CATASTROPHIC RISK MANAGEMENT IN NON-LIFE INSURANCE

Valéria Skřivánková, Alena Tartaľová

1. Introduction

At the end of the second millennium, a view of catastrophic events during the last thousand years was published (see [13]). By this source, the worst natural catastrophes in terms of fatalities were floods in China in 1887 (900 000 victims). From the point of financial losses, the most costly insurance events in the 20th century were earthquakes in Japan (1995), in USA (California, 1994), and in Turkey (1999), further hurricane in USA (Florida, 1992) and snow-storms in Europe (1990). Besides the natural catastrophes man-made disasters (e.g. airplane catastrophes and terrorism) may have extremal insurance losses, too.

The increasing trend of the catastrophic event frequency and of the claim amount connected with them, forces the insurance companies to search effective risk management methods. In practice, it means to have better reinsurance and to use cat-bonds, weather derivates and other modern financial instruments (see Zmeškal [12]). In the theoretical sphere one can follow an intensive development of extreme value theory (see monographs from Beirlant et al. [1] and Embrechts et al. [2] and plenty of other publications).

The quantifiability of non-life insurance claims makes the mathematical modelling more tractable. From the viewpoint of the reinsurance company, the estimation of the upper tail of the claim size distribution is the major interest. It is necessary for determining the net premium of a reinsurance contract and for standing the optimal level of reinsurance.

This paper deals with some theoretical aspects of modelling extremal insurance claims and with practical searching the optimal reinsurance level. First, we formulate the problem of generalization of classical Cramér-Lundberg collective risk model. Then using some well known extreme value results we study two suitable methods for extremal claims registration. In the end, detailed statistical analysis follows using real fire insurance claims.

2. Formulation of the Problem

The classical collective risk model for non-life insurance claims was introduced by Filip Lundberg in 1909 and developed by Harold Cramér in 1930´s. They showed that under some assumptions the homogeneous Poisson process is the key model for non-life insurance claims. The Cramér-Lundberg assumptions are:

1. The individual insurance claims X_i are independent, identically distributed (iid) random variables with distribution function $F(x)$, finite mean $E(X_i) = \mu < \infty$ and variance $D(X_i) = \sigma^2 < \infty$.
2. The claim times $0 < T_1 < T_2 < \ldots$ are random variables.
3. The inter-arrival times between two following claims $Y_1 = T_1$, $Y_i = T_i - T_{i-1}$, for $i = 2, 3, \ldots$ are independent, exponentially distributed random variables with finite mean $E(Y_i) = 1/\lambda$, for $\lambda > 0$.
4. The sequences $\{X_i\}_i$ and $\{Y_i\}_i$ are independent.
5. The number of insurance claims in the interval $[0,t]$ is a random process $\{N_t; t \geq 0\}$ where
$$N_t = \sup\{n \geq 1; T_n \geq t\}$$
The assumption 3 forces that $\{N_t; t \geq 0\}$ is a homogeneous Poisson process with intensity λ.
6. The risk reserve of an insurance company depending on time, is a random process $\{R_t; t \geq 0\}$, where
$$R_t = r + C_t - S_t$$
and $r = R_0$ is the initial reserve in time $t=0$, $C_t = c.t$ is the total insurance premium until time t, $S_t = \sum_i X_i, i = 1,2,\ldots,N_t$ stands for the total claims amount until time t.

This model works only for small insurance claims. If we allow the occurrence of extremal claims with non-zero probability, we must modify the classical C-L model and change assumptions 1, 3, 5. Then instead of an exponential type distribution we consider heavy-tailed subexponential distribution and the claim number process $\{N_t; t \geq 0\}$ will be a renewal process.

The second problem is, that heavy-tailed subexponential distributions may have infinite variance, so the classical central limit theorem (CLT) doesn't hold. The normal approximation has tendency to undervalue losses on the upper tail of the distribution which is very undesirable because the largest claims have significant financial impact. Modifying the CLT conditions, about finite mean and variance, as limit distribution we get stable distribution.

