Erasmus University Rotterdam
Erasmus School of Economics

Bachelor Thesis: Econometrics and Operations Research

Testing for Heavy Tails using Empirical Distribution Shape: The Tail Ratio Test

Author: Alessandro Kowsoleea, Student ID: 410165

Supervisor: A. J. Koning
Second assessor: P. J. F. Groenen

Final date: 2 July 2017

Contents

Abstract

We look at goodness-of-fit tests for heavy tailed distributions. As these distributions are found in numerous fields, adequate goodness-of-fit testing helps to correctly model them. We inspect the power of four different tests to find the best one: the Kolmogorov-Smirnov test, the Anderson-Darling test, the Estimated Score test and the Tail Ratio test. The latter is developed in this paper. We use simulation to find critical values of these tests and use simulation once again to find their power. Our results show that the Tail Ratio test has the highest power of the four tests.

1 Introduction

In various fields, variables of interest exhibit more tail observations than accounted for in conventional models. If the number of tail observations decreases to zero slower than exponentially as we get further in the tail, we speak of a heavy tailed distribution. More formally, a distribution F is called heavy tailed with tail index α if:

$$\lim_{t\to\infty} \frac{1 - F(tx)}{1 - F(x)} = x^{-\alpha}, \forall x > 0, \alpha > 0 \tag{1}$$

Heavy tailed distributions find applications in various practical fields. Rizzo (2009) mentions applications in economics, finance and actuarial science such as modeling income, losses or claim sizes. Abidin et al. (2014) mention modeling the likelihood of floods and the occurrence of drought using heavy tailed distributions. Finally, Reittu and Norros (2004) use heavy tailed distributions to model Internet traffic size.

Most of the literature on heavy tailed distributions has been on fitting a distribution and estimating its parameters. On the other hand much less has been written about goodness-of-fit testing for heavy tailed distributions. The relevance of development in modeling heavy tailed data is clear: it has an explanatory or predictive function. The use of these models however warrants the relevance of goodness-of-fit testing: to find a well fitting model with better explanatory or predictive power, adequate goodness-of-fit tests are necessary. Therefore this research is directly relevant to all the applications of heavy tailed modeling.

Hüsler and Peng (2008) show an overview of results in both modeling and testing univariate and multivariate heavy tails. Other examples of research in this direction include Beirlant et al. (2006) who introduce a kernel statistic for Pareto-type behavior and Goegebeur and Guillou (2010) who introduce Lewis- and Jackson-type test statistics for Weibull-type behavior. Whereas both their null-hypotheses assume a certain distribution, Koning and Peng (2008) conduct goodness-of-fit tests under a null-hypothesis with no assumed distribution.

This paper continues in line with Koning and Peng (2008): We look at several tests for the following null-hypothesis and its alternative:

$$H_0 :\text{"}F \text{ is heavy tailed", versus}$$
$$H_a :\text{"}F \text{ is not heavy tailed"}$$

in which F is a distribution function. This null-hypothesis has two implications: the tail index parameter remains unknown and thus needs to be estimated and the tail index only describes tail behavior of the distribution, giving only a partial description of the underlying distribution.
The aim of this paper is to distinguish the power of four different tests. The first test is the Kolmogorov-Smirnov test (KS). Details of this test can be found in Kolmogorov (1933) and Smirnov (1939). The second test is the Anderson-Darling test (AD), presented in Anderson and Darling (1952). This can be seen as a weighted version of the KS test. In Koning and Peng (2008), the estimated score test is presented (SC) and its performance is evaluated and compared to the KS test, among others. The derivation of this test can be found in Hjort and Koning (2002). We extend on the paper of Koning and Peng (2008) by adding the AD test and "tail ratio" test (TR) to the comparison group. This final test, the tail ratio test, is developed below. It is an intuitive test based on the ratio between a sample's largest order statistic and other upper order statistics.
Koning and Peng (2008) find that the integrated version of the SC test is more powerful than the SC test, the (integrated) KS test and (integrated) Berk-Jones statistic for most of the chosen tail lengths. However, we do not take the integrated tests into account. Their paper shows comparable power between the SC and KS tests. Their performance has to our knowledge not been compared to that of the AD test. Since this test is often used, we think a comparison of their power is interesting. On the other hand the tail-ratio test has no set track record. This raises the research question:

How does the power of the TR test compare to the powers of the SC, KS and AD tests when testing H_0 against H_a?

