Small sample estimation and testing for heavy tails

Fabián, Zdeněk (1st author)
Academy of Sciences of the Czech Republic, Institute of Computer Science
Pod vodárenskou věží 2
City (with Postcode),Czech Republic
E-mail: zdenek@cs.cas.cz

Stehlík, Milan (2nd author, corresponding)
Johannes Kepler University in Linz, Department of Applied Statistics
Freistädter Straße 315
Linz 4040, Austria
E-mail: Milan.Stehlik@jku.at

Střelec, Luboš (3rd author)
Mendel University of Agriculture and Forestry, Department of Statistics and Operational Analysis,
Zemědelská 1
Brno 613 00, Czech Republic
E-mail: lubos.strelec@mendelu.cz

1 Introduction

If the model F, underlying the data, is a regularly varying function with index $-1/\alpha, \alpha > 0$ it is usually supposed that the top scaled order statistics $\frac{X_{n-j+1,n}}{X_{n-k,n}}, 1 \leq i \leq k$, are Pareto distributed. Hill (1975) derived a procedure of Pareto tail estimation by the MLE. Later on, many authors tried to robustify the Hill estimator, but they still rely on maximum likelihood (Fraga Alves (2001) has introduced a new lower bound, Gomes and Oliveira (2003), Li et al. (2010) introduced powers of original statistics). However, the influence function of Hill estimator is slowly increasing, but unbounded. Hill procedure is thus no robust and many authors tried to make the original Hill robust (see Beran and Schell (2010), Vandewalle et al. (2007), ...). In Fabián (2001) a new score method of score moment estimators has been proposed. It appeared that these score moment estimators are robust for a heavy tailed distributions (see Stehlík et al.(2010)). In this paper we understand "*The Hill estimator*" as a specific procedure for studying of the tail of Pareto distribution. Instead of implementing "The Hill estimator" procedure, we implement the score moment procedure. For the case of Pareto distribution, the Hill estimator procedure with the score moment estimator has been investigated in Stehlík et al.(2011) for optimal testing for normality against Pareto tail. Since the score moment estimator is simple, it is easy to implement it to the Hill procedure.

In literature, mainly the asymptotical properties of tail estimators are studied. However, in many situations, asymptotics is simplifying the underlying process too much (see e.g. Huisman et al. (2001)). We may illustrate this fact by a severe bias of the Hill based estimators or by a distributional insensitivity of asymptotical estimators. In many practical situations, such as operational risk assessment, data are sparse (with n often bellow 50) and therefore robust estimator with good small sample properties is needed.

Thus the aim of the paper is to introduce the distribution sensitive tail estimation procedure, which is easy to implement for various distributions. Here we construct distribution-sensitive estimators based on the Hills procedure using score moment estimators for other heavy-tailed distributions (Pareto, Fréchet, Burr, log-gama, inverse gamma, etc.), and to study their statistical properties, both theoretically (e.g consistency, asymptotical distribution) and by means of simulation experiments (e.g.

comparisons of exact effectiveness of our method for different heavy tailed distributions). A specific problem is finding of an optimal threshold k, say, yielding a trade off in between of variance and bias of the Hill estimator. Simulation results by Embrechts et al. (1997) showed that the Hill estimator and its alternatives work well over large ranges of values for k in the case of Pareto distribution. However, Hill estimator is often giving wrong results for distributions different from the Pareto one. Their "Hill horror plots" actually show deviations of the Hill estimates trending farther away from the true value of the tail index as k is increased.

When MSE is employed to study the quality of estimation, then we are getting a bath-tube shape of MSE against threshold k, since higher order statistics did not see the different underlying distribution, however, first order statistics are very different (e.g. Frechet distribution as a Pareto tail distribution, see Gomes and Oliveira (2003)).

In this paper we illustrate t-Hill. We also quantify the robustness and compare efficiency with other competitors. The paper is organized as follows. In the next section we recall the theory of scalar score. In section 3 we discussed the t-Hill estimator, introduced firstly in Stehlík et al.(2011). The section comparing t-Hill and Hill estimators follows. Therein contamination of underling data is controlled by means of score variance of Pareto distribution. Comparisons show that t-Hill estimator outperforms Hill estimator. In section 5 we introduce the t-Hill plot. We end with powers of selected tests for normality against Pareto distribution.

