

IDENTIFICATION OF PROBABILITY DISTRIBUTION USING SKEWNESS-KURTOSIS GRAPH IN INSURANCE

VÁCLAV SLÁDEK, IVANA MALÁ, ADAM ČABLA

University of Economics, Prague, Faculty of Informatics and Statistics,
Department of Statistics and Probability,
W. Churchill Sq. 4, Prague, Czech Republic
e-mails: xslav20@vse.cz, malai@vse.cz, adam.cabla@vse.cz

Abstract

Insurance company has to estimate its liability (present or future) from its business. Estimation could be based on parametric or non-parametric approaches. Correct identification of probability distributions in parametric approach could lead to accurate inferences about liability. It could lead to greater reliability in estimation and in the best case, to reduce the costs of capital. Skewness-kurtosis graph is a very useful tool for an identification of probability distribution. It is based on intra-relations between moments of the probability distribution. Probability distributions typical for insurance are lognormal, Pareto, gamma distribution and many others. A lot of them could be classified as heavy-tail or long-tail distribution, i.e. extreme values have high probability of being selected in a random sample. Estimates of product moments of skewness and kurtosis are sensitive to the extreme values presence and are limited by a sample size. L-moments are a robust alternative to product moments. This contribution is focused on skewness-kurtosis graph based on L-moments, so-called L-skewness, and L-kurtosis. Its application is shown on Monte Carlo simulation and also an empirical study is presented. A distribution of the amount of Motor third-party liability claims is identified.

Keywords: skewness-kurtosis graph, L-moments, MTPL claims, loss distribution

JEL Codes: G22, C13

1. Introduction

An insurance company has to estimate liability from its business by law. Companies use statistical methods to calculate the estimate of liability. Parametric methods are usually based on the assumption of probability distribution of analyzed variable. Unfortunately, proper method for identification of distribution family is not often used in practice and experience plays the main role in choosing the distribution. Inappropriate choice of the distribution family leads to biased inferences and confusing or even misleading results, even if statistical methods are correctly applied. For this reason, methods of identification of acceptable distributions are extremely important, as the correct identification of probability distribution could lead to accurate inference about liability, to greater reliability in estimation and, in the best case, to reduction in costs of capital. In general, identification could be graphic or numeric. Graphic analysis of data was popular in the past, but it is one of the most topical issues at present.

The histogram is the basic graph used for the task, it is the non-parametric estimate of the density of distribution. A big disadvantage of the histogram is the choice of rules for determining the number and width of intervals, this problem may be overcome by a kernel density estimator. A graph comparing the theoretical values of skewness and kurtosis

measured by product moments (third and fourth normalized product moment) to their estimates is sometimes preferred. This graph is called skewness-kurtosis graph, it is based on intra-relations (within distribution) between moments of the distribution. It compares estimated skewness and kurtosis with theoretical values for selected distributions. In insurance, statistical methods often work with variables having a so-called heavy-tail or long-tail distribution, i.e. extreme values have high probability of being selected. Distributions typical for a continuous variable in insurance, so-called loss distributions, are generalized Pareto, lognormal, Gamma and Weibull distribution. Estimates of product moments of skewness and kurtosis are not robust in presence of extreme values in a random sample and they are also limited by a size of the sample. These estimates are known to be highly biased with large variance for small samples (Vogel and Fennessey, 1993).

L-moments represent a robust alternative to product moments defined in (Hosking, 1990). The only assumption for all L-moments to be finite is a finite expected value of the distribution and we can evaluate coefficients of skewness and kurtosis even for distributions without finite variance. As the censored observations in the analysed datasets are a problem of insurance data, product moments are not suitable for the description of such datasets. L-moments for censored data are used in (Wang *et al.*, 2010; Vogel *et al.*, 1998). Skewness-kurtosis graph based on L-moments were introduced in (Hosking, 1990) and used predominantly on river or flood data, for example (Vogel and Fennessey, 1993; Peel *et al.*, 2001; Hosking and Wallis, 1997). Study on insurance data does not yet exist.

The aim of the contribution is to promote the skewness-kurtosis graph based on L-moments for identification of distribution in insurance. In the text, we discuss positive properties of the graph in comparison with the use of classic product moments and we introduce it as a useful tool for distribution identification in insurance. We apply the concept in a case study of motor third-party liability (MTPL) claims in the Czech Republic, but it can enrich the spectrum of commonly used graphical tools in general. We are interested only in the suitable family of distributions, not in an estimation of parameters or statistical inference. Skewness-kurtosis graph based on L-moments are described in Section 2. Section 3 contains application on a real dataset, where different approaches are used.

All computations and simulations are performed in program R (Rmetrics Core Team *et al.*, 2017) using packages *lmomco* (Rmetrics Core Team and Asquith, 2018) and *moments* (Komsta and Novomestky, 2015).

