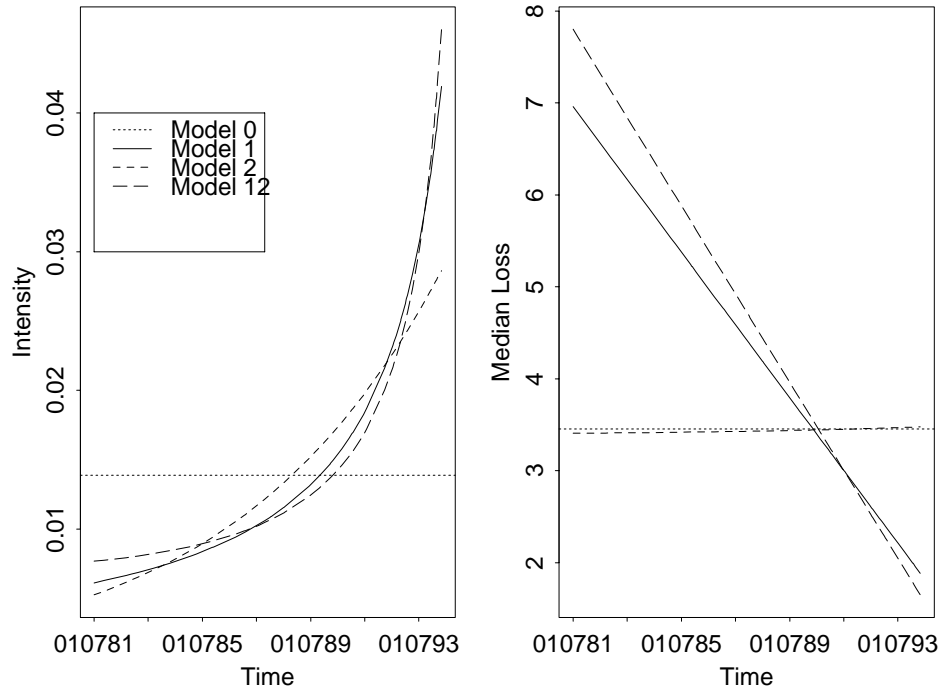


Figure 8: Implications of the four fitted models for changing intensity of the exceedances and changing median loss size.

($\beta_1 > 0$) must also model a decreasing trend in size of losses. Model 2 if it is to give an increasing intensity of losses ($\beta_2 > 0$), cannot at the same time give a decreasing scaling of losses. Perhaps for this reason Model 2 appears inferior to Model 1 in likelihood comparisons. Model 12 with its additional parameter, possesses sufficient flexibility to model all possible combinations of increasing and decreasing intensities and loss sizes.

5 Developing Scenarios

Our fitted models are the starting points for developing scenarios. We are interested in examining how the losses which exceed the threshold might develop in the next three years beginning in May 1994 when the data come to an end.

Our scenarios all have the following form. For each future year $k = 1, 2, 3$ we consider the time interval $[365(k - 1), 365k]$ and we define N_k to be the number of losses in this period. We suppose that N_k has a Poisson

distribution with mean

$$\int_{365(k-1)}^{365k} \lambda(t) dt,$$

so that $\lambda(t)$ is the (possibly) time-dependent intensity of a Poisson process. The parameters of this process will be chosen with regard to the fitted models.

The N_k new losses in the k th year are $Y_{k,1}, \dots, Y_{k,N_k}$ and will be such that the excesses $Y_{k,i} - u$ have a GPD distribution with parameters ξ and ϕ where these parameters are again chosen with reference to the fitted models. Thus the loss process will be a compound non-homogeneous Poisson process giving rise to a total loss volume in year k of

$$Z_k = \sum_{i=1}^{N_k} Y_{k,i}.$$

5.1 Frequency Scenarios

We consider the following scenarios for loss frequency. Note that we do not consider Model 1 itself; although Model 1 provides the best description of past losses, it quickly leads to unrealistic results when extrapolated into the future. By the third year of simulation the intensity of Model 1 is greater than one so that natural catastrophe losses occur every day. We do consider Model 2 which gives a gentler development of future loss numbers.

- Scenario A: Future losses occur according to the basic POT model Model 0. They form a homogeneous Poisson process with intensity

$$\lambda = (1 + \hat{\alpha}_3(1 - \hat{\alpha}_1) \exp(-\hat{\alpha}_2))^{-\hat{\alpha}_3^{-1}} = 0.014,$$

where $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ are the parameter estimates in Model 0, as summarized in Table 2.

