Department for Applied Statistics
Johannes Kepler University Linz

IFAS Research Paper Series
2010-51

Small sample robust testing for Normality against Pareto tails

Milan Stehlík, Zdeněk Fabián[a] and Luboš Střelec[b]

November 2010

[a]Academy of Sciences of the Czech Republic
[b]Mendel University in Brno

Abstract

The aim of this paper is to introduce the general form (so called RT class) of the robust and classical Jarque-Bera (JB) test based on the location functional. We introduce the two step procedure which is optimal for testing against the individual or contaminated Pareto alternative. As a reference for such a contamination we consider different Pareto distributions. We also give practical guidelines for robust testing for normality against short and heavy tailed alternatives. We concentrate mainly on simulation results for moderate and small samples. However, we also prove consistency and asymptotic distribution for introduced tests. We show that as the suitable measure of nominal level of Pareto tail parameter we may take the t-Hill estimator introduced in the paper. To guarantee the consistency of the whole procedure, we also prove the consistency of t-Hill estimator. The introduced general class of robust tests of the normality is illustrated at the selected datasets of financial time series.

KEY WORDS: Hill estimator, t-Hill estimator, testing for normality, robust tests for normality, returns, power comparison, Pareto tail, consistency, location functional

1 Introduction

Problem of testing for normality plays central role by many financial decisions. Financial time series typically results in volatility clustering, leptokurtosis probability function of returns with fat tails and a higher peak at the mean than the normal distribution. In the majority of cases of relevant analysis is expected that returns derived from financial time series is Gaussian normal distributed random variable with constant expected value and constant variance. But actually this is not true in many practical situations and a lot of tests has been developed to test for normality. The absence of exact solutions for the sampling distributions generated a large number of simulation studies exploring the power of these statistics as both directional and omnibus tests (see for example (Gel and Gastwirth (2008)), (Gel et al. (2007)), (Brys et al. (2008)), among others). In this paper we introduce the general class of robust tests, the so called RT, based on robustification of JB and generalization of robust JB test. In particular we provide remedies for some weaknesses of JB and robust JB tests. Several examples aim to convince the reader that construction of RT class has a natural basis.

Two issues typically enter the testing for normality against heavy tails: small samples and contaminated heavy tailed data ((Brys et al. (2008)) and (Stehlík et al. (2010))). In this paper we provide a guidelines how to efficiently test for a normality against European Pareto distribution, possibly contaminated, with the density

$$\frac{\alpha c^{\alpha}}{x^{\alpha+1}}, x > c. \tag{1}$$

This distribution we denote for convenience as Pareto (α, c). We also provide a guidelines how to efficiently test for normality against short tailed and heavy tailed alternatives. We will concentrate especially on a small samples. Large sample behavior of tests for normality has been already rather extensively studied in the literature (see e.g. (Locke and Spurrier (1977)) for U statistics based testing or (Saniga and Miles (1979)) for a normality testing against stable alternatives). However, to the best knowledge of authors, small sample situations are not covered satisfactory. One exception is paper by (Spiegelhalter (1980)), which however, provide results only for sample size $n = 20$ and does not consider Pareto alternatives. Roughly saying, small samples almost always pass a normality test. Normality tests have little power to tell whether or not a small sample of data comes from a Gaussian distribution. To overcome this problem we introduce the simple two-step procedure to test efficiently against Pareto tails:

Step 1) To estimate the nominal value of Pareto tail α under alternative.

Step 2) Based on the result in Step 1 to choose appropriate (preferably the most powerful) test for normality in the reasonable subclass of RT class.

The paper is organized as follows. In section 2 the general robustification of JB test is introduced and main theoretical results are given. There exists a large amount of tests included in the RT class and it is not possible to discuss them entirely. For our purposes we introduce the RT_{JB} and RT_{RJB} classes of tests. The RT class power sensitivity on the tail index is given by the means of simulation in section 3. In section 4 we introduce the suitable estimator of Pareto tail, based on t-score, preferable to the other estimators under the contamination of Pareto tail. To guarantee consistency of the whole procedure we also prove consistency of t-estimator of Pareto tail parameter. In section 5 the given methods are illustrated on the real data example. Discussion concludes the paper. To maintain the continuity of explanation proofs and technicalities are put into Appendix.

2 General approach to robust JB test: RT class

Interestingly enough (see (Urzua (1996))), the classical JB test (see (Bera and Jarque (1981)), with the well known test statistics $JB = \frac{n}{6}\left(\frac{\hat{\mu}_3}{\hat{\mu}_2^{3/2}}\right)^2 + \frac{n}{24}\left(\frac{\hat{\mu}_4}{\hat{\mu}_2^2} - 3\right)^2$) has been known among statisticians since the work of (Bowman and Shenton (1975)). They derived it after noting that, under normality, the asymptotic means of $\sqrt{b}_1$ and b_2 are 0 and 3, the asymptotic variances are $6/n$ and $24/n$, and the asymptotic covariance is 0. Yet, there are few instances in the statistics literature where the Bowman-Shenton-Jarque-Bera test has been studied. As one author states in a comprehensive survey of tests for normality "Due to the slow convergence of b_2 to normality this test is not useful" (see (D'Agostino (1986)), p. 391). However, JB test has gained grate popularity among economists. Later on, various modifications has been made (e.g. test of Urzua (Urzua (1996))) and for a robustification see robust Jarque-Bera (RJB) test (see (Gel and Gastwirth (2008))).

In our paper we introduce the general class of robust tests, the so called RT, based on robustification of JB and generalization of robust JB test. In particular we provide remedies for some weaknesses of JB and robust JB tests. RT class is a flexible class of robust tests for normality based on a location functional. A location functional has been introduced in a seminal paper by (Bickel and Lehman (1975)) and it looks to be playing a crucial role also by robust testing. For a recent discussion on location functional importance see (Wilcox (2005)). As it will be seen later, the power of RT class test mimics the effectiveness of location estimator in typical cases. Thus trade off between power and robustness is a typical issue here.

The examples of location functional, relevant to our paper, are mean $T_{(0)} = \int x dF_n(x)$, median $T_{(1)} = F_n^{-1}(1/2)$, trimmed mean $T_{(2)}(s) = \frac{1}{n-2s}\sum_{i=s+1}^{n-s} X_{i:n}$, $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$ be the order statistics, and pseudo-median, $T_{(3)} = H_n^{-1}(1/2)$, where $H_n(y) = \int F_n(2y-x)h(x)dx$.

Now let us define the RT class of normality tests. For that reason we relax the form of j-th theoretical moment $\mu_j = E(X - E(X))^j$ estimator by taking $M_j(r, T(F_n)) = \frac{1}{n-2r}\sum_{m=1+r}^{n-r} \varphi_j(X_{m:n} - T(F_n))$, $j \in \{0,1,2,3,4\}$ and φ_j is tractable and continuous function $\varphi_0(x) = \sqrt{\pi/2}|x|$, $\varphi_1(x) = x$, $\varphi_2(x) = x^2$, $\varphi_3(x) = x^3$ and $\varphi_4(x) = x^4$. The RT class is defined by

$$RT = \frac{k_1(n)}{C_1}\left(\frac{M_{j_1}^{\alpha_1}(r_1, T_{(i_1)}(s_1))}{M_{j_2}^{\alpha_2}(r_2, T_{(i_2)}(s_2))} - K_1\right)^2 + \frac{k_2(n)}{C_2}\left(\frac{M_{j_3}^{\alpha_3}(r_3, T_{(i_3)}(s_3))}{M_{j_4}^{\alpha_4}(r_4, T_{(i_4)}(s_4))} - K_2\right)^2. \quad (2)$$

The following theorems justify the feasibility of RT class.

Under the null hypothesis, for $T(F_n)$ being mean, median, pseudo-median or trimmed mean we have

$$\lim_{n\to\infty} E(M_j(r, T(F_n)) = E(X - E(X))^j := \mu_j, j = 0,1,2,3,4, \quad (3)$$

where $\mu_0 = \sigma$. In other words, $M_j(r, T(F_n))$ is a consistent estimator of μ_j.

It can be seen from (2) there exist a vast amount of RT class tests, which we can obtain for a different settings of $r_i, T_{(i)}$, etc. Already known special cases of RT class are classical Jarque-Bera test, test of Urzua, robust Jarque-Bera (RJB), among others. The RT class approach may serve well as an testing instrument for the type of stock return distribution (see (Yu (2001))).

2.1 RT_{JB} subclass

In this subsection we focus on robustification of the Jarque-Bera test, given by RT_{JB} class defined as follows

$$RT_{JB} = \frac{k_1(n)}{C_1}\left(\frac{M_3(r_1, T_{(i_1)}(s_1))}{M_2^{3/2}(r_2, T_{(i_2)}(s_2))} - K_1\right)^2 + \frac{k_2(n)}{C_2}\left(\frac{M_4(r_3, T_{(i_3)}(s_3))}{M_2^2(r_4, T_{(i_4)}(s_4))} - K_2\right)^2. \quad (4)$$

The RT_{JB} subclass is a special case of RT test statistics for $k_1(n) = n$, $k_2(n) = n$, $\alpha_1 = 1$, $\alpha_2 = 3/2$, $\alpha_3 = 1$, $\alpha_4 = 2$, $j_1 = 3$, $j_2 = 2$, $j_3 = 4$, $j_4 = 2$.

Let $X_1, ..., X_n$ be iid $N(\mu, \sigma^2)$, i.e. null hypothesis holds. Then

$$\sqrt{n}\begin{pmatrix} \left(M_3(r_1, T_{(i_1)}(s_1))/M_2^{3/2}(r_2, T_{(i_2)}(s_2))\right) - K_1 \\ \left(M_4(r_3, T_{(i_3)}(s_3))/M_2^2(r_4, T_{(i_4)}(s_4))\right) - K_2 \end{pmatrix} \to N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} C_1 & 0 \\ 0 & C_2 \end{pmatrix}\right) \quad (5)$$

where $T_{(i)}$ for $i \in \{0,1,2,3\}$ denotes using of arithmetic mean $T_{(0)}$, median $T_{(1)}$, trimmed mean $T_{(2)}(s)$ or pseudo-median $T_{(3)}$, respectively.

Corollary 1 The RT test statistic from (4) asymptotically follows χ_2^2.