2 Scalar score

Basic inference function of mathematical statistics, the score function, is a vector function. We have introduced a scalar inference function, called a scalar score, which reflects main features of a continuous probability distribution and which is simple. Its simplicity have made it possible to introduce new relevant numerical characteristics of continuous distributions.

The scalar score has been introduced in three steps.

i) Let $\mathcal{X}$ be support of distribution F with density f, continuously differentiable according to $x \in \mathcal{X}$ and let $\eta : \mathcal{X} \to \mathbb{R}$ be given by

$$(1) \qquad \eta(x) = \begin{cases} x & \text{if} \quad \mathcal{X} = \mathbb{R} \\ \log(x-a) & \text{if} \quad \mathcal{X} = (a, \infty) \\ \log \frac{x}{1-x} & \text{if} \quad \mathcal{X} = (0,1). \end{cases}$$

Then the *transformation-based score* or shortly the *t-score* is defined by

$$(2) \qquad T(x) = -\frac{1}{f(x)} \frac{d}{dx} \left(\frac{1}{\eta'(x)} f(x) \right).$$

(2) expresses a relative change of a 'basic component of the density', the density divided by the Jacobian of mapping (1).

ii) As a measure of central tendency of distribution $F(x)$ has been suggested the zero of the t-score,

$$x^* : \qquad T(x) = 0,$$

called the *transformation-based mean* or shortly the *t-mean.*

iii) Function

$$(3) \qquad S(x;\theta) = \eta'(x^*)T(x;\theta),$$

called the *scalar score*, has been suggested a scalar inference function of distribution F.

For a particular class of distributions with support $\mathbb{R}$ and location parameter μ, (3) is identical with the score function of distribution F for μ. For a particular class of distributions with "partial" support $\mathcal{X} \neq \mathbb{R}$ and "transformed location parameter" $\tau = \eta^{-1}(\mu)$ (distributions on $\mathcal{X} \neq \mathbb{R}$ are taken as transformed "prototypes" with support $\mathbb{R}$"), (3) was proved to be identical with the score function of distribution F for this parameter. For other distributions, (3) is a new function. The t-mean of distributions with support $\mathbb{R}$ is the mode (if the distribution has the location parameter, the t-mean is its value), the t-mean of distributions with partial support is the transformed mode of the prototype.

Instead of the ordinary moments, the score moments were introduced for any $k \in \mathcal{N}$ by relation

$$M_k(\theta) = ES^k = \int_{\mathcal{X}} S(x;\theta)^k f(x;\theta)\,dx, \tag{4}$$

existing if f satisfies the usual regularity requirements. It appeared that the score moments are often expressed by elementary functions of parameters. $M_1 = 0$. The value $M_2 = ES^2$ of location and transformed location distributions is, respectively, Fisher information for the location and transformed location parameter. Accordingly, ES^2 is the Fisher information for the t-mean. The reciprocal value

$$\omega^2 = \frac{1}{ES^2}, \tag{5}$$

the score variance, appeared to be a natural measure of the variability (dispersion) of the distribution even in cases in which the usual variance does not exist.

For parametric distributions with vector parameter θ, $x^* = x^*(\theta)$. Given data $x_1, ..., x_n$ and a model family $\{F_\theta, \theta \in \Theta\}$, the sample characteristics of central tendency ("center") and dispersion (square of' 'radius") can be obtained as functions of the estimated parameters: the sample t-mean $\hat{x}^* = x^*(\hat{\theta})$ and the sample score variance $\hat{\omega}^2 = \omega^2(\hat{\theta})$. Estimates $\hat{\theta}$ of θ are usually the maximum likelihood estimates or some robust M-estimates in cases of heavy-tailed distributions or if considering gross errors models. We introduced the score moment estimate as the solution of equations

$$\hat{\theta}_{SM} : \qquad \frac{1}{n}\sum_{i=1}^{n} S^k(x_i;\theta) = \mathrm{E}_\theta S^k, \qquad k = 1, ..., m, \tag{6}$$

derived from (4) using the substitution principle.