2. Skewness-kurtosis graph based on L-moments

Skewness-kurtosis graph based on product moments is mentioned in (McCuen, 1985). The graph compares estimated value of skewness and kurtosis with their theoretical values. Values of skewness are on the horizontal axis and values of kurtosis on the vertical axis. If skewness and kurtosis of distribution are independent of any parameter, the distribution is represented by one point on the graph. For example, coefficients of skewness and kurtosis of all normal distributions are equal to 0 and 3, regardless of the mean and standard deviation and these distributions are represented by a single point (0,3) in the graph. When skewness and kurtosis depend on one parameter, all combinations of skewness and kurtosis for different values of the parameter are represented by a curve substituting possible values of the parameter. For example, skewness and kurtosis of the lognormal distribution depend only on variance (not on μ):

$$\begin{aligned} \text{skewness} &= \left(e^{\sigma^2} + 2 \right) \sqrt{e^{\sigma^2} - 1}, \\ \text{kurtosis} &= e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 3. \end{aligned}$$

Both functions of σ^2 are unbounded, this property limits the use of the skewness-kurtosis graph for this distribution. We will show, that all values of L-skewness and L-kurtosis are bounded and the mentioned problem can be overcome.

For these two situations, we could use two-dimensional graph. In the situation of the vector parameter with more parameters, the multi-dimensional graph should be used. We only use two-dimensional graphs in this contribution.

2.1. L-moments

L-moments are the robust alternative to product moments. They are defined as linear combinations of expected values of ordered statistics

$$\lambda_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} EX_{r-k:r}, \quad (1)$$

where r is the order of L-moment (integer value) and $EX_{r-k:r}$ is expected value of the $(r-k)$ the order statistic of a sample size r . It is recommended to use standardized higher L-moments, so-called L-moment ratios, for description of the distribution, which are defined as follows

$$\tau_r = \frac{\lambda_r}{\lambda_2}, \quad r = 3, 4, \dots \quad (2)$$

τ_3 and τ_4 describe skewness and kurtosis of a distribution, they are called L-skewness and L-kurtosis. To obtain all L-moments λ_r finite, we have to assume only the expected value of the distribution to be finite. For this reason, we can use L-skewness and L-kurtosis for more probability distributions than the classical product moments. Moreover, values of L-moment ratios are bounded, so it is easier to interpret their values than values of product moments. According to (Hosking and Wallis, 1997), $|\tau_r| < 1$ for $r \geq 3$ and $\frac{1}{4}(5\tau_3^2 - 1) \leq \tau_4 < 1$. In the literature, it is possible to find the exact formulas for L-moments and L-moment ratios for frequently used distributions (Hosking, 1990; Delicado and Goriab, 2008; Mudholkar and Hutson, 2000; Hosking *et al.*, 1985). The formulas are used in Figures 2-4 to plot the lines for theoretical L-moments ratios. Note, that in the conventional skewness-kurtosis graphs we have a unique point (0,3) for all normal distributions, in case of L-moments we obtain a point (0,0.1226) (Hosking, 1990).

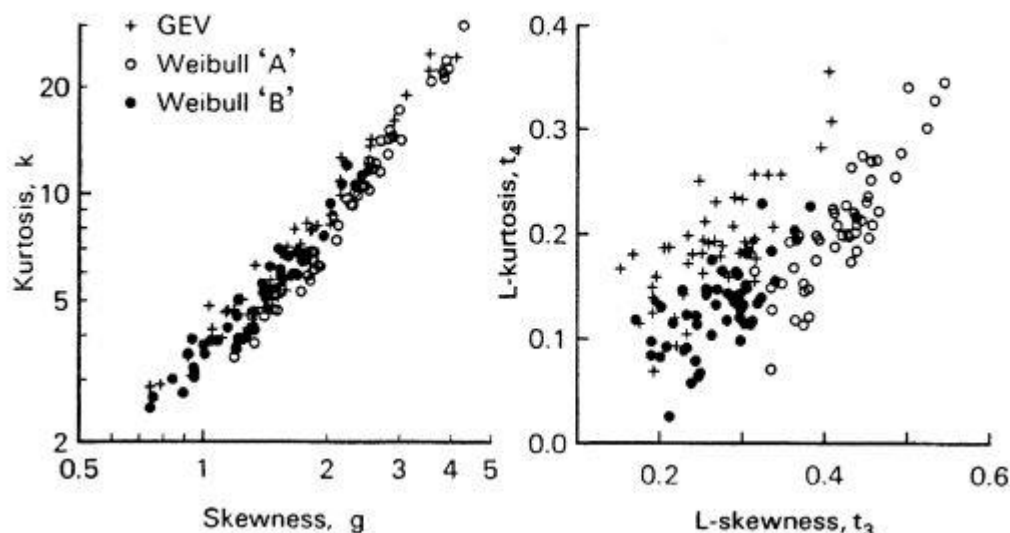
We use notation l_r , $r = 1, 2, \dots$ for sample counterparts to L-moments (sample L-moments) and $t_r = \frac{l_r}{l_2}$, $r = 3, 4, \dots$ for sample L-moments ratios (Hosking, 1990). The points (t_3, t_4) are used in the graph.

