ISSN 1645-6726

INSTITUTO NACIONAL DE ESTATÍSTICA
PORTUGAL

REVSTAT
Statistical Journal

Volume 2, No. 1
June 2004

INDEX

REVSTAT – Statistical Journal
Volume 2, Number 1, June 2004

EXPONENTIALITY VERSUS GENERALIZED PARETO — A RESISTANT AND ROBUST TEST

Authors: M.F. Brilhante
– Department of Mathematics, University of Azores and C.E.A.U.L., Portugal (fbrilhante@notes.uac.pt)

Received: October 2003 Revised: December 2003 Accepted: January 2004

Abstract:

- Using resistant and robust methods we propose the statistic $T_n = (F_U - M)/(M - F_L)$ for testing exponentiality versus generalized Pareto, where F_U, F_L and M are, respectively, the upper and lower fourths and the median of a random sample of size n. The statistic T_n is based on the statistic $V_n = (X_{n:n} - M)/(M - X_{1:n})$ used by Gomes (1982) to discriminate extremal models in a similar context but with a higher breakdown point.
  The simulated power of T_n is compared with the simulated power of $U_n = X_{n:n}/M$ and V_n, which can also be used to test the exponential behaviour of the sample data. Although we observe that the power of T_n is lower than the power of U_n and V_n, we show that the performance of the first test is better than the performance of the two other tests when compared to broadened situations and mixtures commonly used to evaluate resistance and robustness.

Key-Words:

- *generalized Pareto distribution; breakdown point; resistance; robustness; broadened situations; mixtures.*

1. INTRODUCTION

Given the importance of the generalized Pareto distribution in Statistics (*e.g.*, analysis of POT data) we propose a test for testing exponentiality versus generalized Pareto which, although it is not the best one among possible tests, is however resistant to disturbing data, and robust in the sense that it is not sensitive to some departures of the assumptions inherent to a chosen probabilistic model.

Some statistics have been proposed to test the exponential behaviour of samples from generalized Pareto populations, specially for the von Mises–Jenkinson parametrization of the distribution, *i.e.*,

$$F_\beta(x) = 1 - \left(1 + \beta\,\frac{x}{\delta}\right)^{-1/\beta}, \qquad 1 + \beta\,\frac{x}{\delta} > 0, \quad x > 0\ ,$$

where $-\infty < \beta < \infty$ is a shape parameter and $\delta > 0$ a scale parameter. One test that can be used is based on the statistic $U_n = \frac{X_{n:n}}{M}$ (Gomes and van Monfort, 1987). Another possible test is based on $V_n = \frac{X_{n:n}-M}{M-X_{1:n}}$ which was used by Gomes (1982) to discriminate extremal models in a similar context. However, since U_n and V_n are both functions of extreme order statistics, they possess a disadvantage, a zero breakdown point, in the sense of Hampel, as defined below:

Definition 1.1. A statistic T has an α breakdown point $(0 \le \alpha \le 1)$ if the proportion of the sample data that can be replaced by arbitrarily other data with T remaining bounded approaches α.

As an alternative to the tests mentioned above we propose the test statistic

$$T_n = \frac{F_U - M}{M - F_L}\ ,$$

where F_U and F_L denote the upper and lower fourths and M the median of a random sample of size n, with a higher breakdown point (approximately equal to 0.25).

In section 2 we obtain the sample distribution of T_n under the null hypothesis $\beta = 0$ (*i.e.*, exponential behaviour) as well as the limiting distribution. In section 3 the power of the tests T_n, U_n and V_n are compared and the performance of each one is evaluated under broadened situations and mixtures in order to determine their resistance and robustness qualities.

2. SAMPLE DISTRIBUTION OF T_n UNDER THE HYPOTHESIS OF AN EXPONENTIAL PARENT

Let $(X_1, ..., X_n)$ be a random sample from an exponential distribution with distribution function

$$F_0(x) = \left(1 - e^{-x/\delta}\right) I_{]0,+\infty[} \ ,$$

$(\delta > 0)$ and let $(X_{1:n}, ..., X_{n:n})$ be the vector of ascending order statistics associated with the sample.

In order to preserve the ranking symmetry of the fourths from the extremes of the sample, we use the following definition for the $(100p)^{\text{th}}$ sample percentile

$$\xi_p = \begin{cases} X_{\{np\}:n} & \text{if } \ p < 0.5 \,, \\ X_{n-\{n(1-p)\}+1:n} & \text{if } \ p > 0.5 \ , \end{cases}$$

where $\{a\}$ denotes the number a rounded to the nearest integer in the usual way (cf. Casella and Berger, 2002).

Therefore, when n is odd

$$T_n = \frac{X_{n-\{\frac{n}{4}\}+1:n} - X_{\frac{n+1}{2}:n}}{X_{\frac{n+1}{2}:n} - X_{\{\frac{n}{4}\}:n}} \ ,$$

and when n is even

$$T_n = \frac{X_{n-\{\frac{n}{4}\}+1:n} - \frac{1}{2}\left(X_{\frac{n}{2}:n} + X_{\frac{n}{2}+1:n}\right)}{\frac{1}{2}\left(X_{\frac{n}{2}:n} + X_{\frac{n}{2}+1:n}\right) - X_{\{\frac{n}{4}\}:n}} \ .$$

The independence of the spacings of the exponential model yields the independence of the generalized spacings $X_{n-\{\frac{n}{4}\}+1:n} - X_{\frac{n+1}{2}:n}$ and $X_{\frac{n+1}{2}:n} - X_{\{\frac{n}{4}\}:n}$ when n is odd, and therefore the probability density function of T_n was obtained using standard techniques in this case.

When n is even we no longer have independence between numerator and denominator of T_n, and hence the expression that defines the probability density function was obtained calculating the marginal distribution of T_n from the joint probability distribution of the random vector $(X_{\{\frac{n}{4}\}:n},\ X_{n/2:n} - X_{\{\frac{n}{4}\}:n},\ X_{n/2+1:n} - X_{n/2:n},\ T_n)$.

Consequently, if n is odd, the density function is defined by

$$f(t) = \frac{\left(n-\{\frac{n}{4}\}\right)!}{\left(\{\frac{n}{4}\}-1\right)!\,\left(\frac{n-1}{2}-\{\frac{n}{4}\}\right)!^{2}}$$

$$\times \sum_{i=0}^{\frac{n-1}{2}-\{\frac{n}{4}\}} \binom{\frac{n-1}{2}-\{\frac{n}{4}\}}{i} (-1)^i\, B\left(\frac{n+1}{2}+\left(i+\{\tfrac{n}{4}\}\right)t\,,\ \frac{n+1}{2}-\{\tfrac{n}{4}\}\right)$$

$$\times \sum_{j=1}^{\frac{n+1}{2}-\{\frac{n}{4}\}} \frac{1}{n-\{\frac{n}{4}\}+1-j+\left(i+\{\frac{n}{4}\}\right)t}\,, \qquad t>0\,,$$

where $B(\cdot\,,\cdot)$ represents the beta function; and, if n is even,

$$f(t) = \frac{\left(n-\{\frac{n}{4}\}\right)!}{\left(\{\frac{n}{4}\}-1\right)!\,\left(\frac{n}{2}-\{\frac{n}{4}\}-1\right)!^{2}}$$

$$\times \sum_{i,j=0}^{\frac{n}{2}-\{\frac{n}{4}\}-1} \binom{\frac{n}{2}-\{\frac{n}{4}\}-1}{i}\binom{\frac{n}{2}-\{\frac{n}{4}\}-1}{j}$$

$$\times (-1)^{i+j}\,\frac{2\left(n+2j+2\right)t \;+\; \left(n+2i-2j-2+2\{\frac{n}{4}\}\right)t^2}{\left[\frac{n}{2}+j+1+\left(i+\{\frac{n}{4}\}\right)t\right]^2\left[\frac{n}{2}+j+1+\left(\frac{n}{2}-j-1\right)t\right]^2}$$

if $0<t<1$, and

$$f(t) = \frac{\left(n-\{\frac{n}{4}\}\right)!}{\left(\{\frac{n}{4}\}-1\right)!\,\left(\frac{n}{2}-\{\frac{n}{4}\}-1\right)!^{2}}$$

$$\times \sum_{i=0}^{\frac{n}{2}-\{\frac{n}{4}\}-1} \binom{\frac{n}{2}-\{\frac{n}{4}\}-1}{i} (-1)^i\,\frac{2\,B\left(\frac{n}{2}+\left(i+\{\frac{n}{4}\}\right)t+1\,,\ \frac{n}{2}-\{\frac{n}{4}\}\right)}{n-\{\frac{n}{4}\}-i+\left(i+\{\frac{n}{4}\}\right)t}$$

$$\times \left\{\sum_{j=1}^{\frac{n}{2}-\{\frac{n}{4}\}} \frac{1}{n-\{\frac{n}{4}\}-j+1+\left(i+\{\frac{n}{4}\}\right)t} + \frac{1}{n-\{\frac{n}{4}\}-i+\left(i+\{\frac{n}{4}\}\right)t}\right\},$$

if $t\geq 1$.

For larger sample sizes we can use the normal distribution as an approximation to the distribution of T_n. In order to prove that T_n has a limiting normal distribution under the hypothesis of an exponential parent we consider the following lemma (cf. Chernoff *et al.*, 1967).

Lemma 2.1. *Let $Z_{i:n}$ be the i^{th} ascending order statistic of n i.i.d. standard exponential random variables $Z_1, ..., Z_n$. Then,*

$$P\left\{\tau_n^{-1} \sum_{i=1}^{n} a_{i:n}(Z_{i:n} - \mu_{i:n}) \le t\right\} \xrightarrow[n\to+\infty]{} \Phi(t) ,$$

for every t, if and only if,

$$\max_{1\le j\le n} \tau_n^{-1} |b_{j,n}| \xrightarrow[n\to+\infty]{} 0 ,$$

where $\mu_{i:n} = E(Z_{i:n})$, $b_{j,n} = (n-j+1)^{-1} \sum_{i=j}^{n} a_{i:n}$ *and* $\tau_n^2 = \sum_{i=1}^{n} b_{i,n}^2$.