The practical problems in non-life insurance are connected with searching an optimal reinsurance level, with standing the correct reinsurance net premium and with calculating the ruin probability. In this paper we concentrate on determining the optimal retention level. The other problems are studied e.g. in Fecenko [3], Horáková [4], Pacáková [6], [7], Urbaníková [11].

3. Subexponential and Stable Distributions

Subexponential distributions have heavy-tail denoted by:

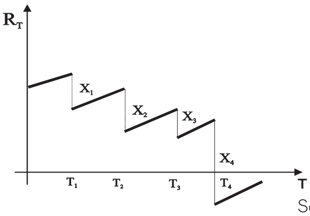
$$\bar{F}(x) = 1 - F(x) = P(X > x),$$

which converges to zero more slowly for $x \to \infty$ than the exponential tail. Subexponential distributions have defining property (see [2]):

$$\lim_{x\to\infty} \frac{P(X_1 + \ldots + X_n > x)}{P(\max\{X_1,\ldots,X_n\} > x)} = 1,\ n \geq 2. \quad (1)$$

Property (1) means that the tail of the sum of iid random variables $S_n = X_1 + \ldots + X_n$ and the tail of the maximum $M_n = \max\{X_1,\ldots,X_n\}$ are asymptotically of the same order. So, the largest claim has strong influence on the total claim amount, even can cause an insurance company ruin (see Fig.1).

Fig. 1: Realization of the risk process



Source: own

Because the distribution of the sum of iid random variables $X_1,\ldots,X_n$ is given by the n-th convolution $F^{*n}(x)$ and

$$P(\max\{X_1,\ldots,X_n\} > x) = P\left[\bigcap_{i=1}^{n}(X_i > x)\right] = \prod_{i=1}^{n} P(X_i > x) = \bar{F}^n(x)$$

we get another version of property (1):

$$\lim_{n\to\infty} \frac{\bar{F}^{*n}(x)}{\bar{F}^{n}(x)} = 1 \quad (2)$$

We can show that some well known distributions as lognormal, loggamma, Pareto and Weibull have property (2). These distributions are in detail studied e.g. in Embrechts et al. [2] or Pacákova [6].

Stable distributions in general, preserve the given type of distribution for the sum of iid random variables. In this sense, the normal, Poisson and gamma distribution are stable (see Tartaľová [10]).

Before we define the stable distributions precisely, we express two terms: right endpoint and non-degenerate distribution:

We define the ***right endpoint*** of a distribution F(x) as

$$X_F = \sup\{x \in R; F(x) < 1\}.$$

A distribution function F(x) is called ***degenerate df*** if $F(x) = 0$, for $x < x_F$ and $F(x) = 1$, for $x > x_F$.

Definition 1. A non-degenerate df *F(x)* of random variable *X* is called ***stable distribution*** if for arbitrary sequence of iid random variables $\{X_n\}_n$ there exist norming constants $a_n \in R$, $b_n > 0$ such that for the sum $S_n = X_1 + X_2 + \ldots + X_n$ holds

$$b_n^{-1}(S_n - a_n) \xrightarrow{d} X, \text{ for } n \to \infty. \quad (3)$$

Definition 2. A non-degenerate df F(x) of random variable X is called ***max-stable distribution*** if for arbitrary sequence of iid random variables $\{X_n\}_n$ there exist norming constants $c_n \in R$, $d_n > 0$ such that for the maximum $M_n = \max\{X_1, X_2,\ldots,X_n\}$ holds

$$d_n^{-1}(M_n - c_n) \xrightarrow{d} X, \text{ for } n \to \infty. \quad (4)$$

4. Limit Distribution for Maxima

The following well known theorem specifies the type of the possible limit distribution for normalized maxima.