The skewness-kurtosis graph compares sample distribution described by the point (*sample skewness*, *sample kurtosis*) with one or more theoretical distributions specified by curves or surfaces given by (*theoretical skewness*, *theoretical kurtosis*) as functions of parameters of the distribution. Unfortunately, estimates of product moments are sensitive to the presence of outliers in a sample and also, sample skewness and sample kurtosis are limited by the size of the sample. For example, estimated skewness could not be higher than $(n-1)/\sqrt{n-2}$. Hosking in (1990) compares the use of standard skewness-kurtosis graph based on product moments (left in Figure 1 with the sample skewness coefficient g on the horizontal axis and the sample coefficient of kurtosis k on the vertical axis in Hosking's notation) with the same graph based on L-moments instead of on the classical moments (right in Figure 1) to show advantage of the use of L-moments. Data from three distributions with the same mean and variance (two Weibull because of its flexibility and one generalized extreme value (GEV)

distributions) were generated and the sample skewness and kurtosis were evaluated based on the product and L-moments. Product moments for all three distributions lie close to the single line but the use L-moments finds three relatively well separated clusters and enables distribution identification. Identification of the distribution using L-moments in this example is more accurate than using product moments.

If we compare properties of product and L-moments, it is not possible to derive general comparison, we know that L-moments and L-moments ratios are asymptotically unbiased and normally distributed and they perform well even in case of small samples. Simulation studies are available for different distributions. For example, small sample properties are studied in (Delicado and Goriab, 2008) for the asymmetric exponential power distribution. Distributions of interest are usually skewed and heavy-tailed such as skewed Normal, skewed Laplace, Pareto or GEV distribution.

Figure 1: Comparison of skewness-kurtosis graph based on product moments and L-moment ratios for similar distributions



Source: Hosking (1990, p. 118).

Let us consider a sample with 100 observations of random variable with generalized Pareto distribution (*GP*) with parameters $\mu = 0$, $\sigma = 1$ and $\xi = 0$. Table 1 compares properties of sample L-moment ratios to sample skewness and kurtosis based on product moments. Properties of moments are estimated using Monte Carlo simulation (5,000 iterations). Sample L-moment ratios are better estimates (with respect to the coefficient of variation) than sample skewness and kurtosis. Values of coefficients of variation of sample L-moments ratios are lower than values of sample product moments, we use a relative variability because of different theoretical value, theoretical values are given in (Hosking, 1990).

Table 1: Comparison of sample L-moment ratios to sample product moments (100 observations)

	theoretical value	expected value	standard error	MSE	coefficient of variation
τ_3	0.333	0.330	0.048	0.002	0.147

skewness	2.000	1.783	0.544	0.343	0.305
τ_4	0.166	0.165	0.047	0.002	0.286
kurtosis	9.000	7.144	3.775	17.697	0.528

Source: The authors' work.

2.2. Simulation study

The aim of this simulation study is to show properties of L-skewness-kurtosis graph. The presented study has two parts. In the first part we illustrate identification of considered distribution based on skewness-kurtosis graph leads. Let us assume that a random variable X come from one of these distributions:

- generalized Pareto (GP),
- lognormal distribution (LN),
- Gamma distribution (G),
- Weibull distribution (W),

and consider random samples drawn from these distributions with parameters:

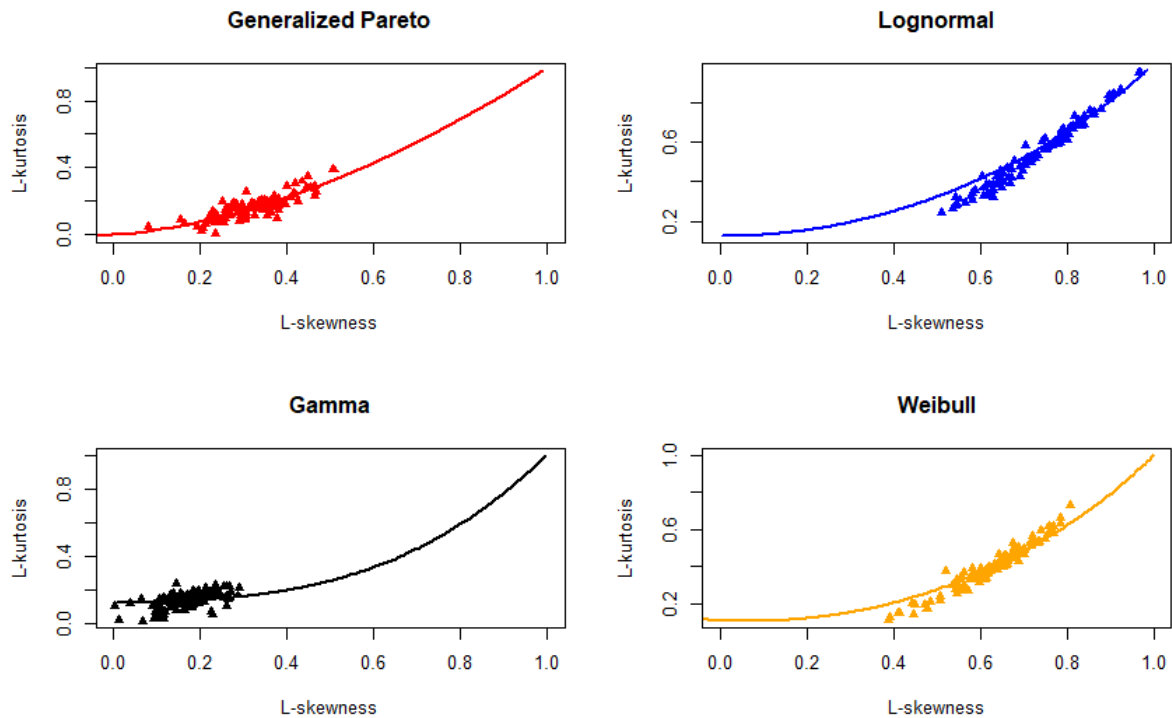
- $GP(\mu = 0, \sigma = 1, \xi = 0)$,
- $LN(\mu = 0, \sigma = e^2)$,
- $G(\alpha = 4, \beta = 1)$,
- $W(\alpha = 1, c = 0.5)$.

