On claim size fitting and rough estimation of risk premiums based on Estonian traffic insurance example

MEELIS KÄÄRIK
University of Tartu
Institute of Mathematical Statistics
J.Liivi 2-513
ESTONIA
meelis.kaarik@ut.ee

MERILI UMBLEJA
University of Tartu
Institute of Mathematical Statistics
J.Liivi 2-513
ESTONIA
a31965@ut.ee

Abstract: Financial and actuarial mathematics offer various problems related to estimation of distributions. Classical models for premium calculations usually require some estimates for both the distribution of individual claim size and also the number of claims. In this work we mainly consider the problem of estimation of individual claim size, but also some basics on the fitting of the distribution of claim number and tools to find rough estimates for risk premium are provided to complete the model. Most of the ideas are applied to a real-life data from Estonian traffic insurance from mid 2006 to mid 2007. The research was initiated by Estonian Traffic Insurance Fund and therefore is of practical importance. The first four sections of the article focus on the distribution of the individual claim size, we search answers for questions like:

- what candidate distributions to use for fitting the data?
- what fitting techniques to use?
- how to measure which of the proposed candidates is best?

We choose five commonly used distributions as possible estimates: lognormal, Pareto, Weibull, beta and gamma. The fitting techniques are based on moment matching or maximum likelihood estimators. For testing goodness of fit (GOF) several classical tests including Chi-square test and Kolmogorov-Smirnov test are used. The accuracy of our approach is evaluated by matching the first and second moments and by plotting PDF-s and CDF-s. The last section of the article focuses on estimation of the the claim amount for the whole portfolio and also describes a simple idea how the standard deviance principle can be used to find a first rough estimate for risk premium when the available history is limited. The estimates for risk premiums are found by the classical collective risk model. Several simplifications are made due to the lack of information, turns out that the resulting estimates are comparable with those used in practice by insurance companies.

Key–Words: estimation of distributions, heavy-tailed distributions, goodness-of-fit tests, collective risk model, premium calculation

I. INTRODUCTION

To offer insurance with reasonable price for both the insurer and the insured, the insurer must somehow model the overall behaviour of the claims. Important aspect to notice is that although by description a classical distribution fitting problem, the focus and topics related to insurance data are somewhat different to those of ”regular” mathematical statistics. This work can be viewed as a preliminary work to introduce the specifics of insurance data and to open the problems related to classical distribution fitting tests and techniques. We propose a simple method consisting of few classical tests and, during a case study, point out the limitations of this approach and the possible directions to work further. We refer to [6],[7] and [9] for most of the techniques used here. In this paper we study the Estonian traffic insurance claims data from period 01.07.2006 - 30.06.2007 (obtained from Estonian Traffic Insurance Fund) and propose some methods to find out which distribution suits best to describe the individual claim sizes. Fitting by well-known (and suitable for insurance data) distributions like lognormal, Pareto, Weibull, gamma and beta distributions is observed. The goodness of fit of a distribution is determined by analyzing the statistical tests, probability plots, quantiles and estimating suitability by visual estimation. The fitting of distributions is done separately for whole dataset and for vehicle types with most claims. In the visual estimation all histograms of data showed the best fit to the lognormal distribution. Unfortunately the statistical tests rejected the null hypothesis (that theoretical distribution suits to describe the data) in most cases. Test rejection is most likely caused by the large sample size and large-scale deviation. The probability plots and the table of quantiles showed the same result. Although all distributions were rejected, the values of test-statistics for lognormal distribution were again best compared to other distributions (in some cases also the beta distribution showed good fit for tail). In conclusion, we choose lognormal distribution to describe the claim size data and are aware that one must be careful while making decisions concerning the tail probabilities.

In the final section of the article we also address the problems related to the number of claims, to the aggregated claims of a portfolio, and to risk premiums. We make use of the well-known collective risk model and apply it to same set of data that was used before for estimation of individual claim size distributions. The main problem we are focusing is the case when the history is very limited and the estimation of claim numbers is therefore not straightforward and needs additional assumptions. The resulting rough estimates for risk premiums are found using standard deviation principle and provide an acceptable starting point for more thorough research.

II. Problems

A. Distributions

It is worth mentioning that the fitting problem for claims size distributions (and generally in insurance mathematics) is quite different from the fitting problem in classical statistics, since the importance in insurance mathematics lies on the tail [2], [7], [15]. Therefore good fit near expectation is not enough, we need to extra carefully study the tail fit. Most of the claim size distributions in insurance mathematics usually belong to the class of heavy-tailed distributions (particularly those with heavy right tail), or, even more specifically, to the class of subexponential distributions. Note also that besides the insurance related topics, heavy-tailed distributions are also used in a wide range of areas, most common examples are survival analysis and reliability analysis, but they can also be used in internet traffic models [13] or even automatic music generation [1]. We will now briefly recall the definitions of these classes of distributions [7].

