SIMULATIONS OF THE EXTREME LOSSES IN NON-LIFE INSURANCE

Viera Pacáková, Bohdan Linda

Introduction

The analyses of insurance risks are an important part of the project of Solvency II preparing of European Commission. Modelling the size of loss is of crucial importance for an insurer. Particular attention is paid to studying the right tail of the distribution, since it is important to not underestimate the size (and frequency) of large losses.

The modelling of loss distributions in non-life insurance is one of the problem areas, where obtaining a good fit to the extreme tails of a distributional model is of major importance. The objective of this paper is to call attention to a new approach to statistical modelling using quantile functions. The use of models based on quantile methods provides an appropriate and flexible approach to the distributional modelling needed to obtain well-fitted tails. Modern computer simulation techniques open up a wide field of practical applications for this theory concept, without requiring the restrictive assumptions and sophisticated mathematics, of many traditional aspect of insurance risk theory.

The conditions under which claims are performed (and data are collected) allow us to consider the claim amounts in non-life insurance to be samples from specific heavy-tailed probability distributions. The Pareto distribution is often used as a model for insurance losses needed to obtain well-fitted tails.

1. Pareto Distribution in Non Life Insurance

Pareto distribution is commonly used to model claim-size distribution in insurance for its convenient properties.

Pareto random variable X has survival function $P(X>x)=1-F(x)=\left(\frac{\lambda}{\lambda+x}\right)^{\alpha}$ with positive parameters α and λ and density function $f(x)=\frac{\alpha\lambda^{\alpha}}{(\lambda+x)^{\alpha+1}}$ that are very flexible. Pareto quantile function $Q(p)$ or the inverse of the Pareto distribution function $F^{-1}(x)$ by [1, p.40] has the form

$$Q(p) = F^{-1}(x) = \lambda\,[(1-p)^{-1/\alpha}-1],\ 0<p<1. \tag{1}$$

When X is Pareto (α,λ) it is readily determine the mean $E(X)=\frac{\lambda}{\alpha-1}$ (when $\alpha>1$) and variance $D(X)=\frac{\alpha\lambda^2}{(\alpha-1)^2(\alpha-2)}$ (when $\alpha>2$). Then method of moments to estimate parameters α, λ is easy to apply. To equate the first two population and sample moments we find estimates:

$$\tilde{\alpha}=\frac{2s^2}{s^2-\bar{x}^2} \qquad \tilde{\lambda}=(\tilde{\alpha}-1)\bar{x} \tag{2}$$

The estimates $\tilde{\alpha}$, $\tilde{\lambda}$ obtained in this way tend to have rather large standard errors, because s^2 has a very large variance. We will obtain estimates of α and λ using maximum likelihood method.

We denote as α,λ the maximum likelihood estimates given data $x_1,x_2,\ldots x_n$ from the Pareto $(\alpha,\ \lambda)$ Solving equation $f(\lambda)=0$ using the initial estimate $\tilde{\lambda}$, where

$$f(\lambda)=A-B=\frac{\sum_{i=1}^{n}\frac{1}{\lambda+x_i}}{\sum_{i=1}^{n}\frac{x_i}{\lambda(\lambda+x_i)}}-\frac{n}{\sum_{i=1}^{n}\ln\left(1+\frac{x_i}{\lambda}\right)} \tag{3}$$

we obtain $\hat{\lambda}$. Substituting $\hat{\lambda}$ in A or B we find $\hat{\alpha}$.

The above mentioned definition of the Pareto distribution is the common used in America. The Pareto distribution with the distribution function at the form

$$F(x)=1-\left(\frac{c}{x}\right)^{\alpha}$$

is the common used definition of the Pareto distribution in Europe. By [8, p. 202] if X is "European" Pareto distributed with parameters c, α, then X-c is "American" Pareto distributed with parameters $(\lambda,\ \alpha)$.