It was shown that $\hat{x}^*$ is consistent and asymptotically normal with variance given in Fabián (2008) (see Fabián (2007)). The t-score moment estimators take into account the assumed form of the distribution, similarly as the maximum likelihood (ML) ones. However, since x_i enters into estimation equations by means of $S(x_i;\theta)$ only and scalar scores of heavy-tailed distributions are bounded, the score moment estimates are in cases of heavy-tailed distributions robust, or, in other words, the t-score estimates of all parameters of heavy-tailed distributions are protected against outliers.

3 Hill-like estimator in case of Pareto distribution

In some cases, the first equation of (6) has a form

$$\hat{x}^*_{SM} : \qquad \sum_{i=1}^{n} S(x_i;x^*) = 0. \tag{7}$$

This is the case of the Pareto distribution $P(\alpha)$ with support $\mathcal{X} = [1, \infty)$ and density

$$f(x) = \frac{\alpha}{x^{\alpha+1}}. \tag{8}$$

Using the mapping $\eta = \log(x-1)$, $\eta'(x) = 1/(x-1)$, the t-score (2) is

$$T(x) = -1 - (x-1)f'(x)/f(x) = \alpha(1 - x^*/x)$$

where the t-mean $x^* = (\alpha+1)/\alpha$. From (7), $\hat{x}^* = \bar{x}_H$ $\bar{x}_H = n/\sum_1^n 1/x_i$ is the harmonic mean, and

$$\hat{\alpha} = 1/(\hat{x}^* - 1).$$

It suggests to introduce a variant of the Hill estimator as

$$\hat{\gamma}_k = \frac{1}{\hat{\alpha}_k} = H^*_{k,n} = \frac{1}{\frac{1}{k}\sum_{i=1}^{k} \frac{X_{n-k,n}}{X_{n-j+1,n}}} - 1, \qquad (9)$$

where harmonic mean is taken from the last k observed values with threshold $X_{n-k,n}$.

Let us call the estimator (9) the Hill score moment (sm) estimator. Since it is based on harmonic mean, it is expected to be to a certain extent resistant to large observations so that it could yield more realistic values than the ordinary Hill estimator.

4 Comparisons

The score variance of the Pareto distribution is $\omega^2 = (\alpha+2)/\alpha^3$. If $\omega = 1$, we have $\gamma = 1/\alpha = 0.657$. Pareto distribution with score variance ω^2 will be denoted by $P(\omega)$.

Hill and sm-Hill plots, denoted by H an H^*, respectively, for one random sample from Pareto $P(1)$ distribution are shown in Fig.1. The length of the sample is 1001 points. It is apparent that H^* in his first part oscillates more than the ordinary Hill. The reason is that H^* is sensitive to an abrupt change of the threshold value. It could be, however, suppressed by a suitable smoothing.

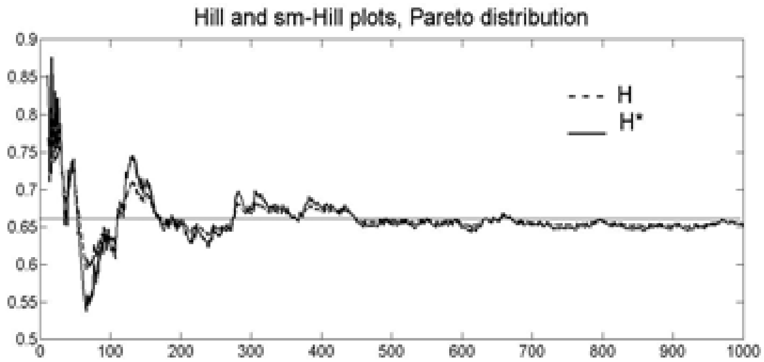

Fig. 1

Fig. 2 shows variance bounds for H_{nk} and H^*_{nk} plots obtained as the average over 100 random samples from $P(1)$.

