Modelling of extreme losses in natural disasters

P. Jindrová, V. Pacáková

Abstract—The aim of this paper is to describe parametric curvefitting methods for modelling extreme historical losses of natural catastrophes in the world. Article summarizes relevant theoretical results Extreme value theory (EVT) and Excess over Threshold Method (EOT) and results of their application to the data about amounts of damages in world catastrophe events in time period 2010-2014, published by Swiss Re studies Sigma. We aim to develop the models for extreme catastrophic losses by selecting a particular probability distributions through statistical analysis of empirical data with the best possible estimate of the upper tail area.

Keywords—Block maxima models, excess over threshold method, extreme value distribution, generalized Pareto distribution.

I. INTRODUCTION

Catastrophic events affect various regions of the world with increasing frequency and intensity (Fig.1 and Fig.2). Many regions are threatened by catastrophic risks large range, where extensive disruptions are frequently, sometimes more than once a year. Large catastrophic events can be caused by natural phenomena or are caused by man (Fig.1). It should be noted that many events of natural character are to a large extent influenced by human activity. Serious events in recent years are often the result of terrorist acts.



Fig.1 Number of catastrophic events 1970-2014

P. Jindrová is with Institute of Mathematics and Quantitative Methods, Faculty of Economics and Administration, University of Pardubice, Pardubice, Studentská 84, 532 10 Pardubice, Czech Republic (e-mail: Pavla.Jindrova@upce.cz).

V. Pacáková is with Institute of Mathematics and Quantitative Methods, Faculty of Economics and Administration, University of Pardubice, Pardubice, Studentská 84, 532 10 Pardubice, Czech Republic (e-mail: Viera.Pacakova@upce.cz). According to the latest sigma study, global insured losses from natural catastrophes and man-made disasters were USD 35 billion in 2014, down from USD 44 billion in 2013 and well below the USD 64 billion-average of the previous 10 years. There were 189 natural catastrophe events in 2014, the highest ever on sigma records, causing global economic losses of USD 110 billion. (Swiss Re sigma No 2/2015, p. 4)



Fig.2 Total catastrophe losses 1970-2014

Catastrophe modelling is a risk management tool that uses computer technology to help insurers, reinsurers and risk managers better assess the potential losses caused by natural and man-made catastrophes. Natural catastrophe models use historical disaster information to simulate the characteristics of potential catastrophes and to determine the potential losses cost. The modelling process evolved in the late 1980s as companies became increasingly aware of their exposure to catastrophic risks. After hurricane Andrew in 1992 and the Northridge earthquake in 1994, use of catastrophe models took off as companies sought to more accurately analyze, write and price for natural catastrophe risk.

We are interested in probability modelling of extreme catastrophe losses, specifically the tails of loss severity distributions. Thus is of particular relevance in reinsurance if we are 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.

In the modelling of catastrophe events statistical methods are commonly used for inference from historical data. Extreme Value Theory (EVT) [2], [8], [9] emerges as a basic tool in modelling such risks. It began with the paper by Dodd in 1923,

followed by the paper Fisher and Tippett in 1928, after by the papers by de Finetti in 1932, by Gumbel in 1935 and by von Mises in 1936, to cite the more relevant; the first complete frame in what regards probabilistic problems is due to Gnedenko in 1943. Following the theoretical developments of the extreme value theory many scholarly papers, as (Zhongxian, 2003), (Skřivánková, Tartaľová, 2008), (Jindrová, Sipková, 2014), (Jindrová, Pacáková, 2016), (Pacáková, Kubec, 2012), (Pacáková, Brebera, 2015) dealing with the variety of practical applications of the theory were published. The Generalized Extreme Value (GEV), Gumbel, Frechet, Weibull, and the Generalized Pareto (GPD) distributions are just the tip of the iceberg of an entirely new and quickly growing branch of statistics. Various authors have noted that this theory is relevant to the modelling of extreme values [3], [8], [10].