We conduct two simulation studies to answer this question. In the first simulation we simulate samples under the null-hypothesis. We use these samples to find the empirical critical values of the various tests. The second simulation concerns calculating the power of the different tests. This is done by simulating samples under the alternative hypothesis and calculating how well the tests can distinguish these from samples under the null-hypothesis. Both simulations are conducted for multiple numbers of order statistics, as the choice of this number influences both critical values and power of the tests.
The rest of this paper is structured as follows: Section two discusses the KS, AD, SC and TR statistics. Section three presents an outline of the simulation study conducted to find both the critical values and powers of the tests. Sections four gives an overview of the simulation results. Finally, in section five we provide our answers to the research questions and make concluding remarks.

2 Methodology: Tests

The KS, AD and SC test have a common base, namely the comparison of the empirical density function to a theoretical density function. In this application, we use a transformation of the observations which follows a certain empirical density. Then, under the null hypothesis and two other conditions, we can compare it to a theoretical density.

Suppose we have n independent identically distributed (i.i.d.) observations $X_1, X_2, \ldots, X_n$. We order them to form the ordered sample $X_{n,1} \leq X_{n,2} \leq \ldots X_{n,n}$. Dividing the last k observations by $X_{n,n-k}$ gives the ordered sample $\frac{X_{n,n-k+1}}{X_{n,n-k}}, \frac{X_{n,n-k+2}}{X_{n,n-k}}, \ldots, \frac{X_{n,n}}{X_{n,n-k}}$. We will further refer to $\frac{X_{n,n-k+i}}{X_{n,n-k}}$ as Y_i, which has the following empirical density function:

$$G_k(r) = \frac{1}{k}\sum_{i=1}^{k} \mathbb{I}(\frac{X_{n,n-k+i}}{X_{n,n-k}} \leq r) = \frac{1}{k}\sum_{i=1}^{k} \mathbb{I}(Y_i \leq r). \tag{2}$$

Under the following two conditions, $G_k(r)$ behaves as the empirical distribution function derived from a random sample from the Pareto distribution with a shape parameter α and a scale parameter 1 (see de Haan and Ferreira (2006) and de Haan and Stadtmüller (1996)). The density of this distribution is given by:

$$G(r; \alpha) = 1 - r^{-\alpha}, \ \forall r > 1. \tag{3}$$

The first condition, which ensures convergence of the distribution, is as follows: suppose there exists a function $A(t) \to 0$, as $t \to \infty$ such that

$$\lim_{t\to\infty} \frac{\frac{1-F(tx)}{1-F(x)} - x^{-\alpha}}{A(t)} = x^{-\alpha}\frac{x^{\rho} - 1}{\rho}, \tag{4}$$

for all $x > 0$, where $\rho \leq 0$. Then, if the second condition about the convergence speed,

$$\lim_{n\to\infty} \sqrt{k}A(n/k) = 0, \tag{5}$$

holds we can indeed compare the above empirical and theoretical densities.

