Department for Applied Statistics
Johannes Kepler University Linz

IFAS Research Paper Series
2008-32

Some notes on the favorable estimation for fitting heavy tailed data

Milan Stehlík [a], Rastislav Potocký [b], Helmut Waldl[c], Zdeněk Fabián[d]

February 2008

[a]Department of Applied Statistics and Econometrics, Johannes Kepler Universität Linz, Austria
[b]Department of Applied Mathematics and Statistics, Comenius University, Bratislava, Slovakia
[c]Department of Applied Statistics and Econometrics, Johannes Kepler Universität Linz, Austria
[d]Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague, Czech Republic

Abstract

Assessment of compound sums has many applications in insurance, auditing and operation risk capital assessment among others. We study the behavior of the total claim amount with claims taken from a homogeneous portfolio. Actuaries distinguish several types of distributions to fit loss data: gamma, log-gamma, log-normal, gamma + log-gamma, gamma + log-normal and Pareto being the most important. We discuss some problems that one can encounter when misemploying the log-normal assumption based methods supported by Basel II framework. The compound sums are demonstrated to be highly sensitive on the individual claims distributions and thus a robust approach is needed. New estimators based on a robustified Johnson score are introduced and compared with the classical estimators (maximal likelihood and moment estimators) and with recently introduced robust estimators of "generalized median" and "trimmed mean" type. We derive the exact distribution of the likelihood ratio tests of homogeneity and tail index of the two-parameter Pareto model which support the assessment of performance of estimators. The real data example illustrates the concepts.

KEY WORDS: heavy-tailed distribution, claims, robust approach, Johnson estimator, insurance, Basel II

This paper is dedicated to Alexander Nagaev, who died on 10th February, 2005.

1 Introduction

Assessment of compound sums has many applications in insurance, auditing and operation risk capital assessment among others. Operational risk has been defined by the Basel Committee as 'the risk of financial loss resulting from inadequate or failed internal processes, people and systems or from external events' (see (Basel II initiative (2004)) and also (Voit 05), p. 349). Recent global trends in financial markets have increased many types of operational risks (see (Alexander 03)). In operation risk capital assessment there are many good reasons to use a lognormal distribution (see (Alexander 03)). Although this will not capture well the extremely high impact losses, these are by definition, very rare indeed. But exact assessment of these upper quantiles of risk, for example calculation of operational risk capital at a 99.9% percentile is of interest for Basel II related frameworks (see (Alexander 03)). This is a complex task and we would like to discuss the favorable estimators of the individual and aggregated heavy-tailed claims. If we are interested only in the extreme quantile estimation for heavy tailed data, then there are several methods, e.g. Blocks method, Peaks-over-threshold (POT) method and Quantile-based methods. In practice, probably the main drawback of the POT method is that for each different choice of a threshold the other parameters and quantile estimates are needed. Typically quantile methods are Hill estimator, Pickands estimator and moment estimator. For maximum likelihood (ML) method improving Q-based methods see (Beirlant et al. (1999)). However, in practice one could be interested in both aspects: in estimating parameters and in drawing conclusions on extreme quantiles behavior. Particularly, statistical extreme value models concern only the ultimate tail section of the distribution while a practitioner faced for instance with reinsurance rating (see (Beirlant et al. (2001))) will need to model also more central areas of the distribution in order to handle the different layers in a flexible way. Several recent papers treated robust and efficient estimation of tail index parameters for Pareto models, for large and small samples (see e.g. (Brazauskas and Serfling (2003))). Robustness in this setup is not satisfactorily covered by the recently arisen literature, however there is a rather good coverage in the estimation (see e.g. (Resnick (2006))). Particularly, in (Cantoni and Ronchetti 04) robust statistical procedures are presented for the analysis of skewed and heavy tailed outcomes as they occur in health care data. The importance of the new approaches supporting the regulation related to Basel II can be seen also from recent papers, e.g. for the case of the capital requirements and capital adequacy ratios for banks (Fouche et al. 06).

In this paper we compare classical ML and method of moments (MM) estimators with newer ones, based on Johnson score (see (Fabián 07)) and those provided by (Brazauskas and Serfling (2003)). As estimations based on Johnson score do not depend on the existence of moments, they may be interesting especially for the class of heavy-tailed distributions. The paper is organized as follows. In the 2nd section we shortly describe estimation based on Johnson score and robust "generalized median" and "trimmed mean" type of estimation. In the 3rd section of the paper we discuss and compare some recent results for compound sums when an individual distribution is from the subexponential family. In the 4th section we derive the distribution of exact likelihood ratio tests of the homogeneity and Pareto-tail index for the Pareto sample. This method, instead of classical KS test heavily dependent on estimation of parameters, is used to compare the performance of the various estimators. In the 5th section we demonstrate why robust estimates are needed in the analysis of sums of claims, discussing both the individual and aggregated claim sensitivity for the underlying distribution. The last section illustrates the proposed approach. In the first example we analyze a real data set consisting of 96 nonlife insurance payments observed in one year. In the 2nd example the Wind catastrophes data (1977) taken from (Hogg and Klugman 84) is analyzed. The proofs and technicalities are put into the Appendix to maintain a better discussion.

2 Estimators based on Johnson score, of "generalized median" and "trimmed mean" type

The maximum likelihood estimation is very sensitive to deviation from the theoretical distributions, also in the class of heavy-tailed distributions. Not surprisingly (see e.g. (Alexander 05)), the maximum likelihood estimator (which is the limiting case of minimum density power divergence estimator, MDPDE, see (Juárez and Schucany 04)) of parameters failed to provide a reasonable estimation. More robust alternatives to MLE approach have been proposed by e.g. (Juárez and Schucany 04) and (Marazzi and Ruffieux 95). (Juárez and Schucany 04) recommends to use MDPDE to estimate the parameters of the Generalized Pareto distribution.

2.1 Estimators based on Johnson score

A new method proposed by Fabián (2006, 2008) is based on a different idea. Data are considered to be taken from a non-contaminated, but possibly heavy-tailed distribution, and are treated by means of the Johnson score, which is a scalar inference function characteristic for the given distribution. It turns out that Johnson scores of heavy tailed distributions are bounded so that the estimates are 'naturally' robust in these cases.

Let $\mathcal{X} = (a, b) \in \mathbb{R}$. The *Johnson score* of distribution F with support $\mathcal{X}$ and with density $f(x)$ continuously differentiable according to $x \in \mathcal{X}$ is defined by

$$T(x) = \frac{1}{f(x)} \frac{d}{dx} \left(-\frac{1}{\eta'(x)} f(x) \right). \tag{1}$$

where $\eta : \mathcal{X} \to \mathbb{R}$, given by

$$\eta(x) = \begin{cases} x & \text{if} \quad (a,b) = \mathbb{R} \\ \log(x-a) & \text{if} \quad -\infty < a < b = \infty \\ \log \dfrac{(x-a)}{(b-x)} & \text{if} \quad -\infty < a < b < \infty \\ -\log(b-x) & \text{if} \quad -\infty = a < b < \infty, \end{cases} \tag{2}$$

is the *Johnson transformation* adapted for arbitrary interval support.

The philosophy behind this concept is the following. The score function of distribution G with support $\mathbb{R}$ and density g,

$$Q(y) = -\frac{g'(y)}{g(y)}, \tag{3}$$

equals to the likelihood score for location, the most important parameter characterizing the mode of G. However, (3) do not characterize distributions F with support $\mathcal{X} \neq \mathbb{R}$. The generalization (1) of (3) for F with support $\mathcal{X} \neq \mathbb{R}$ has the following meaning: F is viewed as a transformed '*prototype*' G with support $\mathbb{R}$, that is, $F(x) = G(\eta(x))$. Denoting by g the density of G, the density of F is

$$f(x) = g(\eta(x))\eta'(x), \quad x \in \mathcal{X}, \tag{4}$$

where $\eta'(x) = d\eta(x)/dx$ is the Jacobian of the transformation. By setting $y = \eta(x)$, we obtain from (1) and (4)

$$T(x) = \frac{1}{g(y)\eta'(x)} \frac{d}{dy}(-g(y)) \frac{dy}{dx} = Q(\eta(x)). \tag{5}$$

Johnson score of F is thus the transformed score function of its prototype.

A unique solution x^* of equation

$$T(x) = 0 \tag{6}$$

is called a *Johnson mean* (the solution is unique if the prototype G of F is unimodal). Johnson mean thus characterizes a typical value of unimodal heavy-tailed distributions, the mean of which does not exists (and is a value near the mean of light-tailed distributions).

Johnson score moments

$$E_f T^k = \int_{\mathcal{X}} T^k(x) dF(x), \qquad k = 1, 2, ... \tag{7}$$

exist if F satisfies the usual regularity requirements, and $E_f T = 0$. Let now $m \in \mathbb{N}$ and $x_1, ..., x_n$ be a random sample from F_θ with unknown $\theta \in \Theta \subseteq \mathbb{R}^m$. By the substitution principle we obtain equations for the 'Johnson score moment estimate' of θ in the form

$$\hat{\theta}_n : \qquad \frac{1}{n} \sum_{i=1}^{n} T^k(x_i; \theta) = ET^k(\theta), \qquad k = 1, ..., m, \tag{8}$$

which turns out to be consistent and asymptotically normal. Moreover, if F is heavy-tailed, T is bounded and the estimates are 'naturally' robust. Since the Johnson mean of $F_\theta, \theta \in \mathbb{R}^m$ is a function $x^* = x^*(\theta)$, its (robust) estimate can be obtained as $\hat{x}^* = x^*(\hat{\theta})$. It was shown by (Fabián 07) that $\hat{x}^*$ is asymptotically $N(x^*, \omega^2)$ where ω^2 is related with the second Johnson score moment of F. The Johnson estimation for particular distributions is treated in (Fabián 08). For illustration, in the case of the log-normal distribution the score based method brings nothing new in the context of possibility of transformation to the normal distribution. Trimmed version of Johnson score method for log-normal distribution coincides (dependently on trimming) with the transformation to R and Huber trimming for the normal distribution.

2.2 Robust estimators of "generalized median" and "trimmed mean" type

Several recent papers treated robust and efficient estimation of tail index parameters for Pareto models given by cdf $F(x) = 1 - (\frac{\lambda}{x})^\alpha, x > \lambda$, see e.g. (Brazauskas and Serfling (2003)) and (Vandewalle et al. 07). Estimators of "generalized median" and "trimmed mean" type were introduced by (Brazauskas and Serfling (2003)). They have been shown to provide more favorable trade-offs between efficiency and robustness than the several well-established estimators, including those corresponding to methods of ML, quantiles and percentile matching. Generalized median (GM) statistics are defined by taking the median of the $\binom{n}{k}$ evaluations of a given kernel $h(x_1, \ldots, x_k)$ over all k-sets of the data. In (Brazauskas and Serfling (2000a)) such estimators were considered for the parameter α in the case of λ known, with a particular kernel $h(x_1, \ldots, x_k, \lambda) = \frac{k}{C_k \sum_{i=1}^{k} \ln(x_i/\lambda)}$, where C_k is a multiplicative median unbiasing factor. The generalized median estimator of α for k-sets and known λ is $\hat{\alpha}_{GM,k,\lambda} = Median\{h(x_{i_1}, \ldots, x_{i_k}, \lambda)\}$

Asymptotic relative efficiency (ARE) with respect to the MLE is increasing with k, while robustness (largest proportion of the sample observations which may be contaminated without effecting the parameter estimation) decreases with k. For k-values between $k = 5, \ldots, 10$ we have high ARE and still relatively high robustness. In (Brazauskas and Serfling (2000b)) two similar generalized quantile estimators and particular generalized median estimators for the parameter α in the case of unknown λ are introduced by substituting λ in the kernel $h(x_1, \ldots, x_k, \lambda)$ with $\min\{x_1, \ldots, x_k\}$

and $\hat{\lambda}_{ML}$ respectively. Since $\hat{\lambda}_{ML}$ is biased we used the bias-corrected MLE $\hat{\lambda}_{unb}$ instead of $\hat{\lambda}_{ML}$. Also the median unbiasing factors have to be changed slightly. The generalized median estimators of α for k-sets and unknown λ are $\hat{\alpha}_{GM,k,min} = Median\{h(x_{i_1},\dots,x_{i_k},min\{x_{i_1},\dots,x_{i_k}\})\}$ and $\hat{\alpha}_{GM,k,\hat{\lambda}_{unb}} = Median\{h(x_{i_1},\dots,x_{i_k},\hat{\lambda}_{unb})\}$. Similarly to the case of known λ a reasonable trade-off between robustness and efficiency can be found with k-values between $k = 5,\dots,10$. ARE of $\hat{\alpha}_{GM,k,min}$ is slightly smaller than ARE of $\hat{\alpha}_{GM,k,\lambda}$ and $\hat{\alpha}_{GM,k,\hat{\lambda}_{unb}}$ respectively.