Choosing of an appropriate constants C_1 and C_2 is the hardest aspect of the variants of RT class. To obtain the constants C_1, C_2 we need to find the expressions for $E(M^k_{n_1,n_2})$ for a finite sample size. Such calculations are very tedious and therefore we obtain these constants from Monte Carlo simulations. Notice, that the critical constant (for small and mid samples) under the trimming ($r > 0$) are different from critical constants without trimming ($r = 0$), since only asymptotical distribution is normal (see (Stigler (1973))) in this case. Notice that two levels of trimming enter RT class: first trimming (with trimming constant s) enters trimming in the location estimator $T(F_n)$, the latter on trimming (with trimming constant r) enters $M_j(r, T)$. Amazing property of RT class and robust tests in general is, that power of RT class mimics the effectiveness of location estimator. Thus practitioner can tune how much of robustness is needed, of course at price of the power. One should be really careful here: mechanical downweighting of peculiar observations may divert attention from important clues to new discovery. Two typical extremal behaviors occur in robust testing: the tests which are more robust have smaller power (since they are not affected by single outliers) and tests with higher power are typically less robust (because they are affected by single outliers).

Therefore, Table 1 and 2 contain the results of Monte Carlo simulations of these constants for RT_{JB} subclass. Notice, that prevalent number of cases led asymptotically to $C_1 = 6$, exception where $C_1 = 18$ for the case of location functional being median or pseudomedian. In all cases $C_2 = 24$ asymptotically. Also notice that for a large trimming (by r_2 and s_2) we obtain the slow convergence to asymptotical value of the constant.

Table 1: Monte Carlo simulations of C_1 of $M_3^1(r_1, T_{(i_1)}(s_1))/M_2^{3/2}(r_2, T_{(i_2)}(s_2))$

r_1,r_2,i_1,i_2,s_1,s_2	$n = 25$	$n = 50$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	*asympt.*
0,0,0,0,0,0	4.73	5.34	5.41	5.72	5.80	6.01	6
0,0,0,1,0,0	4.35	5.10	5.27	5.65	5.77	6.00	6
0,0,0,2,0,1	4.70	5.33	5.41	5.72	5.80	6.01	6
0,0,0,2,0,5	4.55	5.29	5.40	5.72	5.80	6.01	6
0,0,0,3,0,0	4.33	5.09	5.28	5.65	5.77	6.00	6
0,0,1,0,0,0	15.48	16.02	15.93	16.73	16.72	17.23	18
0,0,1,1,0,0	13.42	14.90	15.32	16.39	16.58	17.16	18
0,0,2,0,1,0	6.52	6.48	6.08	6.14	6.01	6.13	6
0,0,2,2,1,1	6.48	6.47	6.08	6.14	6.01	6.13	6
0,0,2,0,5,0	10.36	9.05	7.70	7.19	6.56	6.47	6
0,0,2,2,5,5	9.93	8.96	7.68	7.19	6.56	6.47	6
0,0,3,0,0,0	14.47	14.33	14.22	14.44	14.37	14.77	18
0,0,3,3,0,0	11.63	12.91	13.49	14.07	14.22	14.69	18
0,1,0,2,0,1	20.01	11.23	8.10	7.17	6.45	6.38	6
0,5,0,2,0,5	1555.40	82.89	23.85	13.27	8.70	7.59	6
1,0,2,0,1,0	1.25	2.08	2.82	3.63	4.52	5.14	6
1,1,2,2,1,1	3.53	3.88	4.08	4.50	5.01	5.45	6
5,0,2,0,5,0	0.06	0.30	0.81	1.55	2.76	3.70	6
5,5,2,2,5,5	3.92	3.18	3.18	3.48	4.10	4.66	6

Table 2: Monte Carlo simulations of C_2 of $M_4^1(r_3, T_{(i_3)}(s_3))/M_2^2(r_4, T_{(i_4)}(s_4))$

r_3,r_4,i_3,i_4,s_3,s_4	$n=25$	$n=50$	$n=100$	$n=200$	$n=500$	$n=1000$	*asympt.*
0,0,0,0,0,0	13.67	18.44	20.06	21.36	22.91	24.05	24
0,0,0,1,0,0	13.60	18.32	19.96	21.31	22.89	24.02	24
0,0,0,2,0,1	13.41	18.36	20.04	21.36	22.91	24.05	24
0,0,0,2,0,5	13.11	18.17	19.98	21.34	22.91	24.05	24
0,0,0,3,0,0	13.49	17.94	19.97	21.24	22.91	24.03	24
0,0,1,0,0,0	19.43	21.70	22.04	22.43	23.35	24.33	24
0,0,1,1,0,0	15.86	19.85	20.99	21.89	23.12	24.19	24
0,0,2,0,1,0	15.68	19.36	20.41	21.49	22.94	24.06	24
0,0,2,2,1,1	15.36	19.27	20.38	21.48	22.94	24.06	24
0,0,2,0,5,0	17.45	20.31	20.84	21.67	23.00	24.08	24
0,0,2,2,5,5	16.19	19.94	20.74	21.65	22.99	24.08	24
0,0,3,0,0,0	20.29	22.11	21.66	22.18	23.15	24.23	24
0,0,3,3,0,0	14.82	19.15	20.41	21.50	22.93	24.10	24
0,1,0,2,0,1	476.55	142.40	70.79	46.50	34.82	31.28	24
0,5,0,2,0,5	480641.50	4330.28	550.24	180.03	77.75	53.36	24
1,0,2,0,1,0	2.05	3.11	4.47	6.62	10.63	13.59	24
1,1,2,2,1,1	7.13	8.09	8.95	10.87	14.04	16.33	24
5,0,2,0,5,0	0.14	0.66	1.37	2.1	3.82	5.96	24
5,5,2,2,5,5	5.68	4.02	4.20	5.32	7.72	9.99	24

2.2 RT_{RJB} subclass

Here we focus also on robustification of the robust Jarque-Bera test introduced by (Gel et al. (2007)). This robustification is given by RT_{RJB} class of statistics defined as follows

$$RT_{RJB} = \frac{k_1(n)}{C_1}\left(\frac{M_3(r_1, T_{(i_1)}(s_1))}{M_0^3(r_2, T_{(i_2)}(s_2))} - K_1\right)^2 + \frac{k_2(n)}{C_2}\left(\frac{M_4(r_3, T_{(i_3)}(s_3))}{M_0^4(r_4, T_{(i_4)}(s_4))} - K_2\right)^2. \tag{6}$$

The RT_{RJB} subclass is a special case of RT test statistics for $k_1(n) = n$, $k_2(n) = n$, $\alpha_1 = 1$, $\alpha_2 = 3$, $\alpha_3 = 1$, $\alpha_4 = 4$, $j_1 = 3$, $j_2 = 0$, $j_3 = 4$, $j_4 = 0$.

Let $X_1, ..., X_n$ be iid $N(\mu, \sigma^2)$, i.e. null hypothesis holds. Then

$$\sqrt{n}\begin{pmatrix} (M_3(r_1, T_{(i_1)}(s_1))/M_0^3(r_2, T_{(i_2)}(s_2))) - K_1 \\ (M_4(r_3, T_{(i_3)}(s_3))/M_0^4(r_4, T_{(i_4)}(s_4))) - K_2 \end{pmatrix} \to N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} C_1 & 0 \\ 0 & C_2 \end{pmatrix}\right) \tag{7}$$

where $T_{(i)}$ for $i \in \{0, 1, 2, 3\}$ denotes using of arithmetic mean $T_{(0)}$, median $T_{(1)}$, trimmed mean $T_{(2)}(s)$ or pseudo-median $T_{(3)}$, respectively.

Corollary 2 The RT test statistic from (6) asymptotically follows χ_2^2.

Choosing of an appropriate constants C_1 and C_2 is the hardest aspect of the variants of RT_{RJB} class, again. To obtain the constants C_1, C_2 we use Monte Carlo simulations.

Therefore, Table 3 and 4 contain the results of Monte Carlo simulations of these constants for RT_{RJB} subclass. Notice, that prevalent number of cases led asymptotically to $C_1 = 6$, exception where $C_1 = 18$ for the case of location functional being median or pseudomedian. In all cases $C_2 = 58$ asymptotically. Also notice that for a large trimming (by r_2 and s_2) we obtain the slow convergence to asymptotical value of the constant.

Table 3: Monte Carlo simulations of C_1 of $M_3^1(r_1, T_{(i_1)}(s_1))/M_0^3(r_2, T_{(i_2)}(s_2))$

r_1,r_2,i_1,i_2,s_1,s_2	$n = 25$	$n = 50$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	*asympt.*
0,0,0,0,0,0	5.91	5.84	6.05	6.03	5.89	5.97	6
0,0,0,1,0,0	6.64	6.17	6.22	6.12	5.93	5.99	6
0,0,0,2,0,1	6.13	5.91	6.07	6.04	5.89	5.98	6
0,0,0,2,0,5	6.41	6.00	6.11	6.05	5.9	5.98	6
0,0,0,3,0,0	5.59	5.65	5.96	5.99	5.88	5.97	6
0,0,1,0,0,0	16.41	15.97	16.89	17.20	16.86	17.05	18
0,0,1,1,0,0	19.10	17.23	17.55	17.56	17.00	17.11	18
0,0,2,0,1,0	8.12	7.09	6.80	6.47	6.10	6.09	6
0,0,2,1,1,0	9.10	7.48	6.98	6.56	6.13	6.11	6
0,0,2,2,1,1	8.42	7.16	6.82	6.47	6.10	6.09	6
0,0,2,0,5,0	12.20	9.79	8.58	7.59	6.66	6.43	6
0,0,2,1,5,0	13.87	10.34	8.82	7.70	6.70	6.45	6
0,0,2,2,5,5	13.35	10.06	8.66	7.62	6.67	6.43	6
0,0,3,0,0,0	15.62	14.86	14.99	14.53	14.45	14.52	18
0,0,3,1,0,0	17.22	15.52	15.32	14.70	14.51	14.55	18
0,0,3,3,0,0	14.15	14.16	14.63	14.36	14.39	14.49	18
0,1,0,2,0,1	19.47	10.40	8.16	7.07	6.32	6.20	6
0,5,0,2,0,5	1126.87	63.09	20.58	11.62	7.91	7.02	6
1,0,2,0,1,0	1.33	2.15	3.10	3.87	4.60	5.07	6
1,0,2,1,1,0	1.53	2.29	3.19	3.93	4.62	5.08	6
1,1,2,2,1,1	3.51	3.63	4.11	4.52	4.92	5.26	6
5,0,2,0,5,0	0.05	0.29	0.84	1.62	2.77	3.64	6
5,0,2,1,5,0	0.06	0.31	0.87	1.65	2.79	3.65	6
5,5,2,2,5,5	3.45	2.52	2.69	3.07	3.71	4.27	6