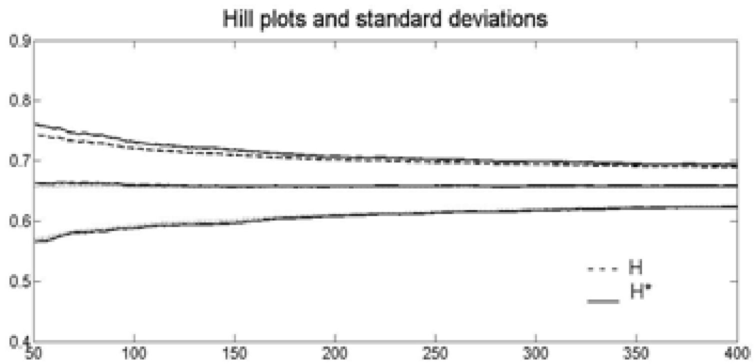

Fig. 2

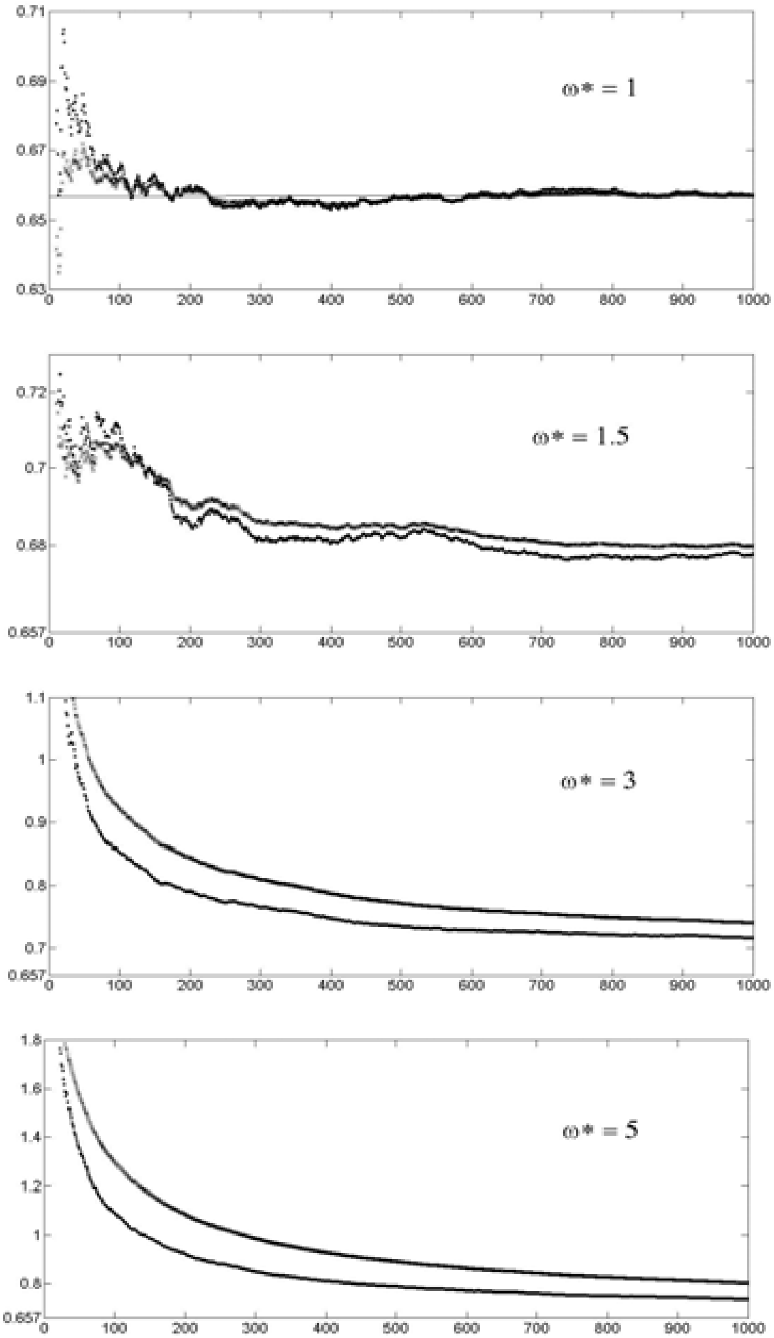

Fig. 3 Average $H_{k,n}(\times)$ and $H^*_{k,n}(\cdot)$ for different ω^*

Fig. 3 documents robustness of H^* estimators. Every graph is an average over 100 random samples from contaminated distribution

$$P_{cont}(1,\omega) = \epsilon P(1) + (1-\epsilon)P(\omega^*)$$

for four different ω^*. Ratios of $\hat{\gamma}(m)/\gamma(1)$ for m=100, 250 and 900 and $n = 1000$ are given for $H_{m,n}$ and $H^*_{m,n}$ in Table 1.