Various tests may be used to assess the fit of a proposed Pareto model, for example Kolmogorov-Smirnoff and χ^2 goodness-of-fit test [4, p. 78-

83]. Other methods are mentioned in the publication [11], which addresses similar issues.

2. Simulation Using Quantile Function

We denoted a set of ordered sampling data of losses by

$x_{(1)}, x_{(2)},..., x_{(y)},..., x_{(n-1)}, x_{(n)}$.

The corresponding random variables are being denoted by

$X_{(1)}, X_{(2)},..., X_{(y)},..., X_{(n-1)}, X_{(n)}$.

Thus $X_{(n)}$ for example is the random variable representing the largest observation of the sample of n. The n random variables are referred as the n order statistics. These statistics play a major role in modelling with quantile distribution $Q(p)$.

Consider first the distribution of the largest observations on $X_{(n)}$ with distribution function denoted $F_{(n)}(x) = p_{(n)}$. By [2, p. 95-96] the probability

$$F_{(n)}(x) = p_{(n)} = P\,(X_{(n)} \leq x)$$

is also probability that all n independent observations on X are less then or equal to this value x, which for each one is p. By the multiplication law of probability

$$p_{(n)}(x) = p^n \text{ so } p = p_{(n)}^{1/n} \text{ and } F(x) = p = p_{(n)}^{1/n}.$$

Inverting $F(x)$, to get the quantile function, we have

$$Q_{(n)}(p_{(n)}) = Q(p_{(n)}^{1/n}).$$

For the general r-th order statistic $X_{(r)}$ the calculation becomes more difficult. The probability that the r-th larges observations is less than some value z is equal

$$p_{(r)} = F_{(r)}(z) = P(X_{(r)} \leq z).$$

This is also probability that at least r of the n independent observations is less or equal to z. The probability of s observations being less than or equal to z is p^s, where $p = F(z)$ is given by the binomial expression

$$P(s \text{ observations} \leq z) = \binom{n}{s} p^s\,(1-p)^{(n-s)}$$

and

$$p_{(r)} = \sum_{s=r}^{n} \binom{n}{s} p^s\,(1-p)^{(n-s)}.$$

This function is the incomplete beta function and is denoted by

$$p_{(r)} = I(p, r, n-r+1).$$

If it can be inverted, then we can write

$$p = BETAINV\,(p_{(r)}, r, n-r+1)$$

From the last two expressions we get

$$Q_{(r)}(p_{(r)}) = Q(BETAINV\,(p_{(r)}, r, n-r+1))$$

- *The order statistics distribution rule*

If a sample of n observations from a distribution with quantile function $Q(p)$ is ordered, then the quantile function of the distribution of the r-th order statistic is given by

$$Q_{(r)}(p_{(r)}) = Q(BETAINV\,(p_{(r)}, r, n-r+1)) \qquad (4)$$

BETAINV (.) is a standard function in packages such as Excel. Thus, the quantiles of the order statistics can be evaluated directly from the distribution $Q(p)$ of the data. A particularly useful application of this result lies in evaluating the medians of the distributions of ordered data. Thus, the median $M_{(r)}$ of the distribution of the r-th order statistics for $p_{(r)} = 0{,}5$ is $Q(BETAINV\,(0{,}99, r, n-r+1))$, 99th percentile of $X_{(r)}$ we get as and so $Q(BETAINV\,(0{,}99, r, n-r+1))$ on.

Probably all spreadsheet software, all statistical software and many pocket calculators provide the user with a simple way of generating random numbers. The basic random number is a number in interval [0, 1] that represents an observation on a continuous uniform distribution. In quantile language the quantile function is

$$S(p) = p,\ 0 \leq p \leq 1.$$

The generating mechanism is designed to produce a steam of approximately independent values $u_1, u_2, ..., u_n$ from the uniform distribution on the interval [0, 1]. We refer to the generation of random variables in such a fashion, and also to the use of such values in the investigation of a model of any type, as simulation. The basis of this is the relevance of the next two rules [9, p. 56-57].