Definition 1 (Heavy-tailed distribution): The distribution of a random variable X is said to have a heavy right tail if

$$\lim_{x\to\infty} e^{\lambda x}\boldsymbol{P}(X>x)=\infty \quad \text{for all } \lambda>0.$$

We may go to more detail and define an important subclass of heavy-tailed distributions, called long-tailed distributions, as follows.

Definition 2 (Long-tailed distribution): The distribution of a random variable X is said to have a long right tail if for all $t>0$,

$$\lim_{x\to\infty}\boldsymbol{P}(X>x+t|X>x)=1.$$

From here we can go even further, obtaining the class of subexponential distributions, where most of the common claim distributions used in insurance practice (Pareto, log-normal, Weibull, but also Burr, log-gamma, transformed beta) belong.

The work is supported by Estonian Science Foundation Grant No 7313.

Definition 3 (Subexponential distribution): Let us have a random variable with distribution function F such that $F(x)<1$ for all $x>0$. Then F is a subexponential distribution function if the following condition holds:

$$\lim_{x\to\infty}\frac{1-F^{n*}(x)}{1-F(x)}=n,\ n\geq 2,$$

where F^{n*} is the n-fold convolution of F.
Subexponential distributions are a subclass of long-tailed distributions.

In the following we shortly introduce the distributions that are used in current study.

- Lognormal distribution, $LnN(\mu,\sigma)$, with probability density function (pdf)
  $$f(x)=\frac{1}{\sigma x\sqrt{2\pi}}e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}$$
  for $x>0$. Lognormal distribution is a subexponential distribution, also very applicable in practice because of the close relations to thoroughly studied normal distribution. The parameters μ $(-\infty<\mu<\infty)$ and σ $(\sigma>0)$ are the expectation and the standard deviation of the corresponding normal distribution, respectively, i.e. if $X\sim LnN(\mu,\sigma)$ then $Y=\ln X\sim N(\mu,\sigma)$.
- Weibull distribution, $W(c,\sigma)$, pdf
  $$f(x)=\frac{cx^{c-1}}{\sigma^c}e^{-\left(\frac{x}{\sigma}\right)^c}$$
  for $x>0$. Weibull distribution is also a subexponential distribution, widely used in survival analysis, reliability analysis and also in insurance mathematics. The parameter c $(c>0)$ is referred to as a shape parameter and the parameter σ $(\sigma>0)$ is a scale parameter.
- Pareto distribution, $Pa(\alpha,\lambda)$, pdf
  $$f(x)=\frac{\alpha\lambda^\alpha}{(x+\lambda)^{\alpha+1}}$$
  for $x>0$. Because of its tail behavior, Pareto distribution is also an obvious choice for crucial and conservative models. The parameter α $(\alpha>0)$ is a shape parameter and the parameter λ $(\lambda>0)$ is a scale parameter. Pareto distribution also belongs to the class of subexponential distributions.
- Beta distribution, $B(\alpha,\beta,\sigma)$, pdf
  $$f(x)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{x^{\alpha-1}(\sigma-x)^{\beta-1}}{\sigma^{\alpha+\beta-1}}$$
  for $0<x<\sigma$. The three parameters offer much variety and also several combinations of those provide distributions suitable for insurance practice, also transformed beta distribution belongs to the class of subexponential distributions. The parameters α $(\alpha>0)$ and β $(\beta>0)$ are shape parameters and the parameter σ $(\sigma>0)$ is a scale parameter.

- Gamma distribution, $\Gamma(\alpha, \sigma)$, pdf

$$f(x) = \frac{1}{\sigma\Gamma(\alpha)} \left(\frac{x}{\sigma}\right)^{\alpha-1} e^{-\frac{x}{\sigma}}$$

  for $x > 0$. Gamma distribution is mainly considered here since some of the previous studies indicated it could have good fit as well, the similarities and differences between gamma, lognormal and Weibull are also studied in [14]. The parameter α ($\alpha > 0$) is a shape parameter and the parameter σ ($\sigma > 0$) is a scale parameter.

B. Estimation

After choosing the class of distributions we focusing on, the next step is to estimate their parameters. Parameters are found mainly by maximum likelihood estimators, using The SAS System and The R project software, the maximum likelihood estimates are found using the iterative Newton-Raphson or Nelder-Mead method [4], [12], [16].

