# COLLECTIVE RISK MODEL IN HETEROGENEOUS PORTFOLIOS OF POLICIES

**Viera Pacáková, Ján Gogola, David Zapletal**

***Abstract***: *The total amount of claims in a particular time period, in actuarial literature named as collective risk, is a quantity of fundamental importance to the proper management of an insurance company. The article aimed to present the possibility and procedure to approximate the collective risk model in a heterogeneous portfolio of policies. The key assumption in all models for aggregate claim amount is that the occurrence of a claim and the amount of a claim can be studied separately. We will show that mixture distributions are convenient as the probability models for claim numbers and for claim amounts in heterogeneous portfolios of policies. We have derived that the negative binomial distribution can be used as a model for claim frequency and the Pareto distribution as a loss distribution model when the portfolios of policies are not homogeneous. The concept of mixture distributions is an important one in insurance, since insurance companies generally deal with heterogeneous risks. The motor compulsory third party liability insurance is an important branch of non-life insurance in many countries; therefore application of the theoretical results is performed on data from this field.*

***Keywords:*** *Collective risk model, Heterogeneous portfolio of policies, Mixture distributions, Negative binomial distribution, Pareto distribution.*

***JEL Classification:*** *C1, C6, C8, G.*

## Introduction

One of the primary goals of actuarial risk theory is the evaluation of the risk associated with a portfolio of insurance contracts. Many insurance contracts (in both life and non-life areas) are short-term, typically one year. Typical automobile insurance, homeowner's insurance and group life and health insurance policies are usually of one year duration.

The primary objective is to model the distribution of total claim costs for portfolios of policies so that business decisions can be made regarding various aspects of the insurance contracts. The total claim cost over a fixed time period is often modelled by considering the frequency of claims and distributions of individual losses separately. A few actuarial publications considered a number of different possible claim size distributions and discussed estimation procedures for them, for example [1], [2], [5], [6], [10], [13].

Insurance companies generally deal with heterogeneous or nonhomogeneous portfolios of policies, shortly with heterogeneous risks, which mean that the distribution of the claim numbers and sometime the distribution of individual claim amounts are differ among policyholders.

In the most of collective risk models practical examples in actuarial literature there is assumption that claim numbers have a Poisson distribution. This is because Poisson distribution generally leads to simple results than other claim number distributions and real data to verify this assumption are often not available. For the Poisson model of number of claims $N$ with parameter $\lambda$ applies equality of $E(N) = D(N) = \lambda$, but expectation the heterogeneity across policies to result in greater variation than the Poisson would give.

Since insurance companies generally deal with heterogeneous risks, in our article we first theoretically derive distribution of the number of claims and distribution of the individual losses in heterogeneous portfolios using mixture distributions. Then we explain the theoretical approach to the approximation of the collective risk model in the case of heterogeneous risks.

Finally we verify the theoretical results on the real data of liability motor insurance, which provides us with an unnamed insurance company operating on the Czech insurance market.

## 1 Risk models in heterogeneous portfolios of policies

As already mentioned, we assume a heterogeneous portfolio of insurance policies. Our aim is to derive the distribution of the number of insurance claims and distribution of individual losses in case of heterogeneous risks and knowledge of these distributions use for collective risk model.

The concept of a mixture of distributions [2, p. 25], [1, p. 58], [10, p. 98] will be useful for our goals.

### 1.1 Modelling the number of claims

The Poisson model for the number of claims is a plausible model for a single risk. We would expect the heterogeneity across policies to result in greater variation than the plain Poisson would give.