Using the lemma's notation for $X_{n-\{\frac{n}{4}\}+1:n} - X_{\frac{n+1}{2}:n}$ we have

$$a_{i:n} = \begin{cases} -1 , & i = \frac{n+1}{2} , \\ 1 , & i = n-\{\frac{n}{4}\}+1 , \\ 0 , & \text{elsewhere} , \end{cases}$$

and

$$b_{j,n} = \begin{cases} 0 , & 1 \le j \le \frac{n+1}{2} , \\ \dfrac{1}{n-j+1} , & \frac{n+1}{2}+1 \le j \le n-\{\frac{n}{4}\}+1 , \\ 0 , & n-\{\frac{n}{4}\}+2 \le j \le n . \end{cases}$$

Hence,

$$\tau_n^2 = \sum_{j=\frac{n+1}{2}+1}^{n-\{\frac{n}{4}\}+1} \frac{1}{(n-j+1)^2} = \sum_{k=\{\frac{n}{4}\}}^{\frac{n-1}{2}} \frac{1}{k^2} .$$

Applying the lemma we get

$$\tau_n^{-1}\Big(X_{n-\{\frac{n}{4}\}+1:n} - X_{\frac{n+1}{2}:n} - \lambda_n\Big) \xrightarrow[n\to+\infty]{d} Z \frown \text{Normal}(0,1) ,$$

where

$$\lambda_n = \mu_{n-\{\frac{n}{4}\}+1:n} - \mu_{\frac{n+1}{2}:n} = \sum_{k=\{\frac{n}{4}\}}^{\frac{n-1}{2}} \frac{1}{k} ,$$

and since $\frac{\ln(3/2)}{X_{\frac{n+1}{2}:n} - X_{\{\frac{n}{4}\}:n}}$ converges in probability to 1, it follows from Slutsky's theorem that

$$\ln(3/2)\; \tau_n^{-1}\left(T_n - \frac{\lambda_n}{\ln(3/2)}\right) \xrightarrow[n\to+\infty]{d} Z \frown \text{Normal}(0,1) .$$

However, simpler normalizing constants can be found. Since $\lambda_n \sim \ln 2$ and $\tau_n^2 \sim \frac{2}{n}$ we have

$$\ln(3/2)\sqrt{\frac{n}{2}}\left(T_n - \frac{\ln 2}{\ln(3/2)}\right) \xrightarrow[n\to\infty]{d} Z \frown \text{Normal}(0,1) .$$

On the other hand, it is quite straightforward to show that

$$\ln 2\; U_n - \ln n \xrightarrow[n\to\infty]{d} Y \frown \text{Gumbel}(0,1)$$

and

$$\ln 2\; V_n - \ln(n/2) \xrightarrow[n\to\infty]{d} Y \frown \text{Gumbel}(0,1) .$$

3. POWER, RESISTANCE AND ROBUSTNESS COMPARISON

The choice of the appropriate statistical test for a particular situation must be guided by a sensible criteriom. Usually, power considerations weight considerably in the decision process.

For inference purposes and comparison of the power functions we register on Table 1 the simulated critical points (based on 4999 simulations) of the sample distributions of T_n, U_n and V_n under the null hypothesis $\beta = 0$.

Table 1: Simulated critical points of T_n, U_n and V_n.

	α							
	.01	.025	.05	.1	.9	.95	.975	.99
T_{10}	.21*	.30*	.40*	.56*	4.90*	6.88*	9.37*	13.65*
U_{10}	1.53	1.70	1.87	2.12	7.62	9.53	11.68	14.97
V_{10}	.59	.79	1.00	1.29	8.00	10.38	12.99	18.07
T_{20}	.43*	.54*	.66*	.83*	3.89*	4.89*	5.97*	7.59*
U_{20}	2.13	2.37	2.65	2.98	8.50	10.08	11.82	13.95
V_{20}	1.18	1.47	1.75	2.11	8.23	10.02	11.95	14.72
T_{30}	.50*	.61*	.72*	.88*	3.30*	4.00*	4.74*	5.80*
U_{30}	2.53	2.91	3.17	3.54	9.11	10.47	12.05	13.87
V_{30}	1.62	1.96	2.25	2.66	8.57	10.10	11.64	13.57
T_{50}	.67	.79	.89	1.03	2.86	3.31	3.80	4.50
U_{50}	3.24	3.49	3.85	4.23	9.55	11.02	12.24	14.12
V_{50}	2.29	2.58	2.91	3.33	8.83	10.35	11.77	13.67
T_{100}	.87	.99	1.09	1.21	2.47	2.75	3.02	3.32
U_{100}	4.19	4.53	4.83	5.24	10.46	11.56	12.63	14.33
V_{100}	3.26	3.57	3.88	4.30	9.58	10.73	11.79	13.46
T_{250}	1.12	1.20	1.28	1.36	2.17	2.31	2.44	2.57
U_{250}	5.49	5.89	6.21	6.68	11.51	12.60	13.71	15.09
V_{250}	4.52	4.92	5.25	5.72	10.58	11.66	12.79	14.22

(* exact critical points)

Figures 1 and 2 show the simulated power functions (5 000 simulations) of T_n, U_n and V_n for a $\alpha = 0.05$ level right one-sided and two-sided tests and $n = 10, 20, 30, 50, 100, 250$.

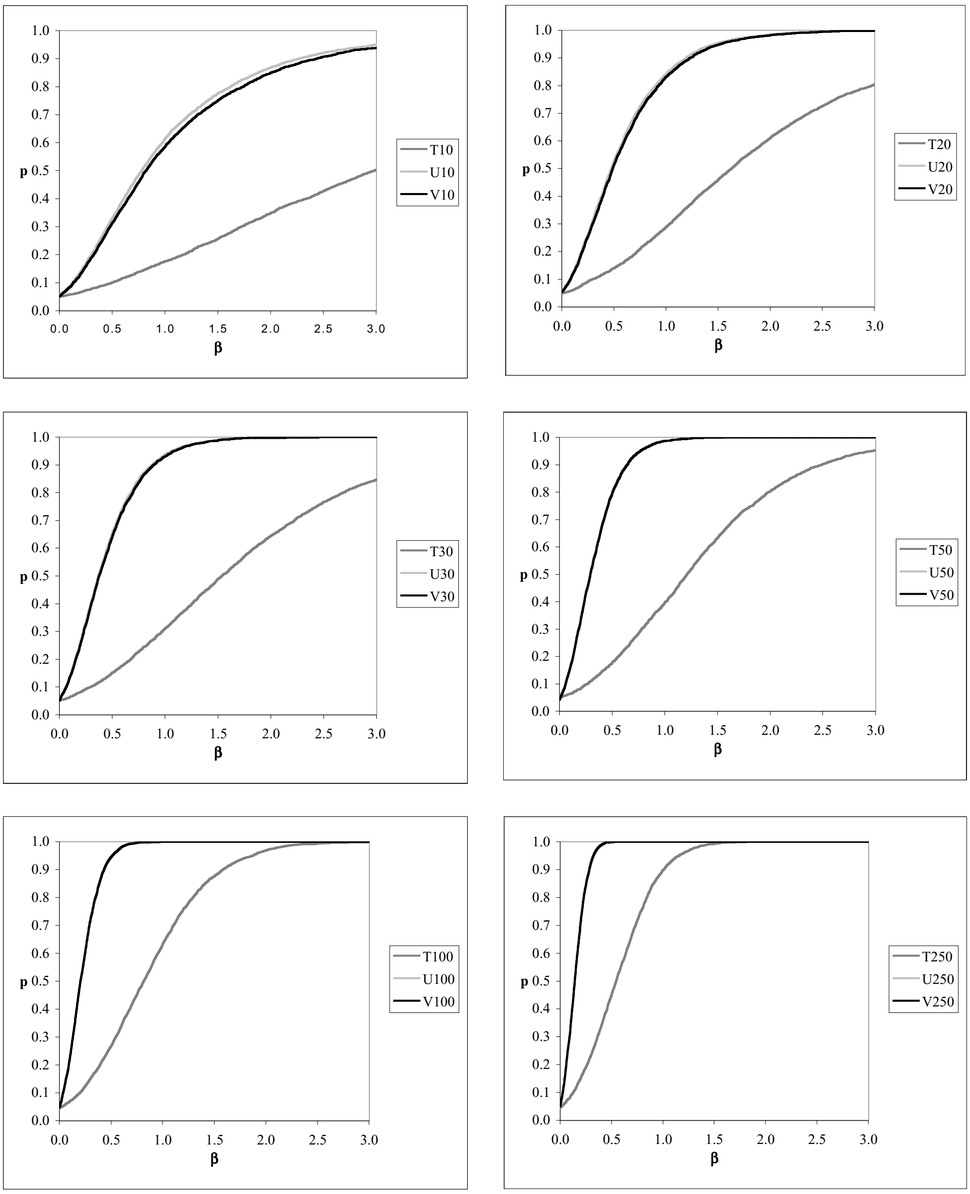


Figure 1: Power functions for a $\alpha = 0.05$ level right one-sided test.

From Figures 1 and 2 we observe that T_n performs quite badly in detecting departures from the exponential behaviour when compared with the other two tests. If we had to choose based exclusively on the power of the test, we would choose for smaller sample sizes ($n \leq 30$) U_n and for larger sample sizes U_n or V_n.

A comparison of the power of the three tests was also made for a $\alpha = 0.01$ level one-sided and two-sided tests, but the results are not presented here because they reveal a similar pattern as in the case $\alpha = 0.05$.