Theorem 1 (Fisher-Tippet):

Let $\{X_n\}_n$ be a sequence of iid random variables. If there exist norming constants $c_n \in R$, $d_n > 0$ and some non-degenerate distribution function H such that (4) holds, then H belongs to the type of one of the following three ***standard extreme value distributions***:

1. Fréchet: $\Phi_\alpha(x) = \exp\{-x^{-\alpha}\}$, $x > 0$, $\alpha > 0$ (else $\Phi_\alpha(x) = 0$).
2. Weibull: $\Psi_\alpha(x) = \exp\{-(-x)^\alpha\}$, $x \leq 0$, $\alpha > 0$ (else $\Psi_\alpha(x) = 1$).
3. Gumbel: $\Lambda(x) = \exp\{-e^{-x}\}$, $x \in R$.

Proof. For the sketch of the proof see Embrechts et al. [2], p.122.

There exists a one-parameter representation of the three standard cases in one family of dfs - the general extreme value distribution H_ξ.

Definition 4. The distribution function $H_\xi(x)$ defined for $1 + \xi x > 0$ by

$$H_\xi = \begin{cases} \exp\left\{-(1+\xi x)^{-1/\xi}\right\}, \text{ if } \xi \neq 0 \\ \exp\left\{-e^{-x}\right\}, \text{ if } \xi = 0 \end{cases} \tag{5}$$

is called ***the generalized extreme value distribution (GEV)***. The parameter ξ is called ***extreme value index (EVI)***. H_ξ corresponds to

- Fréchet distribution for $\xi = \alpha^{-1} > 0$, $x > -\xi^{-1}$,
- Weibull distribution for $\xi = \alpha^{-1} < 0$, $x < -\xi^{-1}$,
- Gumbel distribution for $\xi = 0$, $x \in R$.

The three standard extreme value distributions have different tails and serve as limit distribution for different types of distributions:

Fréchet - long tail (for Pareto, Cauchy, Student and loggamma distributions),

Gumbel - moderately long tail (for exponential, normal, lognormal and gamma distributions),

Weibull - short tail (for uniform and beta distributions).

To model the tail of the underlying distribution *F*, we follow the excesses above sufficiently high threshold *u*. Let X be a random variable with df *F* and with right endpoint x_F. We define the ***excess distribution function*** $F_u(x)$ for $u < x_F$ by

$$F_u(x) = P\left(X - u \leq x \mid X > u\right) = \frac{F(x+u) - F(u)}{\overline{F}(u)}, \ x \geq 0 \tag{6}$$

In an insurance context F_u is an excess-of-loss df. The conditional mean

$$e(u) = E\left(X - u \mid X > u\right), \tag{7}$$

as a function of the chosen threshold u, is called ***mean excess function***. To estimate *e(u)* is an important first step in deciding on the premium.

The following theorem says that the only possible limit distribution for the excesses over high threshold is the generalized Pareto distribution. First we define this function.

Definition 5. The distribution function $G_\xi(x)$ defined for $1 + \xi x > 0$ by

$$G_\xi(x) = \begin{cases} 1-(1+\xi x)^{-1/\xi}, \text{ if } \xi \neq 0 \\ 1-e^{-x}, \text{ if } \xi = 0. \end{cases} \tag{8}$$

Is called ***the generalized Pareto distribution (GPD)***.

For $\xi > 0$ is $x \geq 0$ and for $\xi < 0$ is $0 \leq x \leq \xi^{-1}$. For $\xi = 0$ we get the exponential distribution.

We can extend the family of Pareto distributions adding scaling parameter ß and location parameter γ as

$$G_{\xi,\beta}(x) := G_\xi(x/\beta) \quad \text{or} \quad G_{\xi,\beta,\gamma}(x) := G_\xi\left(\frac{x-\gamma}{\beta}\right).$$

Theorem 2. (Pickands, Balkema and de Haan)

$F_u(x)$ is an excess distribution function if and only if we can find a positive measurable function $\beta = \beta(u)$ for every $\xi > 0$ such that

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

Proof. See Embrechts et al. [2], p.165.

If we want to model the distribution of the maximum of excesses over threshold, we consider a random number of random variables, i.e. $M_N = \max\{X_1, X_2, \ldots, X_{N_t}\}$, where $\{N_t; \geq 0\}$ is a Poisson process. Then we obtain the following result.

Theorem 3.