The algorithm of this part of the study could be described by following steps:

1. Create skewness-kurtosis graph based on L-moments for a considered distribution.
2. Generate random sample (with 50 observations) from the considered distribution with selected parameters.
3. Compute t_3 and t_4 .
4. Repeat step 2 and 3 (100 iterations).
5. Draw estimated points (t_3, t_4) into skewness-kurtosis graph.

Results are shown in Figure 2. We can see that estimated τ_3 and τ_4 lie onto or close to the theoretical curve for each distribution. Unfortunately, the theoretical distribution is unknown and we have to choose between a lot of types of distributions. Many of them have a similar shape of the curve describing intra-relations between moments (see Figures 3 and 4).

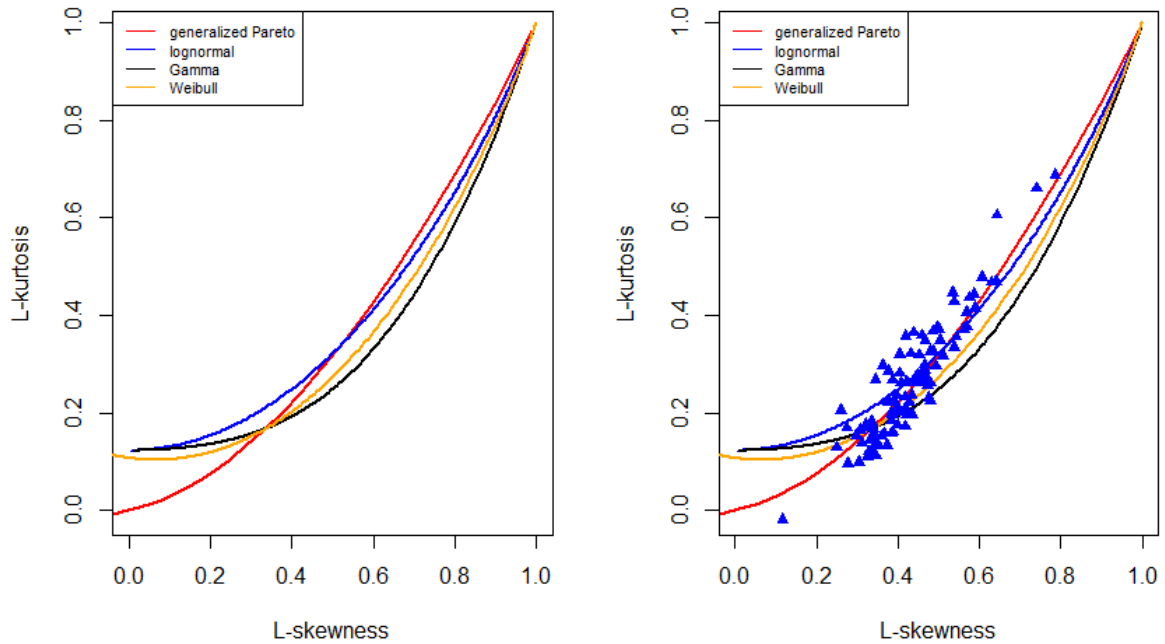
Figure 2: Skewness-kurtosis graphs for selected distributions



Source: The authors' work.

The aim of the second part is to demonstrate an identification of distribution when we are not sure about a distribution of a random variable. Let assume that a positive random variable could follow one of the distributions mentioned in the first part of the study (generalized Pareto, lognormal, Gamma and Weibull distribution). The left graph on Figure 3 shows a skewness-kurtosis graph for all considered distributions. The shape of curves seems similar, especially Gamma and Weibull distribution. Let us generate random samples with 50 observations from a lognormal distribution with parameters $\mu = 0$, $\sigma = e^1$ and show estimated (τ_3, τ_4) points (t_3, t_4) on the graph (right graph on Figure 3). It seems that the points are close to the curve for generalized Pareto distribution even they were drawn from the lognormal distribution. Inaccurate inference could be presented without enough experience in skewness-kurtosis graph method. To avoid this problem, (Peel *et al.*, 2001) recommend not to identify distribution using all points but use the mean of estimated of τ_3 and τ_4 . The point (t_3, t_4) estimates the mean $(E(t_3), E(t_4))$, from the asymptotical unbiasedness of L-moment ratios (Hosking, 1990) we obtain estimate of (τ_3, τ_4) .