A last hurdle remains in the scale parameter α. As it is unknown, we need to estimate it in order to be able to compare the densities. For this purpose we use the well known Hill estimator (see Hill (1975)). In the case of this Pareto distribution it is given by:

$$\hat{\alpha} = \Big[\frac{1}{k}\sum_{i=1}^{k} log(Y_i)\Big]^{-1}. \tag{6}$$

2.1 Kolmogorov-Smirnov test

Of the three tests that compare the empirical and theoretical distribution, the KS test does it in the most straightforward way. We define the following function:

$$KS(r,\alpha) = 1 - G_k(r) - (1 - G(r,\alpha)) = 1 - G_k(r) - r^{-\alpha}. \tag{7}$$

The test statistic would normally be computed as:

$$KS = \sup_{r>1} |\sqrt{k} KS(r, \hat{\alpha})|. \tag{8}$$

However, since $KS(r,\alpha)$ is not smooth and has a multitude of local maximums, finding this supremum can be rather difficult. We therefore opt for a simplified version of this statistic:

$$KS^* = \sqrt{k} \max_{r \in Y} (1 - G_k(r) - r^{-\hat{\alpha}}). \tag{9}$$

2.2 Anderson-Darling test

As stated before, the AD test can be interpreted as a weighted version of the KS test. The test is introduced in Anderson and Darling (1952) as follows:

$$AD = \sup_{r>1} \sqrt{k} |G_k(r) - G(r; \hat{\alpha})| \sqrt{\Psi(G(r;\alpha))}. \tag{10}$$

Herein Ψ is a weighting function based on the theoretical distribution function. For this application, however, we use a weighting function based on the empirical distribution function. We give extra weight to the larger order statistics. Another adaption we make to the original statistic is a discretisation: again due to the non-smooth nature of the function, we take the maximum of a set of function values. The test statistic we use is:

$$AD^* = \sqrt{k} \max_{r \in Y} (1 - \frac{1}{k} \sum_{i=1}^{k} w_i \mathbb{I}(Y_i < r) - r^{-\hat{\alpha}}), \tag{11}$$

where w_i is the weight assigned to the i-th observation, given by:

$$w_i = \frac{i^2}{\sum_{i=1}^{k} i^2}. \tag{12}$$

2.3 Estimated score test

Just like the previous two tests, the SC test compares the theoretical and empirical distribution. This is done using martingale theory. It uses that $G_k(r)$ is a sub-martingale. Then using Doob-Meyer decomposition (see Meyer (1962)) the sub-martingale is transformed into a martingale by subtracting the predictable part of the sub-martingale. In this case, the predictable process is given by $\int_1^r (1 - G_k(s) d\Lambda(s,\alpha)$. For a more formal derivation of this martingale see Hjort and Koning (2002). This motivates the following function introduced in Koning and Peng (2008):

$$SC(r,\alpha) = G_k(r) - \int_1^r (1 - G_k(s))d\Lambda(s;\alpha), \tag{13}$$

where $\Lambda(r,\alpha) = -log(1 - G(r;\alpha)) = \alpha log(r)$ is the cumulative hazard function of $G(r,\alpha)$. Equation 13 then simplifies to:

$$SC(r;\alpha) = G_k(r) - \alpha \int_1^r \frac{1 - G_k(s)}{s} ds. \tag{14}$$

As $1 - G_k(s) = 0$ when $Y_i < s$, the integral can be solved easily. We rewrite equation 13 once more to:

$$SC(r;\alpha) = G_k(r) - \alpha log(\min(Y_i, r)). \tag{15}$$

Finally, the test statistic is computed as:

$$SC = sup_{r>1}|\sqrt{k}SC(r;\hat{\alpha})|. \tag{16}$$

2.4 Tail ratio test

Unlike the other tests discussed in this paper, the tail ratio test is not based on the relation between the empirical distribution and a theoretical distribution. Rather, the tail ratio test only uses information about the shape of the empirical distribution to draw inference about heaviness of the tail.
As the name suggests, the tail ratio tests uses a ratio of tail observations. Specifically, for a tail of length k, the test statistic is given by:

$$TR = (\sum_{i=1}^{k} |\frac{X_{n,n-i}}{X_{n,n}}|)^{-1}. \tag{17}$$

The intuition behind this statistic is quite simple. Assume a distribution is heavy tailed, then it has a higher amount of tail observations. Now as there are more observations in the tail, the ratio between the last tail observation and the second-to-last one should be high. The same holds for the last tail observation and the third-to-last one, the last one and the fourth-to-last one, and so on. The TR statistic is based on this pattern, where we take the sum of all these ratios over a specified tail length. Finally the inverse is taken, to set the statistic in line with most tests: the statistic takes a lower value if it fails to reject the null-hypothesis and grows as it becomes more likely to reject the null-hypothesis.
Due to the test's functional form, we expect the chosen tail length to heavily influence the test's performance. If the tail length is chosen too long, compared to the actual number of tail observations, the test statistic will be computed based on too many regular (non-extreme) values.

3 Methodology: Simulation

3.1 Critical values

In order to find the critical values of the various tests, we simulate under the null-hypothesis. For this purpose we use a Frechet distribution, in line with Koning and Peng (2008). This distribution has the following density:

$$F(x) = exp(-x^{-1}). \tag{18}$$

We create 100.000 samples of size $n = 1000$ by drawing uniformly distributed random variables on the interval $[0, 1]$ and transform these to Frechet distributed random variables using the inverse transformation method. The transformation we use is given by:

$$X = F^{-1}(U) = -\frac{1}{log(U)}. \tag{19}$$

Once these samples are created, we calculate the various test statistics over them. We sort the 100.000 test statistics. For a 5% confidence level, the suitable critical value is then found at the 95-th quantile. We calculate this as the average of the 95000-nd and 95001-st sorted test statistic.
This whole simulation procedure is repeated for $k = 20, 30, \ldots, 200$. That is, we calculate the critical values of the different tests for various numbers of order statistics.

3.2 Power

In order to calculate the power of the different tests, we simulate under the alternative hypothesis. For this purpose we again use the same distributions as used in Koning and Peng (2008). We use the distributions $1 - F(x) = [1 - \delta log(x)]^{-\frac{1}{\delta}}$ where $x > 1$. We pair the same numbers of order statistics as before $k = 20, 30, \ldots, 200$ to various values of δ, such that $\delta\sqrt{k} = 1.5, 2, ..., 3.5$. An exact overview of all the values of δ is presented in the appendix. To create the random variables from these distributions we again use the inverse transformation method, where we use that $1 - F(x)$ is also on a $[0, 1]$ interval. This transformation is given by:

$$X = F^{-1}(U) = exp(\frac{U^{-\delta} - 1}{\delta}). \tag{20}$$

We simulate 10.000 samples from each of the distributions, each with size $n = 1000$. Over each of the samples we calculate the various test statistics and compare these to their respective critical values. The power of the test is then the fraction of the samples for which the null-hypothesis is rejected correctly in favor of the alternative.

4 Results

Our first results are concerned with the critical values of the various tests. They are summarized below in table 1. As our simulations are similar to those in Koning and Peng (2008), the critical values of the KS and SC test should be similar. This is indeed the case for the SC test. The majority of the critical values found in our simulations are up to two decimals the same. The slight deviations are dedicated to the randomness of simulation: although we simulate from the same distribution, the simulation is very unlikely to yield the exact same samples. On the other hand, the critical values for the KS test are all slightly lower. This is likely due to the discretisation of the test.
Two things worth noting from this table are the trend in the critical values of the AD and TR tests. The critical values of the AD test show a positive trend with the tail size. This is likely to be caused by the choice of weights, which also heavily depends on tail size. The critical values of the TR test on the other hand show a negative trend with the tail size. This is likely to be caused by the design of the statistic, which heavily depends on the number of order statistics.