In (Brazauskas and Serfling (2000a)) a trimmed mean estimator $\hat{\alpha}_{TM}$ for α is discussed for the case of known λ. This estimator is similar to the MLE but discards the proportion β_1 of the lowermost and the proportion β_2 of the uppermost observations. The remaining observations are weighted such that $\hat{\alpha}_{TM}^{-1}$ is mean-unbiased for α^{-1}: $\hat{\alpha}_{TM,\beta_1,\beta_2} = (\sum_{i=1}^{n} c_{n,i,\beta_1,\beta_2}\frac{\log(x_{(i)})}{\lambda})^{-1}$ with mean unbiasing factors c_{n,i,β_1,β_2} and $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$.

3 Large deviations of compound sums within the subexponential family

Here we consider the random sum of claims

$$S = \sum_{i=1}^{N} X_i, \tag{9}$$

where N denotes an integer valued counting variable and $\{X_i\}$ constitutes a sequence of independent, identically distributed (iid) non-negative random variables with cdf F independent on counting. S is called the total claim amount or aggregate claim. Such a model is typical in ruin theory. We are interested in probabilities of large deviations of S under the assumption that F is heavy tailed. In particular, we assume that X_i has no finite exponential moments, i.e. standard large deviation theory does not apply (the existence of the moment generating function is crucial for proving Cramér's theorem, for more see (Mikosch and Nagaev 01)). A r.v. X (or its d.f.) is said to be heavy tailed on right-hand if $E(e^{rX}) = +\infty$ for any $r > 0$. Such a situation is typical for many distributions met in insurance, when one is interested in modeling large claims. In such a case distributions with exponentially decaying tail do not form adequate models. In what follows $\bar{F}(x) = 1 - F(x), x \ge 0$ denotes the tail of the cdf F. The most important heavy-tailed subclass is the subexponential class (denoted as $\mathcal{S}$). Provided that $X > 0$ a.s., (Mikosch and Nagaev 98) distinguish the following typical subclasses of $\mathcal{S}$: regularly varying tails RV(α), lognormal-type tails LN(γ), and Weibull-like tails WE(α). Distributions $F \in$ RV(α) are for instance the infinite variance stable distributions (with $\alpha < 2$), including the Cauchy distribution (with $\alpha = 1$), the Fréchet distribution, which is one of the extreme value distributions, the Burr and the loggamma distributions. The lognormal distribution belongs to LN(2). The heavy-tailed Weibull and the Benktander-type-II distributions are members of WE(α). All these distributions do not satisfy Cramér's condition. For a precise definition of these distributions we refer for instance to (Embrechts et al. 03), Chapter 1. As we have already mentioned, distributions in classes RV(α), LN(γ) and WE(α) are subexponential distributions provided that $X > 0$ a.s. The subexponentiality means that the tail of the sum of n random variables becomes large by a dominating large random variable, i.e. F is subexponential if for every $n \geq 2$ and x tending to $+\infty$, $P(\sum_{i=1}^{n} X_i > x) \sim P(\max_{1\leq k\leq n} X_k > x)$ holds. There are two other heavy-tailed subclasses, the class $\mathcal{L}$ of long-tailed d.f.'s and the class $\mathcal{D}$ of d.f.'s with dominatedly varying tails, which are closely related to class $\mathcal{S}$ (see (Ng et al. 02)). It is well-known that $\mathcal{D}\cap\mathcal{L} \subset \mathcal{S} \subset \mathcal{L}$, $\mathcal{D} \not\subset \mathcal{S}$, $\mathcal{S} \not\subset \mathcal{D}$ and $\text{RV}(\alpha) \subset \mathcal{D}$ (see (Ng et al. 02) and (Embrechts et al. 03), p. 50).

In what follows all random variables are positive with infinite support, i.e. $F(x) < 1$ for all $x > 0$. Let $F^{n^\star}$ be the n-fold convolution of F (for the definition see Embrechts et all, p.39). Now let us assume a subexponential individual claim. (Mikosch and Nagaev 01) have proved that for a claim from $T_1 = \{X : X \in \text{RV}(\alpha), \alpha > 1\}$ and under the assumption that the moment generating function of the counting variable N exists in a neighborhood of the origin, we have

$$\Delta(x) := \frac{P(S > x)}{P(X > x)} - E(N) \to 0, x \to +\infty. \quad (10)$$

But without additional conditions the approximation of $P(S > x)$ by $E(N)P(X > x)$ may be very bad, since the rate of convergence in (10) can be arbitrarily slow as it is shown in (Mikosch and Nagaev 01) (see also example 4 in section 5). In the same paper they proved that $\Delta(x) = O(1/x), x \to +\infty$ under some regularity conditions. For instance if $1 - F$ is regularly varying with some $\alpha > 2$, then its integrated tail distribution $F_I(x) = (1/E(X)) \int_0^x (1 - F(u))du, x \geq 0$ with $0 < E(x) < \infty$ satisfies these conditions. Here we relate the set T_1 to the set T_2 of claim r.v.'s, for which Theorem 2.3 in (Ng et al. 02) holds. Assuming $X > 0$ a.s. the next Lemma shows that $T_1 \subset T_2$ (for Proof see Appendix).

Lemma Assume $X > 0$ a.s. Then $T_1 \subset T_2 := \{X : X \in \mathcal{D} \cap \mathcal{L}$ and $E(X) < +\infty\}$.

Coming back to $\Delta(x)$ asymptotics, for instance let $\bar{F}_\alpha(x) = x^{-\alpha}$ be the tail of Pareto distribution, $\alpha > 0, x \geq 1$. (Mikosch and Nagaev 01) proved that even for such a regular distribution with $\alpha > 2$ the rate $O(1/x)$ cannot be improved. (Christoph 05) (see Theorem 2) has proved that for $1 < \alpha < 2$ and supposing $u_\alpha(x) = O(x^{-r})$ as $x \to \infty$ for $1 + \alpha < r \leq 2\alpha$ and $E(N^3) < \infty, \mu = E(X)$ we have

$$\Delta(x) = \frac{\alpha\mu(E(N^2) - E(N))}{x} + O(x^{-(r-\alpha)}),\ x \to \infty. \quad (11)$$

He assumed that $1 - F$ is regularly varying for some $0 < \alpha < 2, \alpha \neq 1$ and we define $u_\alpha(x) := 1 - F(x) - C(\alpha)x^{-\alpha}$ for some $C(\alpha) > 0$. If F is regularly varying with some index $0 < \alpha < 1$ then Proposition 1 in (Christoph 04) shows that the exact rate of convergence of $\Delta(x)$ is usually $x^{-\alpha}$ as $x \to \infty$. In the special case of $\alpha = 1/2$ Theorem 3 in (Christoph 05) says that

$$\Delta(x) = \frac{\mu^\star E(N^2) - \frac{1}{6}E(N^3) - (\mu^\star - \frac{1}{6})E(N)}{2x} + O(x^{-(r-\frac{1}{2})}),\ x \to \infty,$$

where $\mu^\star = \int_0^\infty xd(F(x) - G_{\frac{1}{2},1}(x))$ is the first pseudomoment and we assume $E(N^4) < \infty$, $u_{\frac{1}{2}}(x) = O(x^{-r})$ as $x \to \infty$ for $\frac{3}{2} < r \leq \frac{5}{2}, r \neq 2$. For $r = 2$

see (Christoph 05). The stable law $G_{\frac{1}{2},1}(x)$ is the Lévy-distribution with density $\frac{1}{2\sqrt{\pi}x^{\frac{3}{2}}e^{4x}}$ for $x > 0$.

4 Testing of the quality of fit

Typically, the classical GOF tests are used to compare the quality of different estimators. For instance, (Brazauskas and Serfling (2003)) use Kolmogorov-Smirnov, Cramér-von-Mises and Anderson-Darling statistics to assess the performance of different estimators. However, typically available for those tests are only the asymptotical quantiles and critical constants. The implementation of the exact or approximative Kolmogorov-Smirnov, Cramér-von-Mises and Anderson-Darling GOF tests for the case of estimated parameters of the null distribution is extremely difficult. Therefore in our paper we use the exact procedure to assess the quality of different estimators. The typical model used for a heavy tail is the $P(\lambda, \alpha)$ Pareto model given by cdf

$$F(x) = 1 - (\frac{\lambda}{x})^{\alpha}, x > \lambda, \tag{12}$$

where $\alpha > 0$ is the shape parameter that characterizes the tail distribution and $\lambda > 0$ is the scale parameter. A well known equivalence relation between Pareto model and truncated exponential distribution $E(\mu, \theta)$ having cdf

$$G(z) = 1 - e^{\frac{-(z-\mu)}{\theta}}, z > \mu, \mu \in R, \theta > 0 \tag{13}$$

constitutes a base for further considerations. Specifically, if random variable X has cdf given by (12), then variable $Z = \log X$ has cdf G given by (13) with $\mu = \log \lambda$ and $\theta = \alpha^{-1}$. Here we construct the exact LR tests of the homogeneity and Pareto tail index. The base will be the ELRT (exact likelihood ratio test) procedures given by (Stehlík (2006)). The following theorem gives the ELRT statistics of homogeneity

$$H_0 : \lambda_1 = \ldots = \lambda_N \text{ versus } nonH_0 \tag{14}$$

for the sample from $P(\lambda, \alpha)$. For proof see Appendix.

Theorem 1 *Let $x_1, \ldots, x_N$ be i.i.d. from the Pareto $P(\lambda, \alpha)$ family. Then the LR statistics $-\ln \Lambda_N$ of the hypothesis (14) has the form*

$$N \ln(\sum_{i=1}^{N} y_i) - N \ln N - \sum_{i=1}^{N} \ln y_i \tag{15}$$

where $y_i = \alpha(\log x_i - \log \lambda)$. Under the H_0 it has the same distribution as the random variable

$$-\ln\{N^N u_1 \ldots u_{N-1}(1 - u_1 - \ldots - u_{N-1})\},$$

where the vector $(u_1, \ldots, u_{N-1})$ has a generalized Beta distribution $B(1, \ldots, 1)$ on the simplex

$$\{u : 0 < u_1 < 1, \ldots, 0 < u_{N-1} < 1 - u_1 - \ldots - u_{N-2}\}.$$

The generalized Beta distribution is in the literature also called the Dirichlet distribution or the multivariate Beta distribution. Notice, that the determination of the likelihood ratio statistic is no problem. The very important property of the LR test of homogeneity is that its distribution is under the H_0 independent on the unknown tail index α (this is an advantage against some asymptotical tests and tests dependent on true but unknown value of α). This property, in the case of exponential model the scale invariance, is discussed in (Stehlík (2006)). For other properties of such a test in the case of exponential samples see (Stehlík (2006)) and (Stehlík and Wagner (2008)). Thus the LR test of the homogeneity is dependent only on the location parameter λ. However, both ML and MM estimation of λ brings problems, since in the case of ML estimation we have $\hat{\lambda}_{ML} = \min\{x_1, ..., x_n\}$ and thus $-\ln \Lambda_N$ becomes infinity and in the case of MM estimation we get a negative values of $y_i's$. However, in actuarial literature, the assumption of λ being known is quite typical, because, as for example (Philbrick (1985)) states, "although there may be situations where this value must be estimated, in virtually all insurance applications this value will be selected in advance". See also discussion by (Rytgaard (1990)). When λ is assumed known, the Pareto $P(\lambda, \alpha)$ model is called a single-parameter Pareto model. When the sample is drawn by single-parameter Pareto model one can be interested in testing a hypothesis about the tail parameter α. The following Theorem 2 gives the exact distribution and the power function of the LR test of the hypothesis

$$H_0 : \alpha = \alpha_0 \text{ versus } H_1 : \alpha \neq \alpha_0. \tag{16}$$

For the proof see Appendix.

