EVT methods as risk management tools

Valéria Skřivánková[1], Matej Juhás[2]

Abstract

The paper deals with some techniques and methods from extreme value theory relevant for insurance risk management. Peaks Over Threshold and Peaks Over Random Threshold methodologies, parameter estimation methods and extreme value condition tests are discussed. At the end of the paper the considered graphical and analytical methods for real non-life insurance data analysis are applied.

Key words

Extreme values, POT and PORT methodologies, parameter estimations, statistical testing

JEL Classification: C13, C16, G22

1. Introduction

The (re)insurance industry is undergoing major changes due to increasing occurance of catastrophic losses. In sense of Solvency II the new situation calls for sophisticated risk management tools including extreme value theory (EVT) methods. Extremes in the real world of finance manifest themselves through stock market crashes or catastrophic insurance losses. EVT yields methods for quantifying such events and their consequences in a statistically optimal way (see the monographs Embrechts at al. (1997), Beirlant at al. (2004), Reiss and Thomas (2007)).

The aim of this paper is to present some of the basic techniques and methods from EVT relevant for insurance risk management. Peaks Over Threshold (POT) and its generalized version Peaks Over Random Threshold (PORT) methodologies suitable for special types of reinsurance namely for excess of loss (XL)-reinsurance are considered. Parameter estimation methods, some methods for choosing optimal threshold and extreme value condition (EVC) tests will be discussed. At the end of the paper we will give real non-life insurance data analysis to present how EVT provides better idea for translating management guidelines into actual numbers.

2. Basic terms and know results

Consider the sequence $\{X_n, n \geq 1\}$ of independent identically distributed (iid) random variables (insurance claims) with continuous distribution function $F(x)$. If we are interested in modeling excedances (extreme claims) which are larger than a sufficiently high (deterministic or random) threshold u, we use the POT or PORT method of registration of extremes. The conditional distribution function of excesses $X-u$, $u > 0$ (tail distribution) is of the form

[1]Valéria Skřivánková, Doc. RNDr., CSc, Institute of Mathematics, Faculty of Science, P. J. Šafárik University in Košice, Jesenná 5, 04001 Košice, SR, e-mail: valeria.skrivankova@upjs.sk
[2]Matej Juhás, RNDr. , Institute of Mathematics, Faculty of Science, P. J. Šafárik University in Košice, Jesenná 5, 04001 Košice, SR, e-mail: matej.juhas@student.upjs.sk
Acknowledgement: Research was partially supported by Slovak grant agency under VEGA No. 1/0410/11 and 1/0931/11

$$F_u(x) = P(X - u \le x \mid X > u),\ \ 0 \le x < x_F \tag{1}$$

where x_F = sup{x; $F(x)$ < 1} is right endpoint of the distribution. Denoting $\overline{F}(x) = 1 - F(x) = P(X > x)$ and using the definition of conditional we get

$$\overline{F}(x+u) = \overline{F}(u) \cdot \overline{F}_u(x), \text{for } x > 0, u > 0.$$

By the well know Balkema, de Haan, Pickands theorem (see Embrechts [2]) the limit distribution of the tail $F_u(x)$, $u \to x_F$ in (1) is generalized Pareto distribution (GPD) of the form

$$G_{\xi,\beta}(x) = \begin{cases} 1 - \left(1 + \xi \dfrac{x}{\beta}\right)^{-1/\xi} & \text{if } \xi \ne 0, \\ 1 - \exp(-x/\beta) & \text{if } \xi = 0, \end{cases} \tag{2}$$

where $x \in \langle 0, \infty)$ for $\xi \ge 0$ and $x \in \langle 0, -\beta/\xi \rangle$ for $\xi < 0$.

Parameter ξ is called *extreme value index* (EVI) and characterizes the three special types of possible limit distributions, if $\xi = 0$, GPD in (2) represents exponential law, for $\xi < 0$ GPD is Beta distribution and for $\xi > 0$ we have Pareto distribution. So the estimated value of EVI gives us an important information about the type of the tail distribution. Special methods of estimation will be studied in Section 3.