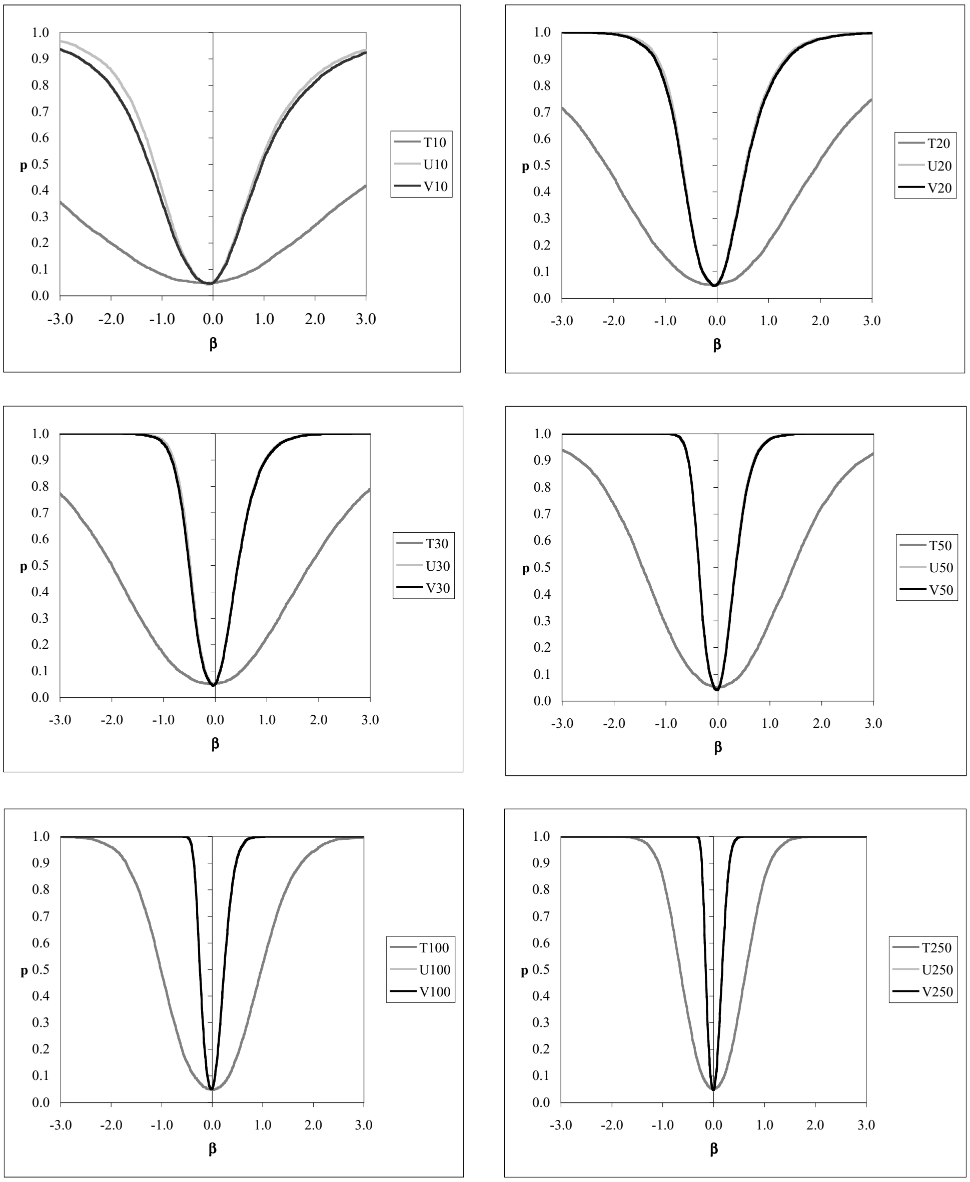


Figure 2: Power functions for a $\alpha = 0.05$ level two-sided test.

The power criteria can be pushed to a second place if we find stronger reasons which can sustain such a decision. In fact, the lesser power of T_n will be in a way compensated when we evaluate its performance after introducing disturbing observations in the sample (*e.g.*, an observation from an exponential population with a larger scale), or when we consider the sample from a mixture of exponentials, which is a variation (or contamination) of the "pure" exponential model. In order to compare the resistance and robustness of the tests we will evaluate the performance of each one in a broadened situation and mixture.

In an one broadened situation, which is the case presented here, we assume that the margins of the random vector $(X_1, ..., X_{n-1}, X^*)$ are independent, $X_1, ..., X_{n-1}$ are standard exponentials and X^* is an exponential variable with distribution function

$$F^*(x) = \left(1 - e^{-\frac{x+K-1}{K}}\right) I_{]1-K,+\infty[} \ .$$

In a mixture situation we assume that the random sample $(X_1, ..., X_n)$ is from a population with distribution function

$$F(x) \;=\; \left[(1-\theta)\left(1-e^{-(x-\theta(1-K))}\right) + \theta\left(1-e^{-\frac{x-\theta(1-K)}{K}}\right)\right] I_{]\theta(1-K),+\infty[} \ ,$$

where $0 < \theta < 1$ (θ is sometimes called the percentage contamination).

In robustness studies it is usual to consider $K = 3, 10$ and $\theta = 0.05, 0.1$ (cf. Hoaglin *et al.*, 1983). However, we will only show the results obtained for $K = 3, 10$ and $\theta = 0.05$, and for the classical level $\alpha = 0.05$.

In Tables 2 to 5 we indicate the probability of rejecting the exponential hypothesis, as well as the standard error of the estimates and the corresponding 95% confidence interval.

Table 2: Right one-sided test in an one broadened situation.

$K = 3$									
n	T_n	*s.e.*	95% C.I.	U_n	*s.e.*	95% C.I.	V_n	*s.e.*	95% C.I.
10	.050	.0031	[.044, .056]	.116	.0045	[.107, .125]	.070	.0036	[.063, .077]
20	.051	.0031	[.045, .057]	.099	.0042	[.091, .107]	.063	.0034	[.056, .070]
30	.048	.0030	[.042, .054]	.093	.0041	[.085, .101]	.067	.0035	[.060, .074]
50	.051	.0031	[.045, .057]	.086	.0040	[.078, .094]	.063	.0034	[.056, .070]
100	.053	.0032	[.047, .059]	.077	.0038	[.070, .084]	.053	.0032	[.047, .059]
250	.049	.0031	[.043, .055]	.079	.0038	[.072, .086]	.053	.0032	[.047, .059]

$K = 10$									
n	T_n	*s.e.*	95% C.I.	U_n	*s.e.*	95% C.I.	V_n	*s.e.*	95% C.I.
10	.055	.0032	[.049, .061]	.248	.0061	[.236, .260]	.186	.0055	[.175, .197]
20	.053	.0032	[.047, .059]	.239	.0060	[.227, .251]	.193	.0056	[.182, .204]
30	.050	.0031	[.044, .056]	.238	.0060	[.226, .250]	.203	.0057	[.192, .214]
50	.050	.0031	[.044, .056]	.226	.0059	[.214, .238]	.196	.0056	[.185, .207]
100	.054	.0032	[.048, .060]	.215	.0058	[.204, .226]	.185	.0055	[.174, .196]
250	.049	.0031	[.043, .055]	.206	.0057	[.195, .217]	.175	.0054	[.164, .186]

Table 3: Two-sided test in an one broadened situation.

$K = 3$									
n	T_n	$s.e.$	95% C.I.	U_n	$s.e.$	95% C.I.	V_n	$s.e.$	95% C.I.
10	.051	.0031	[.045, .057]	.085	.0039	[.077, .093]	.186	.0055	[.175, .197]
20	.052	.0031	[.046, .058]	.073	.0037	[.066, .080]	.273	.0063	[.261, .285]
30	.050	.0031	[.044, .056]	.078	.0038	[.071, .085]	.339	.0067	[.326, .352]
50	.054	.0032	[.048, .060]	.068	.0036	[.061, .075]	.381	.0069	[.368, .394]
100	.054	.0032	[.048, .060]	.070	.0036	[.063, .077]	.428	.0070	[.414, .442]
250	.056	.0033	[.050, .062]	.073	.0037	[.066, .080]	.467	.0071	[.453, .481]

$K = 10$									
n	T_n	$s.e.$	95% C.I.	U_n	$s.e.$	95% C.I.	V_n	$s.e.$	95% C.I.
10	.053	.0032	[.047, .059]	.205	.0057	[.194, .216]	.645	.0068	[.632, .658]
20	.052	.0031	[.046, .058]	.205	.0057	[.194, .216]	.701	.0065	[.688, .714]
30	.050	.0031	[.044, .056]	.208	.0057	[.197, .219]	.711	.0064	[.698, .724]
50	.054	.0032	[.048, .060]	.198	.0056	[.187, .209]	.723	.0063	[.711, .735]
100	.051	.0031	[.045, .057]	.203	.0057	[.192, .214]	.738	.0062	[.726, .750]
250	.056	.0033	[.050, .062]	.194	.0056	[.183, .205]	.750	.0061	[.738, .762]

Table 4: Right one-sided test in a 5% contamination situation.

$K = 3$									
n	T_n	$s.e.$	95% C.I.	U_n	$s.e.$	95% C.I.	V_n	$s.e.$	95% C.I.
10	.055	.0032	[.049, .061]	.142	.0049	[.132, .152]	.086	.0040	[.078, .094]
20	.056	.0033	[.050, .062]	.188	.0055	[.177, .199]	.119	.0046	[.110, .128]
30	.053	.0032	[.047, .059]	.217	.0058	[.206, .228]	.138	.0049	[.128, .148]
50	.058	.0033	[.052, .064]	.268	.0063	[.256, .280]	.182	.0055	[.171, .193]
100	.062	.0034	[.055, .069]	.405	.0069	[.391, .419]	.285	.0064	[.272, .298]
250	.064	.0035	[.057, .071]	.619	.0069	[.606, .632]	.471	.0071	[.457, .485]

$K = 10$									
n	T_n	$s.e.$	95% C.I.	U_n	$s.e.$	95% C.I.	V_n	$s.e.$	95% C.I.
10	.062	.0034	[.055, .069]	.459	.0070	[.445, .473]	.231	.0060	[.219, .243]
20	.069	.0036	[.062, .076]	.685	.0066	[.672, .698]	.382	.0069	[.369, .395]
30	.059	.0033	[.052, .066]	.816	.0055	[.805, .827]	.501	.0071	[.487, .515]
50	.065	.0035	[.058, .072]	.921	.0038	[.914, .928]	.683	.0066	[.670, .696]
100	.074	.0037	[.067, .081]	.992	.0013	[.990, .994]	.887	.0045	[.878, .896]
250	.092	.0041	[.084, .100]	1.000	.0000	—	.994	.0011	[.992, .996]

The analysis of the previous tables show that T_n is by far less sensitive to the disturbing observation, even when it comes from an exponential population with a standard deviation ten times greater than the standard deviation of the standard exponential. In other words this means that with T_n we will be rejecting a true null hypothesis with probability approximately equal to $\alpha = 0.05$. The same can be said when we consider that 5% of the observations are from an exponential population with standard deviation $K = 10$. Therefore the results confirm that T_n is more resistant and robust.