Let $N_i$ is the number of claims made by $i$-th policyholder during the year. The simplest model for claim numbers is the Poisson, so assume that the number of claims $N_i$ made by policyholder $i$ follows a Poisson distribution with mean $\lambda_i$,

$$N_i / \lambda_i \sim Po(\lambda_i)$$

i.e.

$$P(N_i = k) = \frac{\lambda_i^k}{k!} e^{-\lambda_i}, \quad k = 0, 1, 2, ... \tag{1}$$

We know that the values of $\lambda_i$ will vary across the portfolio and so we apply the method of mixtures to the $\lambda_i$. We take $G(\alpha;\beta)$ as the mixing distribution for the Poisson means. So

$$N_i / \lambda_i \sim Po(\lambda_i), \text{ where } \lambda_i \sim G(\alpha;\beta) \text{ for } i = 1, 2, ..., n.$$

Random variable $\lambda$ has a gamma distribution with parameters $\alpha$ and $\beta$, if

$$f(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}, \ \lambda > 0 \tag{2}$$

We will show that the marginal distribution $P(x) = P(N = x)$ of the number $N$ of claims in the whole portfolio of policies during the year is negative binomial $NB(\alpha;\pi)$ with parameters $\alpha, \pi$.

Probability mass function is given by

$$P(N = x/\alpha,\, \pi) = \frac{\Gamma(\alpha + x)}{\Gamma(\alpha)\, x!} \pi^{\alpha} (1-\pi)^{x},\; x = 0, 1, 2, \ldots, \tag{3}$$

The following gives this solution.

$$P(x) = P(N = x) = \int_0^{\infty} f_{N,\lambda}(x, \lambda) d\lambda = \int_0^{\infty} f_{\lambda}(\lambda)\, f_{N/\lambda}(x/\lambda)\, d\lambda =$$

$$= \int_0^{\infty} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \cdot e^{-\lambda} \frac{\lambda^{x}}{x!}\, d\lambda = \frac{\beta^{\alpha}}{\Gamma(\alpha) x!} \int_0^{\infty} \lambda^{\alpha+x-1} e^{-\lambda(\beta+1)}\, d\lambda =$$

$$= \frac{\beta^{\alpha}}{\Gamma(\alpha)\, x!} \cdot \frac{\Gamma(\alpha + x)}{(\beta+1)^{\alpha+x}} = = \frac{\Gamma(\alpha + x)}{\Gamma(\alpha)\, x!} \left(\frac{\beta}{\beta+1}\right)^{\alpha} \left(\frac{1}{\beta+1}\right)^{x}$$

That is by (3) probability function of the negative binomial distribution $NB(\alpha;\pi)$, where $\pi = \dfrac{\beta}{\beta+1}$.

It is possible to derive that [2, p. 31]

$$E(N) = \frac{\alpha(1-\pi)}{\pi} < \frac{\alpha(1-\pi)}{\pi^{2}} = D(N) \tag{4}$$

To conclude, negative binomial distribution is more appropriate model of claims number in heterogeneous portfolios of policies than Poisson distribution, for which $E(N) = D(N) = \lambda$ .

### 1.2 Modelling the claim amounts

The exponential distribution [1, p. 36], [10, p. 56] is one of the simplest models for individual claim amounts with density function

$$f(x) = \lambda_i\, e^{-\lambda_i x}, \quad x > 0,\, \lambda > 0 \tag{5}$$

Suppose that each individual in a large insurance portfolio incurs losses according to an exponential distribution. Our practical knowledge of almost any insurance portfolio (car insurance, home insurance, housing insurance, equipment insurance etc.) tells us that the means of these various distributions will differ among the policyholders. Thus our description of the losses in the portfolio is that each loss follows its own exponential distribution, i.e. the exponential distributions have means which differ from individual to individual [3], [6], [14].

We must now find a description of the variation among the individual means. One way to do this is to assume that the exponential means themselves follow a distribution. In the exponential case, it is convenient to make the following assumption.

Denote the average claim for each policyholder as $\theta_i$, where $i = 1, 2, \ldots, n$ . Let $\lambda_i = \dfrac{1}{\theta_i}$ be the reciprocal of the mean loss for the *i*-th policyholder. We assume that the variation among the $\lambda_i$ can be described by a known gamma distribution $G(\alpha;\beta)$ of the variable $\lambda$, i. e. $\lambda \sim G(\alpha;\beta)$ with probability density function by (2).