Parameter $\beta = \beta(u)$ is the *scale parameter* and it is depending on the threshold u. If the threshold is unknown, we must state it carefully, because if u is too high, we have only a few exceedances and large estimation variance. For too low u the estimate can be bias. Some suitable methods for searching u will be demonstrated in Section 4.

For modeling the distribution of the maximum of exceedances over random threshold we consider the maximum of random number N_t of iid random variables X_i, i = 1, 2, …, N_t. We proved in [8] that in the case $N_t \sim Po(\lambda)$ and $X_i \sim G_{\xi,\beta}(x)$ the distribution of $M_N = \max\{X_1, \ldots, X_{N_t}\}$ is of the form

$$\begin{aligned} P(M_N \le x) &= \exp\left\{-\left(1 + \xi \frac{x - \xi^{-1}\beta(\lambda^{\xi} - 1)^{-1/\xi}}{\beta\lambda^{\xi}}\right)\right\},\ \text{for } \xi \ne 0, \\ P(M_N \le x) &= \exp\left\{-e^{-(x - \beta \ln \lambda)/\beta}\right\},\ \text{for } \xi = 0. \end{aligned} \tag{3}$$

Using formula (3) reinsurance risk managers can easily calculate the risk of insolvency.

3. Parameter estimation methods

The extreme value index ξ is closely related to the tail heaviness of the GPD. The value $\xi = 0$ (exponential tail) can be regarded as a change point: $\xi < 0$ refers to short tails with finite endpoint $x_F < \infty$ and $\xi > 0$ means heavy tailed distribution.

There are some popular ways of detecting heavy tails and estimating the EVI index ξ of GPD (see Embrechts [1], Beirlant [2], Reiss and Thomas [7]). In this section the Pickands, Hill, Decker - Einmahl - de Haan and minimum distance estimators will be discussed.

Pickands estimator

The Pickands estimation is founded on the order statistics $X_{1:n} \le X_{2:n} \le \ldots \le X_{n:n}$ pertaining to the iid random variables $X_1, X_2, \ldots, X_n$. We derived it for $\xi \in R$ (see [4]) in the form

$$\hat{\xi}^{P}_{k,n} = \frac{1}{\ln 2} \ln\left(\frac{x_{n-k:n} - x_{n-2k:n}}{x_{n-2k:n} - x_{n-4k:n}}\right). \tag{4}$$

Hill estimator

The Hill estimation is based on the k upper ordered statistics (similarly as Pickands estimator) and it can be derived for $\xi > 0$ using the maximum likelihood method (see [2] or [5]) in the form

$$\hat{\xi}_{k,n}^{H} = \frac{1}{k}\sum_{i=1}^{k} \ln\left(x_{n-i+1:n}\right) - \ln x_{n-k:n} \quad . \tag{5}$$

Deckers – Einmahl – de Haan estimator

This estimation represents the generalized version of Hill estimation for $\xi \in R$ and it is also called the moment estimation (see [7]). We can calculate it by formula (6)

$$\hat{\xi}^{D} = 1 + H_n^{(1)} + \frac{1}{2}\left(\frac{\left(H_n^{(1)}\right)^2}{H_n^{(2)}} - 1\right)^{-1}, \tag{6}$$

where $H_n^{(1)} = \hat{\xi}_{k,n}^{H}$ a $H_n^{(2)} = \frac{1}{k}\sum_{i=1}^{k}\left(ln\left(x_{n-i+1:n}\right) - ln\, x_{n-k:n}\right)^2$.

Minimum distance estimator

This method requires the definition of some metric distance $d[.]$ between the theoretical distribution $F(x, \theta)$, (θ unknown parameter) and the empirical distribution function $F_n(x)$. If there exists some $\hat{\theta}$ for which

$$d[F(x, \hat{\theta}), F_n(x)] = inf\{d[F(x,\theta), F_n(x)]; \theta\} \tag{7}$$

than $\hat{\theta}$ is called minimum distance estimation. In our case $F(x;\hat{\theta}) = G(x;\hat{\xi},\hat{\beta}) = G_{\hat{\xi},\hat{\beta}}(x)$ and suitable metric distances are presented in the following table 1.