Table 5: Two-sided test in a 5% contamination situation.

$K=3$									
n	T_n	*s.e.*	95% C.I.	U_n	*s.e.*	95% C.I.	V_n	*s.e.*	95% C.I.
10	.050	.0031	[.044, .056]	.106	.0044	[.097, .115]	.074	.0037	[.067, .081]
20	.053	.0032	[.047, .059]	.139	.0049	[.129, .149]	.093	.0041	[.085, .101]
30	.051	.0031	[.045, .057]	.160	.0052	[.150, .170]	.112	.0045	[.103, .121]
50	.052	.0031	[.046, .058]	.212	.0058	[.201, .223]	.143	.0050	[.133, .153]
100	.052	.0031	[.046, .058]	.333	.0067	[.320, .346]	.234	.0060	[.222, .246]
250	.050	.0031	[.044, .056]	.528	.0071	[.514, .542]	.385	.0069	[.372, .398]

$K=10$									
n	T_n	*s.e.*	95% C.I.	U_n	*s.e.*	95% C.I.	V_n	*s.e.*	95% C.I.
10	.052	.0031	[.046, .058]	.545	.0070	[.531, .559]	.205	.0057	[.194, .216]
20	.053	.0032	[.047, .059]	.704	.0065	[.691, .717]	.345	.0067	[.332, .358]
30	.051	.0031	[.045, .057]	.805	.0056	[.794, .816]	.462	.0071	[.448, .476]
50	.054	.0032	[.048, .060]	.913	.0040	[.905, .921]	.645	.0068	[.632, .658]
100	.054	.0032	[.048, .060]	.989	.0015	[.986, .992]	.864	.0048	[.854, .874]
250	.060	.0034	[.053, .067]	1.000	.0000	—	.990	.0014	[.987, .993]

4. FINAL COMMENTS

It is important to use resistant and robust methods given the fact that: (i) classical techniques behave poorly when the general situation departs from the set of initial assumptions; (ii) in practice we never know the exact underlying conditions, specially when it is not so unlikely to admit the existence of disturbing data in the sample.

The conclusions of section 3 reinforce the general idea that resistant and robust methods are the best compromise possible for a large set of scenarios, although not necessarily the best ones for a very specific and limiting situation.

The analysis of the power function shows that the extreme order statistics carry important information for the issue at hand, and therefore trimming out 25% of the sample data may be too drastic. Unfortunately, there is no rule of thumb for an appropriate choice k in $T_{n(k)} = \frac{X_{n-k(n)+1:n}-M}{M-X_{k(n):n}}$ that optimizes results in what concerns power *and* resistance *and* robustness altogether.

ACKNOWLEDGMENTS

This work has been partially supported by FCT/POCTI/FEDER (VEXTRA Project).

The author also acknowledges the suggestions from the referees.

REFERENCES

[1] Abramowitz, M. and Stegun, I.A. (1970). *Handbook of Mathematical Functions*, Dover, New York.

[2] Casella, G. and Berger, R.L. (2002). *Statistical Inference*, 2nd Ed., Duxbury Pacific Grove.

[3] Chernoff, H.; Gastwirth, J.L. and John, M.V. (1967). Asymptotic distribution of linear combinations of order statistics with applications to estimation, *Annals of Mathematical Statistics*, **38**, 52–72.

[4] Gomes, M.I. (1982). A note on statistical choice of extremal models, *Actas IX Jornadas Mat. Hispano-Lusas*, Salamanca, 653–655.

[5] Gomes, M.I. and van Monfort, M.A.J. (1987). Exponentiality versus generalized Pareto – quick tests, *Statistical Climatology*, **87**, 185–195.

[6] Hampel, F.R. (1971). A general qualitative definition of robustness, *Annals of Mathematical Statistics*, **42**, 1887–1896.

[7] Hoaglin, D.C.; Mosteller, F. and Tukey, J.W. (1983). *Understanding Robust and Exploratory Data Analysis*, John Wiley & Sons, New York.

REVSTAT – Statistical Journal
Volume 2, Number 1, June 2004

SIMULTANEOUS TAIL INDEX ESTIMATION

Authors: Jan Beirlant
– Department of Mathematics, Katholieke Universiteit Leuven, Celestijnenlaan 200B, B-3000 Leuven, Belgium
(jan.beirlant@wis.kuleuven.ac.be)

Yuri Goegebeur
– Postdoctoral researcher Fund for Scientific Research – Flanders, Department of Applied Economics, Katholieke Universiteit Leuven, Naamsestraat 69, B-3000 Leuven,
University Center for Statistics, Katholieke Universiteit Leuven, W. de Croylaan 54, B-3001 Heverlee, Belgium
(yuri.goegebeur@econ.kuleuven.ac.be)

Received: November 2003 Revised: January 2004 Accepted: January 2004

Abstract:

- The estimation of the extreme-value index γ based on a sample of independent and identically distributed random variables has received considerable attention in the extreme-value literature. However, the problem of combining data from several groups is hardly studied. In this paper we discuss the simultaneous estimation of tail indices when data on several independent data groups are available. The proposed methods are based on regression models linking tail related statistics to the extreme-value index and parameters describing the second order tail behaviour. For heavy-tailed distributions ($\gamma > 0$), estimators are derived from an exponential regression model for rescaled log-spacings of successive order statistics as described in Beirlant *et al.* (1999) and Feuerverger and Hall (1999). Estimators for $\gamma \in \mathbb{R}$ are obtained using the linear model for *UH*-statistics given in Beirlant *et al.* (2000). In both cases, the optimal number of extremes to be used in the estimation is derived from the asymptotic mean squared error matrix.

Key-Words:

- *extreme-value index; regression; quantile-quantile plot.*

AMS Subject Classification:

- 62G30, 62G32.

1. INTRODUCTION

A central topic in extreme-value theory which continues to receive considerable attention is the estimation of the extreme-value index γ. This index is directly related to the tail of a distribution function F with the tail function $1-F$ becoming more heavy as γ increases. The extreme-value index γ can be estimated from a parametric or a semi-parametric point of view.

Parametric approaches are based on limit theorems which form the core of the extreme-value theory. Consider $X_1, ..., X_n$ independent and identically distributed random variables and let $X_{1,n} \leq ... \leq X_{n,n}$ denote the corresponding ascending order statistics. A first possibility is based on the following result of Fisher and Tippett (1928). If for a distribution function F_X there exist sequences of constants $(a_n > 0)_n$ and $(b_n)_n$ such that

$$\lim_{n\to\infty} P\left(\frac{X_{n,n}-b_n}{a_n} \leq x\right) = \lim_{n\to\infty} F_X^n(a_n\, x + b_n) \;=\; H(x) \tag{1.1}$$

at all continuity points of H, with H a nondegenerate distribution, then F_X is said to belong to the domain of (maximum) attraction of H, denoted $F_X \in \mathcal{D}(H)$. Moreover, it is known that if such a nondegenerate limit distribution H exists, it should be of the form

$$H_\gamma(x) \;=\; \begin{cases} \exp\left(-\left(1+\gamma\,\dfrac{x-\mu}{\sigma}\right)^{-\frac{1}{\gamma}}\right) & 1+\gamma\,\frac{x-\mu}{\sigma} > 0, \quad \gamma \neq 0 \\[2ex] \exp\left(-\exp\left(-\dfrac{x-\mu}{\sigma}\right)\right) & x \in \mathbb{R}, \quad \gamma = 0 \end{cases} \tag{1.2}$$

with $\mu \in \mathbb{R}$ and $\sigma > 0$. This limit distribution is the so-called generalized extreme-value distribution (GEV). Note that the extreme-value index γ appears as a shape parameter in (1.2). Based on this result, γ can be estimated by fitting (1.2) to sample maxima (Gumbel, 1967). A second possibility is based on the generalized Pareto distribution (GPD) given by

$$G_\gamma(x) \;=\; \begin{cases} 1-\left(1+\gamma\,\dfrac{x}{\sigma}\right)^{-\frac{1}{\gamma}} & \gamma \neq 0 \\[2ex] 1-\exp\left(-\dfrac{x}{\sigma}\right) & \gamma = 0 \end{cases} \tag{1.3}$$

with $\sigma > 0$ and with $x > 0$ if $\gamma \geq 0$, $0 < x < -\sigma/\gamma$ if $\gamma < 0$, which is fitted to exceedances over a specified threshold u (Pickands, 1975, Smith, 1985, 1987).

Next to the above described parametric approaches, γ can also be estimated semi-parametrically. Define the tail quantile function U_X as $U_X(x) = \inf\left\{y\colon\, F_X(y) \geq 1-\frac{1}{x}\right\}$, $x > 1$. For Pareto-type or heavy tailed distributions

$(\gamma > 0)$ we have

$$F_X \in \mathcal{D}(H_\gamma) \iff 1 - F_X(x) = x^{-\frac{1}{\gamma}}\,\tilde{l}(x)\,, \quad x > 0 \tag{1.4}$$
$$\iff U_X(x) = x^\gamma\, l(x)\,, \quad x > 1$$

with $\tilde{l}$, l slowly varying functions at infinity i.e. positive functions g such that

$$\frac{g(\lambda x)}{g(x)} \to 1 \quad \text{as} \quad x \to \infty \qquad \forall\, \lambda > 0\ . \tag{1.5}$$

The conditions given in (1.4) characterize completely the first order behavior of F_X. The Pareto quantile plot can be found in the literature as the basis for evaluating the goodness-of-fit hypothesis of strict Pareto behavior. For the strict Pareto distribution $\log U(x) = \gamma \log(x)$ so the log theoretical quantiles stand in linear relationship with $\log(x)$. Replacing the theoretical quantiles $\log U(\frac{n+1}{j})$ by their empirical counterparts $\log U_n(\frac{n+1}{j}) = \log X_{n-j+1,n}$, the coordinates of the points on the quantile plot are given by

$$\left(\log\left(\frac{n+1}{j}\right), \log X_{n-j+1,n}\right) \qquad j = 1, ..., n\ . \tag{1.6}$$

In case of a good fit of the strict Pareto distribution to the data, the points on the Pareto quantile plot should show a straight line pattern. Moreover, the slope of a line through the origin fitted to the Pareto quantile plot will estimate γ. In case the distribution of the data is of Pareto-type, the log-tail quantile function can be written as $\log U(x) = \gamma \log x + \log l(x)$. Since $\log l(x)/\log x \to 0$ as $x \to \infty$ we have that $\log U(x) \sim \gamma \log x$ as $x \to \infty$, so the Pareto quantile plot will be *ultimately* linear. Again, the slope of the linear part will approximate γ. Several well known estimators for γ can be interpreted as estimators of the slope of the linear part of the Pareto quantile plot. For instance, the Hill (1975) estimator given by

$$H_{k,n} = \frac{1}{k}\sum_{j=1}^{k} \log X_{n-j+1,n} - \log X_{n-k,n} \qquad k = 1, ..., n-1 \tag{1.7}$$

clearly measures the average increase of the Pareto quantile plot to the right of the anchor point $\big(\log(\frac{n+1}{k+1}), \log X_{n-k,n}\big)$. Other important estimators for $\gamma > 0$ are the so-called kernel estimators derived by Csörgő *et al.* (1985) and the least squares estimators proposed by Kratz and Resnick (1996) and Schultze and Steinebach (1996) among others.