The marginal distribution of individual claims *X* in a whole portfolio of policies in this case is [2, p. 29], [10, p. 101]:

$$f_X(x)=\int_0^\infty f_{X,\lambda}(x,\lambda)\,d\lambda=\int_0^\infty f_\lambda(\lambda)\,f_{X/\lambda}(x/\lambda)\,d\lambda=$$

$$=\int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}\times\lambda e^{-\lambda x}d\lambda=\frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty \lambda^\alpha e^{-(x+\beta)\lambda}d\lambda=$$

$$=\frac{\beta^\alpha}{\Gamma(\alpha)}\cdot\frac{\Gamma(\alpha+1)}{(x+\beta)^{\alpha+1}}\int_0^\infty\frac{(x+\beta)^{\alpha+1}}{\Gamma(\alpha+1)}\lambda^\alpha e^{-(x+\beta)\lambda}d\lambda=\frac{\alpha\beta^\alpha}{(x+\beta)^{\alpha+1}} \qquad (6)$$

That is density function of the Pareto distribution $Pa(\alpha;\beta)$ [8], [15]. This result gives us a very nice interpretation of the Pareto distribution. $Pa(\alpha;\beta)$ arises when exponentially distributed losses are averaged using a $G(\alpha;\beta)$ mixing distribution.

### 1.3 Modelling the collective risk

When we derived that the probability model for the number of claims in heterogeneous portfolio of policies is negative binomial $NB(\alpha;\pi)$ and the probability model for the individual losses is Pareto $Pa(\alpha;\beta)$, we are going to construct the probability model for the random variable *S*, so called the *collective risk model* as it is defined in [1], [4], [10], [11], [12].

Recall that the random variable *S* denotes the aggregate claims paid by the insurer in the year in respect of this risk. When the variable *N* denotes the number of claims and the random variable $X_i$ denotes the amount of the *i*-th claim, thus

$$S=X_1+X_2+...+X_N \qquad (7)$$

where $X_1,X_2,...,X_N$ are independent, identically distributed variables, $N,X_1,X_2,...,X_N$ are mutually independent, and if $N=0$ than $S=0$.

The distribution of *S* is an example of a compound distribution. When *N* is negative binomial, *S* has a compound negative binomial distribution. If we denote as *G*(*s*) distribution function of *S* and *F*(*x*) the distribution function of $X_i$, so $G(s)=P(S\le s)=F_S(s)$ and $F(x)=P(X_i\le x)$. The *k*-th moment of $X_i$ about zero for $k=1,2,3,...$, will be denoted as $m_k=E(X_i^k)$.

We will focus on determining the approximate collective risk model for *S*, not exact expression for *G*(*s*). For the approximate methods we need to know the moments of *S*. Basic expressions, known in actuarial literature, for example derived in [1], [3], [4], [5], [10], we can write as

$$E(S)=E(N)m_1$$

$$D(S) = E(N)\left(m_2 - m_1^2\right) + D(N)\, m_1^2 \tag{8}$$

$$M_S(z) = M_N\left(\ln M_X(z)\right)$$

By the formulas (8) we can calculate all basic moments of $S$ if we know the relevant moments of $N$ and $X$.

We now determine the basic formulae for the mean, variance and skewness of the compound negative binomial distribution for $S$, which are derived in [1, p. 83]:

$$E(S) = \frac{\alpha \cdot (1-\pi)}{\pi} \cdot m_1 \tag{9}$$

$$D(S) = \frac{\alpha \cdot (1-\pi) \cdot \left(\pi \cdot m_2 + (1-\pi) \cdot m_1^2\right)}{\pi^2} \tag{10}$$

$$skew(S) = \frac{\pi^2 \cdot (1-\pi) \cdot m_3 + 3 \cdot \pi \cdot (1-\pi)^2 \cdot m_1 \cdot m_2 + 2 \cdot (1-\pi)^3 \cdot m_1^3}{\sqrt{\alpha} \cdot \left(\pi \cdot (1-\pi) \cdot m_2 + (1-\pi)^2 \cdot m_1^2\right)^{1,5}} \tag{11}$$