II. EXTREME VALUE THEORY

There are two principal kinds of models for extreme values. The oldest group of models are the *block maxima* models; these are models for the largest observations collected from large samples of identically distributed observations. A more modern group of models are the *peaks-over-threshold* models; these are models for all large observations which exceed a high threshold.

A. Block Maxima Models

By [8] the Fisher-Tippett theorem [1] is the fundamental result in Extreme Value Theory (EVT) and can be considered to have the same status in EVT as the central limit theorem has in the study of sums. The theorem describes the limiting behaviour of appropriately normalized sample maxima.

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

Extreme Value Theorem (Fisher and Tippet, 1928):

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

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

where $M_n = max (X_1, ..., X_n)$, Y is non-degenerate with distribution function G. Then G is of one the following types:

1. Gumbel

$$\Lambda(x) = \exp\{-e^{-x}\}, \quad x \in R \qquad x \in R$$
2. Frechet

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

3. Weibull

$$\Phi_{\alpha}(x) = \begin{cases} \exp\{-(-x)^{\alpha}\} & \text{if } x < 0\\ 0 & \text{if } x \ge 0 \end{cases}$$

These three types of limiting distribution there are in standard form. We can parameterize them within the location and scale families.

1. Gumbel

$$\Lambda(x) = \exp\left\{-\exp\left[-\left(\frac{x-d}{c}\right)\right]\right\}, \quad x \in \mathbb{R}$$

2. Frechet

$$\Phi_{\alpha}(x) = \begin{cases} 0 & \text{if } x \le d \\ \exp\left\{-\left(\frac{x-d}{c}\right)^{-\alpha}\right\} & \text{if } x > d \end{cases}$$

3. Weibull

$$\Psi_{\alpha}(x) = \begin{cases} \exp\left\{-\left(-\frac{x-d}{c}\right)^{\alpha}\right\} & \text{if } x < d\\ 0 & \text{if } x \ge d \end{cases}$$

The generalized Gumbel, Frechet and Weibull families can be combined into a single family of the *generalized extreme value distributions* (GEV) in the form

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

It is straightforward to check the result by letting

$$\alpha = \frac{1}{\xi}, \quad d = \mu - \frac{\sigma}{\xi}, \quad c = \begin{cases} \frac{\sigma}{\xi}, & \text{when } \xi > 0\\ -\frac{\sigma}{\xi}, & \text{when } \xi < 0 \end{cases}$$
(2)

B. Excess over Threshold Method

The modelling using the excess over threshold method follows the assumptions and conclusions in Generalized Pareto Distribution (GPD) Theorem. Suppose $x_1, x_2, ..., x_n$ are raw observations independently from a common distribution F(x). Given a high threshold u, assume $x_{(1)}, x_{(2)}, ..., x_{(k)}$ are observations that exceed u. Here we define the ascendances as $x_i = x_{(i)} - u$ for i = 1, 2, ..., k.

Then for a large enough threshold u by the GPD Theorem (Pickands 1975) the generalized Pareto distribution (3) is the

limiting distribution for the distribution of the excesses, as the threshold tends to the right endpoint.

The conditional distribution function of Y = (X - u / X > u) is approximately

$$H(x) = 1 - \left(1 + \frac{\xi x}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}}$$
(3)

defined on

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

The family of distributions defined by equation (3) is called the *generalized 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.

By GPD Theorem x_i may be regarded as realization of independently random variable which follows a generalized Pareto family with unknown parameters ζ and σ . In case $\zeta \neq 0$, the likelihood function can be obtained directly from (4):

$$L((\xi, \sigma/\mathbf{x})) = \prod_{i=1}^{k} \left\lfloor \frac{1}{\sigma} \left(1 + \frac{\xi x_i}{\sigma} \right)^{-1/\xi-1} \right\rfloor$$
(4)

III. RESULTS OF MODELLING

C. Data and exploratory analysis

The analysis focus on 264 total losses of natural catastrophes ranging from 8 to 210 000 (in USD millions) in time period from January 2010 to December 2014, published in Swiss Re sigma 2011-2015.