Table 4: Monte Carlo simulations of C_2 of $M_4^1(r_3, T_{(i_3)}(s_3))/M_0^4(r_4, T_{(i_4)}(s_4))$

r_3,r_4,i_3,i_4,s_3,s_4	$n = 25$	$n = 50$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	*asympt.*
0,0,0,0,0,0	54.29	54.82	57.51	57.56	57.03	56.87	58
0,0,0,1,0,0	61.45	57.94	59.16	58.31	57.36	57.01	58
0,0,0,2,0,1	56.77	55.46	57.67	57.60	57.04	56.87	58
0,0,0,2,0,5	59.32	56.36	57.96	57.69	57.05	56.88	58
0,0,0,3,0,0	49.97	52.64	56.30	56.66	56.74	56.78	58
0,0,1,0,0,0	65.47	60.03	60.43	58.93	57.61	57.15	58
0,0,1,1,0,0	79.79	65.76	63.19	60.22	58.14	57.39	58
0,0,2,0,1,0	60.08	56.68	58.16	57.75	57.07	56.88	58
0,0,2,1,1,0	68.40	59.98	59.84	58.51	57.40	57.02	58
0,0,2,2,1,1	63.12	57.38	58.32	57.79	57.07	56.89	58
0,0,2,0,5,0	64.60	58.99	59.17	58.11	57.16	56.91	58
0,0,2,1,5,0	75.54	62.75	60.94	58.89	57.49	57.05	58
0,0,2,2,5,5	72.81	60.97	59.68	58.25	57.17	56.92	58
0,0,3,0,0,0	67.14	60.34	60.03	58.81	57.52	57.19	58
0,0,3,1,0,0	76.69	63.97	61.80	59.63	57.86	57.34	58
0,0,3,3,0,0	60.30	57.61	58.54	57.80	57.20	57.09	58
0,1,0,2,0,1	505.67	171.90	107.82	81.83	67.21	62.39	58
0,5,0,2,0,5	283511.75	3276.38	545.07	207.18	106.18	81.29	58
1,0,2,0,1,0	4.02	8.80	15.89	23.46	33.02	38.97	58
1,0,2,1,1,0	4.76	9.45	16.38	23.79	33.23	39.07	58
1,1,2,2,1,1	21.70	22.96	27.32	31.77	38.02	42.13	58
5,0,2,0,5,0	0.06	0.40	1.61	4.78	12.51	20.36	58
5,0,2,1,5,0	0.07	0.43	1.67	4.86	12.59	20.42	58
5,5,2,2,5,5	16.76	11.19	12.01	15.51	21.86	27.77	58

2.3 Finding a tractable group of RT_{JB} and RT_{RJB} tests

Main reason for using RT_{JB} and RT_{RJB} subclasses in this paper is that RT_{JB} generalizes JB and RT_{RJB} generalizes RJB tests. The aim of this section is to find by cluster analysis a group of RT_{JB} and RT_{RJB} tests, which substantially improve the properties of classical JB and RJB tests (e.g. power for symmetric short tailed alternatives like uniform). For that reason we have conducted the power comparison on 61 ordered variants of RT_{JB} and RT_{RJB} tests. By ordered we mean combination of the most efficient mean with robust location estimators (median, pseudomedian, trimmed mean and trim-trim estimator). By trim-trim estimator we mean double trimmed estimator, i.e. one trimming is employed for location estimator, while second trimming is employed for the empirical moment itself. Simulation study has been performed with sample sizes $n = 10$ and $n = 25$, 100000 repetitions and the following alternatives: Cauchy, Laplace, one sided Cauchy, exponential, Weibull, Pareto, uniform and Beta distributions. By clustering based on power values (from all 61 analyzed tests of RT_{JB} class) the following five representatives of RT_{JB} class has been obtained (for detailed definition see Appendix): $RT_{JB}2$, $RT_{JB}9$, $RT_{JB}39$, $RT_{JB}42$ and $RT_{JB}43$. Analogously we conducted the clustering analysis and comparison of powers of RT_{RJB} tests. Consequently, we have obtained from all 61 analyzed tests these four representatives (for detailed definition see Appendix): $RT_{RJB}13$,

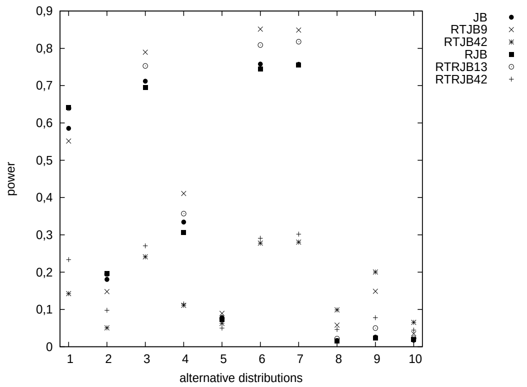

Figure 1: n=10

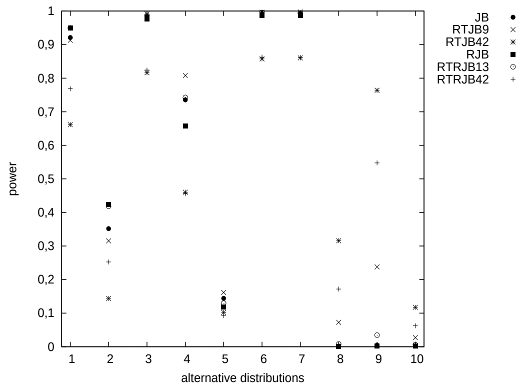

Figure 2: n=25

$RT_{RJB}33$, $RT_{RJB}42$ and $RT_{RJB}59$.

Figures 1 and 2 illustrate the power comparison at the level of significance $\alpha = 0.05$, sample sizes $n = 10$ and $n = 25$, and alternatives: 1: Cauchy, 2: Laplace, 3: one sided Cauchy, 4: exponential, 5: Weibull(2,1), 6: Pareto($\alpha = 1, c = 1$), 7: contaminated Pareto: 0.8 Pareto($\alpha = 1, c = 1$) + 0.2Pareto($\alpha = 2, c = 1$), 8: Uniform, 9: Beta$(0.5, 0.5)$, 10: Beta$(2, 2)$.

From the Figures 1 and 2 we may conclude that:

- For very heavy tailed symmetric alternatives (Cauchy and Laplace) the JB test is less powerful than the RJB test. The RJB test shows less power than the JB test only by detecting the exponential distribution. In other cases JB and RJB tests have comparable power.

- For very small sample size $n = 10$ the $RT_{JB}9$ outperforms JB test for one sided Cauchy, exponential, Pareto, contaminated Pareto, uniform and beta alternatives. For small sample size $n = 25$ the $RT_{JB}9$ outperforms JB test only for exponential, uniform and beta distributions.

- For very small sample size $n = 10$ the $RT_{RJB}13$ outperforms RJB test for one sided Cauchy, exponential, Pareto, contaminated Pareto and beta$(0.5, 0.5)$ alternatives. For small sample size $n = 25$ the $RT_{RJB}13$ outperforms RJB test only for exponential, uniform and beta$(0.5, 0.5)$ distributions.

- For symmetric short tailed alternatives (uniform and beta distributions) we do not recommend JB and RJB tests. We recommend $RT_{JB}42$ and $RT_{JB}9$ tests for very small sample size $n = 10$ or $RT_{JB}42$ and $RT_{RJB}42$ tests for small sample size $n = 25$. These tests improve power significantly.

- For symmetric and asymmetric heavy and moderately heavy tailed alternatives (Cauchy, Laplace, one sided Cauchy, exponential, Pareto and contaminated Pareto distributions) the $RT_{JB}42$ and $RT_{RJB}42$ are less powerful than the other tests, especially for very small sample size $n = 10$. For symmetric very short tailed alternatives (uniform and beta distributions) the $RT_{JB}42$ and $RT_{RJB}42$ tests show the highest power, especially for small sample size $n = 25$.

- Notice, that the discussed tests gain the similar power at alternative Weibull (2,1) for very small and small sample sizes. However, having mid sample $n = 100$ a big range of power differences has been observed already.

3 Sensitivity of power on the Pareto tail parameter

As may be seen from simulations in this section, unknown tail parameter α has a substantial influence on the power of considered tests. More precisely we can conclude that:

A) **Power**
Power of the tests is decreasing with increase of the parameter α.

B) **Parameter c of Pareto (α, c) distribution**
The obtained results of power for Pareto (α, $c = 1$), Pareto (α, $c = 10$) and Pareto (α, $c = 100$) are the same, it means that power does not depend on the parameter c. Therefore, the Tab. 5, 6 and 7 contains the results only for Pareto (α, $c = 1$) distribution.

C) **Comparison of the RT_{JB} and RT_{RJB} subclasses - very small sample size**
For $n = 5$ the most powerful test groups are $RT_{JB}17 - 31$ and $RT_{RJB}17 - 31$, which are tests based on combination mean-trimmed mean for trimming location parameter $s = 1$. For $\alpha \in \{0.1, 0.5, 1, 2, 5\}$ is the most powerful $RT_{JB}17 - 31$ group and for $\alpha = 10$ $RT_{RJB}17 - 31$ group. Notice that the empirical standard deviation of power of $RT_{JB}17 - 31$ and $RT_{RJB}17 - 31$ groups is ≈ 0.01. Despite the relative small deviation of $RT_{JB}17 - 31$ and $RT_{RJB}17 - 31$ tests we have observed relatively large deviations of $RT_{JB}1 - 16$ and $RT_{RJB}2 - 16$ tests. More precisely, the deviation of power in $RT_{JB}2 - 16$ tests is 0.25 for $n = 5$ and $\alpha = 0.1$. For comparison, the deviation of power in $RT_{JB}17 - 31$ and $RT_{RJB}17 - 31$ is only 0.01, uniformly with respect to α. The deviation of power for $RT2 - 16$ is decreasing with α, e.g. for $\alpha = 10$ the deviation reached the level of deviation by $RT_{JB}17 - 31$ and $RT_{RJB}17 - 31$ tests. The higher variability in the group of mean-median may influence the choosing of the most powerful test, however, the power of tests in the mean-median group is in the average better than the power in group of mean-trimmed mean. In contrast, the weakest group for $n = 5$ is $RT_{JB}47 - 61$ group based on combination mean-pseudomedian.

