PRACE NAUKOWE AKADEMII EKONOMICZNEJ WE WROCŁAWIU
Nr 1133 — 2006

Inwestycje finansowe i ubezpieczenia
– tendencje światowe a polski rynek

Viera Pacáková, Bohdan Linda

University of Pardubice

ACTUARIAL MODELLING OF EXTREME LOSSES

In the modelling of extreme events, different approaches had been proposed for certain circumstances. In 1928, Extreme Value Theory (EVT) originated in work of Fisher and Tippett describing the behavior of maximum of independent and identically distributed random variables. Various applications have been implemented successfully in many fields such as: actuarial science, hydrology, climatology, engineering, and economics and finance.

We describe parametric curve fitting methods for modelling extreme historical losses. These methods revolve around the generalized Pareto distribution (GPD) and are supported by extreme value theory.

1. Methods for modelling extreme losses

Extreme events are also called rare events. Extreme events share three characteristics: relatively rareness, huge impact and statistical inexpertness.

We are specifically interested in modelling the tails of loss severity distributions Thus is of particular relevance in reinsurance if we ale 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.

Suppose insurance losses are denoted by the independent, identically distributed random variables $X_1, X_2, \ldots,$ who's common distribution function is $F_X(x) = P(X \le x)$, where $x > 0$.

- *Extreme Value Theorem* [2]

Suppose $X_1, X_2, \ldots$ are *iid* with distribution function (df) $F_X(x)$. If there exist constants $c_n > 0$ and $d_n \in R$ such that

$$\frac{M_n - d_n}{c_n} \to Y, n \to \infty, \tag{1}$$

where $M_n = X_{1,n} = max\ (X_1, \ldots, X_n)$, Y is non-degenerate with distribution function G. Then G is of one the following types:

1. *Gumbel*

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

2. *Frechet*

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

3. *Weibull*

$$\Phi_\alpha(x) = \begin{cases} \exp\left\{-(-x)^\alpha\right\}, & \text{if } x < 0 \\ 0, & \text{if } x \geq 0 \end{cases}.$$

The generalized Gumbel, Frechet and Weibull families can be combined into a single family of distributions in the form

$$G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \text{ where } 1 + \xi\left(\frac{x-\mu}{\sigma}\right) > 0. \tag{2}$$

It is straightforward to check the result by letting

$$\alpha = \frac{1}{\xi}, d = \mu - \frac{\sigma}{\xi},\ c = \begin{cases} \frac{\sigma}{\xi}, & \text{when } \xi > 0 \\ -\frac{\sigma}{\xi}, & \text{when } \xi < 0 \end{cases}. \tag{3}$$

As the GEV describes the limit distribution of normalized maxima, the Generalized Pareto Distribution (GPD) is the limit distribution of scaled excess of high thresholds. The main connection is in the following theorem.

- *GPD Theorem*

Suppose $X_1, X_2, \ldots$ are *iid* with distribution F. Than

$$G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \text{ where } 1 + \xi\left(\frac{x-\mu}{\sigma}\right) > 0,$$

is the limit distribution of the maxima $M_n = X_{1,n} = max(X_1, \ldots, X_n)$. Then for a large enough threshold u, the conditional distribution function of $Y = (X - u / X > u)$ is approximately

$$P\left[X - u < x / X > u\right] \sim H(x) = 1 - \left(1 + \frac{\xi x}{\tilde{\sigma}}\right)^{-1/\xi} \tag{4}$$

defined on $\{x: x > 0 \text{ and } (1+\xi x/\tilde{\sigma}) > 0\}$, where $\tilde{\sigma} = \sigma + \xi(u-\mu)$.

The family of distributions defined by equation (4) is called the General Pareto family (GPD). For a fixed high threshold u, the two parameters are the shape parameter ξ and the scale parameter $\tilde{\sigma}$. For simpler notation, we may just use σ for the scale parameter if there is no confusion. The GPD distribution has many good properties (see for example [1, p. 165].

- *Definition*

Let X is a random variable with distribution function F(x).

$$F_u(x) = P(X-u < x/X > u) = \frac{F(x+u)-F(u)}{1-F(u)} \tag{5}$$

for $x \geq 0$ is the excess distribution of X *over the threshold* u and

$$e(u) = E(X-u/X>u) \tag{6}$$

is called the *mean excess function* of X.