Based on *Swiss re sigma* criteria [15, p. 2] an event is classified as a catastrophe and included in the sigma database when insured claims, total losses or the number of casualties exceed certain thresholds (Table 1).

Table 1: The sigma event selection criteria, 2014				
Insured loss thresholds				
Maritime disasters	19.6 million			
Aviation	39.3 million			
Other losses	48.8 million			
or Total economic loss threshold	97.6 million			
or Casualties				
Dead or missing	20			
Injured	50			
Homeless	2000			
(Sigma No 2/2015, p. 2)				

By Table 1 we have the historical catastrophe data on losses which exceed a certain amount known as a threshold, or the *transacted* losses. The time series plot (Fig.3) allows us to identify the most extreme losses and their approximate times of occurrences. We will start with exploratory analysis of the total catastrophe losses data. Summary statistics in Table 2 and Box and Whisker plot show that there are many small losses and a few very large values of losses. The conclusion is that we need to find some long tail distribution that provides a suitable model for the variation amongst the catastrophe losses data.



Fig.3 Chronologically arranged the total losses of natural catastrophes 2010-2014

Table 2: Summary Statistics for Total Losses

Count	264
Average	2997.82
Median	600.0
Standard deviation	14247.2
Coeff. of variation	475.251%
Minimum	8.0
Maximum	210000
Skewness	12.369
Stnd. skewness	82.0464
Kurtosis	172.897
Stnd. kurtosis	573.434



Fig.4 Box and Whisker plot of total catastrophe losses

D. Fitting lognormal distribution

This analysis shows the results of fitting a lognormal distribution to the data on total losses using Statgraphics Centurion XV statistical package. The estimated parameters of the fitted distribution are shown in Table 3. We have tested whether the lognormal distribution fits the data adequately by

Table 3: Estimated parameters of lognormal distribution

Lognormal
mean = 2033.88
standard deviation $= 5470.41$
Log scale: mean $= 6.56356$
Log scale: std. dev. $= 1.452$

Table 4 shows the results of tests run to determine whether total losses can be adequately modelled by a lognormal distribution. The chi-squared test divides the range of total losses into no overlapping intervals and compares the number of observations in each class to the number expected based on the fitted distribution.

The Kolmogorov-Smirnov test computes the maximum distance between the cumulative distribution of total losses and the CDF of the fitted lognormal distribution. In this case, the maximum distance is 0.0693061.

Table 4: Goodness-of-Fit Tests for total losses

Chi-Squared Test

	Upper	Observed	Expected	Chi-
	Limit	Frequency	Frequency	Squared
at or below	1000.0	164	156.74	0.34
	2000.0	46	44.57	0.05
	3000.0	11	20.40	4.33
	4000.0	11	11.49	0.02
	5000.0	9	7.24	0.43
	6000.0	4	4.91	0.17
	7000.0	5	3.50	0.64
	8000.0	2	2.60	0.14
	10000.0	1	3.53	1.82
above	10000.0	11	9.02	0.44

Chi-Squared = 8.36049 with 7 d.f. P-Value = 0.301882

Kolmogorov-Smiri	nov Test
------------------	----------

	Lognormal
DPLUS	0,0693061
DMINUS	0.051097
DN	0.0693061
P-Value	0.158378

Since the smallest P-value amongst the tests performed is greater than or equal to 0.05, we cannot reject the hypothesis that total losses comes from a lognormal distribution with 95% confidence.

We can also assess visually how well the lognormal distribution fits by Frequency Histogram (Fig.4) and Quantile-Quantile plot against lognormal distribution (Fig.5). The lognormal line is close to the data at the lower end, but deviates away from the data at higher values of losses.



Fig.5 Histogram and fitted lognormal distributions



Fig.6 Quantile-Quantile plot of the fitted lognormal distribution

Evidently, the upper tail of the lognormal distribution is not enough fat. We concluded that we need to find another distribution as a good probability model for extreme natural catastrophe losses.