Table 1.

Criterion	$d(F(x,), F_n(x))$
Kolmogorov–Smirnov	$sup\{\| F(x, \theta) - F_n(x)\|; x\}$
Cramér–von Mises	$\int(F(x, \theta) - F_n(x)^2 dF(x)$

4. Methods for choosing threshold

Various methods for the choice of threshold u are presented in monographs [1], [2] and [7]. We will discuss here four suitable criterions to find the optimal threshold.

Linearity of the mean excess function

We define the mean excess function as conditional mean value, depending on the threshold u, by formula

$$e(u)=E(X-u| X > u). \tag{8}$$

It can be shown for $F(x) \approx G_{\xi,\beta}(x)$ that $e(u)$ is linear function of u. More precisely

$$e(u) = \int_u^{\infty} \frac{1-F(y)}{1-F(u)}dy = \frac{\beta+\xi u}{1-\xi}, \tag{9}$$

for $\beta+\xi u > 0$, $\xi<1$.

Criterion: Choose u from such area where $e(u)$ is approximately linear.

Stability of Pickands plot

The Pickands plot is done by points $[k, \hat{\xi}_{k,n}^{P}]$ where $\hat{\xi}_{k,n}^{P}$ is the Pickands estimation defined by (4).

Criterion: Choose u from such area where the Pickands plot is approximately constant for increasing k.

Stability of Hill's plot

The Hill's plot is given by points [k, $\hat{\xi}^H_{k,n}$] where $\hat{\xi}^H_{k,n}$ is the Hill estimator defined by (5).

Criterion: Choose u from such area where the Hill's plot is stable (the estimation $\hat{\xi}^H_{k,n}$ is approximately constant for increasing k).

Minimization using median

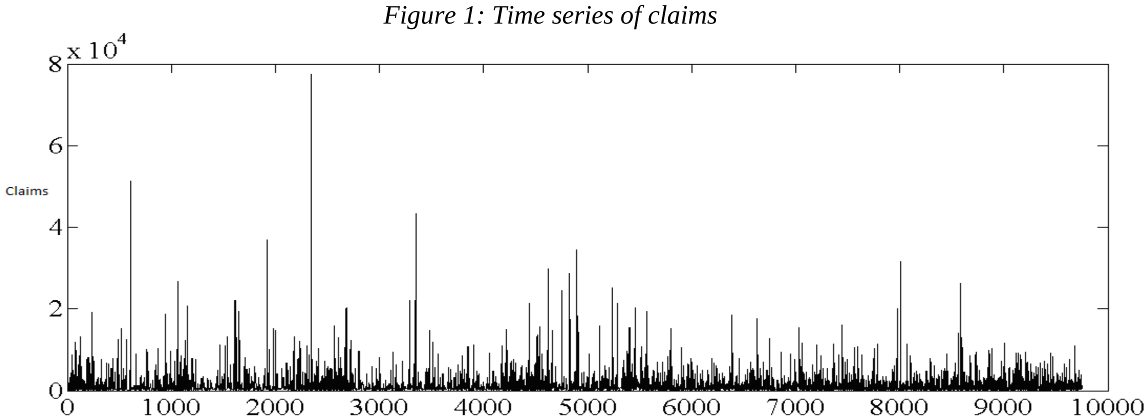
Denote the median of $\xi_{1,n},\xi_{2,n},\ldots,\xi_{k,n}$, with respect to the considered number of exceedances k as $med(\xi_{1,n},\xi_{2,n},\ldots,\xi_{k,n})$, where $\xi_{i,n}$, i = 1, 2, …, k are Hill estimations.
Criterion: The optimal number of exceedances we get by minimizing the following expression

$$\frac{1}{k}\sum_{i\le k} i^c \left|\xi_{i,n} - med(\xi_{1,n},\ldots,\xi_{k,n})\right| \text{ , for } 0 \le c \le 1/2. \qquad (10)$$

5. Statistical analysis of real data

In this section we will analyze 9748 car insurance claims from a Slovak insurance company over the period 1998-2008 (given in Euros). Our main goal will be to model the tail distribution of data, estimate the parameters of considered distribution and state the optimal high of reinsurance level (threshold u). The time series of data, plotted in Figure 1., give us useful information about the occurrence of extreme claims.