Theorem 2 *Let $x_1, \ldots, x_N$ be a sample from the single-parameter Pareto model $P(\alpha)$ (λ is known parameter). Then the LR statistics $-\ln \Lambda_N$ of the hypotheses (16) has the form*

$$-\ln \Lambda_N = G_N\left(\sum_{i=1}^{N} y_i\right) - G_N(N) \tag{17}$$

where $y_i = \alpha(\log x_i - \log \lambda)$. Under H_0 the cumulative distribution function (cdf) of the Wilks statistics $-2\ln \Lambda_N$ has the form (here $\tau > 0$)

$$F_N^{\Gamma}\left\{-NW_{-1}(-e^{-1-\frac{\tau}{2N}})\right\} - F_N^{\Gamma}\left\{-NW_0(-e^{-1-\frac{\tau}{2N}})\right\}. \tag{18}$$

Here we define for $u, x > 0$ the function $G_u(x) = x - u \ln x$. The $W_k,\ k \in \{-1, 0\}$ is k-th branch of Lambert W function (see Appendix) and F_N^{Γ} is gamma cdf with shape parameter N and scale parameter 1.

For the properties of the exact LR test for the case of the (transformed) exponential observations y_i see (Stehlík (2006)).

5 Why robust estimates are needed in the analysis of sums of claims

Actuaries (e.g. see (Hewitt and Lefkowitz 79)) distinguish several types of distributions to fit loss data: gamma, log-gamma, log-normal, gamma + log-gamma, gamma

+ log-normal and Pareto being the most important. They described the method to fit these five types of distributions to loss data and discussed applications of the fitted distributions to estimation problems. Casualty actuaries frequently need to extract information from insurance loss data. Used alone, each of these distributions assumes that the observed losses are generated by a single underlying process. This may not always be the case. For example, a sample of observed losses may contain some that involved litigation and others that did not. In this situation a single distribution may not fit the aggregate data as well as a combination of two or more distributions. The authors do not claim that all insurance loss data can be fitted by the methods described here. However, after many years of actuarial experience (see (Alexander 05; Hewitt and Lefkowitz 79)) one can be convinced that these methods will produce useful results for most practical problems. In operation risk capital assessment there are many good reasons to use a log-normal distribution (see (Alexander 03)). Although this will not capture well the extremely high impact losses, these are by definition, very rare indeed. Moreover, especially when we analyze the sums of claims, robust estimates and procedures are needed. Both of these facts are demonstrated by the following examples. All simulations are made using Matlab 7.1. Here we study empirically the quality of the upper-quantile closeness drawn from classical parametric inference (method of moments and maximum likelihood estimation) and Johnson inference when the individual claim distribution is misspecified. Such studies are typical in various fields of statistics and are sometimes called sensitivity analysis. Example 1 shows the quantiles deviations when log-gamma distribution is misspecified by a log-normal one, in example 2 we particularly study the misspecification of log-gamma by a log-normal claim when MLEs are used. In example 3 the upper quantiles misspecification effect when a mixture of gamma and log-gamma is misspecified by a log-normal distribution is studied. The model of a mixture of gamma and log-gamma is given by (Hogg and Klugman 84), p. 51. Example 4 is motivating the estimation method based on Johnson score given in section 2. In example 5 we demonstrate that the difference between MLE and moment estimation is also large in the case of log-gamma distributed individual claims (compare with example 4). Example 6 shows, that the application of Mikosch Nagaev formula to estimation of upper extreme quantiles is very sensitive. Example 7 discusses the application of Gerd-Christoph formula (11). In all these examples a significant difference in quantiles was observed and plotted by qq-plots. The qq-plot provides a somewhat informal but convenient way of graphical detection of such a difference.

5.1 Example 1

We consider a collective claim arisen from the i.i.d. individual claims log-gamma distributed. For our example we assume, as it is usual in actuarial practice, the individual claims to be log-normal $LN(\mu, \sigma)$. We obtain the ML-estimates of μ and σ by using the values of $z_i = \log(x_i)$ and computing the ML-estimators of the normally distributed data z_i. In our example the expected values of the parameter

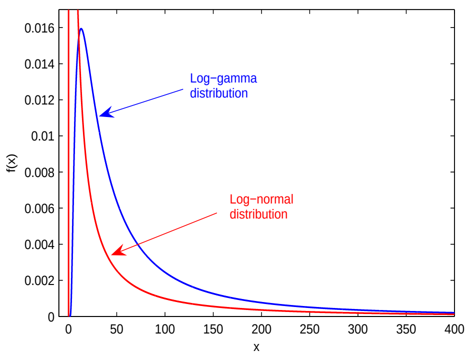


Figure 1: Distributions of the individual claims

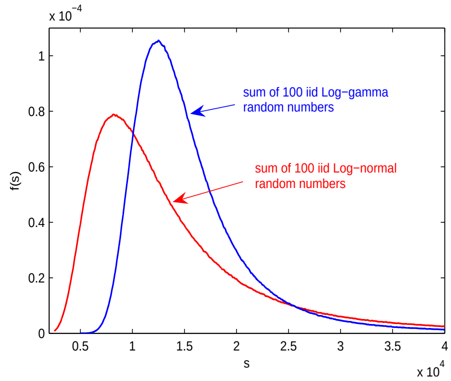


Figure 2: Empirical distributions

estimates are:

$$E(\hat{\mu}) = a\log(\frac{\sqrt{1-\frac{2}{b}}}{(1-\frac{1}{b})^2}), \qquad E(\hat{\sigma}^2) = a\log(\frac{(1-\frac{1}{b})^2}{1-\frac{2}{b}})$$

For the comparison of the distributions of the sums of independent identically distributed individual claims we took these expectations as the parameters of the assumed log-normal distribution.

Comparison of the distributions of the individual claims

For our example we set the parameters of the loggamma distribution to $a = 10$ and $b = 2.5$. For b we chose a value greater than 2 because for the loggamma distribution there exist only the moments of order less than b and we wanted to ensure the existence of the variance for our data. In figure 1 the real distribution (red) of the individual claims (loggamma with $a = 10$ and $b = 2.5$) and the assumed distribution (blue) of the individual claims (lognormal with $\mu = 2.1693$ and $\sigma^2 = 5.8779$) are shown. The two distributions do not look similarly, the lognormal distribution has the heavier tail.

Comparison of the distributions of the sum of claims

We are interested in the sum of iid individual claims $S = \sum_{i=1}^{N} X_i$. The distribution of the sum depends on the distribution of the individual claims. Since the distribution of the sums cannot be derived analytically, we obtained the distributions empirically via simulations.

In our example we set the number of individual claims to be summed up to $N = 100$. 1.000.000 simulations of 100 iid loggamma and lognormal distributed claims respectively were generated and added each. The distributions of these 2.000.000 sums are shown in figure 2. The two distributions clearly differ from each other.

For the computation of ruin probabilities upper thresholds for the sums are used. These thresholds are very sensitive to the distribution of the individual data. This can also be seen in figure 3: If we take the quantiles of the sum of lognormal variables instead of the quantiles of the sum of loggamma variables we will overestimate the ruin probability and vice versa.

We can conclude that the distribution of the sum of claims is highly sensitive to the distribution of the individual claims. Thus for the analysis of operational risk it is necessary to implement robust estimators and robust procedures respectively.

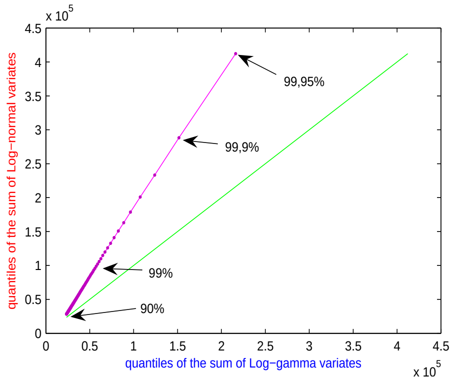


Figure 3: Q-Q-Plot

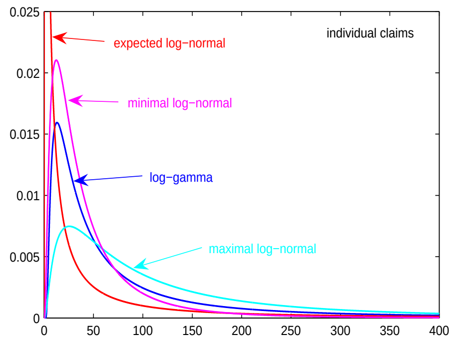


Figure 4: Range of assumed distributions

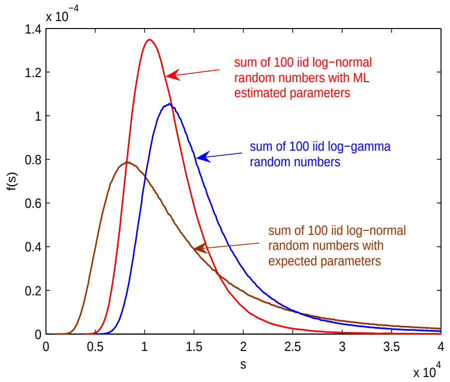


Figure 5: Distribution of the sum of claims based on ML estimation

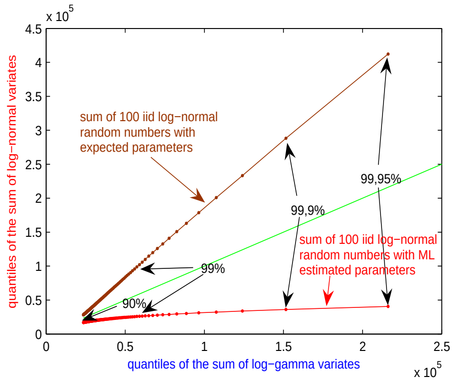


Figure 6: Q-Q plot of the sum of log-gamma and log-normal claims

5.2 Example 2

Again, we consider a collective claim arisen from the i.i.d. individual claims log-gamma distributed and assume that the individual claims obeying a log-normal $LN(\mu, \sigma)$ distribution. In example 1 we have concluded that the distribution of the sum of claims is highly sensitive to the distribution of the individual claims. Here we illustrate the non-robustness of MLE.

For the comparison of the distributions of the sums of independent identically distributed individual claims we took the maximum likelihood estimates (MLE) as parameters of the assumed log-normal distribution. Since we compared the distributions of the sum of claims empirically, we got different MLE in each simulation step. To show how the assumed distributions differ from the log-normal distribution with the expected values of the MLE, we computed the Kolmogorov-Smirnov-test statistic in each simulation step. I.e. we computed the maximal positive and negative difference between the cdf of the log-normal distributions with MLE as parameters and with the expected values of MLE as parameters.

Comparison of the distributions of the individual claims

For our example we set the parameters of the log-gamma distribution to $a = 10$ and $b = 2.5$. For b we chose a value greater than 2 because for the log-gamma distribution only the moments of order less than b exist and we wanted to ensure the existence of the variance. In figure 4 the real distribution of the individual claims

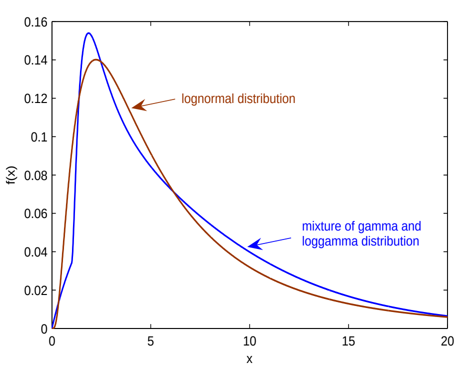


Figure 7: Real (brown) and assumed (blue) distributions of the individual claims

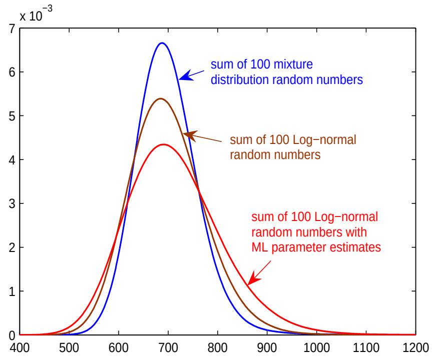


Figure 8: Empirical distributions of the sum of mixture and lognormal distributed random numbers

(log-gamma with $a = 10$ and $b = 2.5$) and the range of assumed distributions of the individual claims (log-normal with expected MLE $\mu = 2.169$ and $\sigma^2 = 5.878$ and distributions with maximal positive and negative cdf deviation from the expected distribution) are shown. The log-gamma and the expected log-normal distribution do not look similarly (the log-normal distribution has the heavier tail) but we are basically interested in the distributions of the sum of N iid claims. We can also see that the variation of the log-normal distributions is not negligible.