Table 1 Values $r(m) = H_{m,n}/\gamma(\omega^*) - 1$ and $r^*(m) = H^*_{m,n}/\gamma(\omega^*) - 1$ for P_{cont} with three different ω^* and $\epsilon = 0.05$ and 0.1, respectively.

ϵ	$\omega*$	$\gamma(\omega*)$	$r(100)$	$r^*(100)$	$r(250)$	$r^*(250)$	$r(900)$	$r^*(900)$
	1.5	0.879	.0363	.0376	.0235	.0187	.0184	.0170
.05	3	1.500	.2027	.1517	.1333	.0982	.0679	.0473
	5	2.165	.4982	.3049	.2792	.1573	.1203	.0656

ϵ	ω^*	$\gamma(\omega^*)$	$r(100)$	$r^*(100)$	$r(250)$	$r^*(250)$	$r(900)$	$r^*(900)$
	1.5	0.879	.0683	.0707	.0502	1.0458	.0339	.0296
.10	3	1.500	.3842	.2828	.2512	.1799	.1340	.0970
	5	2.165	.9298	.6250	.5536	.3352	.2383	.1310

5 Hill-like plot

Under the concept of the Hill estimator we understand the successive averaging of ordered values up to given k. The score moment estimator is usually simple so that it makes possible for many distributions to apply a score moment Hill-like estimator. Consider for instance data generated from the log-gamma distribution $L(c, \alpha)$ with support $\mathcal{X} = (1, \infty)$ and density

$$f(z) = \frac{c^\alpha}{\Gamma(\alpha)}(\log z)^{\alpha-1} z^{-(c+1)}. \tag{10}$$

In this case, a simple scalar score is obtained by the use of mapping $\eta : (1, \infty) \to \mathbb{R}$ in the form

$$\eta(x) = \log(\log x).$$

Since $\eta'(x) = 1/(x \log x)$, by (2)

$$T(x) = \frac{1}{f(x)} \frac{d}{dx} \left(-(\log x)^c \alpha x^{-c}\right) = c \log x - \alpha$$

so that the 'loglog' t-mean is $x^* = e^{\alpha/c}$. As the 'second log-log moment' $ET^2 = E[c^2 \log^2(x/x^*)] = \alpha$, the estimation equations (6) are

$$\sum_{i=1}^{n} c \log x_i - \alpha = 0 \tag{11}$$

$$\sum_{i=1}^{n} (c \log x_i - \alpha)^2 = \alpha \tag{12}$$

By setting $\hat{s}_1 = \frac{1}{k}\sum_{i=1}^{k} \log x_i$ and $\hat{s}_2 = \frac{1}{k}\sum_{i=1}^{k} \log^2 x_i$, it follows from (11) $\hat{\alpha} = \hat{s}_1 \hat{c}$ and from (12) $\hat{c}(\hat{s}_2 - \hat{s}_1^2) = \hat{s}_1$ so that the Hill-like estimate of the tail index (cf. Beirlant et al.(2005)) is given by closed-form expression

$$\hat{\gamma}_k = \frac{1}{\hat{c}\,k} = \frac{\hat{s}_2}{\hat{s}_1^2} - 1.$$

The Hill-like estimates based on log-gamma distribution in Fig. 4 is denoted by Hlg. It is apparent that the log-gamma hill-like estimator estimates the tail index properly whereas both H_{nk} and H^*_{nk}, expecting heavier Pareto tail, show systematic decrease.