The estimation of $\gamma \in \mathbb{R}$ has been studied less extensively. In this paper we will concentrate on the approach based on the generalized quantile plot described in Beirlant *et al.* (1996) and Beirlant *et al.* (2000). For a positive random variable X, consider

$$H_X(x) = \mathrm{E}\Big(\log X - \log U_X(x) \mid X > U_X(x)\Big) \qquad x > 1\ ,$$

the mean residual life function of the log-transformed data, and define the adapted mean excess function UH_X as

$$UH_X(x) = U_X(x)\,H_X(x)$$
$$= U_X(x)\int_1^{\infty}\Big(\log U_X(zx) - \log U_X(x)\Big)\,\frac{dz}{z^2} \qquad x > 1\ . \tag{1.8}$$

In Theorem 1 of Beirlant *et al.* (1996) it is shown that $F_X \in \mathcal{D}(H_\gamma),\ \gamma \in \mathbb{R}$, implies that

$$\log UH_X(x) = \gamma \log x + \log \breve{l}(x) \tag{1.9}$$

with $\breve{l}$ denoting a slowly varying function at infinity. As a consequence, $\log UH_X(x) \sim \gamma \log x$ for $x \to \infty$. Consider $X_1, ..., X_n$ independent and identically distributed positive random variables. Replacing U_X and H_X in (1.8) by their empirical counterparts yields

$$UH_{j,n} = \widehat{UH}_X\left(\frac{n}{j}\right) = X_{n-j,n}\left(\frac{1}{j}\sum_{i=1}^{j}\log X_{n-i+1,n} - \log X_{n-j,n}\right) \tag{1.10}$$

as sample versions for $UH_X(n/j)$, $j = 1, ..., n-1$. For $F_X \in \mathcal{D}(H_\gamma)$, the generalized quantile plot, defined by

$$\left(\log\frac{n}{j},\ \log UH_{j,n}\right) \qquad j = 1, ..., n-1\ , \tag{1.11}$$

should be ultimately linear in the smaller j-values. Further, the slope of the straight line behind the linear part of the generalized quantile plot is the unknown γ. Applying a Hill-type operation on $UH_{j,n},\ j = 1, ..., k$, results in the following estimator for γ, called the adapted Hill estimator $H_{k,n}^{*2}$,

$$H_{k,n}^{*2} = \frac{1}{k}\sum_{j=1}^{k}\log UH_{j,n} - \log UH_{k+1,n}\ . \tag{1.12}$$

Other well known estimators for $\gamma \in \mathbb{R}$ have been proposed by Pickands (1975) and Dekkers *et al.* (1989).

As is clear from the above discussion, the literature on extreme-value methods for a sample of independent and identically distributed data is quite elaborate. However, the problem of combining data from different independent data groups is hardly studied. Nevertheless, the problem is important: consider for instance the combination of earthquake data from different geographical regions such as subduction zones and midocean ridge zones. Often the amount of available data is small and hence the combination of different samples is important in order to gain efficiency.

Of course, regression models with dummy explanatory variables describing the groups, can be used in combination with classical extreme-value models such as the generalized extreme-value distribution (1.2), which is fitted to maxima, or the generalized Pareto distribution (1.3), which is fitted to so-called peaks (or excesses) over threshold data. This approach can be found for instance in Davison and Smith (1990). A major difficulty when working with the GPD in a regression setting is the selection of the threshold. Ideally, the threshold should depend on the covariates in order to take the relative extremity of the observations into account. This issue was also noted by Davison and Smith (1990) and Coles and Tawn (1998). Up to now, solutions seem rather ad hoc and especially designed for the data set at hand. Often the threshold is taken equal over the different groups leading to inefficient use of the data if the scale in the different groups is quite different.

In contrast, the semi-parametric approaches where only the k largest data are used for tail estimation can overcome this problem. In this paper we consider in section 2 the estimation problem of γ in case data on several Pareto-type groups are available. Next, in section 3, we extend the procedure to the general case where the extreme-value index can be positive or negative. The performance of the proposed methods will be illustrated using small sample simulations.

2. LINEAR MODEL FORMULATION, $\gamma > 0$

2.1. Description of the model

Consider independent and identically distributed positive random variables $X_1^{(j)}, ..., X_{n_j}^{(j)}$ with a common distribution function $F_{X^{(j)}}$, $j=1, ..., G$, where G denotes the number of groups. Assume further that the G groups are independent of each other and that the response distributions are of Pareto-type i.e. the tail quantile functions $U_{X^{(j)}}$, $j=1, ..., G$, satisfy

$$U_{X^{(j)}}(x) = x^{\gamma_j}\, l_j(x) \qquad x > 1, \quad \gamma_j > 0 \tag{2.1}$$

where γ_j and l_j denote the extreme-value index respectively the slowly varying function of group j.

In the extreme-value literature one often imposes the so-called slow variation with remainder condition (see section 3.12.1 of Bingham *et al.*, 1987) on the slowly varying function l in (1.4). This second order condition specifies the rate of convergence of the ratio $l(\lambda x)/l(x)$ to its limit as $x \to \infty$.

Assumption *(R_l): There exists a real constant $\rho<0$ and a rate function b satisfying $b(x)\to 0$ as $x\to\infty$, such that for all $\lambda\geq 1$, as $x\to\infty$,*

$$\frac{l(\lambda x)}{l(x)}-1 \sim b(x)\,k_\rho(\lambda)$$

with $k_\rho(\lambda)=\int_1^\lambda v^{\rho-1}dv.$

Note that assumption (R_l) is quite general and is satisfied by the Hall (1982) class of Pareto-type distributions given by

$$U_X(x)=a\,x^\gamma\Big(1+d\,x^\rho+o(x^\rho)\Big)\qquad x>1;\ \ a,\gamma>0;\ \ \rho<0;\ \ d\in\mathbb{R}\ , \tag{2.2}$$

with $b(x)\sim\rho\,dx^\rho$ as $x\to\infty$.

As in a classical one-way ANOVA situation we introduce the parametrization $\gamma_j=\beta_0+\beta_j$, $j=1,...,G$, with $\sum_{j=1}^G\beta_j=0$, so that the parameters β_j denote the difference of the extreme-value index of group j with respect to the global average over all groups. This transformation will now be combined with the following linear model describing the estimation problem of every γ_j, $j=1,...,G$.

Under the slow variation with remainder condition on the l_j, $j=1,...,G$, it can be shown as in Beirlant *et al.* (1999) that the following regression model holds approximately

$$i\Big(\log X^{(j)}_{n_j-i+1,n_j}-\log X^{(j)}_{n_j-i,n_j}\Big)\ \approx\ \left(\gamma_j\ +\ b_j\left(\frac{n_j+1}{k+1}\right)\left(\frac{i}{k+1}\right)^{-\rho_j}\right)F_i^{(j)}\qquad i=1,...,k\ , \tag{2.3}$$

with b_j and ρ_j denoting the function b respectively the parameter ρ of group j and the $F_i^{(j)}$, $i=1,...,k$, are independent standard exponential random variables. In Beirlant *et al.* (2002), the approximation error in (2.3) is shown to be $o_P\big(b_j(\frac{n_j+1}{k+1})\big)$, $j=1,...,G$. Remark that regression model (2.3) is not identifiable when $\rho=0$, for then γ_j and $b_j\big((n_j+1)/(k+1)\big)$ together make up the mean response.

The classical way to estimate the parameters γ_j, $j=1,...,G$, is then given by the Hill (1975) estimates which are obtained as maximum likelihood estimates by omitting the terms $b_j\big(\frac{n_j+1}{k+1}\big)\big(\frac{i}{k+1}\big)^{-\rho_j}$ in model (2.3) (these terms tend to 0 as $n_j\to\infty$ and $k/n_j\to 0$) leading to a simple average of the scaled log-spacings $i\big(\log X^{(j)}_{n_j-i+1,n_j}-\log X^{(j)}_{n_j-i,n_j}\big)$, $i=1,...,k$, as an estimator of γ_j, and hence

$$\hat\beta_0=\frac{1}{G}\sum_{j=1}^G H^{(j)}_{k,n_j}\qquad\text{and}\qquad \hat\beta_j=H^{(j)}_{k,n_j}-\hat\beta_0,\quad j=1,...,G\ , \tag{2.4}$$

in which $H^{(j)}_{k,n_j}$ denotes the Hill estimator for group j

$$H^{(j)}_{k,n_j} = \frac{1}{k}\sum_{i=1}^{k} \log X^{(j)}_{n_j-i+1,n_j} - \log X^{(j)}_{n_j-k,n_j} \; . \tag{2.5}$$

Introducing $\boldsymbol{\Lambda} = \text{Block-diag}\big(\gamma_j^2 I_k;\; j=1,...,G\big)$ and the $k\,G \times G$ matrix

$$\boldsymbol{L} = \begin{bmatrix} \mathbf{1} & \mathbf{1} & \cdots & \mathbf{0} \\ \mathbf{1} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{1} & -\mathbf{1} & \cdots & -\mathbf{1} \end{bmatrix}$$

with $\mathbf{1}$ denoting a k-vector of ones, we find that the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}' = (\hat\beta_0, \hat\beta_1, ..., \hat\beta_{G-1})$ is given by

$$\text{Acov}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{L}'\boldsymbol{\Lambda}^{-1}\boldsymbol{L})^{-1} \; . \tag{2.6}$$

On the other hand the main term of the bias of the estimators (when $n_j \to \infty$ and $k/n_j \to 0$) is given by

$$\text{Abias}(\hat\beta_0) = \frac{1}{G}\sum_{j=1}^{G} \frac{b_j\Big(\frac{n_j+1}{k+1}\Big)}{1-\rho_j} \; , \tag{2.7}$$

$$\text{Abias}(\hat\beta_j) = \frac{b_j\Big(\frac{n_j+1}{k+1}\Big)}{1-\rho_j} - \frac{1}{G}\sum_{l=1}^{G} \frac{b_l\Big(\frac{n_l+1}{k+1}\Big)}{1-\rho_l} \qquad j = 1, ..., G-1 \; . \tag{2.8}$$

Application of the estimators defined by (2.4) and (2.5) involves the selection of the number of extreme order statistics k to be used in the estimation. Remark that we take the tail sample fraction k equal for all groups. If k is chosen too small, the resulting estimators will have a high variance. On the other hand, for larger k values the estimators will perform quite well with respect to variance but will be affected by a larger bias as observations are used which are not really informative for the tail of $F_{X^{(j)}}$, $j = 1, ..., G$. Hence, a good k value should represent a good bias-variance trade-off. Here we will use the trace of the asymptotic mean squared error (AMSE) matrix as optimality criterion.