Comparison of the distributions of the sum of claims

Here we are interested in the sum of iid individual claims $S = \sum_{i=1}^{N} X_i$. We set the number of individual claims to be summed up to $N = 100$. 1.000.000 simulations of 100 iid log-gamma and log-normal distributed claims respectively were generated and added each. The distributions of these 2.000.000 sums are shown in figure 5. The two distributions clearly differ from each other, but the difference is less than the difference in example 1. For the computation of ruin probabilities upper thresholds for the sums are used. These thresholds are very sensitive to the distribution of the individual data. If we look at the Q-Q plot in figure 6 we see that we underestimate the ruin probability if we take the quantiles of the sum of log-normal claims based on ML estimation instead of quantiles of the sum of log-gamma claims. This is totally inconsistent to the results of example 1 where the ruin probability is overestimated when log-normal claims instead of log-gamma claims are assumed.

Also this example shows, that the distribution of the sum of claims is highly sensitive to the distribution of the individual claims. If we use the Johnson estimates, we get the same results as for the MLE, since log-normal distribution can be obtained from normal distribution by the Johnson transformation.

5.3 Example 3

The third example is based on the model given by (Hogg and Klugman 84), p. 51. Actuaries have found (see (Hogg and Klugman 84) and (Hewitt and Lefkowitz 79)), that a mixture of loggamma and gamma distributions is an important model for

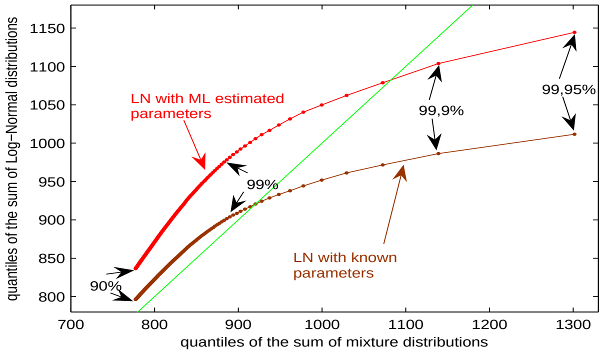


Figure 9: Q-Q-Plot of the distribution of the sum of 100 mixture variables vs the sum of 100 iid lognormal variables

claim distributions. In this example we consider the mixture with p.d.f. of the form

$$f(x) = \begin{cases} (1-p)\dfrac{b_2^{a_2} x^{a_2-1}}{\Gamma(a_2)} \exp(-b_2 x), & 0 < x \leq 1 \\ p\dfrac{b_1^{a_1}(\ln x)^{a_1-1}}{\Gamma(a_1)x^{b_1+1}} + (1-p)\dfrac{b_2^{a_2} x^{a_2-1}}{\Gamma(a_2)} \exp(-b_2 x), & 1 < x, \end{cases}$$

and zero elsewhere. Notice, that the variance is not simply the weighted average of the two variances but also includes a positive term involving the weighted variance of the means. In our example mean and variance of the gamma-distributed component are given by:

$$m_\Gamma = \frac{a_2}{b_2} \qquad v_\Gamma = \frac{a_2}{b_2^2}$$

Mean and variance of the loggamma-distributed component are given by:

$$m_{log\Gamma} = (1 - \frac{1}{b_1})^{-a_1} \qquad v_{log\Gamma} = (1 - \frac{2}{b_1})^{-a_1} - (1 - \frac{1}{b_1})^{-2a_1}$$

Using these parameters, we can compute mean and variance of the mixture distribution as follows:

$$E(x) = p \cdot m_{log\Gamma} + (1-p) \cdot m_\Gamma$$

$$Var(x) = p \cdot v_{log\Gamma} + (1-p) \cdot v_\Gamma + p(1-p)(m_{log\Gamma} - m_\Gamma)^2$$

Assumed distribution of the data

In practice we never know the real distribution of the data. For our example we expect the data to be distributed log-normal (what is a common practice of actuaries). Mean and variance of the lognormal distribution are given by:

$$E(x) = \exp(\mu + \frac{\sigma^2}{2}) \qquad Var(x) = \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$$

For the comparison of the distributions of the sums of independent identically distributed individual claims we took the same means and variances for the mixture-distribution and the assumed lognormal distribution respectively.

Comparison of the distributions of the individual claims

For our example we set the parameters of the loggamma component of the mixture distribution to $a_1 = 3$ and $b_1 = 2.5$. The parameters of the gamma component of the mixture distribution were set to $a_2 = 2$ and $b_2 = 0.25$. Finally we set the mixing probability to $p = 0.3$. In figure 7 the real distribution of the individual claims (mixture of loggamma and gamma) and the assumed distribution of the individual claims (lognormal with $\mu = 1.3393$ and $\sigma^2 = 0.8246$) are shown. The two distributions are quite similar.

Comparison of the distributions of the sum of claims

We are interested in the sum of iid individual claims $S = \sum_{i=1}^{N} X_i$. The distribution of the sum depends on the distribution of the individual claims. Since the distribution of the sums cannot be derived analytically, we obtained the distributions empirically via simulations.

In our example we set the number of individual claims to be summed up to $N = 100$. 1.000.000 simulations of 100 mixture distributions and 100 log-normal distributed claims respectively were generated and added each. The distributions of these 2.000.000 sums are shown in figure 8. The two distributions differ from each other. The sum of the loggamma distributed variables seems to vary more, which is not true, if we look at the empirical variances.

In figure 9 the quantiles of the distribution of the sum of log-normal variables are plotted against the quantiles of the distribution of the sum of the mixture variables. It can be seen that the distribution of the sum of the mixture distributed claims has the heavier tail.

5.4 Example 4

Example 4 is motivating the estimation method based on Johnson score discussed in section 2. Assume for the sake of simplicity that the individual claim distribution is gamma. Then the Johnson score estimation is identical with the method of moments. We have simulated 100 gamma distributed random numbers with parameters $\alpha = 3$ and $\gamma = 0.25$, computed the MLE and moment estimates of α and γ and then simulated 100 gamma random numbers with the MLE and moment estimated parameters each (1.000.000 simulation steps). The distribution of the sum of moment estimated gamma distributions varies more than the sum of MLE gamma distributions (see figure 10). Also the upper quantiles from $(0.9, 0.9995)$ differ as we can see in figure 11. We can conclude a significant difference between maximum likelihood and moment estimation. With another choice of parameter values we get a result that is completely different: with $\alpha = 3$ and $\gamma = 4$ the moment estimation method yields results that are superior to the results of ML estimated parameters.

5.5 Example 5

In this example we demonstrate that the difference between MLE and moment estimation is large also in the log-gamma case (compare with example 2). We simulated 100 log-gamma distributed random numbers with $a = 10$ and $b = 2.5$, computed the MLE and moment estimates of a and b numerically, and then simulated 100 log-gamma random numbers with the MLE and moment estimated parameters each (1.000.000 simulation steps each). The distribution of the sum of MLE log-gamma distributions varies more than the sum of moment estimated log-gamma distributions (see figure 12). Also the upper quantiles (0.9 to 0.9995) differ, as we can see

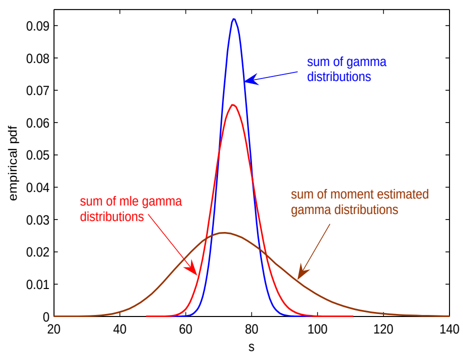


Figure 10: Sum of gamma and estimated gamma distributions

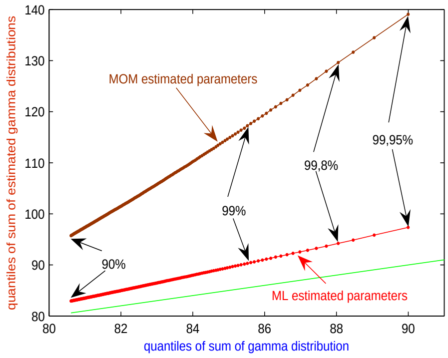


Figure 11: Q-Q plot of the sum of gamma and estimated gamma claims

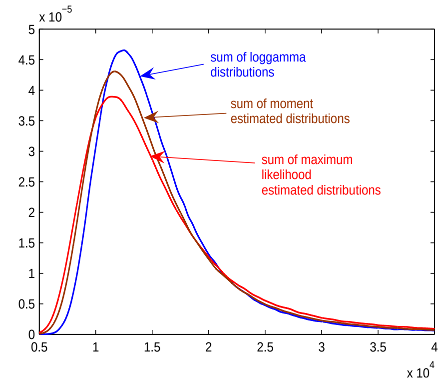


Figure 12: Sum of log-gamma and estimated log-gamma distributions

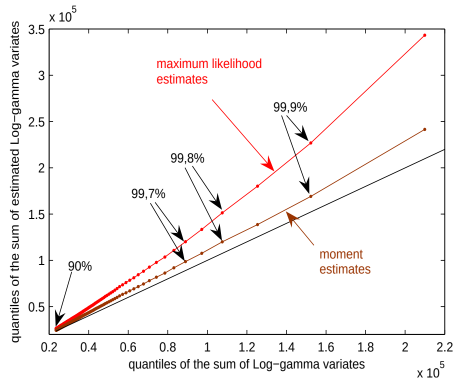


Figure 13: Q-Q plot of the sum of log-gamma and estimated log-gamma claims

in figure 13. Additional details: the expectation of the sum should theoretical be: $100 \cdot (1 - \frac{1}{\gamma})^{-\alpha} = 16538.17$, the mean of the 1.000.000 simulated log-gamma sums was 16577.98, the mean of the 1.000.000 simulated moment estimated log-gamma sums was 16602.03, the mean of the 1.000.000 simulated MLE log-gamma sums was 17408.83 (here moment method is clearly better), the variance of the sum should theoretically be: $100 \cdot ((1 - \frac{2}{b})^{-a} - (1 - \frac{1}{b})^{-2a}) = 9.738 \cdot 10^8$, the variance of the 1.000.000 simulated log-gamma sums was $7.22 \cdot 10^8$, the variance of the 1.000.000 simulated moment estimated log-gamma sums was $1.029 \cdot 10^9$, the variance of the 1.000.000 simulated MLE log-gamma sums was $4.229 \cdot 10^9$.

5.6 Example 6

In this example we demonstrate that application of Mikosch-Nagaev results to estimation of the extreme upper quantiles is rather non-robust and sensitive.

We made 1.000.000 simulations of a Poisson variate N with parameter $\lambda = 10.000$ and then of the sum of N log-gamma variables with parameters $a = 10$ and $b = 2.5$. For a comparison we also simulated the sum of $\lambda = 10.000$ log-gamma variables with parameters $a = 10$ and $b = 2.5$. The two distributions of the sums (with N Poisson distributed and fixed at $N = \lambda$ respectively) seem to be the same.