Let $M_N=\max\{X_1,X_2,...,X_{Nt}\}$, where X_i are iid random variables and $N_t \approx Po(\lambda)$. Then

$$P(M_N \le x) = \exp\left\{-\left[1+\xi\,\frac{x-\xi^{-1}\beta\left(\lambda^{\xi}-1\right)}{\beta\lambda^{\xi}}\right]^{-1/\xi}\right\}, \text{if } \xi \neq 0,$$

$$P(M_N \le x) = \exp\left\{\exp\left[-\left(\frac{x-\beta\ln\lambda}{\beta}\right)\right]\right\}, \text{if } \xi = 0.$$

Proof. For the detailed proof see Skřivánková [9].

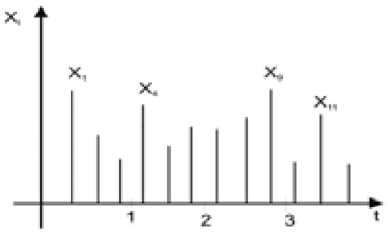
5. Methods of Extreme Value Registration

Different types of non-proportional reinsurance require different approaches to extreme value registration. The most important methods of identification extremes are the method of block-maxima and the peaks-over-threshold method.

5.1. Method of Block-Maxima (BM)

By this method, as extreme values are recorded only the maxima in the block, e.g. annual, monthly or daily maximum. Figure 2 shows the principle of this method. The only three possible limit distributions for block-maxima are by Fisher-Tippett theorem Fréchet, Weibull or Gumbel. To decide for one of them, we estimate the extreme value index ξ using e.g. the maximum likehood method.

Fig. 2: Method of block-maxima: X_1, X_4, X_9, X_{11} are block-maxima.



Source: own

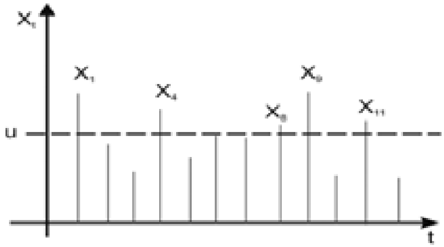
Method of block-maxima has a disadvantage: it uses only one value per block, but in the block can be another big value which we don't consider.

This method can be used in non-proportional Largest Claims Reinsurance (LCR). In general with an LCR (p) treating the reinsurer pays the *p* largest claims. By the BM method we can solve LCR (1).

5.2. Peaks-over-Threshold Method (POT)

Here as extremes we record all exceedances over given threshold *u*. On the Figure 3 is graphically shown the principle of this method. The only possible limit distribution for exceedances is by Pickands-Balkema-de Hann theorem the generalized Pareto distribution $G_\xi(x)$. The problem is to stand the retential level *u* for reinsurance correctly. This method is often used in non-proportional Excess of Loss reinsurance (XL). With XL-reinsurance treaty the reinsurer pays that part of each claim amount which exceeds an agreed limit *u*, the cedant's retention (level).

Fig. 3: Peaks over threshold method: X_1, X_4, X_6, X_9, X_{11} are exceedances over threshold u.



Source: own

6. Real Data Analysis

Considering the short history of Slovak insurance market, there doesn't exist sufficiency of data for reliable statistical analysis of extremal events. For that reason, we realized the practical analysis on fire insurance claims from Copenhagen Re, free downloadable from homepage of Professor Alexander McNeil http://www.ma.hw.ac.uk/~mcneil/ftp/DanishData.txt. For this analysis we used STATGRAPHICS and its procedure Distribution fitting.

The set of data consists of 2167 losses over one million DK (Danish Krones) coming from the time interval 1.3.1980 to 31.12.1990. The minimum is 1 mil. DK and maximum is 263,25 mil. DK. The average is relatively very small, the variance and coefficient of variation are very big. According to skewness parameter, we can assume that data come from positive skewed and heavy-tailed distribution (see Table 1).