Figure 3: Skewness-kurtosis graphs for all distributions

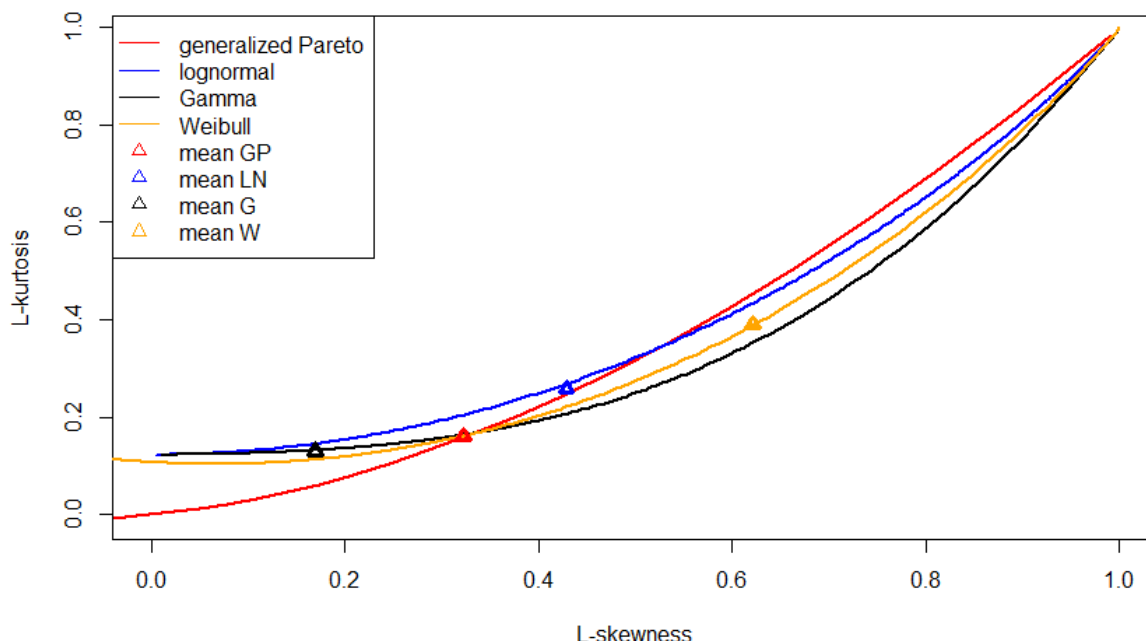


Source: The authors' work.

Figure 4 shows the situation of the mean estimated $\tau_3 \bar{t}_3$ and $\tau_4 \bar{t}_4$ from samples mentioned previously in this section represented by the point (\bar{t}_3, \bar{t}_4) . This approach helps to make graph clearer. Points of the mean lie on the theoretical curve of distribution. The point for lognormal distribution lies between the curve of lognormal and Gamma distribution, but it is closer to lognormal distribution. The point for generalized Pareto distribution lies close to the intersection of generalized Pareto, Gamma and Weibull distributions. The accuracy of this approach depends on the size of samples and number of samples (direction of dependence is the same for both, accuracy increases with increasing sample size and number of samples).

In many cases, we are limited by a number of observations and samples in a real study. Both approaches (single points or means) could lead to inaccurate inferences. We cannot improve estimate using Monte Carlo simulations because we do not know the theoretical distribution. If we know it, there is no need to identify it. The solution would be to use the bootstrap method.

Figure 4: Identification of distribution based on the mean of estimated τ_3 and τ_4



Source: The authors' work.