We studied the speed of convergence of the Mikosch-Nagaev theorem with the

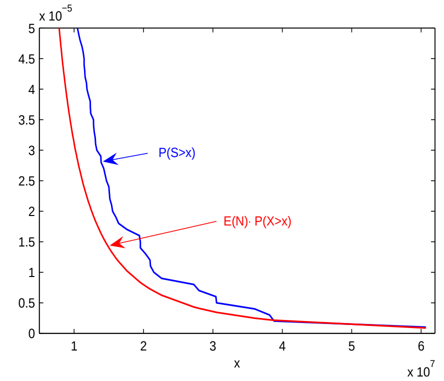


Figure 14: May we approximate $P(S > x)$ by $E(N) \cdot P(X > x)$ for large x?

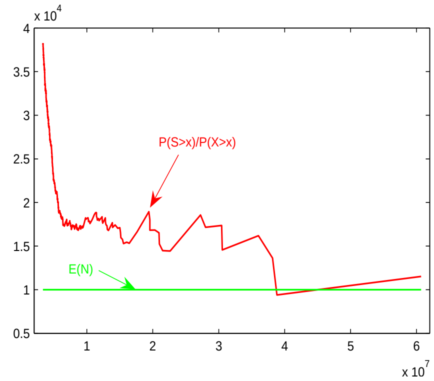


Figure 15: convergence of the Mikosch-Nagaev theorem with log-gamma distributed data

empirical distribution of the sum of N iid log-gamma variates with N Poisson distributed. First we compared $P(S > x)$ with $E(N) \cdot P(X > x)$ which can be seen in figure 14. It seems as if $E(N) \cdot P(X > x) - P(S > x) \to 0$ with $x \to \infty$. Convergence does not look that fast if we plot $\frac{P(S>x)}{P(X>x)}$ against $E(N)$ (see figure 15) but it still seems to be acceptable. The applicability of the theorem can be seen if we plot $\frac{P(S>x)}{P(X>x)}$ for small values of $P(S > x)$ (see figure 16). It seems as if $\frac{P(S>x)}{P(X>x)}$ would converge to $E(N)$ but the theorem is not applicable for probability values used in actuarial practice, convergence is too slow.

5.7 Example 7

In this example we demonstrate that application of Christoph's formula (11) to estimation of the extreme upper quantiles is rather sensitive.

We made 1.000.000 simulations of the sum of N Pareto distributed random variates with $\alpha = 1.5, \lambda = 1$ where N is Poisson distributed with parameter $\lambda = 10.000$. In figure 17 we can see $\Delta(x)$ and $\frac{\alpha\mu(E(N^2)-E(N))}{x}$. It looks as if the Christoph's theorem based asymptotics is sensitive in this setup. This can be observed also in the figures 18 and 19 where we plotted $\varepsilon(x) := \Delta(x) - \frac{\alpha\mu(E(N^2)-E(N))}{x}$ and $x^{\alpha-r}$ for $1+\alpha < r \leq 2\alpha$ against x and $P(S > x)$ respectively. This behavior can be partially explained by the fact that we know $\Delta(x)$ only empirically and that the variation of the empirical $\Delta(x)$ is very high for high x-values.

6 Illustrating Examples

In this section we illustrate the performance of estimators on real data set of Wind catastrophes claims (1977) modeled by Pareto distribution and on real data based on a non-life insurance sample consisting of 96 payments in one year. Here we consider both American and European Pareto parametric model. These models arise as parametric models in actuarial science, economics and reliability as well as in semiparametric modeling of upper observations in samples from distributions which are regularly varying or in the domain of attraction of extreme value distribution.

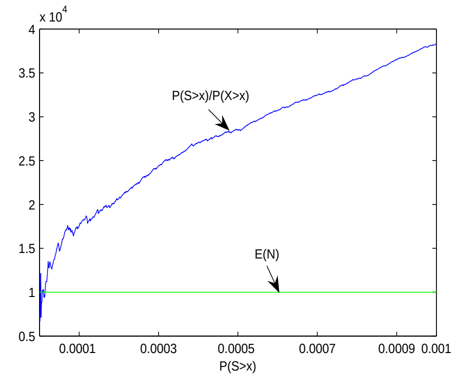


Figure 16: Fit of upper quantiles by Mikosch-Nagaev formula

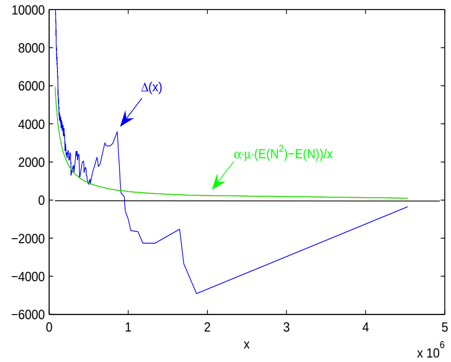


Figure 17: Christoph-Theorem fit

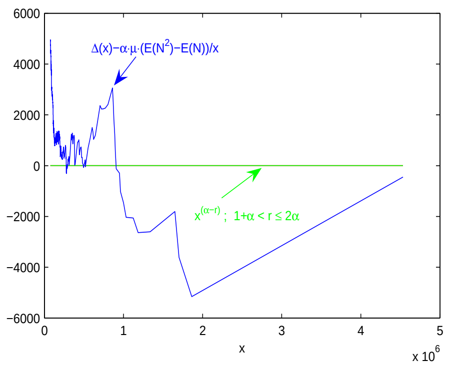


Figure 18: Christoph-Theorem fit

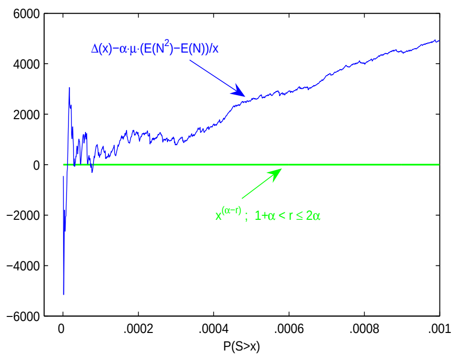


Figure 19: Christoph-Theorem fit

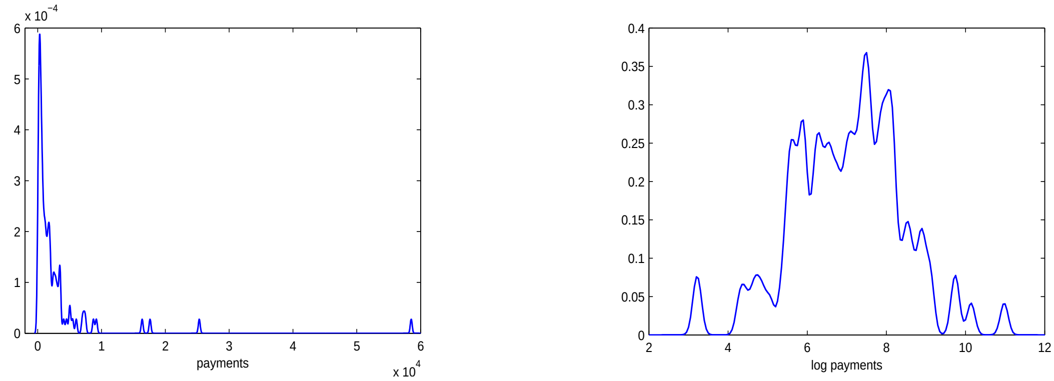


Figure 20: distribution of the payments with linear and the logarithmic scale

6.1 Example 1: Application to non-life insurance data

Here we consider a non-life insurance (real data) sample consisting of 96 payments in one year taken from (Plevová 99). The values are in thousands of Slovak crowns, displayed in table 1, see the data plot in figure 20.

24	26	73	84	102	115	132	159	207	240	241	254
268	272	282	300	302	329	346	359	367	375	378	384
452	475	495	503	531	543	563	594	609	671	687	691
716	757	821	829	885	893	968	1053	1081	1083	1150	1205
1262	1270	1351	1385	1498	1546	1565	1635	1671	1706	1820	1829
1855	1873	1914	2030	2066	2240	2413	2421	2521	2586	2727	2797
2850	2989	3110	3166	3383	3443	3515	3521	3531	4068	4527	5006
5065	5481	6045	7003	7245	7477	8738	9197	16370	17605	25318	58524

Table 1: Payments (in thousands of Slovak crowns)

First, we used the QQ method to show that the data are heavy tailed. Some typical diagnostics like Hill estimator (see figure 21) do not work optimally for these data but the QQ method works pretty well. The Hill estimator is a consistent estimator of $\frac{1}{\alpha}$ as $n, k \to \infty$ and $\frac{k}{n} \to 0$ under the convergence condition on the tail empirical measure (see (Resnick (2006))).

In figure 22 we give a QQ plot of the log-transformed data matched against exponential quantiles. Clearly, we should be looking at exceedances as not all data fall on the line. So we can conclude that the data are heavy tailed. Here the employed QQ method is based on the consideration, that our data have exact or approximate Pareto tail, what is also the base of the POT method, where the approximate distribution of the large values relative to a threshold is replaced by the limiting Pareto distribution. This means $P(X > x) \approx (\frac{x}{\lambda})^{-\alpha}, x > \lambda$ and we have $P(\alpha \ln \frac{X}{\lambda} > y) = e^{-y}, y > 0$, (see (Resnick (2006))) so we should plot $\{(-\ln(1 - \frac{i}{n+1}), \ln X_{i:n}), 1 \le i \le n\}$ and if H_0 of Pareto tail is correct or at least approximately correct, the plot should be roughly linear with slope α^{-1} and intercept $\ln \lambda$. This is an adaptation of the classical QQ plot which for normal data has a robust version RQQ based on a robust standardization of the observed data (see (Gel et al. 05)).

Assuming the data are a random sample from a distribution with a regularly varying tail with index $-\alpha$ the QQ estimator is a weakly consistent estimator of α^{-1} (for $k, n \to +\infty$ and $k/n \to 0$, see (Resnick (2006)), Theorem 4.3). The QQ

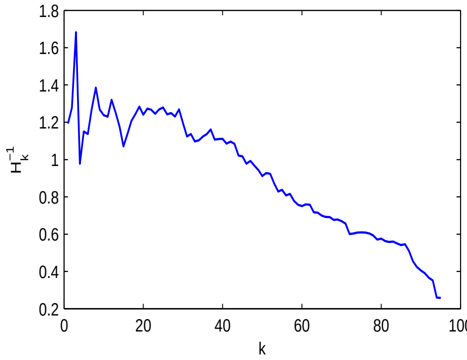


Figure 21: Hill plot

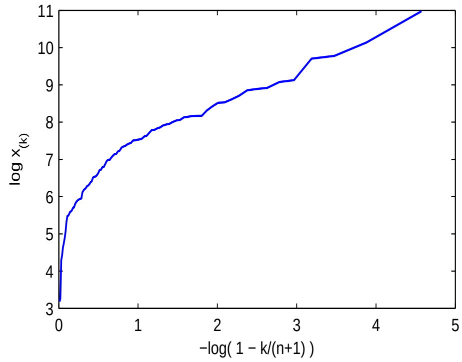


Figure 22: Q - Q plot

estimator, which is the formalization of the idea that the slope of the least squares line fitted to the QQ plot is an estimate of α^{-1}, gave for our data the value of $\hat{\alpha}_{QQ} = 0.7234$.

6.1.1 European Pareto

Here we consider the European Pareto parametrization, $\bar{F}(x) = (\frac{\lambda}{x})^{\alpha}$. We have computed the p-value and estimated Johnson mean $\hat{x}^*$ for the LR test for the Pareto model $P(\alpha, \lambda)$ with fixed (known) λ. We fixed $\lambda = 20$ and used different estimators for α. The results can be seen in table 2.

The p-values and estimated Johnson mean $\hat{x}^*$ for estimated λ parameters can be seen in table 3.

The p-value for the moment estimators cannot be computed because $\hat{\lambda}_{MM} > x_{min}$. The moment estimators $\hat{\alpha}_{MM} = 2.0919$ and $\hat{\lambda}_{MM} = 1560.6$ are inconsistent with $x_{min} > \lambda$. But mean and variance of the Pareto distribution only exist for $\alpha > 2$ which might be an indicator for $\alpha < 2$.