Next step is to evaluate the goodness of fitting, we are using several criteria to find out which of the proposed theoretical distributions suit best. First impression comes from visual estimation of the data, but this is of course unreliable for making any thorough decisions [10].

Secondly, we use two classical goodness-of-fit tests, but keep in mind that they may not suit best to current problems, so more insurance-specific solutions could work better (a subject for later works). More details about these tests can be found, e.g., in [3] and [8].

- Kolmogorov-Smirnov test: can be troublesome with huge amount of data, since it reacts to each small difference. The problem is quite uncharacteristic for usual setup in statistics, where the main problem is to have enough data to get the asymptotic properties to work correctly. In our setup one could say that we have ”too much data”.
- Chi-squared test: because of the nature of the data (very many small claims, rare huge claims) we cannot have classes with equal claim interval, instead we take classes with equal probability.

Lastly, we also use the method of probability plots, where we (graphically) compare the empirical and theoretical quantiles. The closer the empirical and theoretical curves are, the better the estimation.

In short we can describe our proposed simple method for finding the best candidate distributions for further studies as follows:

A. choose a suitable class of distributions (using general or prior information about the specific data);
B. estimate the parameters (by finding maximum likelihood, e.g., with Newton-Raphson method);
C. estimate the goodness of fit:
   1) visual estimation;
   2) classical goodness-of-fit tests (Kolmogorov-Smirnov, chi-squared with equiprobable classes);
   3) probability plots.

Similar ideas can be used in other fields as well, see, e.g., how distribution fitting problems are addressed in the analysis of wind speed data [17], also the theory of extreme values can be used to solve related problems [11].

III. Estimation of claim distributions, a case study

A. Description of the data

The data is obtained from the Estonian Traffic Insurance Fund and it contains all Estonian traffic claims from 01.07.2006 to 30.06.2007.

In the following we give brief overview of the data:

- 39 306 observations (claims) in total;
- average claim per accident is 22 448 EEK;
- most claims lie in region 5 000 - 15 000 (EEK), 15 claims over 1 mln EEK (3.4 % of total claim size);
- 8 different types of vehicles, cars (77.1%), small trucks (6.9%), trucks (4.5%), trailer trucks (3.9%), buses, trolleys, trams (1.8%), tractors (0.6%), motorcycles (0.5%), trailers (0.1%). For 4.6% of the observations the vehicle type was not given.

B. Estimation of the claim distribution for all claims

We will now start to compare how well different distributions fit to our claims data. As it is hard to see the behaviour and shape of the claim distribution from full diagram, we will present a figure of the region where most of the claims lie (but keep in mind that the tail should be treated with extra care).

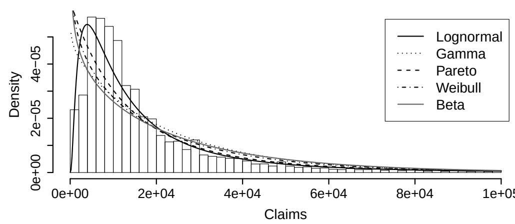

Fig. 1. Fitting different distributions to whole data

From Figure 1 it seems that lognormal distribution follows the data best, it is also suitable for both small and large claims. It is hard to compare the tail fit, but clearly the rest distributions have high discrepancies at small and frequent claims region.

In the following we present the parameters for estimated distributions and the results of goodness of fit tests.

Table 1. Estimates of parameters for different distributions (whole data)

Distribution	Scale parameter	Shape parameter(s)
Lognormal	$\mu = 9.40$	$\sigma = 1.06$
Weibull	$\sigma = 2.05 \cdot 10^4$	$c = 0.89$
Pareto	$\lambda = 6.35 \cdot 10^4$	$\alpha = 4.01$
Beta		$\alpha = 0.83$, $\beta = 176.42$
Gamma	$\sigma = 2.38 \cdot 10^4$	$\alpha = 0.95$

Table 2. Results of K-S and χ^2 tests

Distribution	χ^2-stat.	χ^2 crit. value	K-S statistic	K-S p-value
Lognormal	2 043	14.1	0.0653	$p < 0.001$
Weibull	8 776	14.1	0.1174	$p < 0.001$
Pareto	7 661	14.1	0.1255	$p < 0.001$
Beta	10 772	14.1	0.1477	$p < 0.001$
Gamma	8 074	14.1	0.1278	$p < 0.001$