D) **Comparison of the individual RT_{JB} and RT_{RJB} tests - very small sample size**
For $n = 5$ the most powerful test for small $\alpha \in \{0.1, 0.5\}$ is $RT_{JB}12$ test, for middle $\alpha \in \{1, 2, 5\}$ $RT_{JB}2$ test and for large $\alpha = 10$ $RT_{RJB}24$ test (mean-trimmed mean type for trimming location parameter $s = 1$). The weakest individual test is $RT_{RJB}4$ test, which is even biased.

E) **Comparison of the RT_{JB} and RT_{RJB} subclasses - small sample sizes**
For case $n = 15$ and 25 the results differ from the results obtained for $n = 5$. For $n = 15$ and $n = 25$ the most powerful test group is $RT_{JB}2 - 16$, which are tests

based on combination mean-median. Notice that the empirical standard deviation of power of $RT_{JB}2-16$ group is only ≈ 0.03. The weakest groups for $n=15$ and $n=25$ are $RT_{JB}32-46$ and $RT_{RJB}32-46$ groups based on combination mean-trimmed-trimmed mean for trimming parameters $r=s=1$.

F) **Comparison of the individual RT_{JB} and RT_{RJB} tests - small sample sizes**

For $n=15$ and $n=25$ the most powerful test is the $RT_{JB}9$ test (mean-median type). This test was generally optimal for all asymmetric alternatives (e.g. exponential, log normal, Weibull, χ^2, Burr, Pareto, etc.). In some situations it may happened that $RT_{JB}9$ was not the most powerful test from RT_{JB} class, however the power difference to the most powerful individual test is negligible (the only exception is the case of $n=5$, when $RT_{JB}9$ was not the most powerful test and it lost 10% of the power to the most powerful test). The weakest individual tests are $RT_{JB}41$, $RT_{RJB}42$ (both combination of mean-trimmed mean type) and $RT_{JB}59$ (mean-trim-trim type) tests.

G) **The classical tests**

We consider the following classical tests: the Anderson-Darling test (AD), the Cramer-von Mises test (CM), the Jarque-Bera test (JB), the Jarque-Bera-Urzua test (JBU), the robust Jarque-Bera test (RJB), the Lilliefors (Kolmogorov-Smirnov) test (LT), the Pearson chi-square test (PT), the directed SJ test (SJ_{dir}), the Shapiro-Wilk test (SW), the standardized Geary test (GT), the standardized Uthoff test (UT), the skewness test (SKT), the kurtosis test (KT) and two versions of the Medcouple tests (MC and $MC-LR$). The Shapiro-Wilk test outperforms the other classical tests for all Pareto (α,$c=1$) distributions. The interesting thing is that the Pearson chi-square test is a good competitor for the Shapiro-Wilk test for $n=5$. Comparison between subclasses RT_{JB}, RT_{RJB} and classical tests has shown that the most powerful test is SW test. The only exception is sample size $n=5$, when PT is more powerful than SW test. Selected RT class tests outperforms classical JB, RJB and SJ_{dir} tests. The best overall performance in RT class have tests $RT_{JB}2-16$ and $RT_{JB}17-31$. These classes outperform the $RT_{RJB}2-16$ and $RT_{RJB}17-31$ classes. The most powerful individual tests in RT class are $RT_{JB}9$ and $RT_{JB}12$, however, their powers are lower than the power of SWtest. Notice, that power of $RT_{JB}9$ is comparable with the power of Epps-Pulley test ((Epps and Pulley (1983))).

H) **Contaminations**

We have also considered the contaminated Pareto having cummulative distribution function (CDF) $F=(1-p)\text{Pareto}(\alpha_1,c_1)+p\text{Pareto}(\alpha_2,c_2)$. For our simulation we assume 10000 replications, shapes $\alpha_1=0.1,0.5,1,5$, $c_1=1$, $\alpha_2=10$, $c_2=1$ and $p=0.2$. We have chosen 20% of contamination to be able to explore the extremely small sample sizes ($n=5,15$ and 25). The obtained results more-less mimic the results for the individual Pareto. That means for $n=5$ the most powerful tests are $RT_{JB}2$, $RT_{JB}12$, SW and PT tests (the power of $RT_{JB}2$ and $RT_{JB}12$ tests is even higher than the power of SW and PT tests). For $n=15$ and $n=25$ the most powerful tests are SW and $RT_{JB}9$ (the power of SW test is slightly higher than the power of $RT_{JB}9$ test).

I) **Comparison of JB, RJB, RT_{JB} and RT_{RJB} tests**

The power of JB test is equal or higher than RJB for all analyzed Pareto distributions. On the other hand, we can find some tests from RT_{JB} and RT_{RJB} subclasses, which outperform the classical JB and RJB tests - e.g. $RT_{JB}2$, $RT_{JB}9$, $RT_{JB}39$, $RT_{RJB}13$ and $RT_{RJB}42$ tests.

The following tables list the most important results from our simulations discussed above (A-I). Notice that blank entries in tables mean that datum was not available.

Table 5: Power of analyzed tests against Pareto (α, $c = 1$) distributions for $n = 5$

test	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$	$\alpha = 5$	$\alpha = 10$
$RT_{JB}1 = JB$	0.882	0.597	0.430	0.306	0.220	0.190
$RT_{JB}2-16$	0.802	0.546	0.379	0.270	0.192	0.168
$RT_{JB}17-31$	0.889	0.604	0.437	0.312	0.224	0.192
$RT_{JB}32-46$	0.632	0.357	0.236	0.158	0.116	0.102
$RT_{JB}47-61$	0.373	0.261	0.193	0.144	0.112	0.099
$RT_{JB}1-61$	0.677	0.445	0.313	0.222	0.162	0.141
$RT_{RJB}1-16$	0.843	0.573	0.416	0.289	0.203	0.185
$RT_{RJB}17-31$	0.878	0.588	0.429	0.298	0.213	0.194
$RT_{RJB}32-46$	0.636	0.358	0.243	0.166	0.117	0.108
$RT_{RJB}47-61$	0.649	0.438	0.321	0.219	0.154	0.141
$RT_{RJB}1-61$	0.753	0.491	0.353	0.244	0.172	0.157
JBU	0.879	0.594	0.423	0.302	0.219	0.182
LT	0.898	0.572	0.388	0.266	0.189	0.156
PT	0.949	0.664	0.480	0.355	0.268	0.234
RJB	0.879	0.586	0.412	0.291	0.206	0.173
SJ_{dir}	0.874	0.567	0.391	0.270	0.188	0.158
SW	0.955	0.664	0.468	0.327	0.232	0.196
GT	0.020	0.045	0.048	0.048	0.049	0.050
UT	0.859	0.525	0.348	0.234	0.159	0.136
SKT	0.885	0.606	0.437	0.311	0.226	0.190
KT	0.870	0.564	0.390	0.271	0.189	0.160
MC	.	0.355	0.235	0.169	0.134	0.126
$MC-LR$	.	0.405	0.261	0.186	0.143	0.131

Table 6: Power of analyzed tests against Pareto (α, $c = 1$) distributions for $n = 15$

test	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$	$\alpha = 5$	$\alpha = 10$
$RT_{JB}1 = JB$	1.000	0.980	0.917	0.803	0.650	0.579
$RT_{JB}2 - 16$	1.000	0.986	0.934	0.832	0.685	0.616
$RT_{JB}17 - 31$	1.000	0.980	0.919	0.806	0.655	0.584
$RT_{JB}32 - 46$	0.986	0.896	0.776	0.630	0.473	0.411
$RT_{JB}47 - 61$	0.999	0.966	0.888	0.762	0.598	0.527
$RT_{JB}1 - 61$	0.996	0.958	0.880	0.758	0.603	0.535
$RT_{RJB}1 - 16$	1.000	0.978	0.906	0.784	0.627	0.543
$RT_{RJB}17 - 31$	0.999	0.956	0.863	0.732	0.572	0.491
$RT_{RJB}32 - 46$	0.995	0.913	0.782	0.626	0.460	0.386
$RT_{RJB}47 - 61$	0.997	0.955	0.862	0.728	0.562	0.482
$RT_{RJB}1 - 61$	0.998	0.951	0.854	0.719	0.556	0.477
JBU	0.999	0.966	0.880	0.752	0.591	0.525
AD	1.000	0.997	0.968	0.896	0.766	0.710
CM	1.000	0.996	0.960	0.878	0.730	0.669
LT	1.000	0.990	0.920	0.787	0.611	0.540
RJB	1.000	0.976	0.897	0.767	0.602	0.533
PT	1.000	0.993	0.942	0.821	0.628	0.543
SJ_{dir}	1.000	0.970	0.868	0.702	0.507	0.435
SW	1.000	0.998	0.976	0.921	0.809	0.760
GT	0.985	0.820	0.620	0.429	0.268	0.218
UT	1.000	0.961	0.840	0.656	0.450	0.377
SKT	1.000	0.984	0.932	0.838	0.702	0.641
KT	0.995	0.908	0.764	0.597	0.426	0.367
MC	0.000	0.770	0.550	0.398	0.285	0.259
$MC - LR$	0.000	0.914	0.711	0.530	0.375	0.327