Defining the AMSE matrix Ω of $\hat{\boldsymbol{\beta}}$ as

$$\Omega(k) = (\boldsymbol{L}'\boldsymbol{\Lambda}^{-1}\boldsymbol{L})^{-1} + \boldsymbol{\kappa}\,\boldsymbol{\kappa}' \; , \tag{2.9}$$

with $\boldsymbol{\kappa}$ denoting the G-vector containing the asymptotic bias expressions given by (2.7) and (2.8), the optimal number of extremes to be used in the estimation, k_{opt}, is defined as

$$k_{\text{opt}} = \arg\min \text{tr}\; \Omega(k) \; .$$

Note that $\Omega(k)$ depends on the unknown γ_j, ρ_j, $j=1,...,G$, and $b_j\big(\frac{n_j+1}{k+1}\big)$, $k=1,...,n_j-1$, $j=1,...,G$, which implies that the optimal k has to be derived from an estimate of $\Omega(k)$. The following algorithm is used to estimate k_{opt} and hence γ_j, $j=1,...,G$, adaptively:

1. Obtain initial estimates of $\gamma_j,\ \rho_j,\ j=1,...,G$, together with estimates of $b_j\big(\frac{n_j+1}{k+1}\big),\ k=1,...,n_j-1,\ j=1,...,G$,

2. for $k=2,...,\min(n_j; j=1,...,G)-1$:
compute tr $\hat{\Omega}(k)$ and let
$$\hat{k}_{\text{opt}} \;=\; \arg\min \operatorname{tr} \hat{\Omega}(k) \ ,$$

3. repeat step 2 but with the parameter estimates obtained from using a common k and obtain an update of the parameter estimates.

The initial estimates for the unknown parameters (cf. step 1) are obtained by fitting model (2.3) to the k largest observations of each group using a maximum likelihood method (see Beirlant *et al.*, 1999).

Inference about the regression vector $\boldsymbol{\beta}$ can be drawn using a likelihood ratio test statistic. For $k/n_j,\ j=1,...,G$, sufficiently small, the slowly varying nuisance part of (2.3) can be ignored and hence inference can be based on the reduced model $j(\log X^{(j)}_{n_j-i+1,n_j}-\log X^{(j)}_{n_j-i,n_j}) \approx (\beta_0+\beta_j)\,F^{(j)}_i,\ i=1,...,k$, $j=1,...,G$. As in a 'classical' one-way ANOVA situation the hypothesis of main interest is $H_0\colon\ \beta_1=...=\beta_{G-1}=0$.

2.2. An illustration

The procedure described above will be illustrated on a simulated dataset. The dataset contains observations from two groups with $n_1=n_2=500$. Data were generated from Burr(η,τ,λ) distributions (Burr, 1942). The Burr(η,τ,λ) distribution function, given by
$$F_X(x) = 1-\left(\frac{\eta}{\eta+x^\tau}\right)^{\lambda} \qquad x>0;\quad \eta,\tau,\lambda>0\ ,$$
is clearly of Pareto-type with $\gamma=\frac{1}{\lambda\tau}$ and $\rho=-\frac{1}{\lambda}$. For the simulated data $\lambda=1$ and $\eta=1$ for both groups and $\tau_1=1,\ \tau_2=2$ so $\gamma_1=1,\ \gamma_2=0.5$ and $\rho_1=\rho_2=-1$. Application of the above algorithm results in $\hat{k}_{\text{opt}}=99,\ H^{(1)}_{99,500}=1.007$ and $H^{(2)}_{99,500}=0.560$. In Figure 1 we show the Pareto quantile plot for both groups. On each quantile plot we superimposed the fitted lines passing through the anchor points $\big(\log\big(\frac{501}{100}\big), \log x^{(j)}_{401,500}\big),\ j=1,2$, with respective slopes $H^{(j)}_{99,500},\ j=1,2$. The lines fit the linear part of the Pareto quantile plot quite well. In Figure 2 we plot the trace of the AMSE-matrix (full line) and the trace of the estimated AMSE-matrix (broken line) versus the number of extremes used in the estimation of the regression coefficients, k. Note that around the optimal k value tr $\Omega(k)$ is estimated quite well. The tr $\Omega(k)$ function reaches its minimum at $k=79$ whereas for the estimate the minimum is reached at $k=99$. For this simulated dataset, the observed value of the likelihood ratio test statistic to assess the validity of $H_0\colon\ \beta_1=0$ equals 16.834, leading to a rejection of H_0.

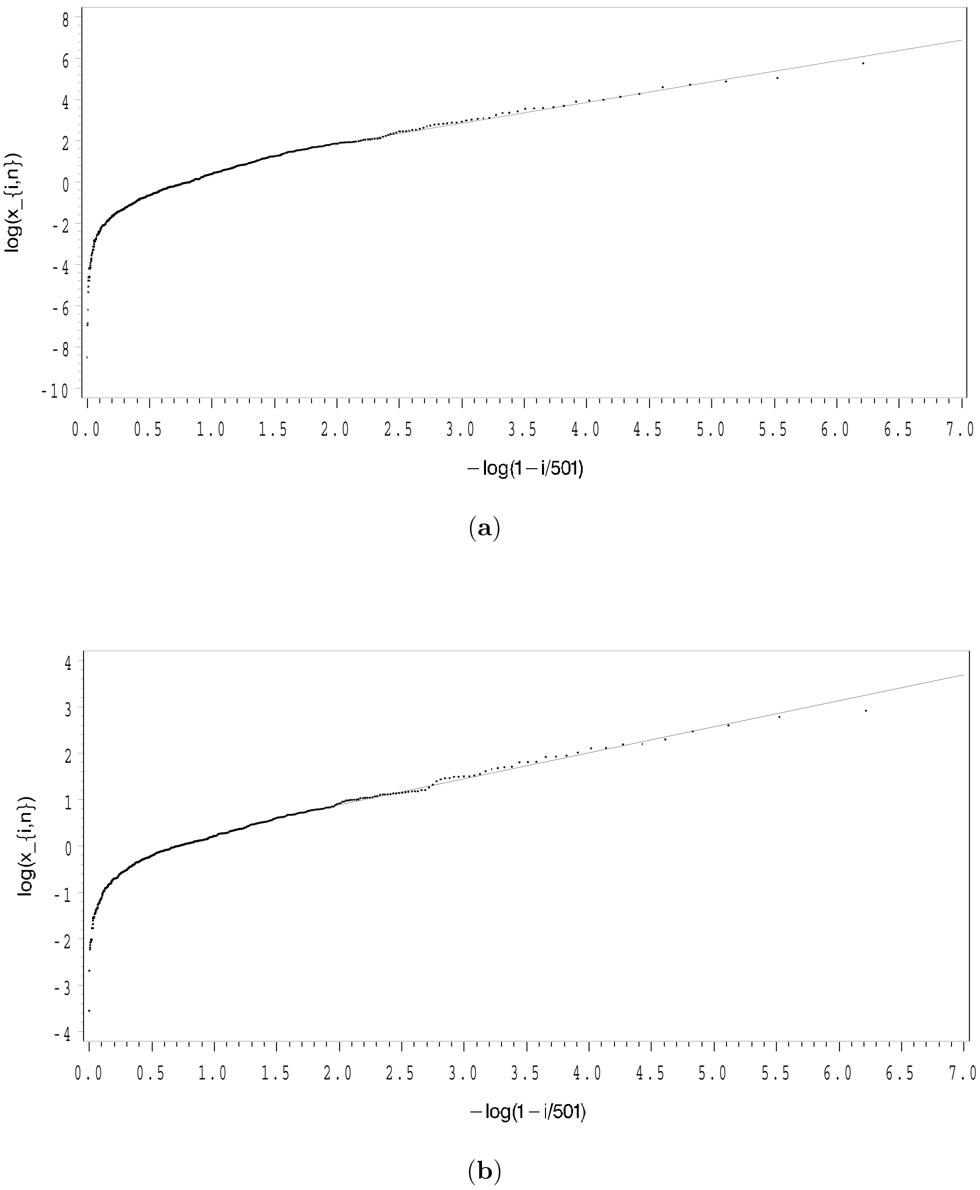


Figure 1: Burr$(1,\tau_j,1)$ simulation with $n_1 = n_2 = 500$, $\tau_1 = 1$ and $\tau_2 = 2$.
(**a**) Pareto quantile plot for group 1 with line through $\left(\log(\frac{501}{100}), \log x^{(1)}_{401,500}\right)$ and slope $H^{(1)}_{99,500} = 1.007$ superimposed;
(**b**) Pareto quantile plot for group 2 with line through $\left(\log(\frac{501}{100}), \log x^{(2)}_{401,500}\right)$ and slope $H^{(2)}_{99,500} = 0.560$ superimposed.