Tab. 1: Characteristics of data

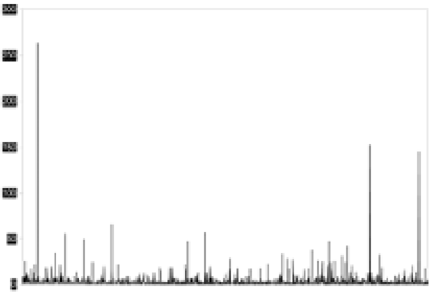
Count	2167
Minimum	1,0
Maximum	263,25
Average	3,385
Variance	72,377
Coeff.of variation	251,321 %
Skewness	18,763

Source: own

6.1. Graphical Analysis

On the Figure 4 is the time series plot from 1.3.1980 to 31.12.1990. It allows us to identify the extremal losses and their approximate time of occurrence. We can also see whether there exists an evidence of clustering large losses which signalize the dependence of data. Our data can be considered as independent.

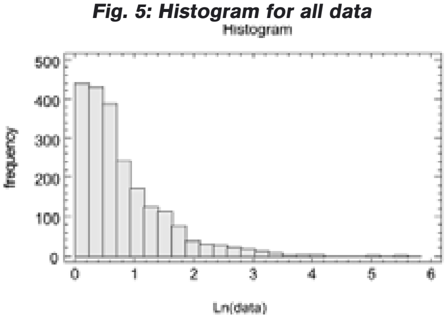
Fig. 4: Time series plot for data



Source: own

The *histogram* on the Figure 5 shows the wide range of the data and also that there are extreme values on the right tail of distribution.

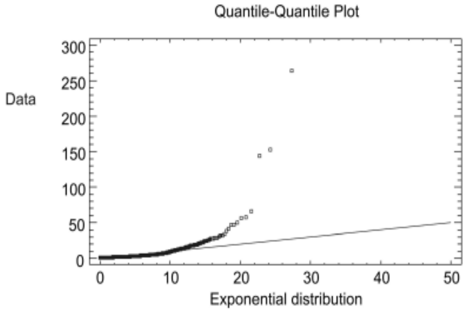
Fig. 5: Histogram for all data



Source: own

The quantile-quantile plot (QQ-plot) of the empirical distribution against the exponential distribution (on the Figure 6) is a very useful tool to describe heavy tails. In the case that the points lead approximately on a line it examines visually the hypothesis that the losses come from an exponential distribution. If there exist a significant deviation upstairs on the right end, the empirical distribution has heavy-tail. Linearity in the graph can be easily checked by eye. The quantiles of the empirical distribution function on the x-axis are plotted against the quantiles of the exponential distribution function on the y-axis. The plot is:

Fig. 6: QQ-plot against exponential distribution

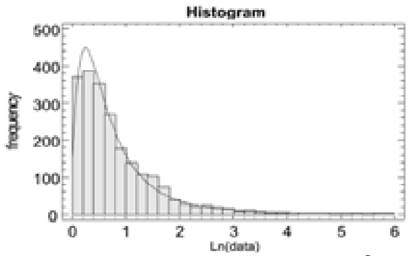


Source: own

6.2. Distribution Fitting

To fit all the data, as a suitable model was used lognormal distribution. The graphical agreement is shown on Figure 7. The maximum likehood estimates of parameters of lognormal distribution are using Statgraphics: shape=0,805278; standard deviation=0,860362 and low threshold=–0,105903. The considered lognormal distribution is the best model for our data, the p-value of goodness-of-fit tests is greater than the given significant level (for χ^2-test and Kolmogorovov-Smirnovov test).

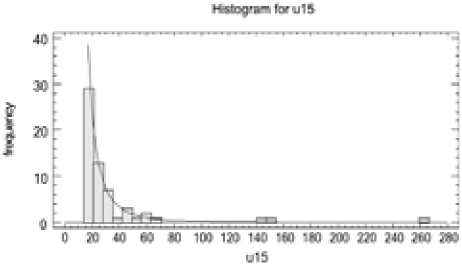
Fig. 7: Histogram and lognormal distribution for all data