Table 7: Power of analyzed tests against Pareto (α, $c = 1$) distributions for $n = 25$

test	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$	$\alpha = 5$	$\alpha = 10$
$RT_{JB}1 = JB$	1.000	1.000	0.993	0.958	0.875	0.817
$RT_{JB}2 - 16$	1.000	1.000	0.994	0.967	0.897	0.846
$RT_{JB}17 - 31$	1.000	1.000	0.993	0.959	0.879	0.822
$RT_{JB}32 - 46$	0.997	0.977	0.932	0.844	0.710	0.642
$RT_{JB}47 - 61$	1.000	0.998	0.984	0.938	0.842	0.780
$RT_{JB}1 - 61$	0.999	0.994	0.976	0.927	0.833	0.773
$RT_{RJB}1 - 16$	1.000	0.999	0.989	0.947	0.850	0.773
$RT_{RJB}17 - 31$	1.000	0.998	0.979	0.916	0.800	0.715
$RT_{RJB}32 - 46$	1.000	0.993	0.959	0.874	0.736	0.647
$RT_{RJB}47 - 61$	1.000	0.996	0.975	0.912	0.792	0.705
$RT_{RJB}1 - 61$	1.000	0.997	0.976	0.913	0.795	0.711
JBU	1.000	0.999	0.986	0.939	0.828	0.774
AD	1.000	1.000	0.999	0.989	0.956	0.928
CM	1.000	1.000	0.998	0.984	0.935	0.896
LT	1.000	1.000	0.994	0.958	0.862	0.791
PT	1.000	1.000	0.996	0.966	0.885	0.837
RJB	1.000	0.999	0.987	0.939	0.825	0.765
SJ_{dir}	1.000	0.998	0.973	0.880	0.705	0.609
SW	1.000	1.000	1.000	0.995	0.976	0.957
GT	1.000	0.965	0.850	0.641	0.424	0.334
UT	1.000	0.998	0.965	0.850	0.652	0.549
SKT	1.000	1.000	0.996	0.976	0.915	0.882
KT	1.000	0.988	0.928	0.801	0.621	0.534
MC	0.000	0.918	0.741	0.570	0.429	0.388
$MC - LR$	0.000	0.981	0.868	0.695	0.519	0.466

4 t-Estimator of Pareto tail

As it was shown in previous section, the power of RT class tests depend on a particular value of the Pareto parameter α. The Pareto-type distribution means that as $x \to \infty$, then survival function $\bar{F}(x) = 1 - F(x)$, where F is the c.d.f., can be written as $\bar{F}(x) = x^{-\alpha} l(x)$, where $\alpha > 0$ and l is a slowly varying function. The parameter $\gamma = 1/\alpha$ is known as the extreme value index or tail index, which helps to indicate the size and frequency of extreme events under F.

Let $X_1, ..., X_n$ be *iid* sample from F. If F is strictly Pareto, $\bar{F}(x) = cx^{-\alpha}, x > x_c$, the distribution of relative excesses $Y_i = X_i/t$ over high threshold t conditionally on $X_i > t$ is Pareto with parameter α and support $[1, \infty)$. Denoting the corresponding order statistics by $X_{1,n} \leq ... \leq X_{n,n}$, Hill (1975) suggested to estimate $\hat{\gamma}$ by

$$\hat{\gamma}_k = H_{k,n} = \frac{1}{k} \sum_{j=1}^{k} \log \frac{X_{n-j+1,n}}{X_{n-k,n}} \tag{8}$$

where $X_{n-k,n}$ is the k-th threshold. The Hill estimator is based on a fact that for a sample $Y_1, ..., Y_n$ from strict Pareto distribution with support $[1, \infty)$ and survival function $\bar{F}(x) = x^{-\alpha}$,

$$\frac{1}{\hat{\alpha}_n} = \frac{1}{n}\sum_{i=1}^{n} \log Y_i$$

is the maximum likelihood estimator of $1/\alpha$. The Hill estimator $H_{k,n}$ was shown by Mason (1982) to be consistent estimator for γ (as $k, n \to \infty, k/n \to 0$) whatever the slowly varying function l may be. Since for every choice of k, one obtains another estimator $\hat{\gamma}_k = H_{n,k}$, results are studied by means of Hill plots $\{k, H_{n,k}\}$ for some range of $k \leq n - 1$. However, maximum likelihood estimators are often not very robust, which makes them sensitive to few particular observations, which constitutes a serious problem even in extreme value statistics. Using maximum likelihood estimator point of view, the assumption that for a Pareto-type distribution, above a certain threshold, the relative excesses behave as ordered data from a strict Pareto distribution is sometimes over-optimistic. This mostly happens when the slowly varying part disappears at a very slow rate in many instances resulting in severe bias.

It is known, that formal heavy-tailed propositions can only be satisfactorily involved for empirical constructs if sample data can be taken as a reasonable representation of the underlying distribution. In practice, distribution data may be contaminated by errors. The point of departure is recent research which has shown that Hill estimator is nonrobust. This means that small amounts of data contamination in the wrong place can reverse unambiguous conclusions. The "wrong place" usually means in the upper tail of distribution. As shown in (Brazauskas and Serfling (2000A)), small errors in the estimation of the tail index can bring large errors in the estimation of quantiles. Robust methods for extreme values have been recently addressed by literature. (Brazauskas and Serfling (2000B)) consider robust estimation in the strict Pareto model. (Vandewalle et al. (2007)) proposed robust tail index estimation procedure for the semi-parametric setting of Pareto-type distributions. As discussed in the paper (Stehlík et al. (2010)), t-estimation is at least competitive estimation technique at presence of heavy tails. In (Fabián and Stehlík (2008)) we have shown that t-estimation is clearly better when contamination is present. In this paper we study the generalization of Hill estimator based on t-estimator for Pareto and we show it to be more robust than the classical one. The main novelty of this approach is distributional sensitivity of the estimator: despite all classical modification of Hill estimator for Pareto regularly varying tails are based on asymptotics $x \to \infty$, our method is more sensitive to the interior of the distribution and thus to the distribution itself. In the recent literature there were some works on robustification of Hill estimator, however, our main aim in this section is to construct the t-Hill like estimator, which is distributional sensitive. However, as it can be seen from this section, as a side effect we get also robustness. For the theory of t-estimation see Appendix and references therein.

Let us consider Pareto distribution with support $\mathcal{X} = [1, \infty)$ and density

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

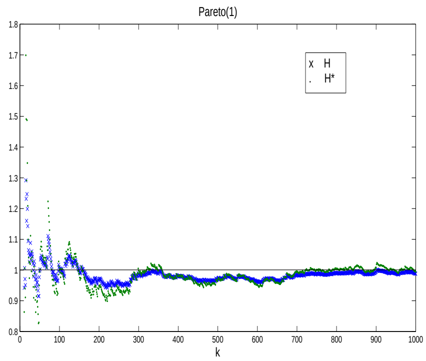

Figure 3:

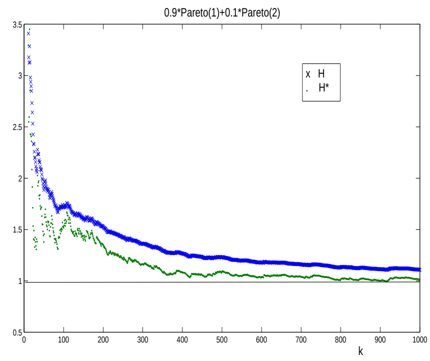

Figure 4:

t-variant of the Hill estimator has the form

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

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

In (Fabián and Stehlík (2009)) we proven the consistency of the t-Hill estimator for Pareto distribution. The consistency of the t-Hill estimator guarantees the consistency of the whole procedure. For the case of unknown location parameter of the Pareto distributions the following lemma provides the estimation technique based on t-estimation.

For European Pareto with density (1) the t-estimator of location parameter c has the form

$$\hat{c}_M = \min(\hat{c}, x_{(1)}),$$

where $x_{(1)}$ is the first order statistic, which is also the maximum likelihood estimate of the threshold parameter.

In the following subsection we illustrate the natural robustness of the t-Hill estimator of the Pareto distributions. For this purpose we compare t-Hill and Hill estimators by the means of simulations. For more detailed discussion on natural robustness of t-estimation see (Stehlík et al. (2010)) or (Fabián and Stehlík (2008)).

4.1 Comparisons

One of problems with the Hill estimator is that it is not sufficiently robust. On the other hand, since the t-Hill estimator is based on harmonic mean, it is resistant to large observations so that it may yield more realistic values for large k. Hill and t-Hill plots for random sample from Pareto $P(1)$ distribution are shown in Figure 3. The length of the sample was 1001 points. It is apparent that t-moment Hill estimator in his first part too much oscillates. The reason is that it is very sensitive to an abrupt change of the threshold value.

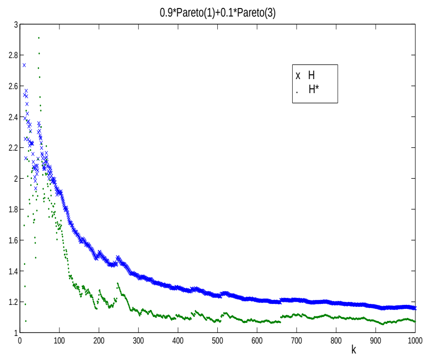

Figure 5:

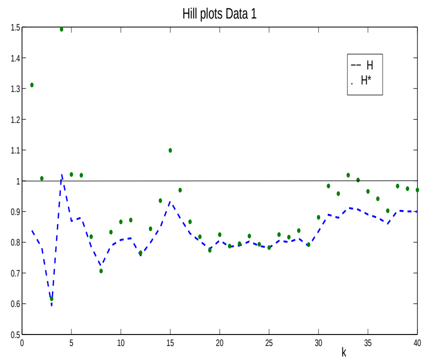

Figure 6:

Figure 4 and Figure 5 show Hill plot $H = \{k, H_{n,k}\}$ and t-Hill plot $H^* = \{k, H^*_{n,k}\}$ for samples generated from the contaminated Pareto distribution

$$F_c = 0.9 * P(1) + 0.1 * P(\delta)$$

with $\delta = 2$ (see Figure 4) and $\delta = 3$ (see Figure 5).

It is apparent that values of t-Hill plots for large k are not too much influenced by large observed values as in ordinary Hill plots.

5 Illustrative examples

5.1 t-Hill estimator

Real data are taken from Example 1 in (Stehlík et al. (2010)). These data consist of 96 payments in one year in non life insurance. Figure 6 illustrates the Pareto tail estimation by t-Hill and Hill estimators for this data.

5.2 Illustrating of step 1 and step 2

Here we illustrate steps 1 and 2 of our procedure. Firstly we generate Pareto distribution with shape parameter α, and secondly we estimate $\hat{\alpha}$ by t-Hill estimator from the generated data. Afterwards we choose optimal tests for normality for both α and $\hat{\alpha}$ and compare their powers. We have chosen Pareto distributions ($\alpha \in \{0.5, 1, 3\}$, $c = 1$) for $n = 5$ and $n = 15$. For t-Hill estimator we have chosen $k = 2$ a $k = 5$.

- Pareto ($\alpha = 0.5$, $c = 1$, $n = 5$, $k = 2$): we recommend tests $RT_{JB}2$, SW and PT. We have $\hat{\alpha} = 0.56$. For such a value we recommend tests $RT_{JB}2$, SW and PT.