Note that the compound negative binomial distribution is always positively skewed. By section 2.2 the individual losses in heterogeneous portfolios of policies are distributed by Pareto $Pa(\alpha;\beta)$ with density function (6). To be able to substitute to formulae (11), (12), (13), the $k$-th moment about zero, $k = 1, 2, 3$ of $X \sim Pa(\alpha;\beta)$ we need. The expressions for these moments by [1], [6] and [9] are

$$m_1 = E(X) = \frac{\beta}{\alpha - 1}, \qquad \alpha > 1 \tag{12}$$

$$m_2 = E(X^2) = \frac{2\beta^2}{(\alpha-1)(\alpha-2)}, \qquad \alpha > 2 \tag{13}$$

$$m_3 = E(X^3) = \frac{6\beta^3}{(\alpha-1)(\alpha-2)(\alpha-3)}, \qquad \alpha > 3 \tag{14}$$

## 2 The translated gamma approximation to *G(s)*

Now we will use the results of Chapter 2 to find approximations for the distribution of collective risk $S = X_1 + X_2 + ... + X_N$.

Because $X_1, X_2, ..., X_N$ are independent, identically distributed variables, when number of claims $N$ is reasonable large, a normal approximation to the distribution of $S$ may be used by Central limit theorem. But the normal distribution can take negative values and is symmetric, while aggregate claims S are always nonnegative and in case of compound negative binomial distribution always positively skewed.

One alternative to the normal approximation which does not have these deficiencies is translated gamma distribution. We denote $G(\alpha;\beta)$ the gamma distribution with mean $\frac{\alpha}{\beta}$, variance $\frac{\alpha}{\beta^2}$, and skewness $\frac{2}{\sqrt{\alpha}}$.

Let $\mu$, $\sigma^2$ and $\gamma$ denote the mean, variance and coefficient of skewness of $S$. We assume $S$ has approximately the same distribution as the random variable $k + Y$ where $k$ is a constant and $Y$ has a gamma distribution $G(\alpha, \beta)$. The parameters $k$, $\alpha$ and $\beta$ are chosen so that $k + Y$ has the same first three moments as $S$.

Equating the coefficients of skewness, variance and means of $S$ with the same characteristics of $k + Y$ gives the following three formulae:

$$\gamma = \frac{2}{\sqrt{\alpha}} \qquad \sigma^2 = \frac{\alpha}{\beta^2} \qquad \mu = k + \frac{\alpha}{\beta} \tag{15}$$

from which $\alpha$, $\beta$ and k can be calculated.

## 3 Methods for practical application

The Chapter 2 and Chapter 3 contain all the necessary information to allow us to find the translated gamma approximation to collective risk $S$ based on the real data of insurance company about the number and the amount of claims for each policyholder.

Such data are required to estimate the parameters of the negative binomial distribution of number of claims and parameters of the Pareto distribution of individual losses in a heterogeneous portfolio of policies. The goodness of fit tests can be used to verify whether data may come from the expected distributions with estimated parameters.

The method of moment is easy to use to estimate parameters of negative binomial distribution $NB(\alpha;\pi)$ from expression of $E(N)$ and $D(N)$ by (4) if we estimate $E(N)$ by sample mean and $D(N)$ by sample variance [7, p. 100], [2, p. 32].. Maximum likelihood estimates [2], [10] provide some statistical software packages, for example Statgraphics Centurion.