6.1.2 American Pareto

Here we consider the American Pareto parametrization, which is considered also in (Plevová 99), $\bar{F}(x) = (\frac{\lambda}{\lambda+x})^{\alpha}$. We computed the estimated Johnson mean $\hat{x}^*$ for the LR test for the American Pareto model $P(\alpha, \lambda)$ with fixed and estimated λ respectively.

The parameter estimates and estimated Johnson mean $\hat{x}^*$ with $\lambda = 3000$ fixed and estimated $\hat{\lambda}$ can be seen in the tables 4 and 5.

The solution of the estimation equations for the Johnson estimates yield improper results when λ is not fixed. The parameter estimates we obtained are in agreement with the numerical results given in (Plevová 99), i.e. $\hat{\alpha}_{ML} = 1.9088$, $\hat{\lambda}_{ML} = 2704.48$, $\hat{\alpha}_{MM} = 2.46968$ and $\hat{\lambda}_{MM} = 4394.25$. See figure (23) for comparing the different sums of Pareto distributed random variables.

Running of the Kolmogorov Smirnov test is not a routine task here, since we are estimating the parameters of the null hypothesis distribution. The comparison of the empirical cdf and estimated cdf is given in Figure 24.

	$\hat{\alpha}$	p-value	$\hat{x}^*$
ML	0.2484	1	100.51
unbiased	0.2458	0.9185	101.363
MM	1.0067	$5.0938 \cdot 10^{-71}$	39.866
QQ	0.7234	$3.4641 \cdot 10^{-37}$	47.646
Johnson	0.0537	$3.4395 \cdot 10^{-33}$	392.43
Trimmed Mean Estimators $\hat{\alpha}_{TM,\beta_1,\beta_2}$			
$\beta_1 = 0,\ \beta_2 = \frac{1}{96}$	0.2400	0.7390	103.32
$\beta_1 = 0,\ \beta_2 = \frac{2}{96}$	0.2335	0.5487	105.66
$\beta_1 = 0,\ \beta_2 = \frac{3}{96}$	0.2278	0.4042	107.78
$\beta_1 = 0,\ \beta_2 = \frac{4}{96}$	0.2228	0.2965	109.75
$\beta_1 = 0,\ \beta_2 = \frac{5}{96}$	0.2180	0.2113	111.74
$\beta_1 = 0,\ \beta_2 = \frac{6}{96}$	0.2180	0.2113	111.74
$\beta_1 = 0,\ \beta_2 = \frac{7}{96}$	0.2094	0.1036	115.53
$\beta_1 = 0,\ \beta_2 = \frac{8}{96}$	0.2054	0.0713	117.37
$\beta_1 = 0,\ \beta_2 = \frac{9}{96}$	0.2017	0.0486	119.18
$\beta_1 = 0,\ \beta_2 = \frac{10}{96}$	0.1980	0.0326	120.99
Generalized Median Estimators $\hat{\alpha}_{GM,k,\lambda}$			
$k = 1$	0.1682	0.0003	138.92
$k = 2$	0.2078	0.0902	116.22
$k = 3$	0.2209	0.2595	110.55
$k = 4$	0.2277	0.4007	107.84
$k = 5$	0.2317	0.5019	106.30
$k = 6$	0.2346	0.5803	105.24
$k = 7$	0.2365	0.6327	104.58
$k = 8$	0.2379	0.6761	104.05
$k = 9$	0.2391	0.7098	103.66
$k = 10$	0.2401	0.7400	103.31
$k = 15$	0.2428	0.8244	102.37
$k = 20$	0.2442	0.8688	101.89
$k = 30$	0.2456	0.9122	101.43
$k = 50$	0.2467	0.9480	101.05
$k = 70$	0.2472	0.9628	100.90
$k = 94$	0.2475	0.9728	100.80
$k = 95$	0.2476	0.9745	100.78
$k = 96$	0.2475	0.9729	100.80

Table 2: Estimates $\hat{\alpha}$, p-values and Johnson mean for $\lambda = 20$ fixed for European Pareto

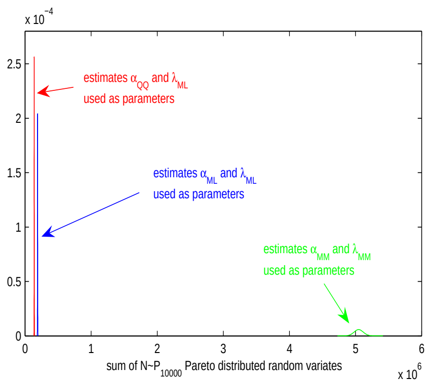


Figure 23: Distributions of the sums of Pareto variates with estimated parameters

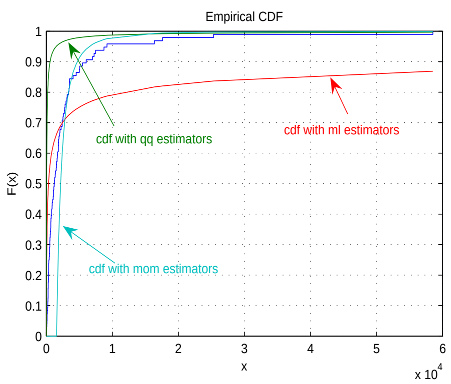


Figure 24: Graphical comparison of the ecdf and estimated cdf of the non-life insurance data

	$\hat{\lambda}$	$\hat{\alpha}$	p-value	$\hat{x}^*$
ML	24	0.2602	1	116.24
unbiased	23.0185	0.2547	0.9185	113.3914
MM	1560.6	2.0919	nan	2306.7
QQ	24 (ML)	0.7234	$1.3117 \cdot 10^{-33}$	57.176
Johnson	24 (ML)	0.0651	$1.7776 \cdot 10^{-28}$	392.43
Generalized Median Estimators $\hat{\alpha}_{GM,k,min}$				
$k=2$	–	0.5265	–	–
$k=3$	–	0.5230	–	–
$k=4$	–	0.5014	–	–
$k=5$	–	0.4823	–	–
$k=6$	–	0.4656	–	–
$k=7$	–	0.4504	–	–
$k=8$	–	0.4367	–	–
$k=9$	–	0.4244	–	–
$k=10$	–	0.4124	–	–
$k=15$	–	0.3706	–	–
$k=20$	–	0.3475	–	–
$k=30$	–	0.2797	–	–
$k=50$	–	0.2600	–	–
$k=70$	–	0.2574	–	–
$k=94$	–	0.2566	–	–
$k=95$	–	0.2566	–	–
$k=96$	–	0.2566	–	–
Generalized Median Estimators $\hat{\alpha}_{GM,k,\hat{\lambda}_{unb}}$				
$k=1$	23.0185 (unb)	0.1715	0.0002	157.25
$k=2$	23.0185 (unb)	0.2127	0.0701	131.26
$k=3$	23.0185 (unb)	0.2262	0.2152	124.80
$k=4$	23.0185 (unb)	0.2332	0.3423	121.72
$k=5$	23.0185 (unb)	0.2374	0.4358	119.96
$k=6$	23.0185 (unb)	0.2404	0.5094	118.76
$k=7$	23.0185 (unb)	0.2423	0.5590	118.01
$k=8$	23.0185 (unb)	0.2439	0.6005	117.41
$k=9$	23.0185 (unb)	0.2450	0.6328	116.96
$k=10$	23.0185 (unb)	0.2461	0.6621	116.56
$k=15$	23.0185 (unb)	0.2489	0.7443	115.50
$k=20$	23.0185 (unb)	0.2504	0.7880	114.95
$k=30$	23.0185 (unb)	0.2518	0.8308	114.43
$k=50$	23.0185 (unb)	0.2530	0.8665	114.00
$k=70$	23.0185 (unb)	0.2535	0.8813	113.83
$k=94$	23.0185 (unb)	0.2538	0.8912	113.71
$k=95$	23.0185 (unb)	0.2539	0.8930	113.69
$k=96$	23.0185 (unb)	0.2538	0.8913	113.71

Table 3: Estimates $\hat{\alpha}$, $\hat{\lambda}$, p-values and Johnson mean for European Pareto

	$\hat{\alpha}$	$\hat{x}^*$
ML	2.38	1260.5
MM	2.0034	1497.5
Johnson	2.0645	1453.2

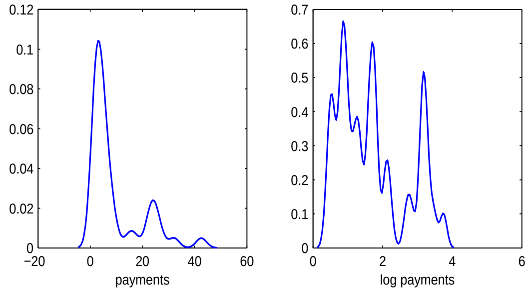
Table 4: Estimates $\hat{\alpha}$, and Johnson mean for $\lambda = 3000$ fixed

	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{x}^*$
ML	2704.3	1.9088	1416.8
MM	4412.5	2.4758	1782.3

Table 5: Estimates $\hat{\alpha}$, $\hat{\lambda}$ and Johnson mean for American Pareto

2	2	2	2	2	2	2	2
2	2	2	2	3	3	3	3
4	4	4	5	5	5	5	6
6	6	6	8	8	9	15	17
22	23	24	24	25	27	32	43

Table 6: Wind catastrophe data



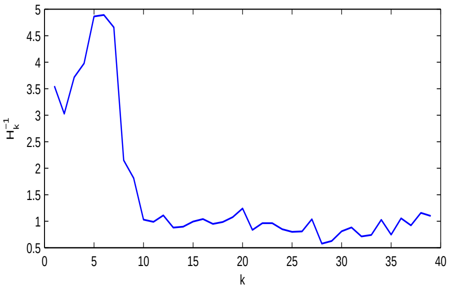


Figure 25: Hill plot

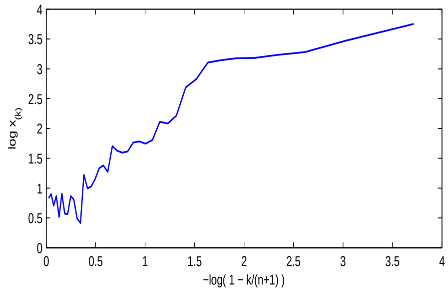


Figure 26: Q - Q plot

6.2 Example 2: Wind catastrophes claims (1977)

The Wind catastrophes data set is taken from (Hogg and Klugman 84), p. 64. It represents 40 losses that occurred in 1977 due to wind related catastrophes. The data were recorded to the nearest \$1.000.000 and include only those loses of \$2.000.000 or more. The table 6 displays the losses in million of dollars.

Since the data were rounded, for the estimation of the parameters of the Pareto distribution (cdf: $F(x) = 1 - (\frac{\lambda}{x})^{\alpha}$) uniformly distributed random numbers ($r \sim \mathbf{U}[-0.5; 0.5)$) were added to the original data. In the actuarial approach this is called data de-grouping.

6.2.1 European Pareto

The p-values and estimated Johnson mean $\hat{x}^*$ for estimated λ are given in Table 7.

The simulated distributions of sums of Pareto distributed random variates are quite different. We simulated 100.000 sums of N Pareto distributed rv with the ML, QQ and Johnson estimated parameters. In figure 27 the 3 pdfs are plotted. The distribution of the sums using the moment estimates is far away from the other two distributions, but also the distributions using $\hat{\alpha}_{QQ}$ and $\hat{\alpha}_{ML}$ respectively are clearly different from each other. In this case the qq-plots in figure 28 in fact are unnecessary, everyone can see the differences also without these plots. The distribution using the moment estimates is obviously wrong, which can be seen from figure 29 where the pdf of a Pareto distribution with parameters $\hat{\lambda}_{MM}$ and $\hat{\alpha}_{MM}$ is plot together with the empirical distribution of the data. The moment estimates for λ and α are inconsistent with the data (all data should be greater than λ, but $\hat{\lambda}_{MM} = 5.3102$). Mean and variance only exist for $\alpha > 2$. This might be an indicator either for $\alpha < 2$ or that we are fitting the wrong distribution.