- *The Q-transformation rule*

If $z = T(x)$ is non-decreasing function of x and $Q(p)$ is a quantile function, then $T(Q(p))$ is a quantile function too.

- *The uniform transformation rule*

If U has a uniform distribution then the variable X, where $x = Q(u)$ has a distribution with quantile function $Q(p)$. Thus data and distributions can be visualized as generated from the uniform distribution by transformation $Q(.)$, where $Q(p)$ is the quantile function.

This rule follows directly from the Q-transformation rule when it is observed that the quantile function of the uniform distribution is just p.

The uniform transformation rule shows that the values of x from any distribution with quantile function $Q(p)$ can be simulated as

$$x_i = Q(u_i),\ i = 1, 2, ..., n$$

where $u_1, u_2, ..., u_n$ are simulated from uniform distribution on the interval [0, 1]. The non-decreasing nature of $Q(.)$ ensures the proper ordering of the x.

The quantile function thus provides the natural way to simulate values for those distributions for which it is an explicit function of p.

3. Simulation of Extreme Values

In a number of applications of quantile functions in non-life insurance interest focuses particularly on the extreme observations in the tails of the data. Fortunately it is possible to simulate the observations in one tail without simulating the central values. We will present here how to do this.

Consider the right-hand tail. The distribution of the largest observation has been shown to be $Q(p^{1/n})$. Thus by [7, p. 96] the largest observation can be simulated as $x_n = Q(u_{(n)})$, where $u_{(n)} = v_n^{1/n}$ and v_n is a random number from interval [0, 1]. If we now generate a set of transformed variables by

$$\begin{aligned} u_{(n)} &= v_n^{1/n} \\ u_{(n-1)} &= (v_{n-1})^{\frac{1}{n-1}} \cdot u_{(n)} \\ u_{(n-2)} &= (v_{n-2})^{\frac{1}{n-2}} \cdot u_{(n-1)} \\ &\vdots \end{aligned} \qquad (5)$$

where the v_i, $i = n, n - 1, n - 2, ...$ are simply simulated set of independent random uniform variables, not ordered in any way. It will be seen from their definitions that u_i, $i = n, n - 1, n - 2, ...$, form a decreasing series of values with $u_{(i-1)} < u_{(i)}$.

In fact, values $u_{(i)}$ form an ordering sequence from a uniform distribution. Notice that $u_{(n)}$ once is obtained, the relations have the general form

$$u_{(m)} = (v_m)^{\frac{1}{m}} \cdot u_{(m+1)},\ m = n-1, n-2, ...$$

The order statistics for the largest observations on X are then simulated by

$$\begin{aligned} x_{(n)} &= Q(u_{(n)}) \\ x_{(n-1)} &= Q(u_{(n-1)}) \\ x_{(n-2)} &= Q(u_{(n-2)}) \\ &\vdots \end{aligned} \qquad (6)$$

In most simulation studies of n observations are generated and the sample analyses m times to give an overall view of their behaviour. A technique that is sometimes used as an alternative to such simulation, is to use a simple of ideal observations, sometimes called a profile. Such a set of ideal observations could be the medians $M_r\, u_{(n)} = v_n^{1/n}$, $r = 1, 2, ..., n$.

4. Demonstration Example

We will present the illustrative example of modelling losses by Pareto probability distribution. We have observed sample of 91 claim amounts (Kč) in a motor hull insurance portfolio. This data set is small relative to many which one may encounter in practice; however, it will provide a useful example of how one might search for a loss distribution to model typical claims and simulate the extreme losses.

The mean and standard deviation of this data have been calculated by $\bar{x} = 47111{,}17$ Kč and $s = 97044{,}05$ Kč.