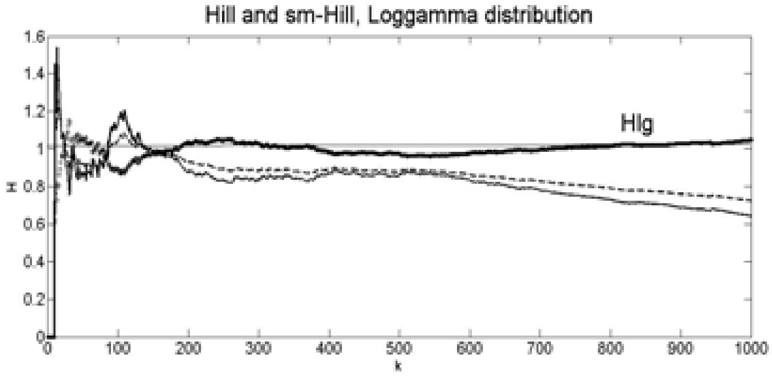

Fig. 4

6 Power of selected tests for normality against Pareto distribution

In this section we present power of selected classical and robust tests for normality against Pareto alternative distributions. For this purpose we assume Pareto (α, c) distribution for $\alpha \in \{0.5, 1, 2, 5, 10\}$, $c = 1$. Simulation study has been performed with sample sizes $n \in \{n = 10, 15, 20, 25\}$, 100000 repetitions and the following tests of normality: the classical Jarque-Bera test (JB), the robust Jarque-Bera test (RJB), the Shapiro-Wilk test (SW), the Anderson-Darling test (AD), the Lilliefors test (LT), directed SJ test (SJ_{dir}), three medcouple tests $(MC1, MC2, MC3)$ and selected RT tests which were introduced in Stehlík et al.(2011), i.e. $RT_{JB}9$, $RT_{JB}39$ and $RT_{JB}42$ tests. Therein we substantially used t-Hill estimator for Pareto tail to classify optimal test again given alternative. Therefore, Table 2 presents the results of Monte Carlo simulations of power of analyzet tests against Pareto $(\alpha, c = 1)$ alternative distributions.

Table 2: Power of analyzed tests against Pareto $(\alpha, c = 1)$ alternative distributions for $n = 10, 15, 20$ and 25

	$n=10$					$n=15$				
test	$\alpha=0.5$	$\alpha=1$	$\alpha=2$	$\alpha=5$	$\alpha=10$	$\alpha=0.5$	$\alpha=1$	$\alpha=2$	$\alpha=5$	$\alpha=10$
JB	0.886	0.757	0.610	0.444	0.403	0.979	0.912	0.799	0.650	0.585
AD	0.958	0.855	0.712	0.545	0.483	0.997	0.968	0.896	0.766	0.710
LT	0.916	0.765	0.585	0.416	0.367	0.990	0.920	0.787	0.611	0.540
RJB	0.885	0.745	0.579	0.414	0.366	0.976	0.897	0.767	0.602	0.533
SJ_{dir}	0.875	0.712	0.530	0.359	0.306	0.970	0.868	0.702	0.507	0.435
SW	**0.966**	**0.870**	**0.741**	**0.578**	**0.518**	**0.998**	**0.976**	**0.921**	**0.809**	**0.760**
$MC1$	0.635	0.445	0.330	0.251	0.218	0.770	0.550	0.398	0.285	0.259
$MC2$	0.217	0.150	0.121	0.107	0.105	0.425	0.253	0.174	0.141	0.121
$MC3$	0.732	0.507	0.361	0.261	0.229	0.914	0.711	0.530	0.375	0.327
$RT_{JB}9$	0.956	0.857	0.707	0.556	0.484	0.994	0.953	0.863	0.721	0.656
$RT_{JB}39$	0.825	0.669	0.501	0.366	0.308	0.973	0.894	0.758	0.588	0.516
$RT_{JB}42$	0.441	0.284	0.193	0.137	0.130	0.748	0.543	0.394	0.286	0.244
	$n=20$					$n=25$				
test	$\alpha=0.5$	$\alpha=1$	$\alpha=2$	$\alpha=5$	$\alpha=10$	$\alpha=0.5$	$\alpha=1$	$\alpha=2$	$\alpha=5$	$\alpha=10$
JB	0.997	0.974	0.908	0.789	0.716	**1.000**	0.993	0.961	0.875	0.830
AD	**1.000**	0.995	0.968	0.894	0.845	**1.000**	0.999	0.989	0.956	0.928
LT	0.999	0.978	0.902	0.753	0.676	**1.000**	0.994	0.958	0.862	0.791
RJB	0.996	0.963	0.879	0.740	0.656	0.999	0.987	0.939	0.825	0.765
SJ_{dir}	0.993	0.941	0.810	0.624	0.528	0.998	0.973	0.880	0.705	0.609
SW	**1.000**	**0.997**	**0.983**	**0.926**	**0.891**	**1.000**	**1.000**	**0.995**	**0.976**	**0.957**
$MC1$	0.891	0.706	0.529	0.399	0.358	0.918	0.741	0.570	0.429	0.388
$MC2$	0.518	0.322	0.212	0.170	0.167	0.557	0.320	0.222	0.169	0.160
$MC3$	0.972	0.845	0.664	0.504	0.453	0.981	0.868	0.695	0.519	0.466
$RT_{JB}9$	0.999	0.991	0.948	0.854	0.801	**1.000**	0.997	0.978	0.919	0.876
$RT_{JB}39$	0.995	0.965	0.879	0.737	0.661	0.999	0.989	0.944	0.838	0.770
$RT_{JB}42$	0.892	0.738	0.579	0.432	0.384	0.959	0.857	0.716	0.574	0.518