Table 1: Critical values of KS, AD, SC and TR tests at 5% significance

k	KS	AD	SC	TR
20	0,952	4,229	1,375	4,600
30	0,980	5,280	1,334	4,125
40	0,989	6,154	1,320	3,842
50	0,998	6,919	1,312	3,660
60	1,005	7,608	1,314	3,521
70	1,010	8,239	1,318	3,412
80	1,012	8,825	1,320	3,325
90	1,015	9,374	1,325	3,252
100	1,017	9,893	1,330	3,195
110	1,023	10,386	1,337	3,143
120	1,026	10,857	1,343	3,094
130	1,030	11,308	1,350	3,052
140	1,036	11,741	1,357	3,013
150	1,040	12,160	1,365	2,982
160	1,046	12,564	1,376	2,952
170	1,052	12,956	1,385	2,925
180	1,057	13,336	1,394	2,899
190	1,061	13,706	1,406	2,876
200	1,066	14,066	1,415	2,855

The other results are concerned with the powers of the various tests. Tables 3 through 6 in the appendix summarize these. Again, we see that the powers we simulated for the SC tests are very close to those simulated in Koning and

Peng (2008). The small discrepancies are likely caused by difference in simulated samples. On the other hand, our discretised version of the KS test shows slightly higher power than the KS test conducted by Koning and Peng. This stands out, as the discretisation is a simplification of the test and one would expect the original to be more powerful.

As the power of each of the tests varies over both the number of order statistics k and δ, we compare general trends of each of the tests, varying these parameters. The KS test range from 0.2 to approximately 0.75. The lowest powers are found for the smallest value of $\delta\sqrt{k}$. The powers in general increase as $\delta\sqrt{k}$ increases. For a $\delta\sqrt{k} = 1.5, 2, 2.5$ the power of the test decreases as k increases. For $\delta\sqrt{k} = 3, 3.5$ there is no obvious decrease. At this level the highest powers are observed for the middle values of k, between $k = 70$ and $k = 130$.

The AD test has powers which start a bit higher than the KS test. Its lowest power is 0.32. On the other end its highest power is approximately 0.72, slightly lower than the highest power of the KS test. The AD test exhibits a similar pattern over the values of $\delta\sqrt{k}$ as the KS test: the power increases as $\delta\sqrt{k}$ increases. The trend over the values of k is different. For the smaller values, $k = 20, 30$, the power of the AD test is lower. Thereafter, the power stays rather constant for the various values of k up to $k = 200$, for a given $\delta\sqrt{k}$.

The powers of the SC test behave as the powers of both the KS and AD test in the sense that they increase in general as $\delta\sqrt{k}$ increases. However, trends in power are all over the place for given values of $\delta\sqrt{k}$. For $\delta\sqrt{k} = 1.5, 2, 2.5$ the test exhibits a negative trend as k increases. For $\delta\sqrt{k} = 3$, the power is rather constant over values of k. Finally, for $\delta\sqrt{k} = 3.5$ the test exhibits a positive trend. Of the tests we have discussed so far, this test has the lowest minimum power, less than 0.15. The highest power of this test is comparable to the other two: approximately 0.73.

The final test we discuss, the TR test, exhibits the greatest powers. They range from 0.37 all the way up to 0.99. The general trend of higher powers for higher values of $\delta\sqrt{k}$ also holds for this test. Given a value of $\delta\sqrt{k}$, the power of this test decreases as k increases. Of all tests that seem to exhibit this, this test has the steepest descent. But then again, its initial power for $k = 20$ is much bigger than the others'.

In figure 1 below, we present the powers of all the tests for $\delta\sqrt{k} = 2.5$. All the tests exhibit the trends as discussed earlier: the power of KS and SC decreases as k increases, the power of AD stays rather constant and the TR test has the highest power, but shows a steep decrease.