- *Excess over Threshold Method*

The modeling using the excess over threshold method follows the assumptions and conclusions in GPD Theorem. Suppose $x_1, x_2, \ldots, x_n$ are raw observations independently from a common distribution $F(x)$. Given a high threshold u, assume $x_{(1)}, x_{(2)}, \ldots, x_{(k)}$ are an observation that exceeds u. Here we define the ascendances as $x_i = x_{(i)} - u$ for $i = 1, 2, \ldots, k$. By GPD Theorem x_i may be regarded as realization of independently random variable which follows a General Pareto family with unknown parameters ξ and σ. In case $\xi \neq 0$, the likelihood function can be obtained directly from (4):

$$L((\xi,\sigma/\mathbf{x})) = \prod_{i=1}^{k}\left[\frac{1}{\sigma}\left(1+\frac{\xi x_i}{\sigma}\right)^{-1/\xi-1}\right].$$

The mean excess function of GPD

$$e(u) = \frac{\sigma+\xi u}{1-\xi} \quad \text{for} \quad 0<\xi<1$$

is linear in u. Therefore we can check the linearity in the plot described blow:

Given a sample of *iid* observations $x_1, x_2, \ldots, x_n$ and a threshold u. Let $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n_u)}$ be the observations that exceeds u, a graph

$$\left\{\left(u, \frac{1}{n_u}\sum_{i=1}^{n_u}\left(x_{(i)}-u\right)\right): x_{\min} \leq u \leq x_{\max}\right\} \tag{7}$$

is called a *mean residual life plot* (*mrl*). The mean residual life plot provides an accessible approximation to the mean excess function.

In practice, the interpretation of a mean residual life plot may not be simple. Often the linearity is vague for small choice of u and for large u, the sparseness of the data available for calculation causes the large variation of the plot toward the right end. For our purposes, the *mlr* plot is used as a graphical tool in distinguishing between light and heavy-tail models.

2. Illustrational Data Analysis

To illustrate the mention above methods, we use the Danish fire loss data from McNeil's [3] study. The goal of this is to show techniques and plotting strategies which can be employed for similar data.

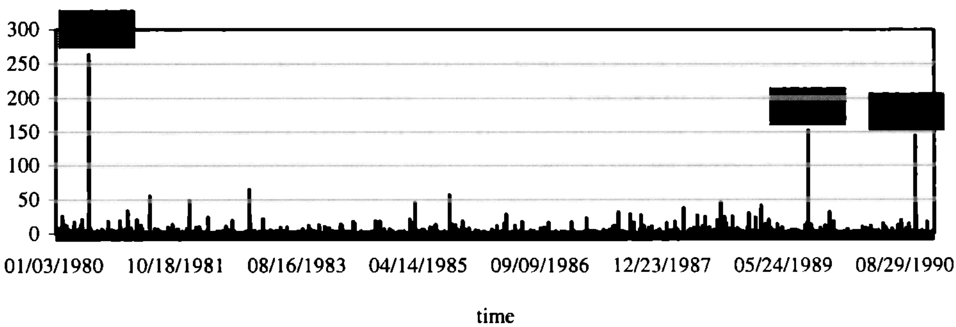

Fig. 1. Time series and log data plots for the Danish data
Source: authors.

The Danish data comprise 2157 losses over one million Danish Krone (DKK) from the years 1980 to 1990 inclusive (fig. 1).

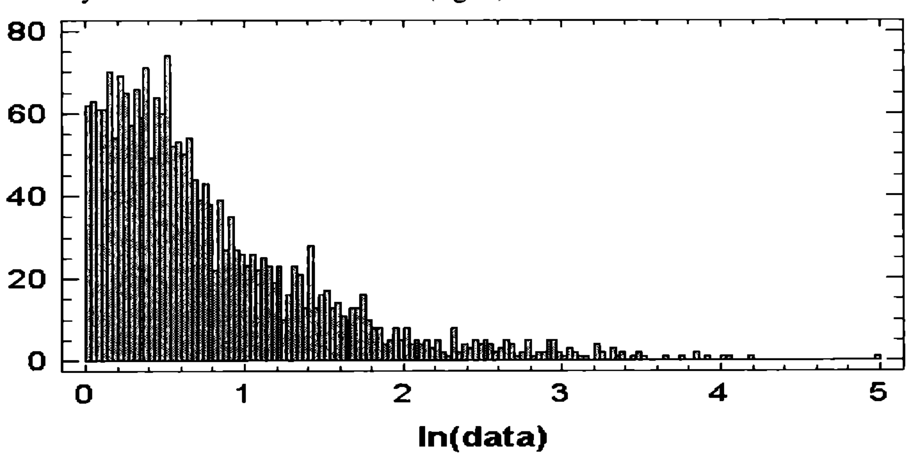

Fig. 2. The histogram on the log scale for the Danish data
Source: authors.

The time series plot (fig. 1) allows us to identify the most extreme losses and their times of occurrence. The histogram (fig. 2) shows the wide range of the data.