The method of moments [1, p. 41] to estimate parameters $\alpha$ and $\beta$ of Pareto loss distribution $Pa(\alpha;\beta)$ is very easy to apply given data $x_1, x_2, ..., x_n$ of individual losses with sample average $\overline{x}$ and sample standard deviation $s$:

$$\tilde{\alpha} = \frac{2s^2}{s^2 - \overline{x}^2} \qquad \tilde{\beta} = (\tilde{\alpha} - 1)\overline{x} \tag{16}$$

The estimates obtained in this way will tend to have rather large standard errors and so the maximum likelihood estimators of parameters of $Pa(\alpha;\beta)$ are preferred [1, p. 41], [10, p. 72]. We denote as $\hat{\alpha}$, $\hat{\beta}$ the maximum likelihood estimates given data $x_1, x_2, ..., x_n$ from the $Pa(\alpha;\beta)$ distribution. If log-likelihood function we denote as $\ell$ we find $\partial\ell/\partial\alpha$ and $\partial\ell/\partial\beta$. We equate the two expressions for parameter $\hat{\alpha}$ and find maximum likelihood estimator $\hat{\beta}$ satisfies $f(\beta) = 0$ where [2, p. 16], [10, p. 72]

$$f(\beta) = A - B = \frac{\sum_{i=1}^{n} \frac{1}{\beta + x_i}}{\sum_{i=1}^{n} \frac{x_i}{\beta(\beta + x_i)}} - \frac{n}{\sum_{i=1}^{n} \ln\left(1 + \frac{x_i}{\beta}\right)} \tag{17}$$

Substituting $\hat{\beta}$ in A or B we can find $\hat{\alpha}$.

Pareto model often gives an excellent fit when we need to find a distribution with rather more weight tail what is typically feature in a heterogeneous portfolios of the policies.

Various statistical tests may be used to check the fit of a proposed model, the most often *Chi-Squared test* [2], [10], [16].

*Chi-Squared test* divides the range of $X$ into $k$ intervals and compares the observed counts $O_i$ (number of data values observed in interval $i$) to the number expected given the fitted distribution $E_i$ (number of data values expected in interval $i$).

Test statistics is given by

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{18}$$

which is compared to a chi-squared distribution with $k - p - 1$ degrees of freedom, where $p$ is the number of parameters estimated when fitting the selected distribution.

## 4 Application of the theoretical results

Practical application of theoretical results mentioned in previous chapters we provide based on data from unnamed Czech insurance company. We know the individual data of the number of claims and of the claim amounts (in CZK) in the portfolio of 25 752 policyholders for compulsory third party liability motor vehicles insurance.

This application at first presents the results of fitting a Poisson and a negative binomial distribution to the data on variable $N$ that is the number of claims made by a policyholder in this portfolio.

***Tab. 1: Estimation of Poisson parameter $\lambda$***

25752 values ranging from 0.0 to 3.0

Fitted Distributions

| *Poisson* |
|---|
| mean = 0.0409289 |

*Source: Output from Statgraphics Centurion XV*

***Tab. 2: Parameters estimation of $NB(\alpha;\pi)$***

25752 values ranging from 0.0 to 3.0

Fitted Distributions

| *Negative Binomial* |
|---|
| event probability = 0.900157 |
| number of successes = 0.369005 |

*Source: Output from Statgraphics Centurion XV*

The output of the procedure *Goodness-of-Fit Tests* in Tab. 3 shows the results of chi-squared test run to determine whether $N$ can be adequately modelled by a Poisson distribution. The chi-squared test divides the range of $N$ into 3 groups and compares the number of observations in each class to the number expected based on the fitted Poisson distribution. It is obvious that the Poisson distribution gives a very poor fit to the higher claim frequencies. Since the smallest $p$-value amongst the tests performed is less than 0.05, we can reject the hypothesis that $N$ comes from a Poisson distribution.