Table 5 Select	ed quantiles	for catastrop	ohe losses
----------------	--------------	---------------	------------

Lower Tail Area (<=)	Lognormal
0.5	708.787
0.75	1887.32
0.9	4556.76
0.95	7722.4
0.99	20773.0

Table 5 presents the values of selected quantiles for the fitted lognormal distribution. The output indicates that the value of the fitted lognormal distribution below which we would find 95% values of catastrophe losses is 7 722,4 USD millions, and the value below which we would find 99% values of catastrophe losses is 20 773 USD millions, so 1% losses could exceed this value.

E. Block Maxima Models – results

Our aim is to obtain a good estimate of the severity distribution in the upper tail, or good fit distributions of the extreme losses exceed some high quantiles.

First we use the oldest group of models known as the *block maxima* models; these are models for the largest observations collected from large samples of identically distributed observations (Part II A).

The catastrophe losses data presented by Fig.3 we have divided into k blocks (Table 7) of essentially equal size n = 11, 16, 21, 24. The maximum of the values in these blocks contains Table 6. For this part of modelling we have used spreadsheet MS Excel.

Table 6: The maxima of the blocks of losse
--

max_11	max_16	max_21	max_24
52400	30000	30000	30000
6400	52400	52400	52400
210000	210000	6400	6400
5000	11000	210000	6150
16600	30000	11000	210000
70000	5000	30000	11000
6800	16600	4000	30000
10300	15000	5000	1560
12500	70000	16600	5000
3700	6800	15000	16600
7000	4960	70000	15000
	12500	2630	8000
	5000	6800	70000
	3700	4820	2630
	5970	10300	6800
	7000	12500	4720
		5000	4960
		3700	12500
		5150	2500
		7000	5000
		1000	3700
			5150
			5970
			7000

We do not have used some special software systems for EVT, but statistical software package Statistica 12 and Extremes Toolkit module of statistical program "R" allow us application of Fisher-Tippet Theorem. This theorem says that generalized extreme value (GEV) distribution is the natural normalized maxima. limit distribution for Module Distributions & Simulation of package Statistica 12 allows maximum likelihood estimation of three parameters and performing goodness of fit test with GEV distribution with distribution function (1) on maxima of blocks in Table 6. The results contain Table 7. By p-values for block maxima modelling by GEV using the Fisher-Tippet Theorem we can see very good fit for maxima of the all 4 blocks, but the best fit in case of n = 16 and n = 11.

Table 7: Results of block maxima modelling

	<i>n</i> = 11	<i>n</i> = 16	<i>n</i> = 21	<i>n</i> = 24
Number				
of blocks	<i>k</i> = 24	<i>k</i> = 17	<i>k</i> = 13	<i>k</i> = 11
MLE (µ)	0.26289	-0.16874	-0.01303	-0.11785
MLE (σ)	1.01372	1.03052	0.79455	0.76895
MLΕ (ξ)	0.15193	0.24680	0.54639	0.12996
SE (μ)	0.35092	0.30472	0.21462	0.18829
SE (σ)	0.27338	0.25096	0.21011	0.14826
SE (ξ)	0.26621	0.25941	0.29545	0.21639
<i>p</i> -value	0.99052	0.99260	0.89667	0.79892

The plots on Fig.7 and Fig.9 are useful for visual examining the fit of GEV distribution based on sample data. We have overlaid a theoretical CDF on the same plot with empirical distribution of the sample to compare them.

The black stair lines show the empirical distribution functions of empirical sample data and the blue curves present the theoretical CDF of the estimated generalized extreme values distributions (EVD) for different blocks of maxima. The red lines are the lower and upper bounds of the 95% confidence interval estimates of the CDF. It can be seen that the estimated EVD models falls inside the bands.



Fig.7 GEV distribution fitted to block maxima for n = 11



Fig.8 QQ-plot against the GEV distribution fitted to block maxima for n = 11



Fig.9 GEV distribution fitted to block maxima for n = 16



Fig.10 QQ-plot against the GEV distribution fitted to block maxima for n = 16

F. Results of Excess over Threshold Method

We have fitted a generalized Pareto distribution using the maximum likelihood method for parameters estimation to the data above threshold of 3000 (Fig.12), above threshold of 5000 (Fig.13) and above threshold 8000 (Fig. 15).