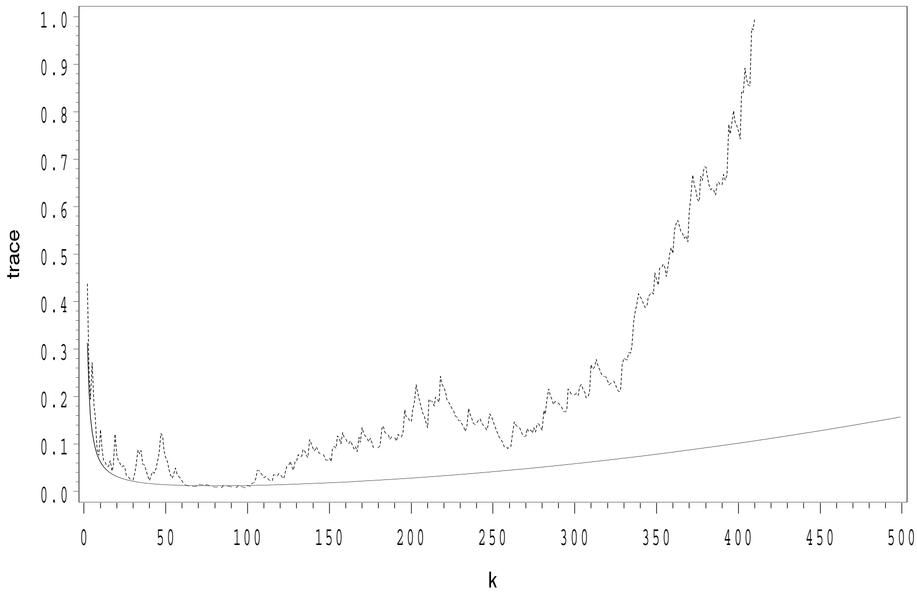


Figure 2: $\mathrm{tr}\,\Omega(k)$ (full line) and $\mathrm{tr}\,\hat{\Omega}(k)$ (broken line) vs k.

2.3. Simulation results and practical examples

2.3.1. Simulation results

We illustrate the small sample behaviour of the weighted least squares estimator $\hat{\boldsymbol{\beta}}_{\hat{k}_{\mathrm{opt}}}$ using a simulation study. Datasets containing observations on 2, 3 and 4 groups were generated from $\mathrm{Burr}(1, \tau_j, \lambda_j)$, $j = 1, ..., G$, distributions. In Tables 1 and 2 we report the sample mean, sample standard deviation, empirical RMSE and the ratio (empirical RMSE under common optimal k)/(empirical RMSE under optimal k for each group separately) for samples of respectively 200 and 500 observations per group. The blocks $\lambda = 0.5$, $\lambda = 1$ and $\lambda = 2$ of both tables report the results in case of a common λ, and hence a common ρ, over the groups. The last λ-block of both tables reports results for the case $\lambda_j = 1/j$, $j = 1, ..., G$, and hence $\rho_j = -j$, $j = 1, ..., G$. Values for the τ-parameters were selected such that $\gamma_j = j$. From the ratio results it is clear that joint estimation of the extreme-value indices with a common k can lead to important gains in empirical MSE compared with a separate analysis. For instance in case $G\!=\!3$ and $\lambda\!=\!2$, joint estimation of β_2 leads to a 30% gain in RMSE. Further, inspection of the first three λ-blocks of both tables indicates that the gains tend to increase with λ.

Table 1: Burr data, 200 observations/group, 500 simulation runs.

		$G=2$		$G=3$			$G=4$			
		β_0	β_1	β_0	β_1	β_2	β_0	β_1	β_2	β_3
value		1.5	−0.5	2	−1	0	2.5	−1.5	−0.5	0.5
$\lambda=0.5$	mean	1.5556	−0.5039	2.0909	−1.0209	0.0021	2.6197	−1.5434	−0.5131	0.5305
	sd	0.1930	0.1596	0.2281	0.2137	0.2373	0.2544	0.2443	0.2697	0.3567
	RMSE	0.2007	0.1595	0.2454	0.2146	0.2371	0.2809	0.2479	0.2697	0.3577
	ratio	0.9531	0.8323	1.0255	0.8837	0.8361	1.0097	0.8905	0.7855	0.8762
$\lambda=1$	mean	1.6694	−0.5386	2.2485	−1.0960	−0.0106	2.8057	−1.6346	−0.5522	0.5450
	sd	0.2787	0.2257	0.3316	0.2706	0.3156	0.3906	0.3449	0.3778	0.4614
	RMSE	0.3259	0.2288	0.4141	0.2869	0.3154	0.4957	0.3699	0.3810	0.4632
	ratio	0.9936	0.8345	0.9957	0.8150	0.7776	1.0058	0.8428	0.7733	0.8475
$\lambda=2$	mean	2.0410	−0.6390	2.6862	−1.2943	−0.0011	3.4326	−1.9971	−0.6735	0.6607
	sd	0.4398	0.3463	0.4951	0.3962	0.4137	0.5749	0.4783	0.5366	0.6491
	RMSE	0.6969	0.3729	0.8459	0.4932	0.4133	1.0952	0.6895	0.5634	0.6681
	ratio	0.9695	0.7929	0.9685	0.7994	0.7061	0.9729	0.8206	0.7830	0.7343
$\lambda_j=1/j$	mean	1.6079	−0.4277	2.1199	−0.8654	−0.0088	2.6173	−1.3233	−0.4795	0.4147
	sd	0.2094	0.1807	0.2251	0.1931	0.2342	0.2203	0.2039	0.2548	0.3257
	RMSE	0.2354	0.1944	0.2548	0.2352	0.2342	0.2494	0.2696	0.2554	0.3363
	ratio	0.9840	0.8974	1.0892	1.0082	0.8329	1.1020	1.0170	0.8230	0.9015

Table 2: Burr data, 500 observations/group, 500 simulation runs.

		$G=2$		$G=3$			$G=4$			
		β_0	β_1	β_0	β_1	β_2	β_0	β_1	β_2	β_3
value		1.5	−0.5	2	−1	0	2.5	−1.5	−0.5	0.5
$\lambda=0.5$	mean	1.5443	−0.5044	2.0412	−1.0094	−0.0043	2.5617	−1.5185	−0.4996	0.5020
	sd	0.1380	0.1147	0.1769	0.1528	0.1868	0.1838	0.1775	0.2069	0.2613
	RMSE	0.1448	0.1147	0.1815	0.1529	0.1867	0.1937	0.1783	0.2067	0.2610
	ratio	1.0223	0.8638	1.1600	0.9581	0.9118	1.1233	0.9360	0.8576	0.8929
$\lambda=1$	mean	1.5915	−0.5126	2.1547	−1.0720	0.0112	2.7099	−1.6101	−0.5293	0.5192
	sd	0.2361	0.1741	0.2474	0.2115	0.2329	0.2580	0.2405	0.2891	0.3657
	RMSE	0.2530	0.1744	0.2915	0.2232	0.2330	0.3324	0.2643	0.2903	0.3658
	ratio	1.0454	0.8382	0.9870	0.8447	0.7551	1.0001	0.8506	0.8173	0.8216
$\lambda=2$	mean	1.8470	−0.5969	2.4778	−1.2054	0.0009	3.1297	−1.8401	−0.5969	0.5744
	sd	0.3471	0.2575	0.3909	0.3110	0.3424	0.4022	0.3442	0.3804	0.4884
	RMSE	0.4905	0.2749	0.6171	0.3724	0.3421	0.7470	0.4837	0.3922	0.4936
	ratio	1.0089	0.8265	0.9869	0.8235	0.7558	0.9995	0.8697	0.7440	0.7543
$\lambda_j=1/j$	mean	1.5648	−0.4519	2.0722	−0.9158	−0.0330	2.5736	−1.3565	−0.5005	0.4668
	sd	0.1578	0.1391	0.1481	0.1367	0.1849	0.1523	0.1533	0.1912	0.2406
	RMSE	0.1704	0.1470	0.1647	0.1604	0.1877	0.1690	0.2099	0.1910	0.2427
	ratio	1.0395	0.9922	1.0576	0.9993	0.9823	1.1274	1.1048	0.8420	0.9789

Also, we performed a small sample simulation study to assess whether the likelihood ratio test for the hypothesis of no factor effects in the reduced model

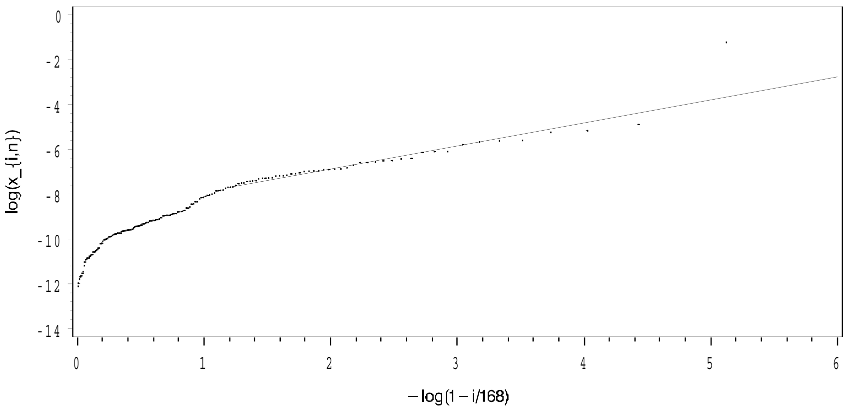
$j\big(\log X^{(j)}_{n_j-i+1,n_j} - \log X^{(j)}_{n_j-i,n_j}\big) \approx (\beta_0+\beta_j)F^{(j)}_i,\ i=1,...,k,\ j=1,...,G,$ satisfies the proposed significance level. Datasets containing 3 groups of 500 observations each were generated from $\text{Burr}(1,\tau_j,\lambda_j)$ distributions. Concerning the parameter λ_j (and hence ρ_j), 4 cases were considered: $\lambda_j = 0.5$, $\lambda_j = 1$, $\lambda_j = 2$ and $\lambda_j = 1/j$, $j = 1,...,G$. The parameters τ_j, $j = 1,...,G$, were selected such that all $\gamma_j = 1$ (simulation under H_0) and the significance level was set at $\alpha = 0.05$. Table 3 reports the empirical significance level for each setting of ρ for different k-values. As can be seen, the empirical significance levels are slightly below 0.05 for the cases with common ρ. In case the ρ parameter varies over the groups, the test performs only well at the smaller values of k, a result that could be expected.