3. Distribution of MTPL claims

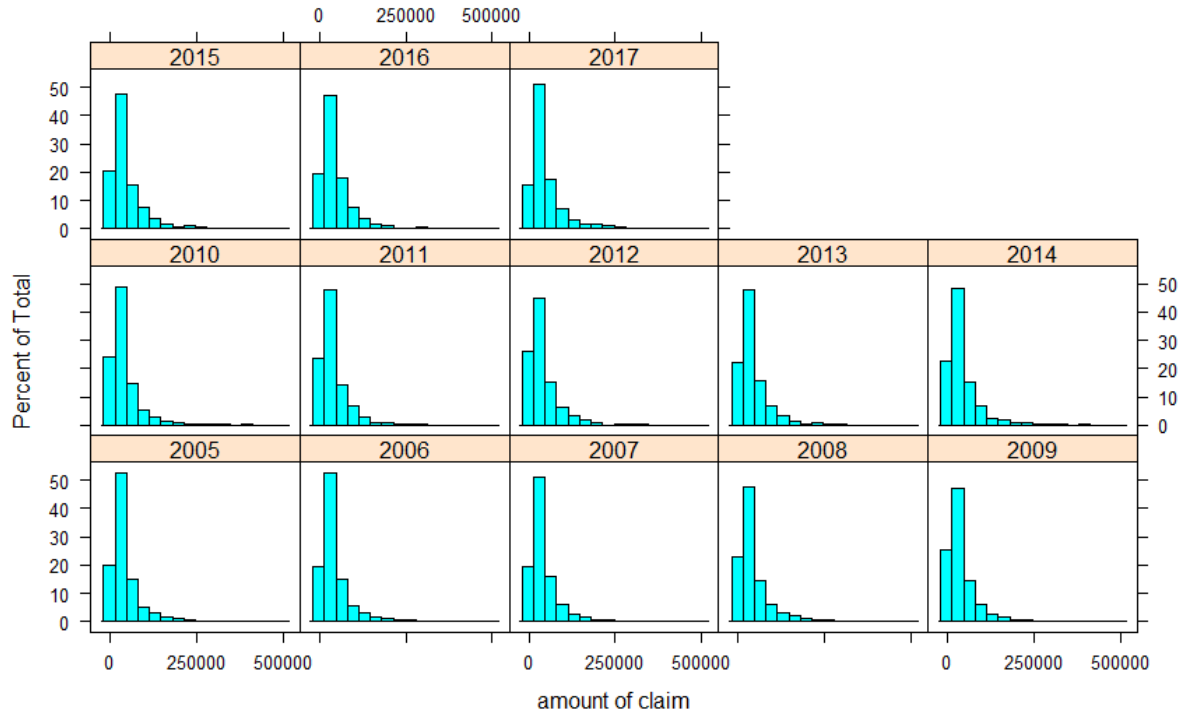
3.1. Data

Dataset was provided by the Czech Insurers' Bureau, an organization of insurance companies that are authorized by providing liability insurance for claims caused by vehicles. The dataset contains 31,123 material claims, which occurred in period 2005 – 2017 and were reported during this period. Amount of claims consists of cumulative paid until 31.12.2017 and value of reserve on 31.12.2017. 6 % of considered claims are opened, we identify them as claims with a non-zero value of reserve. Censored observations are concentrated into last accident years (22 % and 40 % of claims occurred in 2016 and 2017 are still open). The amount of these claims may change in the future, these claims are so-called right-censored observations. Figure 5 shows estimated shape of the distribution of amount according to an accident year. It seems, that type of distribution does not change in time, probably there is no reason to identify distribution for each accident year separately. After removing censored observations, the shape of the distributions does not change. The parameters of location and variation change slightly but it should not have an influence on the type of distribution.

Table 2 contains all computed samples L-skewness and L-kurtosis for our dataset according to an accident year. t_3^{all} and t_4^{all} are characteristics computed from all data, t_3^{cen} and t_4^{cen} represents censored L-skewness and L-kurtosis. Czech Insurers' Bureau recommends modeling liabilities from MTPL separately for small claims and large claims, where large claims are claims with an amount higher or equal to a one million. These large claims are considered as outliers. There could be a high probability of influencing the estimate of the shape of distribution although L-moments are robust than product moments. t_3^{lim} and t_4^{lim} represent values of these L-skewness and L-kurtosis computed only for small claims. Values in Table 2 show that large claims have minimal impact on estimated L-moment ratios (their share in the dataset is small), but censoring has, especially for last accident years (from 2015 to 2017), because their share in dataset are higher than in previous years – we observe heavy

censoring. Estimates of last accident years are also influenced by a small number of claims, because many claims have not been reported yet.

Figure 5: Histograms of claims amount according to an accident year



Source: The authors' work.

For identification a distribution, let us consider only t_r^{all} and t_r^{cen} . Skewness-kurtosis graphs are shown in Figure 6. Left graph represents results obtained using t_3^{all} and t_4^{all} , right graph using censored L-moment ratios. Blank points represent values for each accident year, filled point value for all period. Censored values for last accident years are much different compare to others, but they lie close to theoretical curves than in a situation without censoring. It appears censoring gives accurate results for this dataset. We can see that points are close to lognormal or generalized Pareto distributions. It seems that nearest curve for points in Figure 6, is probably the curve for generalized Pareto distribution.

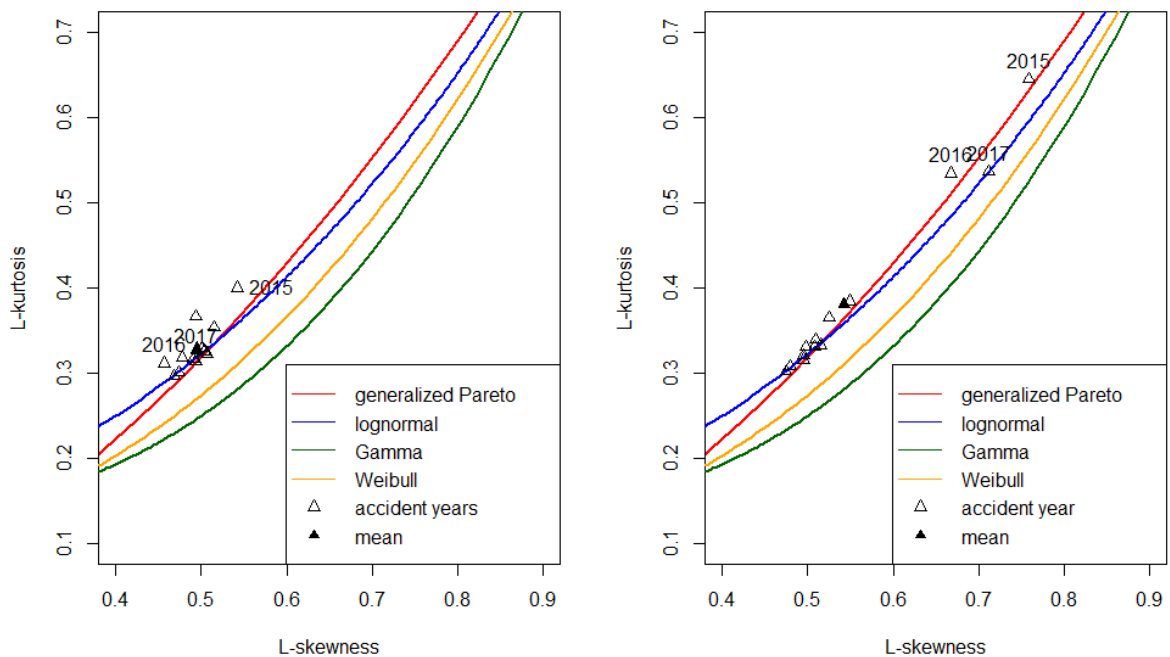
Using bootstrap method we tested variability of estimates for accident years with a high share of censored claims in the dataset (accident years 2015, 2016, 2017). For each year we generate 50 bootstrap samples and estimate L-skewness and L-kurtosis (consider censoring). Figure 7 shows obtained results. Location of points for these three years shows high variability of estimate. Points of accident years 2015 and 2016 lie near to generalized Pareto distribution curve. Points of the accident year 2017 are more concentrated than others years but they are still close to the lognormal distribution. It seems that amount of MTPL claims occurred during 2017 follows lognormal distribution instead generalized Pareto. But, estimates of L-moment ratios could be biased. Without assuming censoring, an estimated combination of L-skewness and L-kurtosis is close to generalized Pareto instead lognormal distribution (Figure 6).