We suppose the random variable X that is claim amount, is Pareto with distribution function

$$p = F(x) = 1 - \left(\frac{\lambda}{\lambda + x}\right)^{\alpha} \qquad (7)$$

Using the method of moments to estimate the parameters α, λ of a Pareto distribution to solve the equations (2) we get estimators that are $\tilde{\alpha} = 2{,}6167$, $\tilde{\lambda} = 73163{,}76$.

Of course, asymptotically maximum likelihood estimators are preferred. The maximum likelihood estimator $\hat{\lambda} = 37277{,}81$ is a solution of equation (3) by numerical method using tool Goal seek of Excel and $\hat{\alpha} = 1{,}7394$ has then be found as A or B from equation (3).

Now we can check whether the Pareto distribution with maximum likelihood estimators provides an adequate fit to the data using $\chi 2$ goodness-of-fit test. The $\chi 2$ statistic is computed as usual by

$$\chi 2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

with k-1-p degrees of freedom (k is the number of intervals and p is the number of estimated parameters). As O_i we denote the observed frequencies and as E_i expected frequencies supposed Pareto distribution. The result of goodness of fit test presents Table 1.

The Pareto model gives an excellent fit because of calculated value $\chi^2 = 0{,}63138$ is less than critical value $\chi^2_{0,95} = 9{,}48772$ with $4 = 7\text{-}1\text{-}2$ degrees of freedom.

The quantile function of Pareto distribution that is fit on our data set has by (1) the form

Tab.1: Observed and fitted values for the Pareto model

x	O_i	P_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$
- 5000	42	0,429683	39,6714	0,14021
- 10000	17	0,205483	18,4934	0,12060
- 15000	9	0,113084	10,1776	0,13625
- 20000	6	0,068441	6,1597	0,00414
- 30000	6	0,074673	6,7206	0,07726
- 450000	5	0,049385	4,4447	0,06938
Above 45000	6	0,059251	5,3326	0,08353
	91	1,000000	91,00000	**0,63138**

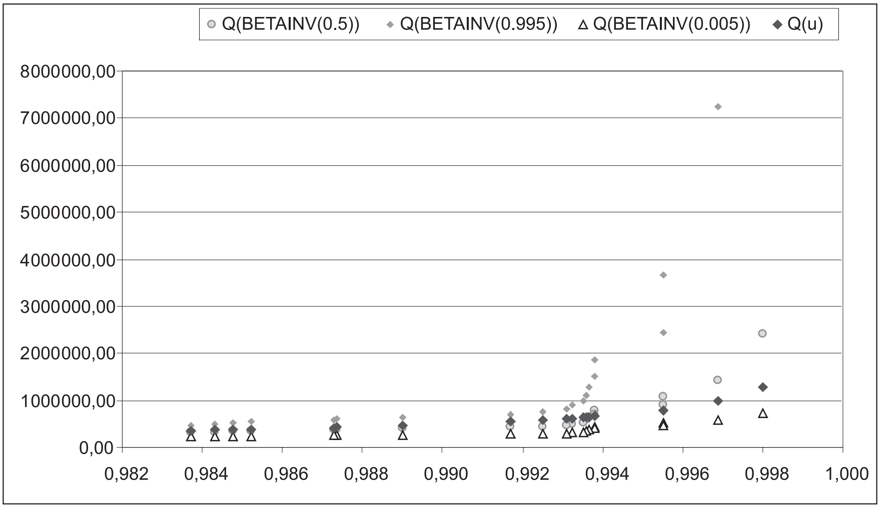
Source: Own calculations.