- Pareto ($\alpha = 1$, $c = 1$, $n = 5$, $k = 2$): we recommend tests $RT_{JB}2$, $RT_{RJB}13$, SW and PT. We have $\hat{\alpha} = 0.93$. For such a value we recommend tests $RT_{JB}2$, $RT_{RJB}13$, SW and PT.

- Pareto ($\alpha = 3$, $c = 1$, $n = 5$, $k = 2$): we recommend tests $RT_{JB}2$, $RT_{RJB}13$, SW and PT. We have $\hat{\alpha} = 3.31$. For such a value we recommend tests $RT_{JB}2$, $RT_{RJB}13$, SW and PT.

- Pareto ($\alpha = 0.5$, $c = 1$, $n = 15$, $k = 5$): we recommend tests $RT_{JB}9$, $RT_{RJB}13$ and SW. We have $\hat{\alpha} = 0.36$. For such a value we recommend tests $RT_{JB}9$, $RT_{RJB}13$ and SW.

- Pareto ($\alpha = 1$, $c = 1$, $n = 15$, $k = 5$): we recommend tests $RT_{RJB}13$ and SW. We have $\hat{\alpha} = 1.22$. For such a value we recommend tests $RT_{RJB}13$ and SW.

- Pareto ($\alpha = 3$, $c = 1$, $n = 15$, $k = 5$): we recommend tests SW and AD which show higher powers than classes RT_{JB} and RT_{RJB}. We have $\hat{\alpha} = 3.28$. For such a value we recommend SW test.

- Contaminated Pareto distribution $F = 0.8\,\text{Pareto}(\alpha = 0.5, c = 1) + 0.2\text{Pareto}(\alpha = 10, c = 1)$ for $n = 15$ and $k = 5$. We have $\hat{\alpha} = 0.65$. For such a value we recommend tests $RT_{JB}9$, $RT_{JB}13$, $RT_{RJB}13$, SW and AD.

- Contaminated Pareto distribution $F = 0.8\text{Pareto}(\alpha = 1, c = 1) + 0.2\text{Pareto}(\alpha = 10, c = 1)$ for $n = 15$ and $k = 5$. We have $\hat{\alpha} = 0.78$. For such a value we recommend tests $RT_{JB}9$, SW and AD.

- Contaminated Pareto distribution $F = 0.8\text{Pareto}(\alpha = 3, c = 1) + 0.2\text{Pareto}(\alpha = 10, c = 1)$ for $n = 15$ and $k = 5$. We have $\hat{\alpha} = 3.09$. For such a value we recommend tests $RT_{JB}9$, SW and AD.

5.3 Testing for normality

Here we consider two illustrative examples.

Firstly we consider the illustrative example of claims for the mandatory, non-funded 1^{st} (pay-as-you-go) pillar given by (Potocký and Stehlík (2005)). Therein is considered a closed group of Slovakian people, all aged 50 in the year 1998, and interest is in the estimation of the total claim amount for this group in the year 2010 when the members are supposed to retire. We have $n = 10$ of the salaries given from (SLOVSTAT on-line) (see Labour Market, III.3-10, Structure of average gross nominal monthly wage of employees in the economy of the SR). Fig. 7, 8 and 9 present histogram, normal Q-Q plot and boxplot, respectively. In all cases normality is not rejected. The smallest p-value is 0.149 of KT test and the highest one is 0.976 of directed SJ test.

Secondly we consider the illustrative example of absolute logarithmic returns of USD/EUR exchange rate for $n = 25$ (Q2/2003 - Q2/2009). The Pareto fit to the data was checked by KS test gaining p-value of 0.14. Fig. 10, 11 and 12 present histogram, normal Q-Q plot and boxplot, respectively. Selected tests from the classical tests and RT_{JB} and RT_{RJB} classes show p-value greater than 0.05. In particular we consider tests CM, LT, directed SJ, SW, GT, UT, MC, $MC-LR$, $RT_{JB}32$, $RT_{JB}36-38$, $RT_{JB}40$, $RT_{JB}44$, $RT_{RJB}32$, $RT_{RJB}35-38$, $RT_{RJB}40$, $RT_{RJB}42$, $RT_{RJB}44-46$. Selected tests have a smaller power for the considered Pareto alternatives. Moreover,

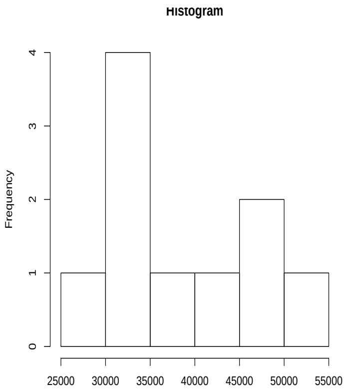

Figure 7: Histogram Salaries

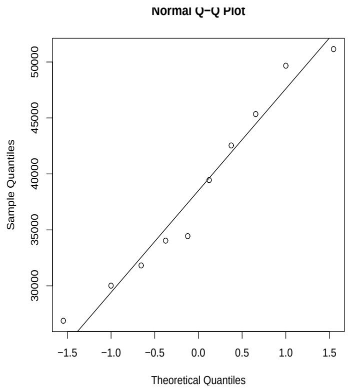

Figure 8: Normal Q-Q plot for Salaries

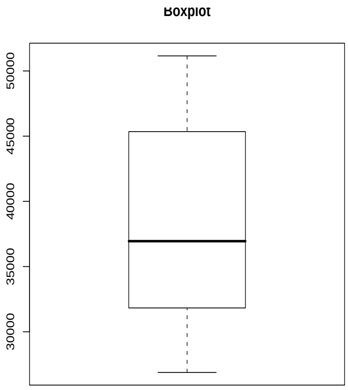

Figure 9: Box plot for Salaries

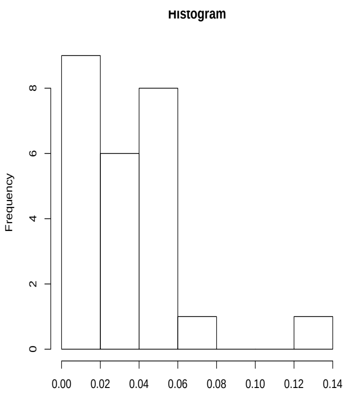

Figure 10: Histogram for USD/EUR

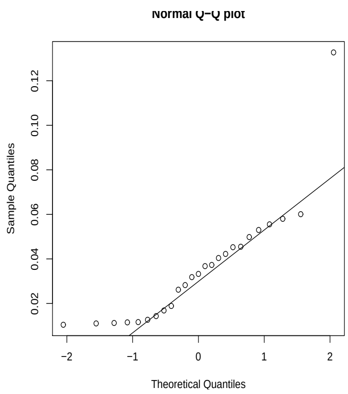

Figure 11: Normal Q-Q plot for USD/EUR

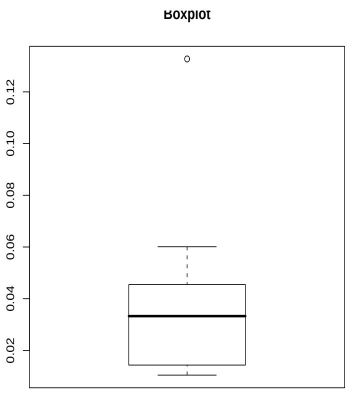

Figure 12: Box plot for USD/EUR

these tests are not able to detect asymmetric alternatives. The smallest p-value is 0 of selected tests and the highest one is 0.634 of MC test.

The selected tests have a high power against symmetric heavy tailed alternatives. The distribution of absolute logarithmic price changes could be considered to be asymmetric (see also histogram and boxplot). The outlier presence is a reason why some tests did not rejected normality.

6 Discussion and conclusions

This paper introduces the general class of robust tests for normality and discuss their properties. The paper also deals with applications of these normality tests on datasets of selected financial time series. In the simulation study we have focused on the power study of selected tests from RT_{JB} and RT_{RJB} subclasses. We have compared these tests with the selected classical tests for the normality on the large scale of alternatives.

General guidelines. Based upon our experience we can recommend the following general guidelines for normality testing.

- For heavy tailed symmetric or asymmetric alternatives: there exist no better tests than SJ_{dir} or RJB. Comparable powers are obtained also for $RT_{JB}9$, $RT_{JB}43$, $RT_{RJB}13$ and $RT_{RJB}59$ tests.

- For asymmetric light tailed alternative there is no better test than SW. The best mimic from RT_{JB} and RT_{RJB} subclasses is $RT_{JB}9$ (with a small decrease in power)

- For symmetric short tailed alternative the best test is KT test. For Beta$(0.5, 0.5)$ SW outperforms KT test. RT_{JB} subclass outperforms RT_{RJB} subclass. The best test from RT_{JB} subclass is $RT_{JB}42$ test. For uniform alternative, $n = 25$, power of KT test is 0.447, SW test power is 0.290 and $RT_{JB}42$ test power is 0.316. For Beta$(0.5, 0.5)$ alternative, $n = 25$, power of KT test is 0.823, SW test power is 0.861 and $RT_{JB}42$ test power is 0.763

- For bimodal alternatives, the best power is gained by Geary and Uthoff tests. The JB, JBU, RJB, KT and SJ_{dir} are even biased, especially for small samples. In contrast some robust modifications of JB and RJB tests ($RT_{JB}39$, $RT_{JB}42$ and $RT_{RJB}42$) show nonzero power and are comparable with the Geary and Uthoff tests. Power increase of the selected tests from RT_{JB} and RT_{RJB} subclass in comparison with the classical JB and RJB tests is as follows: e.g. for the bimodal mixture $F = 0.5N(0,1)+0.5N(5,1)$ and sample size $n = 24$ JB, JBU, RJB tests and directed SJ test gain almost zero powers. However selected tests from RT_{JB} and RT_{RJB} subclass gain relatively good power, i.e. $RT_{JB}39$ has power 0.81, $RT_{RJB}42$ has power 0.95 and $RT_{JB}42$ power is even 0.97. The latter values of power are comparable with the best tests for normality against such alternatives, e.g. Geary and Uthoff tests with power 0.99.