Table 2: Computed sample L-moment ratios according to accident year

accident year	t_3^{all}	t_3^{cen}	t_3^{lim}	t_4^{all}	t_4^{cen}	t_4^{lim}
2005	0.504	0.508	0.504	0.326	0.331	0.326
2006	0.507	0.515	0.498	0.323	0.333	0.310
2007	0.495	0.498	0.478	0.327	0.330	0.305
2008	0.493	0.495	0.484	0.314	0.316	0.302
2009	0.491	0.496	0.491	0.316	0.320	0.316
2010	0.515	0.525	0.494	0.353	0.365	0.325
2011	0.501	0.510	0.478	0.329	0.339	0.300
2012	0.473	0.479	0.464	0.301	0.308	0.289
2013	0.469	0.475	0.459	0.296	0.302	0.284
2014	0.477	0.550	0.468	0.319	0.384	0.307
2015	0.543	0.759	0.466	0.400	0.645	0.304
2016	0.456	0.667	0.439	0.311	0.535	0.290
2017	0.493	0.712	0.458	0.367	0.536	0.324
all period	0.495	0.543	0.477	0.328	0.381	0.306

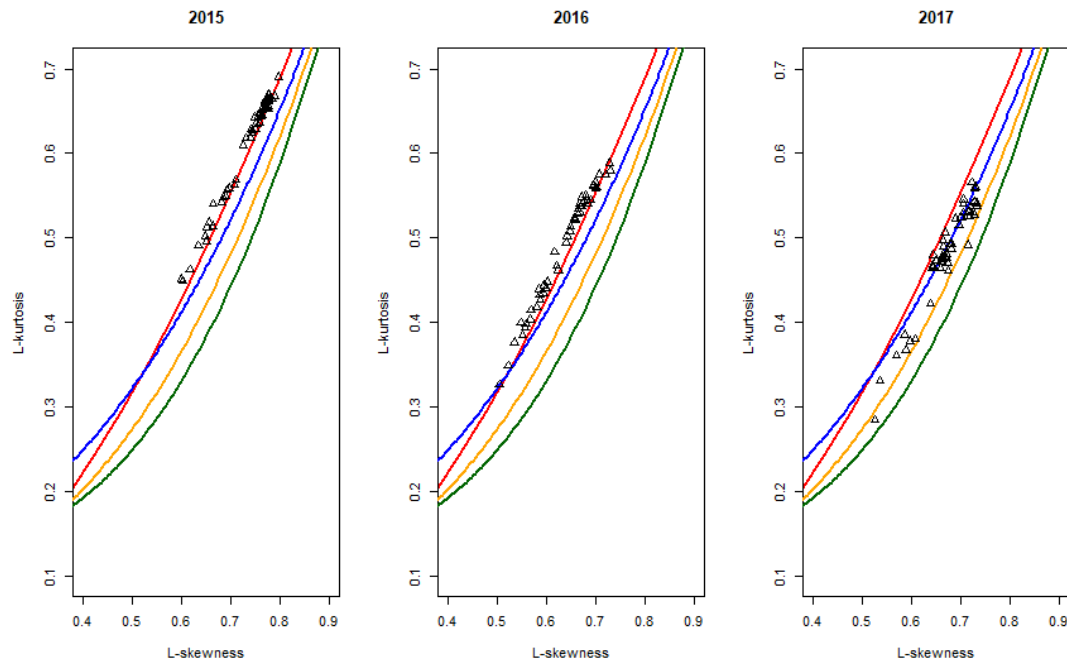
Source: The authors' work.

Figure 6: Skewness-kurtosis graph for t_3^{all}, t_4^{all} (left) and t_3^{cen}, t_4^{cen} (right)



Source: The authors' work.

Figure 7: Variability of L-moment ratios for the accident year 2015-2017, using bootstrap



Source: The authors' work.