Running of the Kolmogorov Smirnov test is not a routine task here, since we are

	$\hat{\lambda}$	$\hat{\alpha}$	p-value	$\hat{x}^*$
ML	1.5118	0.7668	1	3.4832
unbiased	1.4599	0.7281	0.8736	3.4647
MM	5.3102	2.3559	$1.9962 \cdot 10^{-24}$	7.5642
QQ	1.5118 (ML)	0.9294	0.2100	3.1384
Johnson	1.5118 (ML)	0.6492	0.3064	3.8405
Generalized Median Estimators 1				
$k = 2$	–	0.7410	–	–
$k = 3$	–	0.7244	–	–
$k = 4$	–	0.7473	–	–
$k = 5$	–	0.7546	–	–
$k = 6$	–	0.7550	–	–
$k = 7$	–	0.7556	–	–
$k = 8$	–	0.7550	–	–
$k = 9$	–	0.7547	–	–
$k = 10$	–	0.7541	–	–
$k = 11$	–	0.7530	–	–
$k = 12$	–	0.7519	–	–
$k = 15$	–	0.7497	–	–
$k = 25$	–	0.7460	–	–
$k = 35$	–	0.7424	–	–
$k = 36$	–	0.7418	–	–
$k = 37$	–	0.7420	–	–
$k = 38$	–	0.7418	–	–
$k = 39$	–	0.7392	–	–
$k = 40$	–	0.7413	–	–
Generalized Median Estimators 2				
$k = 1$	1.4599 (unb)	0.5452	0.0592	4.1374
$k = 2$	1.4599 (unb)	0.6216	0.2611	3.8083
$k = 3$	1.4599 (unb)	0.6573	0.4299	3.6809
$k = 4$	1.4599 (unb)	0.6781	0.5489	3.6126
$k = 5$	1.4599 (unb)	0.6870	0.6031	3.5849
$k = 6$	1.4599 (unb)	0.6927	0.6393	3.5673
$k = 7$	1.4599 (unb)	0.6976	0.6706	3.5525
$k = 8$	1.4599 (unb)	0.7009	0.6917	3.5428
$k = 9$	1.4599 (unb)	0.7038	0.7109	3.5341
$k = 10$	1.4599 (unb)	0.7060	0.7253	3.5276
$k = 11$	1.4599 (unb)	0.7075	0.7351	3.5233
$k = 12$	1.4599 (unb)	0.7091	0.7458	3.5186
$k = 15$	1.4599 (unb)	0.7122	0.7665	3.5096
$k = 25$	1.4599 (unb)	0.7176	0.8024	3.4942
$k = 35$	1.4599 (unb)	0.7198	0.8173	3.4880
$k = 36$	1.4599 (unb)	0.7199	0.8178	3.4877
$k = 37$	1.4599 (unb)	0.7205	0.8219	3.4860
$k = 38$	1.4599 (unb)	0.7208	0.8237	3.4853
$k = 39$	1.4599 (unb)	0.7202	0.8199	3.4869
$k = 40$	1.4599 (unb)	0.7219	0.8316	3.4820

Table 7: Estimates $\hat{\alpha}$, $\hat{\lambda}$, p-values and Johnson mean for European Pareto

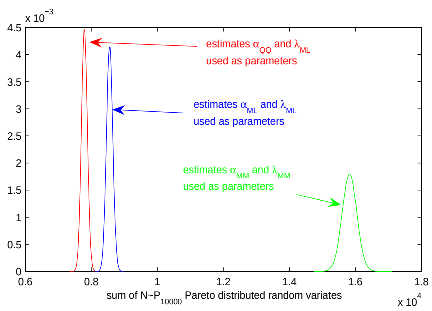


Figure 27: distribution of the simulated sum of 10.000 Pareto distributed rv

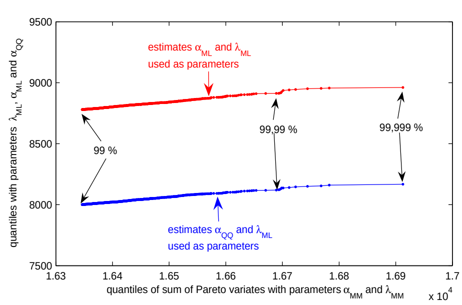


Figure 28: QQ - plot of simulated sums of Pareto using MM estimates and ML and QQ estimates respectively

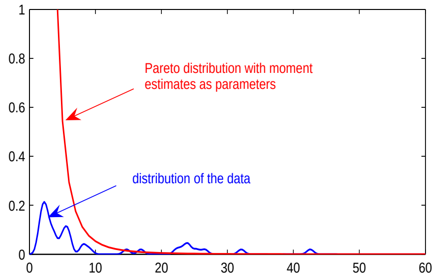


Figure 29: the Pareto distribution using moment estimates as parameters does not fit the data

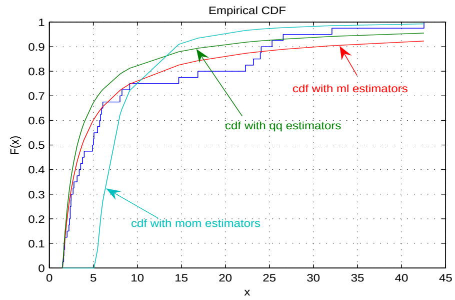


Figure 30: Graphical comparison of the ecdf and estimated cdf of the Wind Catastrophes Data

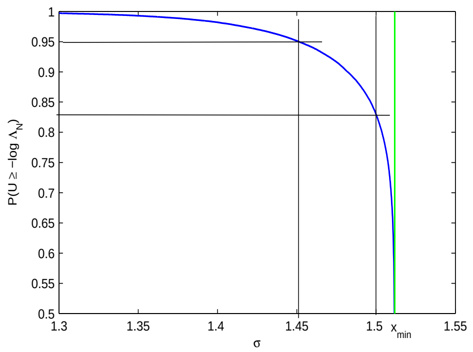


Figure 31: p-values for exact LR homogeneity test

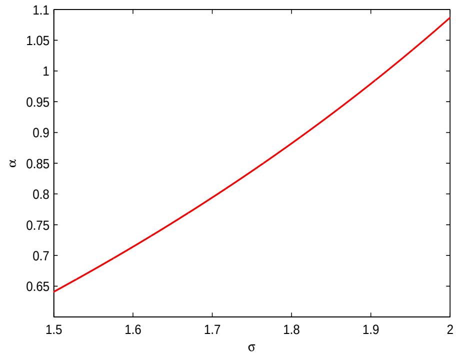


Figure 32: $\hat{\alpha}_J$ against λ

estimating the parameters of the null hypothesis distribution. The comparison of the empirical cdf and estimated cdf is given in Figure 30.

Now, let us consider the situation mostly typical in actuarial practice: the parameter λ is known. We have used the exact LR test of homogeneity (see Theorem 1) to test the goodness of fit for various values of λ. The figure 31 plots the p-values against λ. Particularly, for $\lambda = 1.5$ we obtain 0.8299 and the λ-value for which $P(U >= -\ln\Lambda) = 0.95$ is 1.4513.

For the case of Johnson estimation we have $\frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}} = \lambda\frac{\alpha+1}{\alpha}$. The dependence of $\hat{\alpha}_J$ values on λ is given by Figure 32. For $\lambda = 1.5$ we obtain $\hat{\alpha}_J = 0.6409$.

We computed the p-value for the LR test for the Pareto model $P(\alpha, \lambda)$ with fixed (known) λ. We fixed $\lambda = 1.5$ and used different estimators for α and estimated also the Johnson mean $\hat{x}^*$ (see table 8).

The ML estimators are biased but we can get unbiased estimates for λ and α on the basis of the ML estimators:

$$\hat{\alpha}_{unb} = \hat{\alpha}_{ML}\frac{n-1}{n} \qquad \hat{\lambda}_{unb} = \hat{\lambda}_{ML}(1 - \frac{1}{n\alpha})$$

(see also discussion in (Rytgaard (1990)) for the case of $\hat{\alpha}_{unb}$ with the known λ). For both estimates the second parameter has to be known respectively. So we computed the unbiased estimates recursively starting with the ML estimators. The iterations converge very fast (2 or 3 steps).

	$\hat{\alpha}$	p-value	$\hat{x}^*$
ML	0.7623	1	3.4678
unbiased	0.7432	0.8736	3.5183
MM	1.1941	0.0022	2.7561
QQ	0.9294	0.1957	3.1140
Johnson	0.6409	0.2874	3.8405
Trimmed Mean Estimators			
$\beta_1 = 0,\ \beta_2 = 0.025$	0.7271	0.7673	3.5630
$\beta_1 = 0,\ \beta_2 = 0.05$	0.7043	0.6220	3.6299
$\beta_1 = 0,\ \beta_2 = 0.075$	0.6868	0.5178	3.6841
$\beta_1 = 0,\ \beta_2 = 0.1$	0.6743	0.4481	3.7246
$\beta_1 = 0,\ \beta_2 = 0.125$	0.6657	0.4030	3.7534
$\beta_1 = 0,\ \beta_2 = 0.15$	0.6613	0.3812	3.7681
$\beta_1 = 0,\ \beta_2 = 0.175$	0.6608	0.3785	3.7700
$\beta_1 = 0,\ \beta_2 = 0.2$	0.6641	0.3950	3.7587
$\beta_1 = 0,\ \beta_2 = 0.225$	0.6658	0.4038	3.7528
$\beta_1 = 0,\ \beta_2 = 0.25$	0.6687	0.4186	3.7432
Generalized Median Estimators			
$k = 1$	0.5787	0.0964	4.0920
$k = 2$	0.6545	0.3480	3.7919
$k = 3$	0.6906	0.5400	3.6719
$k = 4$	0.7119	0.6700	3.6069
$k = 5$	0.7208	0.7265	3.5810
$k = 6$	0.7265	0.7638	3.5646
$k = 7$	0.7315	0.7962	3.5506
$k = 8$	0.7347	0.8176	3.5415
$k = 9$	0.7377	0.8372	3.5333
$k = 10$	0.7399	0.8519	3.5272
$k = 11$	0.7414	0.8615	3.5232
$k = 12$	0.7430	0.8724	3.5187
$k = 15$	0.7462	0.8931	3.5103
$k = 25$	0.7515	0.9289	3.4959
$k = 35$	0.7537	0.9434	3.4901
$k = 36$	0.7538	0.9438	3.4900
$k = 37$	0.7544	0.9480	3.4883
$k = 38$	0.7547	0.9498	3.4876
$k = 39$	0.7541	0.9458	3.4892
$k = 40$	0.7559	0.9580	3.4843

Table 8: Estimates $\hat{\alpha}$, p-values and Johnson mean for $\lambda = 1.5$ fixed for European Pareto

	$\hat{\alpha}$	$\hat{x}^*$
ML	4.1332	12.0971
MM	6.4191	7.7893
Johnson	6.3728	7.8459

Table 9: Estimates $\hat{\alpha}$ and Johnson mean for $\lambda = 1.5$ fixed for American Pareto

	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{x}^*$
ML	55.1486	6.9473	7.9381
MM	104.9797	12.3779	8.4812
Johnson	43.1091	5.5988	9.8501

Table 10: Estimates $\hat{\alpha}$, $\hat{\lambda}$ and Johnson mean for American Pareto

The value of QQ estimator (0.9294) tend to confirm that a typical tail parameter value for property is 1.0 (see (Philbrick (1985))).

6.2.2 American Pareto

Also for the wind catastrophe data we estimated the parameters and the Johnson mean $\hat{x}^*$ for the American Pareto model $P(\alpha, \lambda)$ with λ fixed at $\lambda = 50$ and unknown λ respectively. These estimates can be seen in the tables 9 and 10.

The p-values for the LR test for the Pareto model cannot be computed because all λ-estimates are $> \min x_i$

7 Conclusions

In recent decades the field of financial and insurance risk management has undergone explosive development. This paper discusses the favorable estimation for fitting individual heavy-tailed data or the aggregated claims of such heavy tailed individuals. The main novelty of the paper is that

(1) we compare the method of Johnson score based estimators to the other ones, recently discussed in e.g. (Brazauskas and Serfling (2003)) and show their advantages in some cases but also mention their drawbacks in others. We think they may serve as a good substitute where the so-called classical methods fail.