From the Table 2 we may conclude that:

- The Shapiro-Wilk test outperforms the other tests for normality.
- For small sample sizes $n = 10, 15, 20$ and 25 the $RT_{JB}9$ outperforms JB test for all analyzed Pareto alternatives. For example, power of the classical Jarque-Bera test against Pareto ($\alpha = 1,\ c = 1$) alternative and very small sample size $n = 10$ is 0.757. In comparison, power of $RT_{JB}9$ test (test based on "mean-median" robustification of the classical Jarque-Bera test - see Stehlík et al.(2011)) against the same alternative is 0.857 and is comparable with the Shapiro-Wilk test, which has power of 0.870
- Power of the tests is decreasing with increase of the parameter α. For example, power of the Shapiro-Wilk test against Pareto ($\alpha = 1,\ c = 1$) alternative and small sample size $n = 15$ is 0.976 and against Pareto ($\alpha = 10,\ c = 1$) alternative and the same sample size is 0.760.
- The smallest power show the medcouple tests and $RT_{JB}42$ (test based on "trimm-trimm" robustification of the classical Jarque-Bera test - see Stehlík et al.(2011)).

Based upon our experience we can recommend using of the Shapiro-Wilk test and $RT_{JB}9$ test instead of the classical Jarque-Bera test (JB) and robust Jarque-Bera test (RJB), especially for small and very small sample sizes.

References

Beirlant J., Goegebeur Y., Segers J. and Teugels, J. (2005). *Statistics of Extremes.* Theory and applications. Wiley.

Beran J. and Schell D. (2010). On robust tail index estimation, *Computational Statistics and Data Analysis,* doi:10.1016/j.csda.2010.05.028

Fabián Z. (2001). Induced cores and their use in robust parametric estimation, *Communication in Statistics, Theory Methods*, **30** (2001), pp.537-556.

Fabián Z. (2007). Estimation of simple characteristics of samples from skewed and heavy-tailed distribution, in *Recent Advances in Stochastic Modeling and Data Analysis* (ed. Skiadas, C.), Singapore, World Scientific (2007), pp.43-50.

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

Fraga Alves, M.I. (2001). A location invariant Hill-type estimator. *Extremes* 4, 199-217.

M. Ivette Gomes and Oliveira O. (2003). Censoring estimators of a positive tail index, *Statistics & Probability Letters* 65(3): 147-159

Huisman, R. et al. (2001). Tail-Index Estimates in Small Samples, *Journal of Business & Economic Statistics*, 19(2): 208-216.

Li J, Peng Z and Nadarajah S. (2010). Asymptotic normality of location invariant heavy tail index estimator, *Extremes* 13 (3):269-290.

Stehlík M., Potocký R., Waldl H. and Fabián Z. (2010) On the favourable estimation of fitting heavy tailed data, *Computational Statistics*, 25:485-503.

Stehlík M., Fabián Z. and Střelec L. Small sample robust testing for Normality against Pareto tails, accepted for *Journal of Statistical Planning and Inference.*

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