Figure 1: Time series of claims



We can see from Table 2, that the kurtosis of the distribution is essentially larger than 3 (which is the kurtosis of standard normal distribution), so the data have heavy tailed distribution.

Table 2: Basic characteristics of data

Count	Min	Max	Mean	Median	Variance	Skewness	Kurtosis
9748	150.2	77384.5	1348.458	571.015	6831693	8.5	134.997

By the Balkema, de Haan and Pickands theorem we need to estimate the parameters of generalized Pareto distribution $G_{\xi,\beta}$. Using the well know maximum likelihood method the estimated parameters are $\hat{\xi}_{MLE} = 0.345$, $\hat{\beta}_{MLE} = 3501.22$.
To calculate Pickands estimates we use program Maple and the following procedure.

```
PickandsOdhad := proc( )
global data, odhad;
local k, o, n;
odhad := [ ];
n := nops(data);
for k from 1 to n/4 do
    if (data[2·k] − data[4·k]) ≠ 0 then
        o := evalf( 1/ln(2) · ln( (data[k] − data[2·k]) / (data[2·k] − data[4·k]) ) );
        odhad := [op(odhad), [k, o]];
    end if;
 end do;
 end proc;
```

In Figure 2 and 3 we can see dependence of Pickands estimation on the number of considered extremes (exceedances). With respect to the stable part of the detailed graph we chose the number of extremes k = 1300. The corresponding value of the estimated parameter is $\hat{\xi}^{P}_{k,n} = 0.734498$.