By visual observation, the best-fitting distribution is lognormal, but as seen from the table, both Kolmogorov-Smirnov and χ^2 test reject all proposed distributions. This result is somewhat expected, as we have a very large dataset, and K-S test reacts to each small discrepancy. It is now suitable to recall the famous quote by prof. George E. Box: *"All models are wrong, but some are useful"* [5]. So the goal that remains is to take the best choice out of bad choices. Note that the values of test statistics are by far best for lognormal distribution (χ^2-statistic is about 4 times smaller than the rest and Kolmogorov-Smirnov statistic is 2 times smaller). We now also measure the performance of proposed theoretical distributions by comparing the corresponding probability plots, best fit is again by lognormal distribution, with gamma, Pareto and Weibull distributions behaving slightly worse. The corresponding probability plots are shown on figures 2-5.

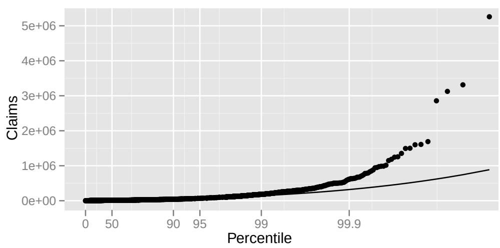

Fig. 2. Probability plot for lognormal distribution

It can be seen that large amount of the sample suits quite well with the proposed theoretical distributions, but after 0.99-quantile the theoretical tail is too light (estimate is too optimistic), fitting by Pareto distribution gives more conservative estimate for tail, but it is also too conservative for most of the data. For more information, also corresponding quantile tables are included in Appendix 1.

These results would actually be really good for most situations, but in insurance we have to be extra careful

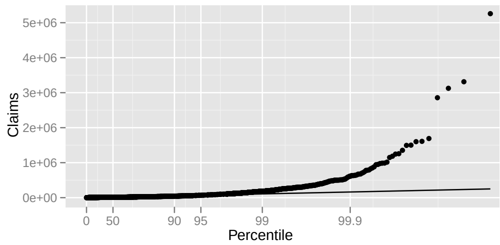

Fig. 3. Probability plot for gamma distribution

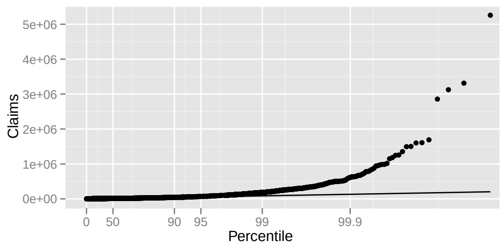

Fig. 4. Probability plot for Weibull distribution

with the tail estimation.

To get more similar claim behaviour and to achieve more homogeneous data structure we also conduct the same study by vehicle type. The results are presented in the next section.

C. Estimation of the claim distributions for different vehicle types

Similarly to the previous section we follow the same general procedure for each vehicle type separately. In this paper we only give the results for three most common vehicle types: cars, small trucks and trucks.

The corresponding parameter estimates and test statistics for Kolmogorov-Smirnov and χ^2 tests are given in the following tables 3-8. Also the corresponding quantile tables are given in appendices 2-4.

Table 3. Estimates of parameters for different distributions (cars)

Distribution	Scale parameter	Shape parameter(s)
Lognormal	$\mu = 9.38$	$\sigma = 1.05$
Weibull	$\sigma = 1.98 \cdot 10^4$	$c = 0.89$
Pareto	$\lambda = 7.06 \cdot 10^4$	$\alpha = 4.51$
Beta		$\alpha = 0.87$, $\beta = 118.03$
Gamma	$\sigma = 2.17 \cdot 10^4$	$\alpha = 0.98$

Table 4. Results of K-S and χ^2 tests (cars)

Distribution	χ^2-stat.	χ^2 crit. value	K-S statistic	K-S p-value
Lognormal	1 267	14.1	0.0684	$p < 0.001$
Weibull	6 201	14.1	0.1178	$p < 0.001$
Pareto	5 631	14.1	0.1273	$p < 0.001$
Beta	7 190	14.1	0.1422	$p < 0.001$
Gamma	5 458	14.1	0.1242	$p < 0.001$

Table 5. Estimates of parameters for different distri-

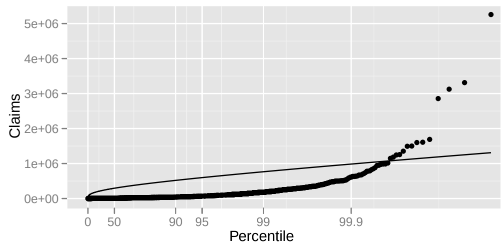

Fig. 5. Probability plot for Pareto distribution

butions (small trucks)