Tab. 2: Steps of simulation of the 20 largest losses

v	n	1/n	$v^{1/n}$	u	Q(u)
0,135493	1000	0,001	0,9980032	**0,9980032**	**1291697,514**
0,331321	999	0,001001	0,9988948	**0,9969002**	**994804,452**
0,253843	998	0,001002	0,9986272	**0,9955316**	**799110,676**
0,993465	997	0,001003	0,9999934	**0,9955251**	**798406,978**
0,180922	996	0,001004	0,9982849	**0,9938177**	**656697,334**
0,997123	995	0,001005	0,9999971	**0,9938148**	**656511,689**
0,855881	994	0,001006	0,9998434	**0,9936592**	**646673,091**
0,919813	993	0,001007	0,9999158	**0,9935756**	**641539,780**
0,943984	992	0,0010081	0,9999419	**0,9935178**	**638057,160**
0,76104	991	0,0010091	0,9997245	**0,9932441**	**622188,114**
0,865165	990	0,0010101	0,9998537	**0,9930988**	**614169,640**
0,561498	989	0,0010111	0,9994166	**0,9925194**	**584666,674**
0,436941	988	0,0010121	0,9991623	**0,991688**	**548102,839**
0,068052	987	0,0010132	0,9972808	**0,9889915**	**460784,822**
0,198585	986	0,0010142	0,9983619	**0,9873713**	**422982,371**
0,905523	985	0,0010152	0,9998993	**0,9872719**	**420910,870**
0,130303	984	0,0010163	0,9979311	**0,9852293**	**383336,723**
0,624701	983	0,0010173	0,9995215	**0,9847579**	**375807,461**
0,64864	982	0,0010183	0,9995593	**0,9843239**	**369193,331**
0,554228	981	0,0010194	0,9993986	**0,9837319**	**360622,519**

Source: Own calculations.

Fig. 1: Graphical result of simulation of the 20 extreme losses



Source: Graphical result of own calculations.

$$Q_{(p)} = 37277{,}81\,(1-p)^{-1/1{,}7394} - 37277{,}81 \qquad (8)$$

Table 2 contains the results of simulation of th e 20 largest values in sample of 1000 Pareto distributed losses with quantile function (8) step by step part 4 using terms (5) and (6).

On the Figure 1 we can see simulated values $x_i = Q(u_i)$, $i = 1000,999,...,991$ and the quantiles $x_{0,5}$, $x_{0,995}$, $x_{0,005}$ of the order statistics $x_{(1000)}$, $x_{(999)}$,... , $x_{(981)}$. Quantiles $x_{0,005}$ and $x_{0,995}$ give the bounds which the 20 largest values of Pareto distributed losses would exceed with probability only 0,01.

Simulation of *p* the largest claim amount in non-life insurance portfolio is useful in case of re-insurance. We can use this information in non proportional reinsurance of the types of LCR(*p*), when insurance company cedes p the largest amounts of loss to reinsurer, and ECOMOR, when reinsurance company pay losses that exceed p-th largest value in decreasing sequence of claim amounts [2, p. 128-129].

Other methods based on high order statistics, known as *Extreme Value Theory methods*, such as *methods of block –maxima, peaks-over-threshold approach*, or methods based on *Generalized Pareto distribution* are used in actuarial practice to model catastrophe risks. This approach is discussed for example in articles [6] or [10].

The paper is product part of the research project GAČR 402/09/1866: *Modelling, simulation and management of insurance risks.*

References

[1] BOLAND, P. J. *Statistical and Probabilistic Methods in Actuarial Science.* 1st ed. New York: Chapman&Hall/CRC, 2007. ISBN 1-58488-695-1.

[2] CIPRA, T. *Zajištění a přenos rizik v pojišťovnictví.* 1. vyd. Praha: GRADA Publishing, 2004. ISBN 80-247-0838-8.

[3] GILCHRIST, W. G. *Statistical Modelling with Quantile Functions.* 1st ed. London: Chapman & Hall/CRC, 2000. ISBN 1-58488-174-7.

[4] PACÁKOVÁ, V. *Aplikovaná poistná štatistika.* 3. prepracované a doplnené vydanie. Bratislava: IURA EDITION, 2004. ISBN 80-8078-004-8.

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

[6] PACÁKOVÁ, V., LINDA, B. Actuarial Modelling of Extreme losses. *Prace naukowe* Akademii Ekonomicznej im. Oskara Langego we Wrocławiu. Wocław: Wydawnictwo Akademii Ekonomicznej, 2006, pp. 369 -375. PL ISSN 0324-8445.