- Scenario B: Future losses essentially follow Model 2. They form a non-homogeneous Poisson process with intensity

$$\lambda(t) = (1 + \hat{\alpha}_3(1 - \hat{\alpha}_1) \exp(-\hat{\alpha}_2 - \hat{\beta}_2 t))^{-\hat{\alpha}_3^{-1}},$$

where $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\beta}_2$ and $\hat{\alpha}_3$ are the parameter estimates in Model 2.

- Scenario C: Losses occur as a homogeneous Poisson process with a constant intensity equal to to the maximum intensity attained by Model 1 over the period until end 1993

$$\lambda = (1 + \hat{\alpha}_3(1 - \hat{\alpha}_1) \exp(-\hat{\alpha}_2))^{-\hat{\alpha}_3^{-1}} = 0.0419,$$

where $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ are the parameter estimates in Model 1.

- Scenario D: Losses occur as a non-homogeneous Poisson process with a linearly increasing intensity given by

$$\lambda(t) = l + mt$$

where $l = 0.0419$ the maximum intensity attained by Model 1 and $m = (0.0419 - 0.0061)/4986 = 7.64e^{-06}$, the average change in intensity per day in Model 1.

5.2 Severity Scenarios

Although we have found a trend of decreasing average losses above the threshold, we do not extrapolate this trend into the future. We wish to construct stress scenarios where we consider the effect of adverse scenarios; thus we use scenarios which suppose that severities stay constant, rather than decreasing.

- Scenario 1: Excess losses over the threshold have the GPD model implied by Model 0. Parameter values are $\xi = 0.92$ and $\phi = 2.53$. These values imply a mean loss size of 44.12 and a median loss of 3.45. This is essentially the same as the median loss size in Model 2, where loss size is to all intents and purposes also modelled as constant. The mean is almost undefined and does not provide a reliable guide to an average loss.
- Scenario 2: Excess losses have a stressed GPD distribution with parameters $\xi = 1.17$ and $\phi = 3.17$. These values are the estimates from Model 0 increased by one standard error. They imply a median loss size of 4.38 but mean loss size is undefined since $\xi > 1$.

We combine frequency and severity scenarios to get the full set of stress scenarios: A1, A2, B1, B2, C1, C2, D1, D2.

5.3 Predicting Loss Numbers

Because of our Poisson assumption we can easily calculate for each scenario the expected number of losses in future year periods and the associated standard deviation (which is the square root of the expected number.) We obtain the results of Table 4.

Note in the years 1992 and 1993 we observed 10 and 7 large losses respectively; in the year 1994 there had been 7 losses by the end of April alone. The data ended at this point but it is possible we were heading for a record loss frequency in 1994.

Against this background it is interesting to compare the expected numbers of losses under the different scenarios. Under scenario A we expect the modest number of about five losses per year in future years, which could

	Scenario							
	A1	A2	B1	B2	C1	C2	D1	D2
N_1	5.1	(2.3)	11.2	(3.3)	15.3	(3.9)	15.8	(4.0)
N_2	5.1	(2.3)	12.7	(3.6)	15.3	(3.9)	16.8	(4.1)
N_3	5.1	(2.3)	14.4	(3.8)	15.3	(3.9)	17.8	(4.2)

Table 4: Expected numbers of losses over the threshold ($u = 1$) in the next three years

be an underestimate. Under scenario B, which extrapolates model 2, we see a steady increase beginning at 11 losses per year; this would follow on smoothly from the losses observed in 1992 to April 1994. Scenario C assumes that the intensity has climbed to a new constant regime in the period of the data so that the expected number in future years will be about 15. Scenario D, which allows a linear increase in intensity to crudely extrapolate Model 1, gives the most extreme estimates.

5.4 Predicting Loss Volumes

	Scenario							
	A1	B1	C1	D1				
Z_1	33.1	(197)	92.7	(525)	133.8	(744)	139.4	(666)
Z_2	32.3	(240)	107.8	(582)	140.8	(675)	146.8	(766)
Z_3	32.3	(228)	118.9	(522)	139.5	(782)	157.4	(866)

	Scenario							
	A2	B2	C2	D2				
Z_1	49.3	(670)	161.1	(1780)	232.5	(2379)	232.4	(2093)
Z_2	51.8	(1016)	176.1	(1701)	221.6	(2453)	278.9	(2682)
Z_3	47.7	(614)	217.4	(2311)	238.2	(2110)	291.7	(3032)

Table 5: Median and 95th percentile of simulated loss volume distributions for losses over a threshold $u = 1$; these quantiles are derived from 1000 simulations under each scenario.