***Tab. 3: Results of Chi-squered test for Poisson distribution of N***

| | Lower Limit | Upper Limit | Observed Frequency | Expected Frequency | Chi-Squared |
|---|---|---|---|---|---|
| at or below | | 0.0 | 24773 | 24719.28 | 0.12 |
| | 1.0 | 1.0 | 909 | 1011.73 | 10.43 |
| | 2.0 | 3.0 | 70 | 20.99 | 114.44 |

Chi-Squared = 124.984 with 1 d.f. $p$-value = 0.0

*Source: Output from Statgraphics Centurion XV*

Tab. 4 shows the results of fitting a negative binomial distribution to the data on $N$. The estimated parameters of the fitted distribution are shown in Tab. 2. It means the results of tests run to determine whether $N$ can be adequately modelled by a negative binomial distribution with estimated parameters.

***Tab. 4: Results of Chi-squered test for negative binomial of N***

| | Lower Limit | Upper Limit | Observed Frequency | Expected Frequency | Chi-Squared |
|---|---|---|---|---|---|
| at or below | | 0.0 | 24773 | 24771.61 | 0.00 |
| | 1.0 | 1.0 | 909 | 912.65 | 0.01 |
| | 2.0 | 2.0 | 65 | 62.37 | 0.11 |
| | 3.0 | | 5 | 5.37 | 0.03 |

Chi-Squared = 0.150921 with 1 d.f. $p$-value = 0.697657

*Source: Output from Statgraphics Centurion XV*

The chi-squared test in Tab. 4 divides the range of 4 classes and compares the number of observations in each class to the number of expected based on the fitted negative binomial distribution. The fit of the negative binomial to the higher frequencies is very good. Since the $p$-value = 0.697657 the test performed is greater than 0.05, we do not reject the hypothesis that $N$ comes from a negative binomial distribution. These results correspond with the sub-chapter 2.1.

According to derived equation (6) can be assumed that the individual losses are Pareto distributed. We have used the data about 979 claim amounts in our portfolio and we have fitted the Pareto model to the data. The calculations of chi-squared test procedure by (18) for fitted Pareto distribution have completely done in Excel and results of calculations present Tab. 5.

Maximum likelihood method gives us the estimators of two parameters of Pareto distribution by solution equation (17) with help of *Solver*:

$$\hat{\alpha} = 5.608 \text{ and } \hat{\beta} = 7511.3 \qquad (19)$$

Now we can perform chi-squared test on the Pareto distribution with maximum likelihood estimators of parameters by (18).

***Tab. 5: Chi-squared test for the Pareto model***

| ***UL*** | $O_i$ | ***F*(*UL*)** | $p_i$ | $E_i$ | $\chi^2$ |
|---|---|---|---|---|---|
| 2000 | 741 | 0.733906 | 0.7495805 | 733.8393 | 0.069873 |
| 4000 | 155 | 0.908755 | 0.1748483 | 171.1765 | 1.528706 |
| 6000 | 46 | 0.962843 | 0.0540884 | 52.95259 | 0.912863 |
| 8000 | 20 | 0.982867 | 0.0200241 | 19.60361 | 0.008015 |
| 10000 | 9 | 0.991322 | 0.0084542 | 8.276653 | 0.063218 |
| over 10000 | 8 | | 0.0086785 | 8.496214 | 0.028981 |
| **Total** | **979** | | | | **2.611656** |

*Source: own calculations*

Tab. 5 contains the chi-squared test procedure based on real data. In the headings there are: *UL* –upper limit of the intervals, $O_i$ – observed frequencies, $E_i$ – expected frequencies, *F*(*UL*)- value of the Pareto distribution function in upper limit *UL* of the intervals, $p_i$ - probability that random amount of claim is just from *i*-th interval under Pareto model with parameters estimated by maximum likelihood method. The $\chi^2$ statistic is computed by (18), and we get

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(O_i - E_i\right)^2}{E_i} = 2.612$$

We have found *critical value* $\chi^2_{0.95}$ = 7.815 as a 95th percentile of $\chi^2$ distribution with 6-1-2 = 3 degrees of freedom and *p*-value = 0.45545, using the *Insert Function* in Excel, which indicate good fit for the Pareto distribution, because $\chi^2$ = 2.612< 7.815 and *p*-value is greater than 0.05. This result is also consistent with theoretical findings in sub-chapter 2.2.