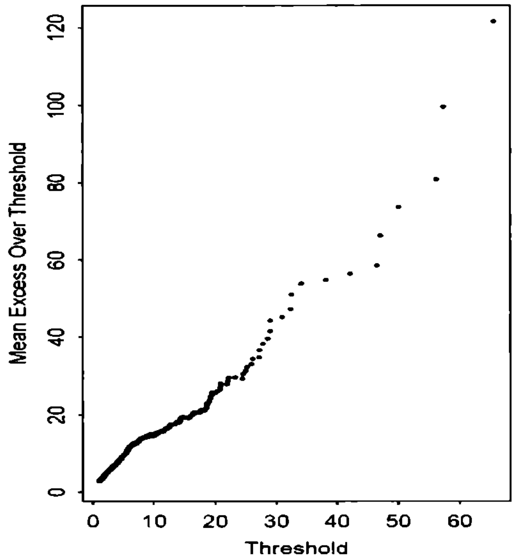

Fig. 3. The sample means excess function
Source: [3].

Useful graphical tool is the plot of the sample mean excess function (fig. 3) which is the plot $\left\{\left(u, e_n\left(u\right)\right), X_{1:n} < u < X_{n:n}\right\}$, where $X_{1:n}$ and $X_{n:n}$ are the first and nth order statistics and $e_n\left(u\right)$ is the sample mean excess function defined by

$$e_n\left(u\right) = \frac{\sum_{i=1}^{n}\left(X_i - u\right)^+}{\sum_{i=1}^{n} 1_{\{x_i > u\}}} \tag{8}$$

i.e. the sum of excesses over the threshold u divided by the number of data points which exceed the threshold u . The sample mean excess function $e_n\left(u\right)$ is an empirical estimate of the mean excess function defined by (6).

If empirical plot seems to 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 (GPD) with positive shape parameter in the tail area above u. This is precisely the kind of behavior in the Danish data (fig. 3). There is evidence of a straitening out 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 DPD might provide a reasonable fit to the entire dataset.

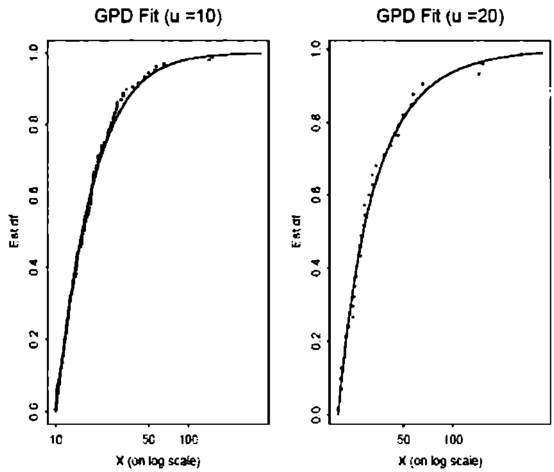

Fig. 4. Fitting GPD to data on exceedances of high thresholds
Source: [3].

We use three parameter forms 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. In left plot (figure 4) GPD is fitted to 109 exceedances of the threshold 10. The parameter estimates are $\xi = 0,497$ and $\sigma = 6,98$. In right plot GPD is fitted to 36 exceedances of the threshold 20. The parameter estimates are $\xi = 0,684$ and $\sigma = 9,63$.

References

[1] Embrechs P., Kluppelberg C., Mikosch T., *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin 1997.

[2] Fisher R.A., Tippet L.H.C., *Limiting Forms of the Frequency Distribution of the Largest or Smallest Number of the Sample*, Proc. Cambridge Philos. Soc. 24 199928, 180-190.

[3] McNeil A., *Estimating the Tails of Loss Severity Distributions using Extreme Value Therory*, ETH Zentrum, Zürich 1996.

[4] Pacáková V., *Aplikovaná poistná štatistika*, IURA EDITION, Bratislava 2004.

[5] Parisi F., *Extreme Value Theory and Standard & Poor's Rating*, ABS Research, New York 2000.

[6] Zhongxian H., *Actuarial Modelling of Extremal Events Using Transformed Generalized Extreme Value Distributions and Generalized Pareto Distributions*, The Ohio State University 2003.

MODELOWANIE AKTUARIALNE STRAT EKSTREMALNYCH

Streszczenie

W niniejszym artykule zaprezentowano wykorzystanie parametrycznych metod dopasowywania krzywych do modelowania ekstremalnych historycznych strat. Streszczono odpowiednie wyniki teoretyczne pochodzące z teorii wartości ekstremalnych (*extreme value theory* (EVT)) i metody przekroczeń (*excess over threshold method* (EOT)) oraz przedstawiono przykład zastosowania tych wyników do danych A. McNeila (1996) dotyczących wysokich strat spowodowanych pożarami w Danii.