Figure 2: Pickands plot

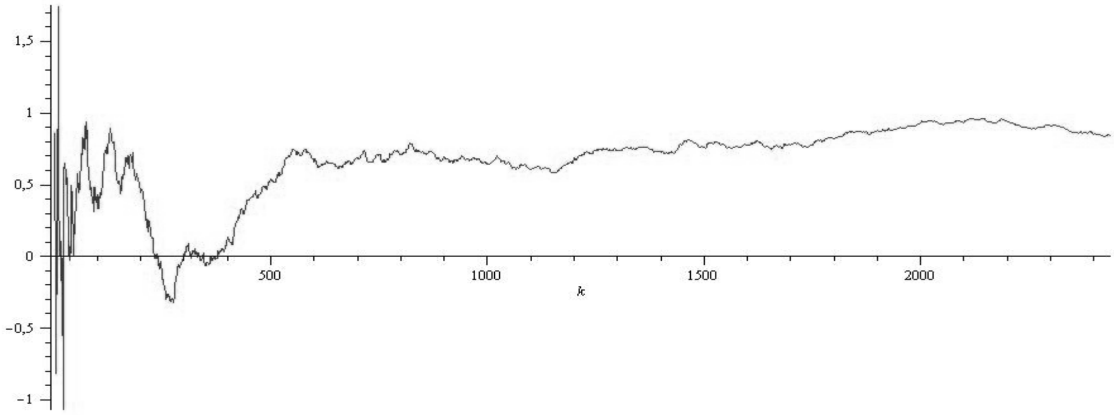
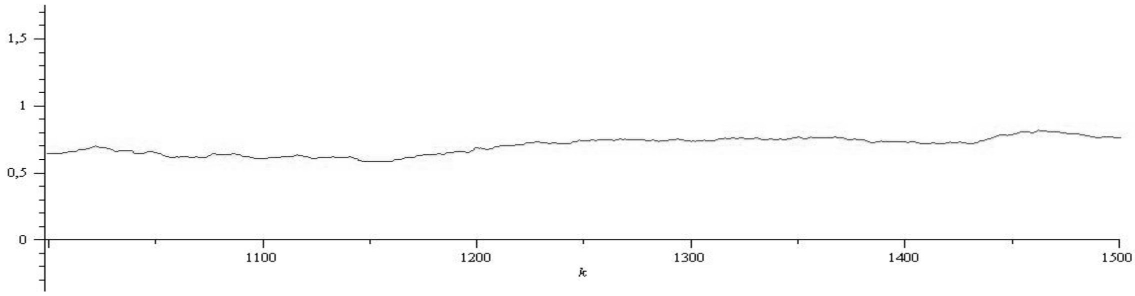
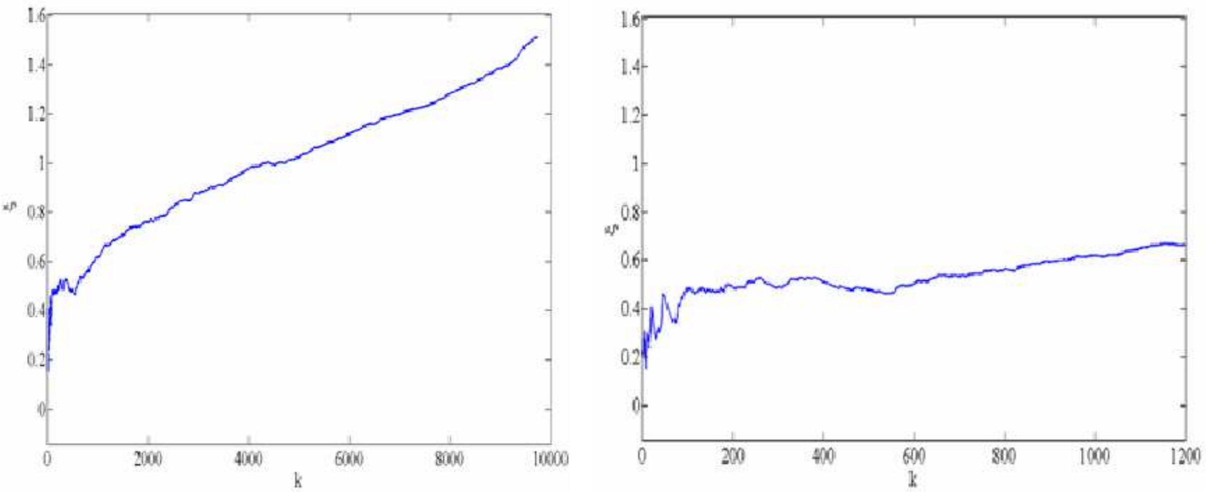


Figure 3: Pickands plot detail



The Hill estimation $\hat{\xi}^{H}_{k,n}$ was calculated by (5). In Figure 4 we have plotted the dependence of Hill's estimation on the number of extremes. Similarly as in the case of Pickands estimation we found the stable part of the graph for k = 380 and stated the value of estimated parameter $\hat{\xi}^{H}_{k,n} = 0.529$.

Figure 4: Hill's plot (left) and it's detail (right)