Source: own

We can see that the tail of lognormal distribution converges to zero very fast, so don‘t fit the largest values. Thus we use the peaks-over-threshold method to fit only the tail of the distribution. At first we must choose an appropriate threshold *u*. Our modelling is based on a limit theorem which applies above high thresholds. So if we choose too low threshold, we can get biased estimates because the theorem does not apply. On the other hand, if we set too high threshold we will have only a few data and our estimates will be prone to high standard errors. The QQ-plot on the Figure 6 signalized that the data about 15 are dispersed. So we choose threshold *u*=15, which corresponds to 60 values. By the Pickands - Balkema- de Haan theorem the best model for exceedances over threshold is the generalized Pareto distribution, for our data with two parameters: shape=1,61727 and location=10,011. The graphical agreement with two-parameter Pareto distribution $G_{\xi,\beta}$ is on the Figure 8.

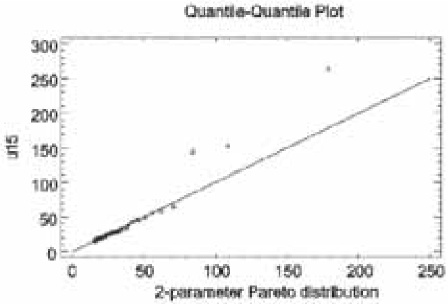
Fig. 8: The tail histogram and Pareto distribution



Source: own

On the Figure 9 we see the noticeable agreement between the quantiles of Pareto distribution and theoretical distribution of 60 exceedances over threshold *u*=15.

Fig. 9: QQ-plot against Pareto distribution



Source: own

We have calculated (see Tab.2 below) the number of exceedances and the p-value of Kolmogorov-Smirnov goodness-of-fit test for some selected threshold. The best fit is for u=15, when p-value is the largest. A similar result (0,927368) of p-value gives also the Chi-squared test. So we choose the optimal retention level *u*=15 mil. DK.

Tab.2: Comparisons for the different threshold u

Distribution	u	Number of excesses	P-Value of K-S test
Lognormal	0	2167	0,0908965
GPD	5	254	0,444256
GPD	10	109	0,767173
GPD	15	60	0,962803
GPD	20	36	0,913898

Source: own

7. Conclusion

In this paper we have made the familiar assumption of independent, identically distributed random variables. In practice we may be confronted with clustering, trends, seasonal effects and other kinds of dependencies. In that case the considered methods must be modified. Furthermore we can consider records as extreme values and follow their distribution and waiting time for the next record. An interesting problem arises if we modify the assumption 6 in the Cramér-Lundberg model and suppose that Ct is a random variable.

Acknowledgements

The first author was supported by the grants 1/2630/05, 1/0724/08 and the second one by the grant 1/3815/06 of the Grant Agency of Slovak Republic (VEGA).

References:

[1] BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J. and TEUGELS, J. *Statistics of Extremes: Theory and Applications*. New York: Wiley, 2004. ISBN 0-471-97647-4.

[2] EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. *Modelling extremal events for insurance and finance*. Berlin: Springer, 1997. ISBN 3-540-60931-8.

[3] FECENKO, J. and ŠELENG, M. Niektoré aspekty XL-zaistenia. *Zborník zo seminára Poistná*

matematika v teórii a praxi. Bratislava, 2005, p. 12-24, ISBN 80-225-2106-X.

[4] HORÁKOVÁ, G. and KOZUBÍK, A. Calculations of the expected claims for catastrophic events. *Journal of Information, Control and Management Systems*, 2005, Vol. 3., No.1, pp. 9-16, ISSN 1336-1716.

[5] MCNEIL, A. J. *Estimating the tails of loss severity distributions using extreme value theory*. ASTIN Bulletin, 27: pp.117-137. [online] [cit. 2007-8-20]. <http://www.casact.org/library/astin/vol27no1/117.pdf>.

[6] PACÁKOVÁ, V. Modelovanie a simulácia poistných rizík v neživotnom poistení. *E+M Ekonomie a Management*, 2007, roč. 10, č. 3, s. 109-121. ISSN 1212-3609.