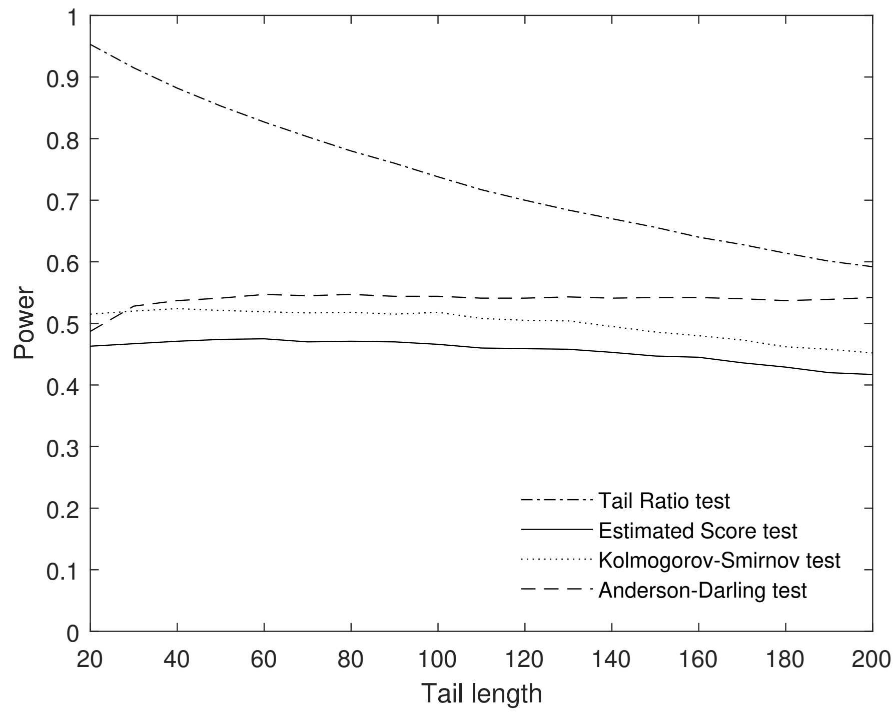

Figure 1: Power of KS, AD, SC and TR tests for $\delta\sqrt{k} = 2.5$

5 Conclusion

Heavy-tailed distributions find applications in various fields. Research into them has focussed on estimating these distributions, rather than testing for them. We discussed a framework for comparing empirical and theoretical distribution functions. In this framework we discussed three tests for the purpose of testing for a heavy tail: the Kolmogorov-Smirnov, Anderson-Darling and estimated score test. Then we discussed a test for the same purpose based only on the empirical distribution. We simulated under the null-hypothesis to find critical values for each of the tests. Finally, we simulated under the alternative to determine the tests' respective powers.

Now we have determined these powers, we can answer our research question: *How does the power of the TR test compare to the powers of the SC, KS and AD tests when testing H_0 against H_a?*

Our results showed that the TR test is more powerful than the SC, KS and AD tests for all configurations we examined. For the larger amounts of order statistics incorporated into the the test, the test is slightly more powerful than the others. However, this tests excels in small samples, as it exhibits powers of

up to 0.99.
Another interesting finding is the improvement of the AD test. By discretising the test, we have not only improved computability, but we also improved the test's power relative to the original.
We suggest further research to look into the small sample performance of the TR test: what causes the test's high power? Another question is whether the the TR test is a proper goodness-of-fit test. In this paper we have used the test only against a single alternative. A goodness-of-fit test should be an omnibus test: a test that works against all alternatives. The test can be used in a same setting, using another alternative distributions, to find out whether it is just powerful against this specific alternative or the alternative in general.

6 Appendix

Table 2: All δ used in various pairs of $\delta\sqrt{k}$

		$\delta\sqrt{k}$				
		1,5	2	2,5	3	3,5
	20	0,335	0,447	0,559	0,671	0,783
	30	0,274	0,365	0,456	0,548	0,639
	40	0,237	0,316	0,395	0,474	0,553
	50	0,212	0,283	0,354	0,424	0,495
	60	0,194	0,258	0,323	0,387	0,452
	70	0,179	0,239	0,299	0,359	0,418
	80	0,168	0,224	0,280	0,335	0,391
	90	0,158	0,211	0,264	0,316	0,369
	100	0,150	0,200	0,250	0,300	0,350
k	110	0,143	0,191	0,238	0,286	0,334
	120	0,137	0,183	0,228	0,274	0,320
	130	0,132	0,175	0,219	0,263	0,307
	140	0,127	0,169	0,211	0,254	0,296
	150	0,122	0,163	0,204	0,245	0,286
	160	0,119	0,158	0,198	0,237	0,277
	170	0,115	0,153	0,192	0,230	0,268
	180	0,112	0,149	0,186	0,224	0,261
	190	0,109	0,145	0,181	0,218	0,254
	200	0,106	0,141	0,177	0,212	0,247