Distribution	Scale parameter	Shape parameter(s)
Lognormal	$\mu = 9.48$	$\sigma = 1.00$
Weibull	$\sigma = 2.16 \cdot 10^4$	$c = 0.89$
Pareto	$\lambda = 7.29 \cdot 10^4$	$\alpha = 4.31$
Beta		$\alpha = 0.58,\ \beta = 43.53$
Gamma	$\sigma = 2.34 \cdot 10^4$	$\alpha = 0.99$

Table 6. Results of K-S and χ^2 tests (small trucks)

Distribution	χ^2-stat.	χ^2 crit. value	K-S statistic	K-S p-value
Lognormal	127	14.1	0.0539	$p < 0.001$
Weibull	664	14.1	0.1247	$p < 0.001$
Pareto	599	14.1	0.1345	$p < 0.001$
Beta	1 721	14.1	0.2072	$p < 0.001$
Gamma	605	14.1	0.1301	$p < 0.001$

Table 7. Estimates of parameters for different distributions (trucks)

Distribution	Scale parameter	Shape parameter(s)
Lognormal	$\mu = 9.63$	$\sigma = 1.01$
Weibull	$\sigma = 2.54 \cdot 10^4$	$c = 0.94$
Pareto	$\lambda = 9.62 \cdot 10^4$	$\alpha = 4.74$
Beta		$\alpha = 0.58,\ \beta = 43.53$
Gamma	$\sigma = 2.47 \cdot 10^4$	$\alpha = 1.06$

Table 8. Results of K-S and χ^2 tests (trucks)

Distribution	χ^2-stat.	χ^2 crit. value	K-S statistic	K-S p-value
Lognormal	59	14.1	0.0476	$p < 0.001$
Weibull	321	14.1	0.0981	$p < 0.001$
Pareto	292	14.1	0.1208	$p < 0.001$
Beta	532	14.1	0.1527	$p < 0.001$
Gamma	275	14.1	0.1214	$p < 0.001$

Unfortunately the results for different vehicle types do not differ much from the results obtained for whole data. As cars form 77% of the data, the similarity to previous results was quite expected, but turns out that lognormal distribution is best fit for small trucks and trucks as well (although only 6.9% of the claims were related to small trucks and 4.5% to trucks). Also, the values of test-statistics for gamma, Weibull and Pareto are quite close (but noticeably worse than lognormal) and the fitting by beta distribution is worst in all cases. The goodness-of-fit test results together with visual estimation and probability plots lead us to same conclusion as for whole data: the lognormal distribution fits best from these choices, but not too well. The fit on the main part is generally tolerable, but the tail fit needs to be revised.

IV. Rough estimation of the total claim amount and the risk premium

While the individual claim distribution gives as valuable information about the risk behaviour, even more important question is how well can we describe the behaviour of total claims for whole portfolio, and, eventually, how to find a reasonable price for risk, i.e. how to calculate the risk premium. The usual way is to apply certain aggregate risk models, e.g. the collective risk model. Thus we need (beside the estimates for individual claims) also some estimates for the number of claims. We refer to [6] for the setup used here.

By the collective risk model the total claim amount can be calculated as

$$S = \sum_{i=1}^{N} X_i,$$

where X_i are individual claims and N is the number of claims. It is assumed that X_i are iid random variables and also independent of N. It is known that in this case the expectation and variance of total claim amount S can be calculated as:

$$ES = ENEX_1 \tag{1}$$

and

$$VarS = ENVarX_1 + VarN(EX_1)^2. \tag{2}$$

We are particularly interested in question how to find some first "rough" estimates in case when the history of claims is very limited, which obviously makes the situation more complicated. In our example we have only data from one year available, which makes it impossible to estimate the distribution of the number of claims unless we bring in additional assumptions or simplifications. Therefore we assume that

- each vehicle can have only one loss per year (this restriction is used since we do not have corresponding information);
- claim number N is binomially distributed, $N \sim Bin(n, p)$ (follows from the first assumption, but Poisson or negative binomial can also be used for more conservative estimates);
- total claim amount is approximately normally distributed $S \sim N(\mu, \sigma)$ (is justified if we have huge homogeneous portfolio, in practice we need to be careful as the estimates might be too optimistic).