We have confirmed a good fit with negative binomial distribution with parameters in Tab. 2 of variable *N* and a good fit with Pareto distribution with parameters (19) estimated by maximum likelihood methods of variable *X*, so we can calculate the *k*-th moments about zero, $k = 1, 2, 3$ for $Pa(\alpha;\beta)$ with parameters (19) by expressions (12), (13), (14). Tab. 6 contains the results of calculations.

***Tab. 6: The moments about zero for the Pareto model***

| | |
|---|---|
| m1 | 1630.0288 |
| m2 | 6786732.5496 |
| m3 | 58636257773.6073 |

*Source: own calculations*

Now we know all the necessary values to be used for the formulas (9), (10), (11) to calculate three basic characteristics of collective risk *S* in Tab. 7.

***Tab. 7: The basic characteristics of collective risk model S***

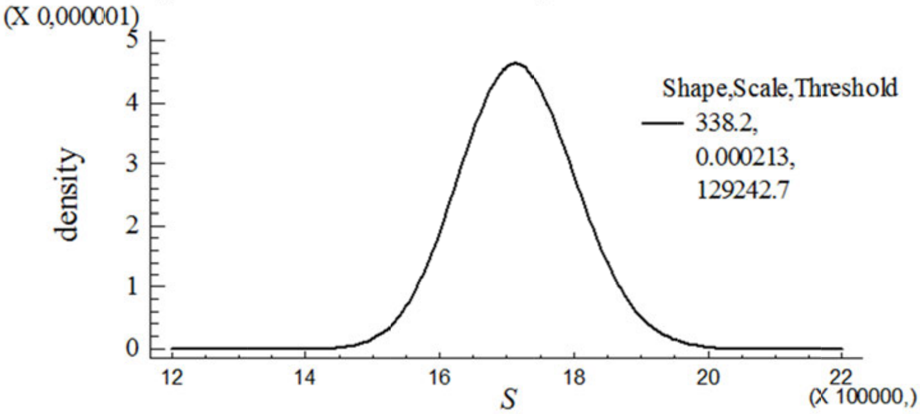
| | |
|---|---|
| $E(S)$ | 1718058.0962 |
| $D(S)$ | 7463870619.6988 |
| $skew(S)$ | 0.1088 |

*Source: own calculations*

Now we can calculate parameters of the translated gamma distribution (Fig. 1) which has the three relevant basic characteristics same as distribution of collective risk *S*. From (15) we get:

$$\alpha = 338.2 \qquad \beta = 0.000213 \qquad k = 129242.7$$

***Fig. 1: Density of translated gamma distribution***




*Source: own processing in the Statgraphics*

## 5 Discussion

The reason for approximating the distribution of collective risk *S* in insurance company by a translated gamma distribution is that it may be easier to obtain quantiles (percentiles) for $(k+Y)$ such as for *S*, or values for quantities such as $P(a < k+Y < b)$ than for $P(a < S < b)$. The values of cumulative distribution function and quantiles for gamma distributions are readily obtained from most statistical computer packages or from Excel.

The some high percentile, the most often 95-th $S_{0,95}$ of the collective risk model determines for an insurance company so called *risk premium RP*. The risk premium in our heterogeneous portfolio of policies we get as

$$S_{0,95} = k + Y_{0,95} = 129\,242.7 + 1\,733\,540.578 = 1\,862\,783.284 \text{ CZK}$$

## Conclusion

The theoretical results in Chapter 2 and demonstration of their application to real data in Chapter 5 confirmed that the number of claims *N* in a heterogeneous portfolio of policies has a negative binomial distribution and so collective risk *S* has a compound negative binomial distribution. We have shown procedure of approximation of collective risk model by translated gamma distribution in case when individual losses are Pareto distributed according to the results in subchapter 2.2. These results are very useful for insurance practice since insurance companies deal mainly with heterogeneous risks. Knowledge of the collective risk model for the insurance company is particularly useful for determining the risk premium and for reserves estimation.