[7] PACÁKOVÁ, V. and ŠOLTÉS, E. Modelling and simulations of extreme losses using quantiles function, *Prace Naukowe Akademii Ekonomicznej im. Oskara Langego we Wroclawiu Nr 1037*. Wydawnictvo AE im. Oskara Langego we Wroclawu, 2004, pp. 91-99. ISSN 0324-8445.

[8] SKŘIVÁNKOVÁ, V. Problém stability pri modelovaní extrémnych udalostí. *Zborník zo seminára Poistná matematika v teórii a praxi*. Bratislava, 2005, pp. 78-84. ISBN 80-225-2106-X.

[9] SKŘIVÁNKOVÁ, V. Štatistická analýza extrémnych hodnôt a metódy ich registrácie v neživotnom poistení. Sborník z 3.mezinárodní konference *Řízení a modelování finančních rizík*. Ostrava, 2006, pp. 270-277. ISBN 80-248-0618-5.

[10] TARTAĽOVÁ, A. *Modelovanie extrémnych poistných udalostí*. Diplomová práca na PF UPJŠ, Košice, 2006.

[11] URBANÍKOVÁ, M. *Wienerov proces a rezerva poisťovne*. Vedecké práce MtF STU, 2000, zv. 9, pp. 119-124. ISBN 80-227-1476-3.

[12] ZMEŠKAL, Z. Oceňování weather derivátu na bázi simulace. *Sborník z 2.mezinárodní konference Řízení a modelování finančních rizík*. Ostrava, 2004. ISBN 80-248-0618-5.

[13] Munich Re. [online]. <http://www.munichre.com>.

[14] Swiss Re. [online]. <http://www.swissre.com>.

Doc. RNDr. Valéria Skřivánková, CSc.
P.J.Šafárik University of Košice
Faculty of Sciences
Institute of Mathematical Sciences
Jesenná 5, 041 54 Košice
valeria.skrivankova@upjs.sk

Mgr. Alena Tartaľová
Technical University of Košice
Faculty of Economics
Department of Applied Mathematics and Business Informatics
Nemcovej 32, 040 01 Košice
alena.tartalova@tuke.sk

Doručeno redakci: 19. 11. 2007
Recenzováno: 11. 1. 2008
Schváleno k publikování: 7. 4. 2008

ABSTRACT

CATASTROPHIC RISK MANAGEMENT IN NON-LIFE INSURANCE

Valéria Skřivánková, Alena Tartaľová

The paper deals with some aspects of modelling catastrophic risk and with its application to non--life insurance claims. First, we formulate the problem of generalization of classical Cramér-Lundberg collective risk model. Then using some well-known extreme value results we study two methods for extremal claims registration. Finally, we apply the theoretical results for real insurance data.

As suitable mathematical models for large insurance claims are used heavy-tailed distributions (subexponential, stable and max-stable distributions).The main reason why we are interested in stable distributions is, that for the extreme value distributions the classical central limit theorem (CLT) condition (finite mean and variance) doesn't hold. Instead of CLT we use the Fisher-Tippett theorem which specifies the limit laws for maximum of independent identically distributed (iid) random variables as Generalised Extreme Value (GEV) distribution.

For recording extreme insurance claims we use two approaches. The first one is based on modelling maximum of the sample and called method of block-maxima. This method is based just on the Fisher-Tippett theorem and in non-life insurance we can use it for non-proportional Largest Claim Reinsurance (LCR). The second approach is based on modelling excess values over the chosen threshold. This approach is called Peaks Over Threshold method and is based on the Picands theorem which specifies the limit law for the exceedances as Generalised Pareto Distribution (GPD). This method is used in non-proportional Excess-of-Loss Reinsurance (XL).

In the end, we apply these methods for modelling real fire insurance claims. We find an optimal exceedance level for reinsurance and identify lognormal distribution for all data and Pareto distribution for the tail. The empirical data are compared with considered theoretical distribution using chi-squared and Kolmogorov-Smirnov goodness-of-fit tests. For detailed statistical analysis of data we use STATGRAPHICS and its procedure Distribution fitting.

Key Words: *extremal insurance claims, limit distributions, methods of extreme registration, statistical analysis of extremes.*

JEL Classification: *C13, C16, G22*