[7] PACÁKOVÁ, V., ŠOLTÉS, E. Modelling and Simulations of the Extreme Losses Using Quantile Functions. *Prace Naukowe* Nr 1037, Wydawnictwo AE im. Oskara Langego we Wrocławiu, 2004, pp. 91 – 99. PL ISSN 0324-8445.

[8] RYTGAARD, M. Estimation in the Pareto Distribution. *ASTIN Bulletin*, Vol. 20, No. 2, pp. 201-216, [online] [cit. 2008-12-10], <http://www.actuaries.org/LIBRARY/ASTIN/vol20no2/201.pdf/>.
[9] SIPKOVÁ, Ľ., SODOMOVÁ, E. *Modelovanie kvantilovými funkciami*. 1. vyd. Bratislava: Vydavateľstvo EKONÓM, 2007. ISBN 978-80-225-2346-2.
[10] SKŘIVÁNKOVA, V., TARTAĽOVÁ, A. Catastrophic Risk Management in Non-life Insurance. *E+M Ekonomie a Management*, 2008, roč. 11, č. 2, s. 65-72. ISSN 1212-3609.
[11] STEHLÍK, M., POTOCKÝ, R., FABIAN, Z., WALDL, H. *Some notes on the favourable estimation of fitting heavy tailed data*. Linz: IFAS res. Reports, Johannes Kepler University, 2008, No. 32.

prof. RNDr. Viera Pacáková, PhD.
Univerzita Pardubice
Fakulta ekonomicko-správní
Ústav matematiky
vpacakova@gmail.com

doc. RNDr. Bohdan Linda, CSc.
Univerzita Pardubice
Fakulta ekonomicko-správní
Ústav matematiky
Bohdan.Linda@upce.cz

Doručeno redakci: 25. 2. 2009
Recenzováno: 15. 4. 2009; 20. 7. 2009
Schváleno k publikování: 14. 9. 2009

ABSTRACT

SIMULATIONS OF EXTREME LOSSES IN NON-LIFE INSURANCE

Viera Pacáková, Bohdan Linda

The analyses of insurance risks are an important part of the project of Solvency II preparing of European Commission. The risk theory is the analysis of the stochastic features of non-life insurance business. The field of insurance risk theory has grown rapidly. There are now many papers and textbooks, which study the foundations of risk processes along strictly theoretical lines. On the other hand there is a need to develop the theories into forms suitable for practical purposes and to demonstrate their application. Modern computer simulation techniques open up a wide field of practical applications for risk theory concepts, without requiring the restrictive assumptions and sophisticated mathematics, of many traditional aspect of insurance risk theory..

Modelling the size or loss is of crucial importance for an insurer. Particular attention is paid to studying the right tail of the distribution, since it is important to not underestimate the size (and frequency) of large losses. The method of maximum likelihood is often used to estimate parameters of possible distributions, and various tests may be used to assess the fit of a proposed model (for example Kolmogorov-Smirnoff, and $\chi 2$ *goodness-of-fit. Often one may find that a mixture of various distributions may be appropriate to model losses due to varying characteristics of both the policies and policyholders.*

The objective of this article is to call attention to a new approach to statistical modelling using quantile functions. This approach can deal with many issues associated with the steps of the statistical modelling process based on quantile methods.

The definition and modelling of loss distributions in non-life insurance is one of the problem areas, where obtaining a good fit to the extreme tails of a distributional model is of major importance. It is a thesis of this article that the use of models based on quantiles provides an appropriate and flexible approach to the distributional modelling needed to obtain well-fitted tails.

We are specifically interested in modelling and simulations the tails of loss 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 losses.

Key Words: *loss distribution, quantile function, order statistics, simulation, Pareto distribution.*

JEL Classification: *C13, C15, C16, G22.*