**References**


[1] BOLAND, P. J. *Statistical and Probabilistic Methods in Actuarial Science*. New York: Chapman&Hall/CRC, 2007, pp. 35-61.

[2] CURRIE, I. D. *Loss Distributions*. Institute of Actuaries and Faculty of Actuaries, London and Edinburgh, 1993.

[3] DAYKIN, C. D., PENTIKÄINEN, T, PESONEN, M. *Practical Risk Theory for Actuaries*, Chapman & Hall, London, 1996.

[4] DICKSON, D.C.M., WATERS, H. R. *Risk models*. Institute of Actuaries and Faculty of Actuaries, London and Edinburgh, 1993.

[5] GRAY, R. J., PITTS, S. M. *Risk Modelling in General Insurance*. Cambridge University Press, Cambridge, 2012.

[6] Hogg, R.V., KLUGMAN, S. A. *Loss Distributions*. John Wiley& Sons, New York, 1984.

[7] HOSSACK, I. B., POLLARD, J. H., ZEHNWIRTH, B. *Introductory statistics with applications in general insurance*. Cambridge University Press, Cambridge, 1992.

[8] JINDROVÁ, P., SIPKOVÁ, Ľ. Statistical Tools for Modeling Claim Severity. In: *European Financial Systems* 2014. Proceedings of the 11th International Scientific Conference. Lednice, June 12-13, 2014. Brno: Masaryk University, 2014, pp. 288-294. ISBN 978-80-210-7153-7.

[9] KRISHNAMOORTHY, K., *Handbook of Statistical Distributions with Applications*. Chapman & Hall/CRC, USA, 2006, ISBN 1-58488-635-8.

[10] PACÁKOVÁ, V. *Aplikovaná poistná štatistika*. Iura Edition, Bratislava, 2004.

[11] PACÁKOVÁ, V. Modelovanie a simulácia rizík v neživotnom poistení. *E+M – Ekonomie a Management*, Technická univerzita v Liberci, 3/2007, ročník X, s. 109-121, ISSN 1212-3609.

[12] PACÁKOVÁ, V., BALCÁRKOVÁ V. Možnosti aproximace rozdělení kolektivního rizika. *Scientific Papers of the University of Pardubice*. Series D, Fakulta ekonomicko-správní 18(3/2010). s. 138-144. ISSN 1211 – 555X, MK ČR E 19548.

[13] PACÁKOVÁ, V. a kol. *Modelování a simulace pojistných rizik*. Vydavatelství Univerzity Pardubice, Pardubice, 2012.

[14] PACÁKOVÁ, V., ZAPLETAL, D. Mixture Distributions in Modelling of Insurance Losses. In: Proceedings of the 2013 International Conference on *Applied Mathematics and Computational Methods in Engineering*. (AMCME 2013). Rhodes Island, Greece, July 16-19, 2013, pp. 122-125. ISBN: 978-1-61804-200-2

[15] PACÁKOVÁ, V., GOGOLA, J. 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. ISBN 978-80-248-3172-5.

[16] PACÁKOVÁ, V., BREBERA, D. Loss Distributions and Simulations in General Insurance and Reinsurance. *International Journal of Mathematics and Computers in Simulation.* NAUN, Volume 9, 2015, pp. 159-167.

## Contact Address

**prof. RNDr. Viera Pacáková, PhD.**
**RNDr. Ján Gogola, PhD.**
**Mgr. David Zapletal, PhD.**
University of Pardubice, Faculty of Economics and Administration
Institute of Mathematics and Quantitative Methods
Studentská 84, 532 10 Pardubice, Czech Republic
Email: Viera.Pacakova@upce.cz, Jan.Gogola@upce.cz, David.Zapletal@upce.cz

Received: 01. 05. 2016
Reviewed: 01. 06. 2016, 17. 07. 2016
Approved for publication: 08. 09. 2016