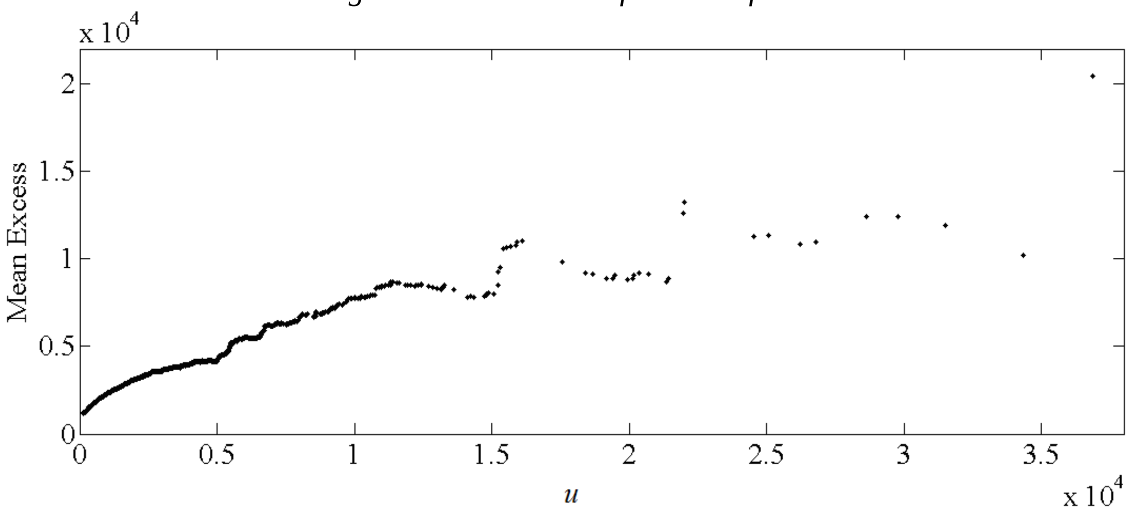


Using Hill estimations we obtain by (6) the Deckers – Einmahl – de Haan estimation for our data $\hat{\xi}^{D} = 0.398991$.

The calculation of minimum distance estimation was realized by Kolmogorov-Smirnov criterion using program Excel and its package EasyFitXL.
We get d[$G_{\hat{\xi},\hat{\beta}}(x)$, $F_n(x)$] = 0.03996 and the estimated parameter values are $\hat{\xi} = 0.30528$ and $\hat{\beta} = 3682.4$.

To state the optimal level of the threshold u, we consider the mean-excess plot [u, $e_n(u)$], where $e_n(u)$ is the empirical version of mean excess function (see Figure 5). We find here the linear (stable) part of the graph and choose the underlying level u = 5520.51 (k = 380).

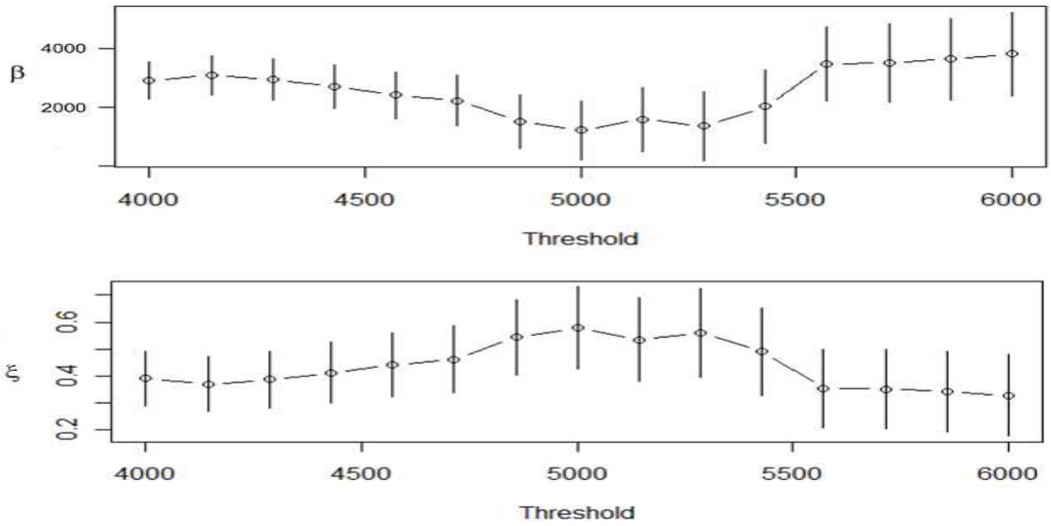
Figure 5: Mean excess function of data



Finding the optimal threshold by minimization of the expression in (10) leads to k = 634 and the overlying level will be u = 4331.71.

Except of the considered methods we could use for visualization of the evolution of parameters ξ, β as functions of u realized by program R and its package extremes. Again the parameter estimation will be from area were the graph is stable. By Figure 6 we choose $\hat{\xi} = 0.35$ and $\hat{\beta} = 3200$.

Figure 6: Evolution of parameters



The results of the parameter estimations are summarized in the following table.