Table 3: Power for KS test with 5% significance level using simulated critical values

		$\delta\sqrt{k}$				
		1,5	2,0	2,5	3,0	3,5
	20	0,284	0,404	0,515	0,613	0,700
	30	0,275	0,393	0,520	0,628	0,722
	40	0,271	0,396	0,524	0,642	0,738
	50	0,268	0,393	0,521	0,645	0,747
	60	0,260	0,389	0,519	0,645	0,749
	70	0,260	0,386	0,517	0,644	0,752
	80	0,251	0,383	0,518	0,644	0,751
	90	0,250	0,382	0,515	0,643	0,753
	100	0,243	0,376	0,518	0,644	0,751
k	110	0,233	0,368	0,508	0,641	0,752
	120	0,232	0,367	0,505	0,642	0,755
	130	0,233	0,361	0,504	0,636	0,753
	140	0,228	0,348	0,495	0,633	0,747
	150	0,222	0,344	0,486	0,623	0,739
	160	0,219	0,339	0,480	0,619	0,737
	170	0,208	0,328	0,473	0,610	0,730
	180	0,204	0,325	0,462	0,608	0,726
	190	0,201	0,322	0,458	0,603	0,726
	200	0,200	0,313	0,452	0,595	0,720

Table 4: Power for AD test with 5% significance level using simulated critical values

		$\delta\sqrt{k}$				
		1,5	2,0	2,5	3,0	3,5
	20	0,319	0,420	0,487	0,495	0,389
	30	0,329	0,434	0,528	0,592	0,616
	40	0,335	0,438	0,537	0,623	0,675
	50	0,334	0,441	0,541	0,630	0,697
	60	0,337	0,442	0,547	0,633	0,707
	70	0,336	0,442	0,545	0,636	0,712
	80	0,333	0,440	0,547	0,637	0,713
	90	0,332	0,439	0,544	0,638	0,715
	100	0,329	0,439	0,544	0,637	0,716
k	110	0,328	0,437	0,541	0,637	0,715
	120	0,329	0,436	0,541	0,639	0,718
	130	0,328	0,437	0,543	0,641	0,719
	140	0,326	0,439	0,541	0,640	0,718
	150	0,325	0,437	0,542	0,639	0,719
	160	0,327	0,435	0,542	0,638	0,718
	170	0,326	0,434	0,540	0,636	0,717
	180	0,326	0,434	0,537	0,638	0,717
	190	0,326	0,435	0,539	0,638	0,717
	200	0,327	0,434	0,542	0,637	0,716

Table 5: Power for SC test with 5% significance level using simulated critical values

		$\delta\sqrt{k}$				
		1,5	2	2,5	3	3,5
	20	0,251	0,355	0,463	0,558	0,642
	30	0,238	0,353	0,467	0,577	0,672
	40	0,224	0,345	0,471	0,582	0,686
	50	0,219	0,341	0,474	0,594	0,702
	60	0,214	0,337	0,475	0,604	0,714
	70	0,211	0,333	0,470	0,602	0,715
	80	0,203	0,331	0,471	0,602	0,720
	90	0,201	0,325	0,470	0,604	0,723
	100	0,191	0,320	0,466	0,605	0,725
k	110	0,187	0,315	0,460	0,607	0,726
	120	0,182	0,315	0,459	0,605	0,727
	130	0,183	0,309	0,458	0,603	0,728
	140	0,179	0,305	0,453	0,599	0,725
	150	0,174	0,303	0,447	0,593	0,726
	160	0,174	0,292	0,445	0,590	0,721
	170	0,167	0,286	0,436	0,583	0,713
	180	0,162	0,279	0,429	0,582	0,712
	190	0,154	0,276	0,420	0,574	0,708
	200	0,148	0,267	0,417	0,567	0,702