- For location-contaminated standard normal distribution alternatives having CDF $F = pN(0,1) + (1-p)N(\mu_2, 1)$ with $p = 0.1, 0.3, 0.5, 0.7, 0.9$ and $\mu_2 = 0, 1, 2, 3, 4, 5$, the best tests are JBU, JB, KT, GT, UT and DT tests. Comparable results are also for $RT_{JB}9$ test.

- For scale-contaminated standard normal distribution alternatives having CDF $F = pN(0,1)+(1-p)N(0, \sigma_2^2)$ with$p = 0.1, 0.3, 0.5, 0.7, 0.9$ and $\sigma_2^2 = 1, 4, 9, 16, 25$, the best tests are SJ_{dir} and RJB. Comparable results are obtained also for $RT_{JB}9$, $RT_{JB}43$ and $RT_{RJB}59$ tests.

- For location- and scale-outliers models, the tests with higher power are JBU, JB, DT, AD, CM, SW, $RT_{JB}39$, $RT_{JB}42$ and $RT_{RJB}42$ tests.

- Many results of this paper are in the coherence or are extending the results of previous studies, e.g. (Thadewald and Bunning 2007).

Interpretation of power differences. Some interpretation for the power behavior can be made using the kurtosis as a metric to measure departure from normality in the certain class of distributions (see (Keilson and Steutel 1974)). For instance the power of RT_{JB} tests under Weibull alternative and mid sample $n = 100$ belongs to range $(0.13, 0.76)$ and power of RT_{RJB} tests belongs to range $(0.03, 0.96)$ since kurtosis is 3.14. Thus the departure from normality is less than by uniform mid sample $n = 100$ with kurtosis 1.83, where the robustness of the tests led to the higher spread of the powers, approximately within the range $(0.00, 0.99)$ of RT_{JB} tests and $(0.00, 1.00)$ of RT_{RJB} tests. On the other hand the power of RT_{JB} under Laplace alternative and mid sample $n = 100$ belongs to range $(0.11, 0.83)$ and power of RT_{RJB} tests belongs to range $(0.42, 0.94)$ since kurtosis is 5.32. These relations will be worth further investigation.

Testing against Pareto tail

In the step 1 of our testing procedure we estimate Pareto tail parameter α by t-Hill estimator. There are two main ways of avoiding misleading conclusions due to nonrobust tools in the presence of contaminated data. One is based on statistics that automatically remove from the sample data that are potentially troublesome. The other relies on the specification of parametric models for the distribution of the data and uses robust estimators of the parameters. As can bee seen from this paper, t-Hill estimator of Pareto tail index is distribution sensitive and "naturally" robust. If more accurate fit to the central part of distribution is needed, we suggest to use e.g. combining a Pareto estimate of the upper tail with a non-parametric estimate of the rest of the distribution, as suggested by (Cowell and Victoria-Feser (2007)) and by (Davidson and Flachaire (2007)) with bootstrap methods. In this paper authors used deliberately t-Hill estimator. However we are aware of fact that competitive estimators may exist and further comparisons and investigations should be done in this direction.

The authors are aware about the results in the Pareto tail parameter tests of the form $H_0 : F$ is of Pareto with tail $\alpha \leq \alpha_0$ against the lighter right tail alternative, e.g. (Jurečková (2000)) or (Jurečková and Picek (2001)). There is a possibility to construct adaptive procedures where 1) in the first step consistent test of α is employed to a dataset; and in the second step 2) we took an appropriate robust tests for normality, dependently on α from the 1^{st} step. However, in our paper we considered mainly mid and small samples and therefore more work should be done to validate such a procedures. Therefore we considered this problem to be out of the scope of this paper and further investigations may be of interest.

Acknowledgement
Research was supported by projects AKTION Austria - Czech Republic Nr. 51p7, Nr. 54p21, Nr. 50p14 and Nr. 54p13.

7 Appendix

7.1 Definition of tests

1) The "mean-median" Jarque-Bera test statistic $RT_{JB}2$ defined by

$$RT_{JB}2 = \frac{n}{6}\left(\frac{M_3(0, T_{(0)}(0))}{M_2^{3/2}(0, T_{(0)}(0))}\right)^2 + \frac{n}{24}\left(\frac{M_4(0, T_{(0)}(0))}{M_2^2(0, T_{(1)})0))} - 3\right)^2.$$

2) The "mean-median" Jarque-Bera test statistic $RT_{JB}9$ defined by

$$RT_{JB}9 = \frac{n}{18}\left(\frac{M_3(0, T_{(1)}(0))}{M_2^{3/2}(0, T_{(0)}(0))}\right)^2 + \frac{n}{24}\left(\frac{M_4(0, T_{(0)}(0))}{M_2^2(0, T_{(0)})0))} - 3\right)^2.$$

3) The "mean-trimmed-trimmed mean" Jarque-Bera test statistic $RT_{JB}39$ defined by

$$RT_{JB}39 = \frac{n}{6}\left(\frac{M_3(1, T_{(2)}(1))}{M_2^{3/2}(0, T_{(0)}(0))}\right)^2 + \frac{n}{24}\left(\frac{M_4(0, T_{(0)}(0))}{M_2^2(0, T_{(0)})0))} - 3\right)^2.$$

4) The "mean-trimmed-trimmed mean" Jarque-Bera test statistic $RT_{JB}42$ defined by

$$RT_{JB}42 = \frac{n}{6}\left(\frac{M_3(1, T_{(2)}(1))}{M_2^{3/2}(0, T_{(0)}(0))}\right)^2 + \frac{n}{24}\left(\frac{M_4(1, T_{(2)}(1))}{M_2^2(1, T_{(2)})1))} - 3\right)^2.$$

5) The "mean-trimmed-trimmed mean" Jarque-Bera test statistic $RT_{JB}43$ defined by

$$RT_{JB}43 = \frac{n}{6}\left(\frac{M_3(1, T_{(2)}(1))}{M_2^{3/2}(1, T_{(2)}(1))}\right)^2 + \frac{n}{24}\left(\frac{M_4(0, T_{(0)}(0))}{M_2^2(0, T_{(0)})0))} - 3\right)^2.$$

6) The $RT_{RJB}13$ test statistic defined by

$$RT_{RJB}13 = \frac{n}{18}\left(\frac{M_3(0,T_{(1)}(0))}{M_0^3(0,T_{(1)}(0))}\right)^2 + \frac{n}{58}\left(\frac{M_4(0,T_{(0)}(0))}{M_0^4(0,T_{(0)}(0))} - 3\right)^2.$$

7) The $RT_{RJB}33$ test statistic defined by

$$RT_{RJB}33 = \frac{n}{6}\left(\frac{M_3(0,T_{(0)}(0))}{M_0^3(0,T_{(0)}(0))}\right)^2 + \frac{n}{58}\left(\frac{M_4(1,T_{(2)}(1))}{M_0^4(0,T_{(0)}(0))} - 3\right)^2.$$

8) The $RT_{RJB}42$ test statistic defined by

$$RT_{RJB}42 = \frac{n}{6}\left(\frac{M_3(1,T_{(2)}(1))}{M_0^3(0,T_{(0)}(0))}\right)^2 + \frac{n}{58}\left(\frac{M_4(1,T_{(2)}(1))}{M_0^4(1,T_{(2)}(1))} - 3\right)^2.$$

9) The $RT_{RJB}59$ test statistic defined by

$$RT_{RJB}59 = \frac{n}{18}\left(\frac{M_3(0,T_{(3)}(0))}{M_0^3(0,T_{(3)}(0))}\right)^2 + \frac{n}{58}\left(\frac{M_4(0,T_{(0)}(0))}{M_0^4(0,T_{(3)}(0))} - 3\right)^2.$$

7.2 t-estimation

It was shown in Fabián (2008) that regular continuous distributions with interval support $\mathcal{X} \in R$ can be characterized, besides the cumulative distribution function $F(x)$ and probability density $f(x)$, by its t-score, given by

$$T(x) = \frac{1}{f(x)}\frac{d}{dx}\left(-\frac{1}{\eta'(x)}f(x)\right), \qquad (10)$$

where $\eta : \mathcal{X} \to R$ is an appropriate, strictly increasing continuous mapping. In the case of support $\mathcal{X} = (a, \infty)$, mapping

$$\eta(x) = \log(x - a) \qquad (11)$$

yields often the simplest formulas for t-scores. The t-score is a suitable function for using the generalized moment method for estimation of parameters of heavy-tailed distributions, since it appeared that T is for these distributions bounded, and the moments

$$ET^k = \int_{\mathcal{X}} T(x)^k \, dF(x), \qquad k = 1, 2, ..., \qquad (12)$$

exist and are often given by simple expressions. Let us call them the *t-score moments*. Particularly,

$$ET = 0 \qquad (13)$$

and ET^2 is the Fisher information for x^*, which is the solution of equation

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

called the *t-mean*, which can be considered as a measure of central tendency of distributions (Fabián, 2008).

Let $\theta \in \Theta \subseteq R_m$ and $(X_1, ..., X_n)$ be iid sample from F_θ. The parametric version of (12) yields the generalized moment estimation equations for θ in the form

$$\hat{\theta}_n : \qquad \frac{1}{n}\sum_{i=1}^{n} T(x_i;\theta)^k = ET^k(\theta), \quad 1 \leq k \leq m. \tag{14}$$

Since $\hat{\theta}_n$ is the M-estimate, it is strongly consistent and asymptotically normal with the asymptotic variance-covariance matrix derived by Fabián (2001). Since distributions with heavy tails have bounded t-scores, $\hat{\theta}_n$ of heavy-tailed distributions are robust with respect to large values in observed samples.

7.3 Proofs

Proof of Theorem 1

For mean Under the null, μ is the mean of Normal distribution and $\sqrt{n}(M_n - \mu) \sim N(0, 1/(4f(\mu)^2))$ (see (Casella and Berger 2002), p. 484). Thus $E(\mu - M_n)^k \to 0$ for $n \to \infty$.

For $k = 2$ we have $E(X_i - M_n)^2 = E(X_i - \mu)^2 + E(\mu - M_n)^2 + 2E(X_i - \mu)(\mu - M_n)$. For $k = 3$ we have

$$E(X_i - M_n)^3 = 3E[(X_i - \mu)^2(\mu - M_n)] + E(\mu - M_n)^3 + E(X_i - \mu)^3 + 3E(X_i - \mu)(\mu - M_n)^2$$

Thus $E(\mu - M_n)^3 \to 0$ for $n \to \infty$ and from Cauchy-Schwarz inequality we have $|E(X_i - \mu)^2(\mu - M_n)| \leq \sqrt{E(X_i - \mu)^4 E(\mu - M_n)^2} \to 0$ for $n \to \infty$. Similarly for other mixed terms.