If N follows binomial distribution, then formulas (1) and (2) simplify to:

$$ES = npEX_1 \tag{3}$$

and

$$VarS = npVarX_1 + np^2(EX_1)^2. \tag{4}$$

As we usually not have a fixed number of insured vehicles during period of interest, we could use the information about how long each vehicle was insured during given period and calculate the number of insured vehicles as

$$n = \frac{total \ \ days \ \ insured \ \ in \ \ given \ \ period}{length \ \ of \ \ given \ \ period}. \quad (5)$$

In our example the insurance period is one year (01.07.2006–31.06.2007), so formula (5) simplifies to

$$n = \frac{total \ \ days \ \ insured}{365}. \quad (6)$$

Also, the parameter p describing the probability of a loss event can be found in the usual way as

$$p = \frac{number \ \ of \ \ claims \ \ in \ \ period}{vehicles \ \ insured \ \ in \ \ period}. \quad (7)$$

Applying formulas (1), (2), (6) and (7) to our case study data, we obtain the following results.

Table 9. The expected values and variances by vehicle type for aggregate loss size (in mln EEK)

Vehicle type	ES	$VarS$
Cars	618.7	1 491.3
Small trucks	58.0	133.9
Trucks	45.3	105.8
Trailer trucks	55.5	132.8
Busses, trolleys, trams	15.9	44.1
Tractors	6.7	14.5
Motorcycles	4.5	12.1
All	797.1	1 945.2

Next we try to find a rough estimate for ”reasonable” risk premium. We choose the standard deviance principle for premium calculation as it obtains certain simple intuitive explanation in this example.

The risk premium $H(S)$ is calculated by standard deviation principle as

$$H(S) = ES + \beta\sqrt{VarS},$$

where $\beta > 0$. It is easy to see that if S is normally distributed, then the portfolio premium can be thought of as a solution of the following equation:

$$\boldsymbol{P}\{H(S) - S > 0\} = p. \quad (8)$$

On the other hand, given n policy-holders with premium P, the portfolio premium $H(S)$ can be calculated as nP, thus the formula (8) simplifies to

$$\boldsymbol{P}\{nP - S > 0\} = p$$

from where the individual premium can be found as

$$P = \frac{ES + \Phi^{-1}(p)\sqrt{VarS}}{n}.$$

The corresponding individual premiums with three different probabilities p are given in table 10.

Table 10. Estimated risk premiums

Vehicle type	$p = 0.9$	$p = 0.95$	$p = 0.99$
Cars	1 728	1 765	1 833
Small trucks	2 465	1 609	2 876
Trucks	3 444	3 669	4 081
Trailer trucks	10 252	10 875	12 020
Busses, trolleys, trams	7 137	7 855	9 176
Tractors	619	695	835
Motorcycles	2 028	2 319	2 854
All	1 677	1 709	1 768

Note that these numbers are obtained after serious simplifications and should be treated very carefully, they can be though of as a starting point for more thorough studies. But it is also worth mentioning that the actual numbers that were used by Estonian insurance companies in 2006 were quite close for most vehicle types.

V. Conclusions

1) The advantages of proposed simple method for distribution fitting are that it is easy to understand and apply, the tools used are classical and well known. The results obtained can be used as a starting point for more thorough studies, e.g. using certain tools from the theory of extreme values to estimate the tail behaviour (work in progress).
2) There is a need for specific tests suitable for the nature of insurance claims (where the data volume is huge, distributions are heavy-tailed and the tail behaviour is of key importance).
3) For the particular case study, none of the selected distributions describe the data at hand well, the goodness of fit tests reject everything.
4) By the values of test-statistics lognormal distribution are best for the whole data and for most vehicle types separately, lognormal distribution also fits best by visual estimation and probability plots (good fit up to 0.99-quantile).
5) The proposed method for premium calculation gives reasonably good initial estimate for further applications.

Appendix

The quantile tables given in appendices are produced using SAS software.

Appendix 1. Comparison of quantiles for empirical and fitted distributions, in mln EEK, whole data

```
 Quantiles for Lognormal Distribution
      ------Quantile------
  Percent Observed Estimated
    1.0    0.00050  0.00103
    5.0    0.00229  0.00212
   10.0    0.00400  0.00312
   25.0    0.00673  0.00594
   50.0    0.01180  0.01214
   75.0    0.02209  0.02481
```

90.0	0.04300	0.04722
95.0	0.06863	0.06940
99.0	0.18097	0.14292

Quantiles for Weibull Distribution

	------Quantile------	
Percent	Observed	Estimated
1.0	0.00050	0.00011
5.0	0.00229	0.00069
10.0	0.00400	0.00157
25.0	0.00673	0.00495
50.0	0.01180	0.01349
75.0	0.02209	0.02977
90.0	0.04300	0.05313
95.0	0.06863	0.07174
99.0	0.18097	0.11720