Table 6: Power for TR test with 5% significance level using simulated critical values

		$\delta\sqrt{k}$				
		1,5	2	2,5	3	3,5
	20	0,817	0,906	0,953	0,975	0,988
	30	0,747	0,854	0,915	0,953	0,972
	40	0,689	0,811	0,882	0,928	0,957
	50	0,645	0,774	0,853	0,905	0,939
	60	0,604	0,739	0,827	0,883	0,921
	70	0,570	0,709	0,803	0,863	0,905
	80	0,545	0,682	0,780	0,844	0,889
	90	0,521	0,659	0,760	0,827	0,875
	100	0,502	0,634	0,738	0,811	0,862
k	110	0,483	0,614	0,717	0,795	0,849
	120	0,467	0,596	0,700	0,780	0,835
	130	0,453	0,576	0,684	0,767	0,823
	140	0,437	0,561	0,670	0,753	0,812
	150	0,420	0,547	0,656	0,739	0,802
	160	0,411	0,536	0,640	0,726	0,790
	170	0,400	0,523	0,628	0,712	0,778
	180	0,389	0,513	0,614	0,702	0,770
	190	0,378	0,502	0,601	0,691	0,759
	200	0,368	0,491	0,592	0,679	0,748

References

Abidin, N. Z., Adam, M. B., and Midi, H. (2014). Hypothesis tests of goodness-of-fit for frechet distributions. *Sciences & Technology*, 22(2):419–432.

Anderson, T. and Darling, D. (1952). Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes. *The annals of mathematical statistics*, 23(2):193–212.

Beirlant, J., Goegebeur, Y., and de Wet, T. (2006). Bias-corrected goodness-of-fit tests for pareto-type behavior. *Journal of Computational and Applied Mathematics*, 186(1):99–116.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction.* Springer Series in Operations Research and Financial Enginering. Springer, New York.

de Haan, L. and Stadtmüller, U. (1996). Generalized regular variation of second order. *Journal of the Australian Mathematical Society (Series A)*, 61(03):381–395.

Goegebeur, Y. and Guillou, A. (2010). Goodness-of-fit testing for weibull-type behavior. *Journal of Statistical Planning and Inference*, 140(6):1417–1436.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The annals of statistics*, 3(5):1163–1174.

Hjort, N. and Koning, A. (2002). Tests for constancy of model parameters over time. *Journal of Nonparametric Statistics*, 14(1-2):113–132.

Hüsler, J. and Peng, L. (2008). Review of testing issues in extremes in honor of Professor Laurens de Haan. *Extremes*, 11:99–111.

Kolmogorov, A. (1933). Sulla determinazione empirica delle leggi di probabilita. *Giornale dell'Istituto Italiano degli Attuari*, 4:1–11.

Koning, A. and Peng, L. (2008). Goodness-of-fit tests for a heavy tailed distribution. *Journal of Statistical Planning and Inference*, 138(12):3960–3981.

Meyer, P. A. (1962). A decomposition theorem for supermartingales. *Illinois Journal of Mathematics*, 6(2):193–205.

Reittu, H. and Norros, I. (2004). On the power-law random graph model of massive data networks. *Performance Evaluation*, 55(1):3–23.

Rizzo, M. (2009). New goodness-of-fit tests for pareto distributions. *Astin Bulletin*, 39(2):691–715.

Smirnov, N. (1939). On the deviation of the empirical distribution function. *Rec. Math.[Mathematicheskii Sbornik] NS*, 6:3–26.