For $k = 4$ we have $E(X_i - M_n)^4 = E(X_i - \mu)^4 + 6E[(X_i - \mu)^2(\mu - M_n)^2] + E(\mu - M_n)^4 + 4E(X_i - \mu)^3(\mu - M_n) + 4E(X_i - \mu)(\mu - M_n)^3$. Thus $E(\mu - M_n)^4 \to 0$ for $n \to \infty$ and from Cauchy-Schwarz inequality we have $|E(X_i - \mu)^2(\mu - M_n)^2| \leq \sqrt{E(X_i - \mu)^4 E(\mu - M_n)^4} \to 0$ for $n \to \infty$. Similarly for other mixed terms.

For trimmed mean and median

First, let us consider the trimmed mean, $M_{2(n)}$. We have $P(|M_{2(n)} - \mu| \geq \epsilon) \sim O(1/n)$, $\forall \epsilon > 0$ for $n \to \infty$ because of the normal asymptotical distribution of $M_{2(n)}$ (see (Stigler (1973))) and $E(M_{2(n)}) = \mu$. Particularly for the median $M_{1(n)}$ we obtain $P(|M_{1(n)} - \mu| \geq \epsilon) \sim 1/(4f(\mu)^2 n)$, $\forall \epsilon > 0$ and $\sqrt{n}(M_{1(n)} - \mu) \sim N(0, 1/(4f(\mu)^2))$ (see (Casella and Berger 2002), p. 484). Thus $M_{i(n)}, i = 1, 2$ converge in probability to μ. Since $g_j(u) = (1/n)\sum_{i=1}^{n} \varphi_j(X_i - u)$ is a continuous function for $j = 0, 1, 2, 3, 4$, $g_j(M_{i(n)})$ converge in probability to $g(\mu)$ which is a consistent estimator of μ_j. Therefore also $g_j(M_{i(n)})$ is a consistent estimator of μ_j, where $\mu_0 = \sigma$.

For pseudo-median

We have for a symmetric density h (under null hypothesis density is normal) $M_{3(n)} \sim N(M_3, \tau^2/n), \tau = \frac{1}{\sqrt{12}\int h^2(x)dx}$ (see (Hettmansperger and McKean (1998))).

Proof of Theorem 2 From convergence in probability, we have the following convergence in distribution $M_{3,i} - \hat{\mu_3} \to 0$, $M_{4,j} - \hat{\mu_4} \to 0$, and $M_{2,l}^2 \to \sigma^4$ for $n \to \infty$. The proof is completed by employing of multivariate Slutsky's theorem.

Proof of Theorem 3

From convergence in probability, we have the following convergence in distribution $M_3 - \hat{\mu_3} \to 0$ and $M_4 - \hat{\mu_4} \to 0$. The rest of the proof follows from Theorem 1 and its proof in (Gel and Gastwirth (2008)).

Proof of Lemma 1.

Using the mapping $\eta = \log(x-1)$, $\eta'(x) = 1/(x-1)$ and, by (10), the t-score (10) is

$$T_\alpha(x) = -1 - (x-a)f'(x)/f(x) = \alpha(1 - x^*/x)$$

where $x^* = (\alpha+1)/\alpha$. It follows from (14) and (13) that

$$\sum_{i=1}^{n} T(x_i;\alpha) = 0$$

so that $\hat{x}^* = \bar{x}_H$ where $\bar{x}_H = n/\sum_1^n 1/x_i$ is the harmonic mean, and

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

It suggests to introduce a variant of the Hill estimator as

$$\hat{\gamma}_k = \frac{1}{\hat{\alpha}_k} = H^*_{k,n} = \frac{1}{\frac{1}{k}\sum_{j=1}^{k} \frac{X_{n-k,n}}{X_{n-j+1,n}}} - 1, \tag{15}$$

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

Proof of Lemma 2.

Using mapping $\eta = \log(x-1)$ we obtain from (10) the t-score $T(x;\alpha) = -1 - (x-c)f'(x)/f(x) = \alpha(1 - x^*/x)$ where $x^* = c(\alpha+1)/\alpha$. Since $ET^2 = \alpha/(\alpha+2)$ and $\eta'(x^*) = 1/(x^*-c)$, the measure of variability is

$$\omega^2 = \frac{c^2(\alpha+2)}{\alpha^3}. \tag{16}$$

For a given ω, α is determined from (16).

The t-score moment equations are

$$\begin{aligned} \sum_{i=1}^{n}(1 - x^*/x_i) &= 0 \\ \frac{1}{n}\sum_{i=1}^{n}(1 - x^*/x_i)^2 &= \frac{1}{\alpha(\alpha+2)}, \end{aligned}$$

so that $\hat{x}^* = \bar{x}_H$ is the harmonic mean. Denoting $\bar{x}_{2H} = n/\sum_{i=1}^n 1/x_i^2$ and

$$\rho = \frac{\bar{x}_{2H}}{\bar{x}_H^2 - \bar{x}_{2H}},$$

from the second equation $\hat{\alpha} = \sqrt(1+\rho) - 1$ and

$$\hat{c} = \bar{x}_H \hat{\alpha}/(\hat{\alpha}+1).$$

The t-score moment estimator of the threshold parameter is then

$$\hat{c}_M = \min(\hat{c}, x_{(1)}),$$

where $x_{(1)}$ is the first order statistic, which is also the maximum likelihood estimate of the threshold parameter.

References

Bera A, Jarque C (1981) Efficient tests for normality, heteroskedasticity and serial independence of regression residuals: Monte Carlo evidence. Econ Lett 7:313-318

Bickel P.J. and Lehmann E.L. (1975). Descriptive Statistics for Nonparametric Models II. Location. Ann. Statist. Volume 3, Number 5 (1975), 1045-1069.

Bowman, K.O., Shenton, L.R. (1975). Omnibus contours for departures from normality based on $\sqrt{b_1}$ and b_2. Biometrika, 62, 1975, p. 243-250.

Brazauskas V. and Serfling R. Robust and efficient estimation of the tail index of a single-parameter Pareto distribution. North Amer. Actuar. J. 4, 12-27.

Brazauskas V. and Serfling R. Robust estimation of tail parameters for two-parameter Pareto and exponential models via generalized quantile statistics. Extremes 3:3, 231-249.

Brys, G., Hubert, M., Struyf, A. (2008). Goodness-of-fit tests based on a robust measure of skewness. Computational Statistics (2008) 23:429-442.

Cassela, G., Berger, R.L. (2002). Statistical Inference, 2nd Edition, Duxbury Advanced Series, Thomson Learning.

Cowell F.A. and Victoria-Feser, M.P. Robust Lorenz Curves: A Semiparametric Approach, Journal of Economic inequality, 5, 21-35

D'Agostino, R.B. (1986). Tests for normal distribution. In D'Agostino, R.B. and Shephens, M.A.: Goodness of fit techniques. New York: Marcel Dekker, 1986, p. 367-419.

Davidson R. and Flachaire E. (2007) Asymptotic and bootstrap inference for inequality and poverty measures, Journal of Econometrics 141 (2007) 141-166

Epps, T.W. and Pulley, L.B., (1983). A test for normality based on the empirical characteristic function. Biometrika, 70, 723-726.

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

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

Fabián Z. and Stehlík M. (2008), A note on favorable estimation when data is contaminated, Communications in Dependability and Quality Management 11(2008)4, 36-43.

Fabián Z. and Stehlík M. (2009), "On robust and distribution sensitive Hill like method", IFAS res. report.43

Geary, R.C. (1935). The ratio of the mean deviation to the standard deviation as a test of normality. Biometrika, 27, p. 310-332.

Gel, Y.R., Gastwirth, J.L. (2008). A robust modification of the Jarque Bera test of normality. *Economics Letters*, 99, 30–32.

Gel, Y.R., Miao, W., Gastwirth, J.L. (2007) Robust directed tests of normality against heavy-tailed alternatives. Computational Statistics & Data Analysis, 51, p. 2734-2746.

Hettmansperger T.P. and McKean J.W. (1998). Robust Nonparametric Statistical Methods *Kendalls Library of Statistics 5*, p. 37.

Jurečková J. (2000) Test of tails based on extreme regression quantiles, Statistics & Probability Letters Volume 49, Issue 1: 53-61.

Jurečková J. and Picek J. (2001) A Class of Tests on the Tail Index, Extremes,Volume 4, Number 2: 165-183.

Keilson J. and Steutel F. W. (1974), Mixtures of Distributions, Moment Inequalities and Measures of Exponentiality and Normality, The Annals of Probability, Vol. 2, No. 1., pp. 112-130.

Locke Ch. and Spurrier J.D. (1977) The use of U-statistics for testing normality against alternatives with both tails heavy or both tails light, Biometrika 1977 64(3):638-640

Potocký R. and Stehlík M., 2005: Analysis of pensions in the 1st pilar under the expected demographic development, in Proceedings of 10th Slovak Conference on Demography, Smolenice.

Resnick S.I. Heavy-tail phenomena (2007). Springer.

Saniga E.M. and Miles J.A. (1979) Power of Some Standard Goodness-of-Fit Tests of Normality Against Asymmetric Stable Alternatives, Journal of the American Statistical Association, Vol. 74, No. 368, pp. 861-865

Spiegelhalter D. J. (1980). An omnibus test for normality for small samples, Biometrika 67: 493-496

Statistical database of indicators of economic and social-economic development in the Slovak Republic.

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

Stigler S.M. (1973). The Asymptotic Distribution of the Trimmed Mean, Ann. Statist. Volume 1, Number 3, 472-477.

Thadewald, T., Bunning, H. (2007). Jarque-Bera test and its competitors for testing normality - a power comparison. Journal of Applied Statistics, Vol. 34, No. 1, 87-105.

Urzua, C.M. (1996). On the correct use of omnibus tests for normality. Economics Letters, 53, p. 247-251.

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

Yu, J., Testing for a Finite Variance Stock Return Distributions, in J.L. Knight and S.E.. Satchell (eds.) Return Distributions in Finance., 2001,

Wilcox, R. R. (2005). Introduction to Robust Estimation and Hypothesis Testing. 2nd Edition. San Diego, CA: Academic Press, pages 20-21.