These plots are useful for examining the distribution based on sample data. We have overlaid a theoretical CDF on the same plot with empirical distribution of the sample to compare them. As in part *E* the black stair lines on Fig.11, Fig.13 and Fig.15 show the empirical distribution functions of empirical sample data and the blue curves present the theoretical CDF of the estimated generalized Pareto distributions for different thresholds. The red lines are the lower and upper bounds of the 95% confidence interval estimates of the CDF. It can be seen that the estimated parametric CDF falls inside the bands. These plots confirm the good fit of all three generalized Pareto distributions on total losses of natural catastrophes.

The QQ-plots (Fig.12, Fig.14, and Fig.16) against the generalized Pareto distributions is another way to examine visually the hypothesis that the losses which exceed a very high threshold come from estimated distributions.



Fig.11 GPD fitted to 42 exceedance of the threshold 3000



Fig.12 QQ-plot against the GPD fitted to 42 exceedances of the threshold 3000



Fig.13 GPD fitted to 22 exceedances of the threshold 5000



Fig.14 QQ-plot against the GPD fitted to 22 exceedances of the threshold 5000



Fig.15 GPD fitted to 11 exceedance of the threshold 8000



Fig.16 QQ-plot against the GPD fitte to 11 exceedances of the threshold 8000

Table 8 contains the maximum likelihood estimations of parameters fitted generalized Pareto distributions on the data above three different thresholds. By *p*-values in this table we can state the best fit in the case of threshold u = 5000.

Table 8: Comparisons of estimated GPD for different thresholds

	$u = 3\ 000$	$u = 5\ 000$	$u = 8\ 000$
parametr ζ	2842.322	4195.862	13706.81
parametr σ	-0.67771	-0.75417	-0.19003
<i>p</i> -value	0.850026	0.959389	0.760575

IV. CONCLUSION

The rising trends of number of catastrophic events and of total catastrophe losses in recent decades require that catastrophe models will continue to evolve amid the everchanging risk landscape. The probability models based on databases about the consequences of catastrophic events provide valuable information to the institutions with a focus on risk management and reduction of the impact of these events in the future.

The probability models for extreme catastrophic losses exceeding a certain amount are particularly important for insurance and reinsurance companies to price a high-excess layer.

We have shown that the generalized extreme values models and generalized Pareto distributions fit well to extremal losses of natural catastrophes and they are useful tools for estimating the tails of loss severity distributions. This is not altogether surprising. As we have explained in part I, using GVT and GPD models have solid foundations in the mathematical theory of the behavior of extremes and it is not simply a question of ad hoc curve fitting.

Even with a good tail estimate we cannot be sure that future does not hold some unexpected catastrophic loss. The extreme value methods which we have explained and applied do not predict the future with certainty, but they do offer good models for explaining the extreme events.

REFERENCES

- A. Brdar Turk, A Quantitative Operational Risk Management Model, Wseas Transactions on Business and Economics, Issue 5, Volume 6, 2009, pp. 241-253.
- [2] P. Embrechs, C. Kluppelberg, T. Mikosch, *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin 1997.
- [3] R. Fisher, L. Tippett, 'Limiting forms of the frequency distribution of the largest or smallest member of a sample', *Proceedings of the Cambridge Philosophical Society* 24, 1928, pp. 180-190
- [4] B. Gnedenko, 'Sur la distribution limite du terme maximum d'une s_erie al_eatoire', Annals of Mathematics 44, 1943, pp. 423-453.
- [5] P. Jindrová, V. Pacáková, Probability Models of Natural Catastrophe Losses. In *Proceedings of the 4th International Conference on Mathematical, Computational and Statistical Sciences* (MCSS'16), Barcelona, 2016, submitted for publication.
- [6] P. Jindrová, Ľ. Sipková, Statistical Tools for Modeling Claim Severity. Conference proceeding from 11th International Scientific Conference on European Financial Systems 2014. Location: Lednice, Masaryk University Brno, Czech Republic, 2014, pp. 288-294.
- [7] M. Käärik, M. Umbleja, On claim size fitting and rough estimation of risk premiums based on Estonian traffic insurance example, *International Journal of Mathematical Models and Methods in Applied Sciences*, Issue 1, Volume 5, 2011, pp. 17-24.
- [8] A. J. McNeil, Estimating the Tails of Loss Severity Distributions using Extreme Value Theory. ETH Zentrum, Zürich 1996 [online]. Available on:

https://www.casact.org/library/astin/vol27no1/117.pdf

- [9] A. McNeil, Estimating the Tails of Loss Severity Distributions Using Extreme Value Theory. ASTIN Bulletin, 27(1), 1997, pp. 117-138.
- [10] V. Pacáková, Modelling and Simulation in Non-life Insurance. Proceedings of the 5th International Conference on Applied Mathematics, Simulation, Modelling (ASM'11), Corfu Island, Greece, July 14-16, 2011, Published by WSEAS Press, p. 129-133.
- [11] V. Pacáková, L. Kubec, Modelling of catastrophic losses. *Scientific Papers of the University of Pardubice*, Series D, Vol. XIX, No. 25 (3/2012), pp. 125-134.
- [12] V. Pacáková, J. Gogola, Pareto Distribution in Insurance and Reinsurance. Conference proceedings from 9th international scientific conference *Financial Management of Firms and Financial Institutions*, VŠB Ostrava, 2013, pp. 298-306.
- [13] V. Pacáková, D. Brebera, Loss Distributions in Insurance Risk management. In: *Recent Advances on Economics and Business Administration*, Proceedings of the International Conference on Economics and Statistics (ES 2015), INASE Conference, Vienna, March 15-17, 2015, pp. 17-22.
- [14] V. Pacáková, D. Brebera, Loss Distributions and Simulations in General Insurance and Reinsurance. *International Journal of Mathematics and Computers in Simulation*. NAUN, Volume 9, 2015, pp. 159-167.
- [15] J. Pickands, 'Statistical inference using extreme order statistics', The Annals of Statistics 3, 1975, 119-131.
- [16] Natural catastrophes and man-made disasters in 2014, SIGMA No2/2015, Swiss Re [online]. Available on:

http://media.swissre.com/documents/sigma2_2015_en_final.pdf

[17] V. Skřivánková, A. Tartaľová, Catastrophic Risk Management in Nonlife Insurance. In *E&M Economics and Management*, 2008, No 2, pp. 65-72. [18] H. Zhongxian, Actuarial modelling of extremal events using transformed generalized extreme value distributions and generalized Pareto distributions, Doctoral thesis, The Ohio State University, 2003 [online]. Available on: <u>http://www.math.ohio-</u>

state.edu/history/phds/abstracts/pdf/Han.Zhongxian.pdf.

Mgr. P. Jindrová, **Ph.D.** graduated from Mathematical analysis at the Faculty of Science of Palacky University in Olomouc in 1993. She lectures mathematics, insurance and financial mathematics in branch of study of Insurance engineering and Management of financial risks. She finished her dissertation thesis which deals with problems of aggregation and disaggregation in economics and mathematics.

Prof. RNDr. Viera Pacáková, Ph.D. graduated in Econometrics (1970) at Comenius University in Bratislava, 1978 - RNDr. in Probability and Mathematical Statistics at the same university, degree Ph.D at University of Economics in Bratislava in 1986, associate prof. in Quantitative Methods in Economics in 1998 and professor in Econometrics and Operation Research at University of Economics in Bratislava in 2006. She was working at Department of Statistics, Faculty of Economic Informatics, University of Economics and Administration in Pardubice since 2005. She has been concentrated on actuarial science and management of financial risks since 1994 in connection with the actuarial education in the Slovak and Czech Republic and she has been achieved considerable results in this area.