4. Conclusion

In the text, skewness-kurtosis graph based on L-moments is used to insurance data with censored values. The graph was introduced by Hosking in (1990) in the first article dedicated to L-moments as a robust version of frequently used skewness-kurtosis graph. Since 1990, wide range contributions have treated the problem of identification of probability distributions using the skewness- kurtosis graph. Moreover, the graph could be used to compare distributions of subpopulations in case of observed subpopulation membership. In this case we represent subgroups of data with the corresponding sample L-skewness and L-kurtosis points. The concept is common in extreme values analysis with application mainly in the modelling of river flows or flood data. But due to its properties, it has wider use. In the contribution, we discuss the use of the graph in insurance and we promote the graph as a robust graphical tool with well-interpretable results. The graph is able to efficiently deal with problems of insurance data as censored values or highly skewed distributions causing outliers in data. The existence of L-skewness and L-kurtosis for all distributions with a finite expected value is also very useful, as we can use the graph even for distributions with infinite variance (for example Pareto distribution for a small shape parameter).

Identification of distribution using skewness-kurtosis graph based on estimated L-moments provides accurate inferences than graph based on estimated product moments, especially long-tail or heavy-tail distribution situation. The variance of estimates in our simulation study is lower for L-moment than for product moments. In case of presence of censored data, L-moments based on the quantile function can be easily evaluated with the use of distribution-free Kaplan-Meier estimates or its generalizations and incorporation of these data is not a problem. In the literature, we have well described the asymptotical normal distribution of the L-moments ratio characteristics. These moments are robust, they are not sensitive to outliers that are frequent if we use highly skewed and heavy-tailed distributions. The boundedness of L-skewness and L-kurtosis enables interpretation of the skewness-kurtosis graphs and comparison of distributions or even samples.

It seems, that censored observations play an important role in identification of distributions. Unfortunately, there is no way to verify it at this moment and further research should be done.

Based on skewness-kurtosis graphs, it seems that amount of MTPL claims according to accident year follows only one type of distribution, generalized Pareto distribution. Only accident year 2017 is close to lognormal distribution instead of generalized Pareto but many claims for this year have not been reported yet and many claims are open (censored).

The shape of most accident years is similar (similar L-skewness and L-kurtosis points) with exception of accident years 2015, 2016 and 2017. The reason is probably censored observations. Ignoring problems with censored claims leads to inference: MTPL claims follow generalized Pareto distribution regardless of the accident year. Which inference is right could be proved only in future, when more claims will be reported and especially closed.

Unfortunately, general recommendation on the distribution of the amount of MTPL claim to modeling liabilities from MTPL could not be provided because each insurance company has its own portfolio of contracts and their behavior may vary. This contribution shows that skewness-kurtosis graph based on L-moments is a simple tool for identifying a distribution which provides precise inferences not only for flood and river data but also for insurance data. In the best scenario, using this tool could lead to reduction in costs of capital.

References

- [1] Delicado P., Goriab M.N. 2008. A small sample comparison of maximum likelihood, moments and L-moments methods for the asymmetric exponential power distribution.
- [2] Hosking, J. R. 1990. L-Moments: Analysis and Estimation of Distribution Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society*, vol. 50, pp. 105-124.
- [3] Hosking, J. R. M., Wallis, J. R., Wood, E. F. 1985. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* 27, pp. 251-261.
- [4] Hosking, J. R., Wallis, J. R. 1997. *Regional Frequency Analysis*. New York: Cambridge University Press. ISBN 978-0-521-43045-6.
- [5] McCuen, R. H. 1985. *Statistical Methods for Engineers*. Prentice-Hall, pp. 61-63.
- [6] Mudholkar, G. S., Hutson, A. D. 2000. The epsilon-skew-normal distribution for analysing near-normal data. *Journal of Statistical Planning and Inference*, 83, pp. 291-309.
- [7] Peel, M. C., Wang, Q. J., Vogel, R. M., McMahan, T. A. 2001. The utility of L-moment ratio diagrams for selecting a regional probability distribution. *Hydrological Sciences*, vol 46, pp. 147-155
- [8] R Core Team 2013: R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, <http://www.r-project.org/>.
- [9] Rmetrics Core Team, Asquith, W.H. 2018. *lmomco*: L-moments, censored L-moments, trimmed L-moments, L-comoments, and many distributions. R package, version 2.3.1, <https://cran.r-project.org/web/packages/lmomco/index.html>
- [10] Rmetrics Core Team, Komsta L. & Novomestky F. 2015. *moments*: Moments, cumulants, skewness, kurtosis and related tests. R package, version 0.14, <https://cran.r-project.org/web/packages/lmomco/index.html>

- [11] Vogel, R. M., Fennessey, N. M. 1993. L Moment Diagrams Should Replace Product Moment Diagrams. *Water Resources Research*, vol. 29, pp. 1745-1752.
- [12] Vogel, R. M., Zafirakou-Koulouris, A., Craig, S. M., Habermeier, J. 1998. L moment diagrams for censored observations. *Water Resources Research*, vol. 34, pp. 1241-1249.
- [13] Wang, W., Hutson, A.D., Miecznikowski, J.C. 2010. L-moment estimation for parametric survival models given censored data. *Statistical Methodology*, 7, pp. 655-667