(2) we consider both aggregated and individual claims and demonstrate that the distribution of aggregate claims is highly sensitive to the distribution of individual claims. This is in agreement with (Brazauskas and Serfling (2000a)) showing that small errors in estimation of the tail index can already produce large errors in the estimation of quantiles based on the tail index. Hence robust operators and procedures have to be implemented.

We have also derived the exact distribution of the LR tests of homogeneity and Pareto-tail index for the Pareto sample, which can be of some interest for the practitioners (e.g. industry professionals or regulators) and theoreticians active in this area. Such a test has some optimality properties, for instance it could be used for any sample size since no asymptotical considerations are involved. As the simulations and a real-data example shows, the favorable estimation is highly sensitive on the underlying parametric model of the heavy tailed data. Our findings for the Pareto are in accord with conclusions of (Brazauskas and Serfling (2000a)) that the ML estimator is efficient but not robust and should be replaced by a competitor. The generalized median approach dominates the ML, quantiles and percentile matching.

ML is not robust also for the case of misspecification of the heavy-tailed distribution. One of the possible dominators is also the estimator based on robustified Johnson score. However, more research should be conducted to characterize the cases when such an estimation gives favorable trade-offs between efficiency and robustness. For the case of MM estimator and Pareto claims we came to the same conclusions as (Brazauskas and Serfling (2000a)). They have shown that for $\alpha > 1$ the corresponding MM estimator exhibits neither satisfactory robustness nor satisfactory efficiency. Although the MM-estimators can also be defined when $\alpha \leq 1$, they fail to satisfy consistency. It follows from the examples of the sections 5 and 6 that the classical methods of moments and maximum likelihood do not reflect the heavy-tailed character of the data satisfactorily. In this paper we have therefore presented some alternative methods how to treat this problem.

8 Appendix

Proof of Lemma

We have

$$T_1 = \{X : X \in \mathrm{RV}(\alpha), \alpha > 1\},$$

and

$$T_2 = \{X : X \in \mathcal{D} \cap \mathcal{L} \text{ and } E(X) < +\infty\}.$$

Moreover $\mathrm{RV}(\alpha) \subset \mathcal{S} \subset \mathcal{L}$ ($X > 0$ a.s.) and $\mathrm{RV}(\alpha) \subset \mathcal{D}$ (see (Embrechts et al. 03), p. 50). Finally $T_1 \subset \mathcal{D} \cap \mathcal{L}$.

If $E(X)$ exists, it can be expressed by ordinary integrals,

$$E(X) = \int_0^{+\infty} \bar{F}(x)dx - \int_{-\infty}^0 F(x)dx.$$

Conversely, the existence of the integrals on the right-hand side implies the existence of the expectation (see (Rényi 70), p. 215). In our case $\int_{-\infty}^0 F(x)dx = 0$ and

$$\int_0^{+\infty} \bar{F}(x)dx = \int_0^{+\infty} x^{-\alpha} L(x)dx$$

and $L > 0$ vary slowly. Applying Lemma in (Feller 71), p.280, the latter integral converge for $\alpha > 1$, which is our case. We have proved that $E(X)$ exists and together with $T_1 \subset \mathcal{D} \cap \mathcal{L}$ we have proved $T_1 \subset T_2$.

□

Proof of Theorem 1

Let $x_i \sim$ Pareto $P(\lambda_i, \alpha)$ be independent. Then $y_i \sim Exp(\ln \lambda_i, 1/\alpha)$. Let us assume the null hypothesis of homogeneity ($\lambda_1 = \lambda_2 = \ldots = \lambda_N$) holds, then the LR statistics has the form (15). Applying the theorem 5 in (Stehlík (2006)) (under the H_0) it has the same distribution as the random variable $-\ln\{N^N u_1 \ldots u_{N-1}(1 - u_1 - \ldots - u_{N-1})\}$, where the vector $(u_1, \ldots, u_{N-1})$ has a generalized Beta distribution $B(1, \ldots, 1)$ on the simplex $\{u : 0 < u_1 < 1, \ldots, 0 < u_{N-1} < 1 - u_1 - \ldots - u_{N-2}\}$.

□

Proof of Theorem 2

Let $x_1, \ldots, x_N$ be a sample from the single-parameter Pareto model $P(\alpha)$ and λ be known parameter. The LR statistics $-\ln \Lambda_N$ of the hypotheses (16) has the form (17) since $y_i = \alpha(\log x_i - \log \lambda) \sim Exp(1)$. Employing the theorem 1 we have that under H_0 the cumulative distribution function of the Wilks statistics $-2 \ln \Lambda_N$ has the form (18).

□

Lambert W-function

The Lambert W function is defined to be the multivalued inverse of the complex function $f(y) = ye^y$. As the equation $ye^y = z$ has an infinite number of solutions for each (non-zero) value of $z \in \mathbb{C}$, the Lambert W has an infinite number of branches. Exactly one of these branches is analytic at 0. Usually this branch is referred to as the principal branch of the Lambert W and is denoted by W or W_0. The other branches are denoted by W_k where $k \in \mathbb{Z} \setminus \{0\}$. The principal branch and the pair of branches W_{-1} and W_1 share an order 2 branch point at $z = -e^{-1}$. A detailed discussion of the branches of the Lambert W can be found in (Corless 96). Since the Lambert W function has many applications in pure and applied mathematics, the branches of the Lambert W function are implemented to many mathematical computational softwares, e.g. the Maple, Matlab, Mathematica and Mathcad. For more information about the implementation and some computational aspects see (Corless 93).

Acknowledgement

Research was supported by ASO project No. SK-0607-BA-018, by the VEGA grant (Slovak Grant Agency) No 1/3016/06, by Research Plan AVOZ10300504 of ICS ASCR and by Aktion No 50p14.

References

Alexander C. (2005). Assessment of Operational Risk Capital, In: *Risk Management, Challenge and Opportunity*, 2nd revised and Enlarged Edition, Edited by M. Frenkel, U. Hommel, M. Rudolf.

Alexander C. (2003). Technical Document for OpRisk2021 Module 1, www.2021solutions.com

International Convergence of Capital Measurement and Capital Standards, A Revised Framework, The Basel comitee for Banking supervision, Bank of International Settlements, Basel 2004, available from http://www.bis.org

Bingham NH, Goldie CM, Teugels JL (1987). *Regular Variation.* Cambridge University Press, Cambridge.

Beirlant J, Dierckx G, Goegebeur Y, Matthys G 1999. Tail index estimation and an exponential regression model. *Extremes* 2(2): 177-200.

Beirlant J, Matthys G, Dierckx G (2001). Heavy-tailed distributions and rating, *Astin Bulletin*, Vol. 31, No. 1, 2001, pp. 37-58

Brazauskas V, Serfling R (2000). Robust and efficient estimation of the tail index of a single-parameter Pareto distribution, *North American Actuarial journal*, 4(4), p. 12-27

Brazauskas V, Serfling R (2001). Robust estimation of the tail parameters for two-parameter Pareto and exponential models via generalized quantile statistics, *Extremes* 3(3), p. 231-249.

Brazauskas V, Serfling R (2003). Favorable Estimators for Fitting Pareto Models: A Study Using Goodness-of-fit Measures with Actual Data, *Astin Bulletin*, Vol. 33, No. 2, 2003, p. 365-381

Cantoni E, Ronchetti E (2004). A robust approach for skewed and heavy-tailed outcomes in the analysis of health care expenditures, *Journal of Health Economics*, Elsevier, Vol. 25, 2, 198-213.

Christoph G. (2005). Exact rates of convergence to ruin probabilities for regularly varying random variables, In: *Proceedings of the International Symposium on Stochastic Models in Reliability, Safety, Security and Logistics*, Beer Sheva, Israel, February 15-17, 2005 , I. Frenkel (ed.) et al., Beer Sheva - Riga, 2005, pp. 75 - 78. ISBN 9984-668-79-7

Christoph G. (2004). Exact rates of convergence for compound sums of random variables with common regularly varying distribution functions, In: *Longevity, Aging and Degradation Models in Reliability, Public Health, Medicine and Biology*, Vol.2, V. Antonov (ed.) et al., St. Petersburg, 2004, pp. 56 - 67. ISBN 5-7422-0638-0

Corless RM, Gonnet GH, Hare DEG, Jeffrey DJ, Knuth DE, On the Lambert W function, *Advances in Computational mathematics* **5** (1996), 329-359

Corless RM, Gonnet GH, Hare DEG, Jeffrey DJ, Knuth DE, Lambert's W Function in Maple, *Maple Technical Newsletter* **9**, Spring 1993, 12-22

Embrechts P, Kluppelberg C, Mikosch T. (2003). *Modeling extremal events, Applications of mathematics*, 4. Edition, Springer.

Fabián Z. (2001). Induced cores and their use in robust parametric estimation. *Commun. Stat., Theory Methods*, 30, 3, 537-556.

Fabián Z. (2006). Johnson point and Johnson variance. In: *Proc. Prague Stochastics 2006*, (eds. Hušková and Janžura), Matfyzpress, 354-363.

Fabián Z. (2007) Parametric estimates by generalized moment method. *Research report* 1014, Inst. of Computer Sciences, AS CR, 2007.

Fabián Z. (2008). New measures of central tendency and variability of continuous distributions, *Communications in Statistics*, Theory and methods 37, 159-174.

Feller W. (1971). *An introduction to Probability Theory and Its Applications*, 2nd Edition, Volume II, John Wiley & Sons, New York.

Fouche CH, Mukuddem-Petersen J, Petersen MA (2006). Continuous-time stochastic modelling of capital adequacy ratios for banks, *Applied Stochastic Models in Business and Industry*, Wiley, Volume 22, Issue 1, 41-71.

Gel YR, Miao W, Gastwirth JL (2005). The importance of checking the assumptions underlying statistical analysis: graphical methods for assessing normality. *Jurimetrics J.* 46, 326.

Hewitt ChC, Lefkowitz B. (1979). Methods for fitting distributions to insurance loss data. In: *Proceedings of the Casualty Actuarial Society*, 66, 139-160.

Hogg RV, Klugman SA. (1984). *Loss distributions*. New York: Wiley.

Juárez SF and Schucany WR (2004). Robust and Efficient Estimation of the Generalized Pareto Distribution, *Extremes* 7, 237-251.

Marazzi A, Ruffieux C. (1995). Implementing M-estimators of the gamma distribution, In: *Robust Statistics, Data Analysis, and Computer Intensive Methods*, R. Helmut, Ed., vol. 109 of Lecture Notes in Statistics, Springer-Verlag, Berlin, 1996

Mikosch T, Nagaev AV (1998). Large deviations of heavy-tailed sums with applications in insurance, *Extremes* 1, No.1, 81-110.

Mikosch T, Nagaev A. (2001). Rates in approximations to ruin probabilities for heavy-tailed distributions, *Extremes* 4, No.1, 67-78.

Ng, K.W., Tang, Q.H. and Yang, H. (2002). Maxima of sums of heavy-tailed random variables, *Astin Bulletin,* Vol. 32, No. 1, p. 43-55.

Philbrick SW (1985) A practical guide to the single parameter Pareto distribution, In: *Proceedings of the Casualty Actuarial Society LXXII*, 44-84.

Plevová, J. (1999). Risk in Insurance, Master Thesis, Comenius University, Bratislava.

Rényi A. (1970). *Probability Theory*, Nort-Holland Series in Applied Mathematics and Mechanics, North-Holland publishing Company, Amsterdam.

Resnick SI. (2006). *Heavy-tailed Phenomena, Probabilistic and Statistical Modeling*, Springer.

Rytgaard M. (1990). Estimation in the pareto distribution, *Astin Bulletin*, 20(2), 201-216.

Stehlík M. (2006). Exact likelihood ratio scale and homogeneity testing of some loss processes. *Statistics and Probability Letters* **76**, 19-26.

Stehlík M, Wagner H. (2008). Exact likelihood ratio testing of the scale Exponential mixtures, *IFAS Research Report*, 2008.

Vandewalle B, Beirlant J, Christmann A, Hubert M (2007). A robust estimator for the tail index of Pareto-type distributions, *Computational Statistics & Data Analysis*, Volume 51, Issue 12: 6252-6268

Voit J. (2005). *The Statistical Mechanics of Financial Markets*, 3rd Edition, Springer.