In the years 1992 and 1993 the observed total loss volumes were 65 and 22 million respectively; in 1994 until the end of April alone the observed volume was 47 million. None of these were record years; the largest loss volumes

were 162 and 99 in 1985 and 1991 respectively. Against this background we have tabulated the results of a simulation study in Table 5.

Under severity scenario 1 the mean loss volume (the mean of the distribution of Z_k) is only barely defined and under severity scenario 2 it is undefined, we choose to use median loss volume as a guide to the average loss burden under the various scenarios. We use the 95th percentile of the loss volume distribution as a guide to the uncertainty in the prognosis, since the variance of Z_k is undefined under all scenarios considered. To obtain these quantile measures we use Monte Carlo; we simulate 1000 replications of possible loss developments under the various scenarios and estimate the quantiles using sample quantiles.

The most salient finding of this simulation study is the dramatic effect that stressing the severity distribution has on loss volumes; in contrast, stressing the frequency distribution has a more modest impact. If we consider the second column of Table 5, we see that under B1 there is a 50% chance that the loss volumes are no greater than 93, 108 and 119. However, there is a 5% chance that the volumes are greater than about 520, 550, and 520. These results seem broadly in keeping with what we have observed. (Note that the Monte Carlo quantile estimates are inaccurate quantile estimates for only 1000 simulations, which explains why the estimate for year 3 is smaller than the estimate for year 2.)

Under scenario B2 the median loss volumes are 161, 176 and 217 and the 95th percentiles are 1780, 1701 and 2311. The medians are thus greater than the medians under scenario D1, the most extreme of the first set of frequency scenarios, and the 95th percentiles are very large, suggesting the possibility of very damaging loss years. With the amount of data we have, we cannot rule out the GPD excess distribution used in severity scenario 2. It is as well to be aware of the possible consequences of stressing severity in this way.

6 Conclusion

Prognoses such as these can clearly help an insurance company with pricing and reserving decisions. Clearly, forecasting future large insurance losses from natural catastrophes is subject to much uncertainty. By basing forecasts on a sensible limiting model supported by extreme value theory, we seek an honest appraisal of the uncertainty involved by examining various scenarios for future loss burden, all of which may be consistent with what we have seen in the past.

Software

Many of the analyses in this paper were carried out in S-Plus using the first author's own EVIS functions. These functions together with documentation

and some datasets are available to S-Plus users over the World Wide Web at <http://www.math.ethz.ch/~mcneil>

Acknowledgement

The first author is Swiss Re Research Fellow at ETH Zurich and thanks Swiss Re for their generous support of this research.

References

- Beirlant, J., Teugels, J. & Vynckier, P. (1996), *Practical analysis of extreme values*, Leuven University Press, Leuven.
- Cox, D. & Snell, E. (1968), ‘A general definition of residuals (with discussion)’, *Journal of the Royal Statistical Society, Series B* **30**, 248–275.
- 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.
- Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997), *Modelling extremal events for insurance and finance*, Springer Verlag, Berlin.
- 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.
- Leadbetter, M., Lindgren, G. & Rootzén, H. (1983), *Extremes and related properties of random sequences and processes*, Springer, Berlin.
- McNeil, A. (1997), ‘Estimating the tails of loss severity distributions using extreme value theory’, *ASTIN Bulletin* **27**, 117–137.

- McNeil, A. & Saladin, T. (1997), The peaks over thresholds method for estimating high quantiles of loss distributions, *in* 'XXVIIth International ASTIN Colloquium', pp. 23–43.
- Rootzén, H. & Tajvidi, N. (1997), 'Extreme value statistics and wind storm losses: a case study', *Scandinavian Actuarial Journal* pp. 70–94.
- Smith, R. (1989), 'Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone', *Statistical Science* **4**, 367–393.
- Smith, R. (1996), Extreme value analysis of insurance risk. Preprint.
- Smith, R. & Shively, T. (1995), 'Point process approach to modeling trends in tropospheric ozone based on exceedances of a high threshold', *Atmospheric Environment* **29**, 3489–3499.