Table3:

	$\hat{\xi}^{P}_{k,n}$	$\hat{\xi}^{H}_{k,n}$	$\hat{\xi}^{D}$	MDE	MLE
$\hat{\xi}$	0.734	0.529	0.399	0.305	0.345
$\hat{\beta}$	-	-	-	3682.4	3501.222

To verify the correspondence between the theoretical and empirical distribution we use Kolmogorov-Smirnov test. Some results are presented in table 4. A review of further suitable tests is published in [6].

Table4: Kolmogorov - Smirnov test

Sample size 380 Statistic 0.03996 p-Value 0.56502		
α	0.05	0.01
Critical value	0.06966	0.08357
Reject	No	No
Sample size 634 Statistic 0.06319 p-Value 0.0121		
α	0.05	0.01
Critical value	0.05393	0.0647
Reject	Yes	No

6. Conclusion

This paper deals with statistical analysis of extreme claims in non-life insurance. We presented various graphical and analytical tools to identify the tail distribution of data and estimate its unknown parameters. Our goal also was to determine the optimal level of reinsurance (high threshold).

The results show that, all types of parameter estimation methods lead to positive value of EVI, so the considered distribution is heavy tailed. Using graphical methods we estimated the suitable level of threshold u, but these methods are not sufficiently objective. By the results, the Pickands and the Hill estimations have large oscillations for low numbers of exceedances so they are not suitable for our data. In addition, the Hill estimator is very sensible to the dependence in data. In [3] a numerical method of threshold calculation is presented and further methods of estimation for parameter β and the following calculation of threshold u are our subject of research.

References

[1] Beirlant, J. at al, 2004. *Statistics of Extremes: Theory and applications*. Wiley, New York.

[2] Embrechts, P. at al, 1997. *Modelling extremal events for insurance and finance*, Springer-Verlag.

[3] Fecenko, J., Šeleng M, 2005. *Niektoré aspekty XL-zaistenia*. Zborník príspevkov z vedeckého seminára Poistná matematika v teórii a praxi, Bratislava, 12-24.

[4] Juhás, M., Skřivánková, V, 2010. *Analýza extrémnych hodnôt v poistení motorových vozidiel metódou POT*. Forum Statisticum Slovacum 5/2010, 97-102.

[5] Juhás, M., Skřivánková, V, 2011. *Odhady parametrov v modeli extrémnych hodnôt a ich výpočet.* Forum Statisticum Slovacum 7/2011, 77-83.

[6] Neves C., Fraga Alves M.I, 2008. Testing EV condition – an overview and recent approaches. REVSTAT – Statistical journal, Vol 6., No 1, 83-100.

[7] Reiss D.-R., Thomas M, 2007. *Statistical analysis of extreme values with applications to insurance, finance, hydrology and other fields.* Birkhäuser Verlag, Basel, 2007.

[8] Skřivánková V. 2006. *Štatistická analýza extrémnych hodnôt a metódy ich registrácie v neživotnom poistení.* 3. mezinárodní konference Řízení a modelování finančních rizik, Ostrava, 270-277.

[9] Skřivánková V., Tartaľová A. 2008, *Catastrophic risk management in non-life insurance*. E+M Economics and Management, 2/2008, 65-72.

Summary

EVT metódy ako nástroje riadenia rizika

V príspevku sa zaoberáme modelovaním excesov (presahov) nad dostatočne vysoký prah, ako špeciálnych prípadov extrémnych hodnôt. Odpovedajúce metódy POT a PORT sa využívajú pri riadení rizika neproporcionálnych zaistení typu XL (excess-of-loss) a CatXL (catastrophe excess-of-loss) tým, že umožnia stanoviť optimálnu hranicu zaistenia a tak minimalizovať riziko krachu. Venujeme sa metódam vhodným pre odhad parametrov rozdelenia chvosta dát, metódam na určenie optimálneho prahu a testom zhody teoretického a empirického rozdelenia. Uvažované metódy a postupy sú aplikované pri štatistickej analýze reálnych dát, pochádzajúcich z havarijného poistenia motorových vozidiel na slovenskom poistnom trhu.