Quantiles for Gamma Distribution

	------Quantile------	
Percent	Observed	Estimated
1.0	0.00050	0.00018
5.0	0.00229	0.00100
10.0	0.00400	0.00212
25.0	0.00673	0.00609
50.0	0.01180	0.01520
75.0	0.02209	0.03112
90.0	0.04300	0.05241
95.0	0.06863	0.06861
99.0	0.18097	0.10638

Quantiles for Beta Distribution

	------Quantile------	
Percent	Observed	Estimated
1.0	0.00050	0.00011
5.0	0.00229	0.00076
10.0	0.00400	0.00179
25.0	0.00673	0.00578
50.0	0.01180	0.01579
75.0	0.02209	0.03417
90.0	0.04300	0.05934
95.0	0.06863	0.07868
99.0	0.18097	0.12391

Appendix 2. Comparison of quantiles for empirical and fitted distributions, in mln EEK, cars

Quantiles for Lognormal Distribution

	------Quantile------	
Percent	Observed	Estimated
1.0	0.00045	0.00103
5.0	0.00237	0.00211
10.0	0.00400	0.00308
25.0	0.00660	0.00582
50.0	0.01170	0.01180
75.0	0.02119	0.02392
90.0	0.04080	0.04516
95.0	0.06438	0.06607
99.0	0.17209	0.13488

Quantiles for Weibull Distribution

	------Quantile------	
Percent	Observed	Estimated
1.0	0.00045	0.00012
5.0	0.00237	0.00072
10.0	0.00400	0.00160
25.0	0.00660	0.00492
50.0	0.01170	0.01314
75.0	0.02119	0.02851
90.0	0.04080	0.05026
95.0	0.06438	0.06745
99.0	0.17209	0.10906

Quantiles for Gamma Distribution

	------Quantile------	
Percent	Observed	Estimated
1.0	0.00045	0.00020
5.0	0.00237	0.00104
10.0	0.00400	0.00217
25.0	0.00660	0.00601
50.0	0.01170	0.01464
75.0	0.02119	0.02950
90.0	0.04080	0.04922
95.0	0.06438	0.06417
99.0	0.17209	0.09892

Quantiles for Beta Distribution

	------Quantile------	
Percent	Observed	Estimated
1.0	0.00045	0.00013
5.0	0.00237	0.00083
10.0	0.00400	0.00187
25.0	0.00660	0.00575
50.0	0.01170	0.01513
75.0	0.02119	0.03193
90.0	0.04080	0.05465
95.0	0.06438	0.07195
99.0	0.17209	0.11216

Appendix 3. Comparison of quantiles for empirical and fitted distributions, in mln EEK, small trucks

Quantiles for Lognormal Distribution

	------Quantile------	
Percent	Observed	Estimated
1.0	0.00082	0.00128
5.0	0.00303	0.00252
10.0	0.00464	0.00363
25.0	0.00700	0.00666
50.0	0.01203	0.01306
75.0	0.02374	0.02563
90.0	0.04500	0.04700
95.0	0.07080	0.06757
99.0	0.20058	0.13350

Quantiles for Weibull Distribution

	------Quantile------	
Percent	Observed	Estimated
1.0	0.00082	0.00013
5.0	0.00303	0.00078
10.0	0.00464	0.00174
25.0	0.00700	0.00536
50.0	0.01203	0.01434
75.0	0.02374	0.03114
90.0	0.04500	0.05492
95.0	0.07080	0.07371
99.0	0.20058	0.11922

Quantiles for Gamma Distribution

	------Quantile------	
Percent	Observed	Estimated
1.0	0.00082	0.00023
5.0	0.00303	0.00119
10.0	0.00464	0.00244
25.0	0.00700	0.00669
50.0	0.01203	0.01614
75.0	0.02374	0.03231
90.0	0.04500	0.05369

95.0	0.07080	0.06987
99.0	0.20058	0.10746

Quantiles for Beta Distribution

	------Quantile------	
Percent	Observed	Estimated
1.0	0.00082	0.00002
5.0	0.00303	0.00031
10.0	0.00464	0.00104
25.0	0.00700	0.00522
50.0	0.01203	0.01961
75.0	0.02374	0.05128
90.0	0.04500	0.09858
95.0	0.07080	0.13607
99.0	0.20058	0.22490

Appendix 4. Comparison of quantiles for empirical and fitted distributions, in mln EEK, trucks

Quantiles for Lognormal Distribution

	------Quantile------	
Percent	Observed	Estimated
1.0	0.00100	0.00145
5.0	0.00312	0.00289
10.0	0.00499	0.00417
25.0	0.00860	0.00772
50.0	0.01463	0.01529
75.0	0.02832	0.03026
90.0	0.05345	0.05596
95.0	0.08014	0.08085
99.0	0.21562	0.16122

Quantiles for Weibull Distribution

	------Quantile------	
Percent	Observed	Estimated
1.0	0.00100	0.00019
5.0	0.00312	0.00109
10.0	0.00499	0.00234
25.0	0.00860	0.00678
50.0	0.01463	0.01722
75.0	0.02832	0.03589
90.0	0.05345	0.06144
95.0	0.08014	0.08120
99.0	0.21562	0.12807

Quantiles for Gamma Distribution

	------Quantile------	
Percent	Observed	Estimated
1.0	0.00100	0.00033
5.0	0.00312	0.00155
10.0	0.00499	0.00307
25.0	0.00860	0.00800
50.0	0.01463	0.01861
75.0	0.02832	0.03636
90.0	0.05345	0.05955
95.0	0.08014	0.07700
99.0	0.21562	0.11734

Quantiles for Beta Distribution

	------Quantile------	
Percent	Observed	Estimated
1.0	0.00100	0.00012
5.0	0.00312	0.00090
10.0	0.00499	0.00217
25.0	0.00860	0.00729
50.0	0.01463	0.02033
75.0	0.02832	0.04401
90.0	0.05345	0.07534
95.0	0.08014	0.09841
99.0	0.21562	0.14904

REFERENCES

[1] M. Alfonsec, M. Cebrián and A. Ortega, "Testing genetic algorithm recombination strategies and the normalized compression distance for computer-generated music", in *Proceedings of the 5th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases, Madrid, Spain*, 2006, pp. 53–58.

[2] S. Asmussen and K. Binswanger, "Simulation of ruin probabilities for subexponential claims", *ASTIN Volume 27*, No 2, 1997, pp. 297–318

[3] G. Blom, *Probability and statistics: theory and applications*, University of Lund, Lund, 1989.

[4] J. F. Bonnans, J. Ch. Gilbert, C. Lemaréchal and C. A. Sagastizábal, *Numerical optimization, theoretical and numerical aspects (2nd edition)*, Springer, 2006.

[5] G. E. P. Box and N. R. Draper, *Empirical model-building and response surface*, John Wiley & Sons Inc, Canada, 1987.

[6] C. D. Daykin, T. Pentikäinen and M. Pesonen, *Practical risk theory for actuaries*, Chapman & Hall, London, Glasgow, New York, Tokyo, Melbourne, Madras, 1994.

[7] P. Embrechts, C. Klüppelberg and T. Mikosch, *Modelling extremal events for insurance and finance*, Springer, Berlin, 1997.

[8] P. H. Garthwaite, I. T. Jolliffe and B. Jones, *Statistical inference*, Oxford University Press, New York, 2002.

[9] R. V. Hogg and S. A. Klugman, *Loss distributions*, John Wiley & Sons Inc, New York, 1984.

[10] B. Huber, "Fitting distributions to data. Practical issues in the use of probabilistic risk assessment". Available: http://www.quantdec.com/envstats/misc/Fitting%20data%20to%20 distributions.pdf, 1999.

[11] L. Kalyvas, C. Siriopoulos and N. Dritsakis, "Reaching extreme events with conditional and unconditional models", *WSEAS Transactions on Business and Economics*, vol. 1, pp. 156-161, 2004.

[12] J. A. Nelder and R. Mead, "A simplex algorithm for function minimization", *Computer Journal* 7, 1965, pp. 308-313.

[13] A. Pattavina and M. Tornatore, "IP traffic characterization: an overview", in *Proceedings of 6th WSEAS International Multiconference on Circuits, Systems, Communications and Computers, Crete, Greece (CSCC 2002)*, 2002, pp. 4431–4434.

[14] E. L. Silva and P. Lisboa, "Analysis of the characteristic features of the density functions for gamma, Weibull and log-normal distributions through RBF network pruning with QLP", in *Proceedings of the 6th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases, Corfu Island, Greece*, 2007, pp. 223–228.

[15] J. L. Teugels, "Large Claims in Insurance Mathematics", *ASTIN Volume 13*, No 2, 1982.

[16] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S (4th edition)*, Springer, New York, 2002.

[17] A. Zaharim, S. K. Najid, A. Mahir and K. Sopian "The suitability of statistical distribution in fitting wind speed data", *WSEAS Transactions on Mathematics*, Vol. 7, 2008, pp. 718–727.