Table 3: Burr data, 500 observations/group, 500 simulation runs: empirical significance levels.

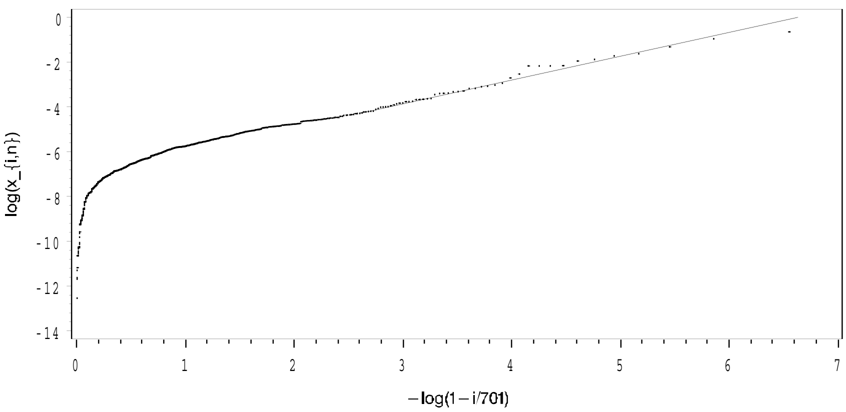
	$k = 10$	$k = 50$	$k = 100$	$k = 200$	$k = 400$
$\rho = -2$	0.034	0.040	0.044	0.044	0.018
$\rho = -1$	0.048	0.036	0.034	0.026	0.044
$\rho = -0.5$	0.040	0.030	0.030	0.020	0.058
$\rho_j = -j$	0.048	0.042	0.094	0.612	1.000

2.3.2. Practical example 1: fire claim data

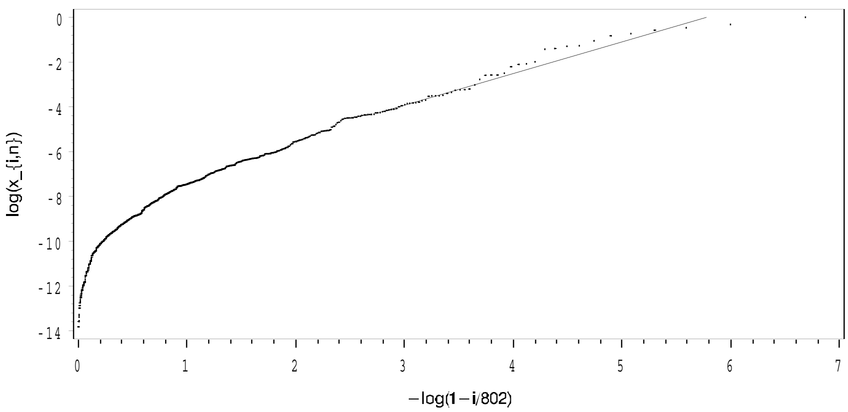
Our first example comes from an actuarial context. The reinsurance broker Aon Re Belgium provided claim data, generated by a fire insurance portfolio, for three types of buildings. The sample sizes are $n_1 = 167$, $n_2 = 700$ and $n_3 = 801$. Application of the proposed procedure results in $\hat{k}_{\text{opt}} = 50$, $H^{(1)}_{50,167} = 1.027$, $H^{(2)}_{50,700} = 1.064$ and $H^{(3)}_{50,801} = 1.413$. The Pareto quantile plots together with the lines passing through $\big(\log(\frac{n_j+1}{\hat{k}_{\text{opt}}+1}), \log x^{(j)}_{n_j-\hat{k}_{\text{opt}},n_j}\big)$ and slopes $H^{(j)}_{\hat{k}_{\text{opt}},n_j}$, $j=1,2,3$, are given in Figure 3. As is clear from this figure, the Pareto quantile plots are almost linear in their extreme values indicating a reasonable fit of the Pareto distribution to the tails of the conditional claim size distributions. Concerning the γ estimate for group 3 (see also Figure 3 (c)) actuaries will find the estimate high. Remark however that other characteristics, such as the sum insured, can have an important influence on the tail index estimates but have been ignored in this analysis. Given an observed value for the likelihood ratio test statistic of 3.152, the null hypothesis of no group effects cannot be rejected.



(**a**)



(**b**)

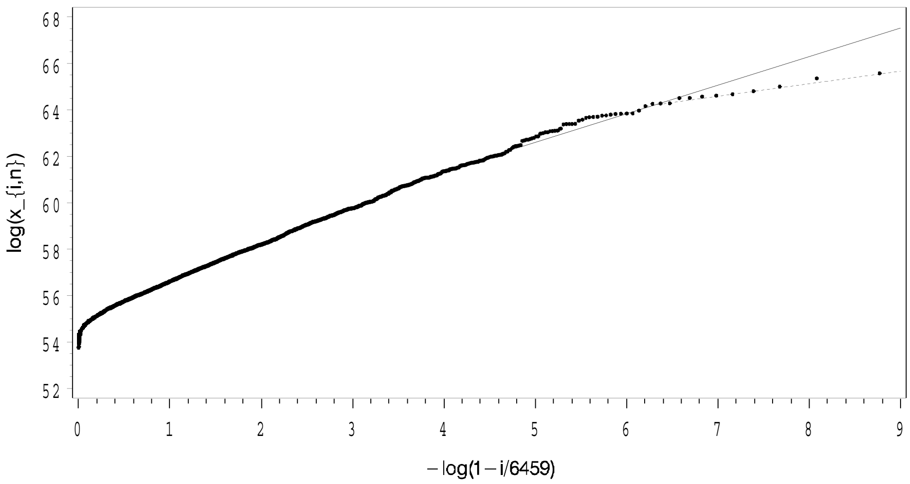


(**c**)

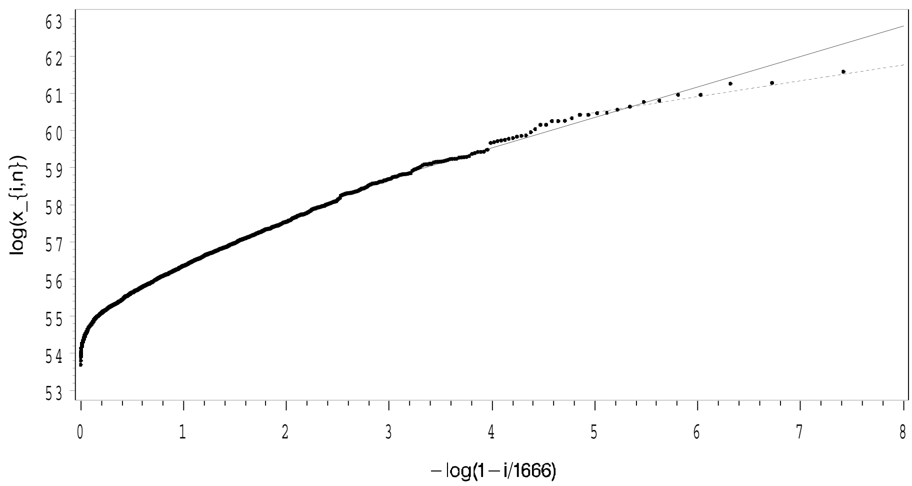
Figure 3: Claim data: Pareto quantile plots.

2.3.3. Practical example 2: earthquake data

As a second example we examined the earthquake data introduced in Pisarenko and Sornette (2001). This dataset is extracted from the Harvard catalog and contains information about the seismic moment (in dyne-cm) of shallow earthquakes (depth < 70 km) over the period 1977–2000. In Pisarenko and Sornette (2001), the tails of the seismic moment distributions for subduction and midocean ridge zones are compared by fitting the generalized Pareto distribution to seismic moment exceedances over 10^{24} dyne-cm. For these data $n_1 = 6458$ (subduction zones) and $n_2 = 1665$ (midocean ridge zones). The procedure described above with $k \geq 20$ yielded $\hat{k}_{\text{opt}} = 97$ with $H^{(1)}_{97,6458} = 1.232$ and $H^{(2)}_{97,1665} = 0.821$. In Figure 4 we show the Pareto quantile plots of the seismic moments for (a) subduction zones and (b) midocean ridge zones on which we superimposed the lines through $\left(\log(\frac{n_j+1}{\hat{k}_{\text{opt}}+1}), \log x^{(j)}_{n_j-\hat{k}_{\text{opt}},n_j}\right)$ with slope $H^{(j)}_{\hat{k}_{\text{opt}},n_j}$, $j = 1, 2$ (solid lines). For the hypothesis test of no difference between the tail heaviness of the seismic moment distribution of subduction and midocean ridge zones a likelihood ratio statistic of 7.92 was obtained, resulting in a rejection of H_0. The GPD based approach described in Pisarenko and Sornette (2001) yielded tail index estimates of 1.51 and 1.02 for subduction respectively midocean ridge zones, so our results are slightly more conservative. Likewise these authors found significant differences in the tail heaviness of the seismic moment distributions. Note that the Pareto quantile plots bend down at the largest observations indicating a weaker behaviour of the ultimate tail of the seismic moment distribution. Nevertheless, these largest observations form more or less a straight line pattern. So, also the ultimate tail could be described by a Pareto-type law. This fact is further illustrated in Figure 5 where we plot $\text{tr}\,\hat{\Omega}(k)$ as a function of $\log(k)$. Relaxation of the constraint that k should be at least 20 results in the global optimum $\hat{k}_{\text{opt}} = 12$ with $\hat{\gamma}_1 = 0.541$ and $\hat{\gamma}_2 = 0.427$. In Figure 4 the resulting optimal fits are plotted with dotted lines. At $\hat{k}_{\text{opt}}$ the null hypothesis of no difference in tail behaviour cannot be rejected on basis of the above described likelihood ratio test statistic. Similarily to the results presented here, Pisarenko and Sornette (2001) also found deviations between the GPD and the ultimate tail of the seismic moment distribution. For plausible explanations of this phenomenon we refer to their paper and the references therein.



(**a**)



(**b**)

Figure 4: Earthquake data: Pareto quantile plots of seismic moments for
(**a**) subduction zones and
(**b**) midocean ridge zones.

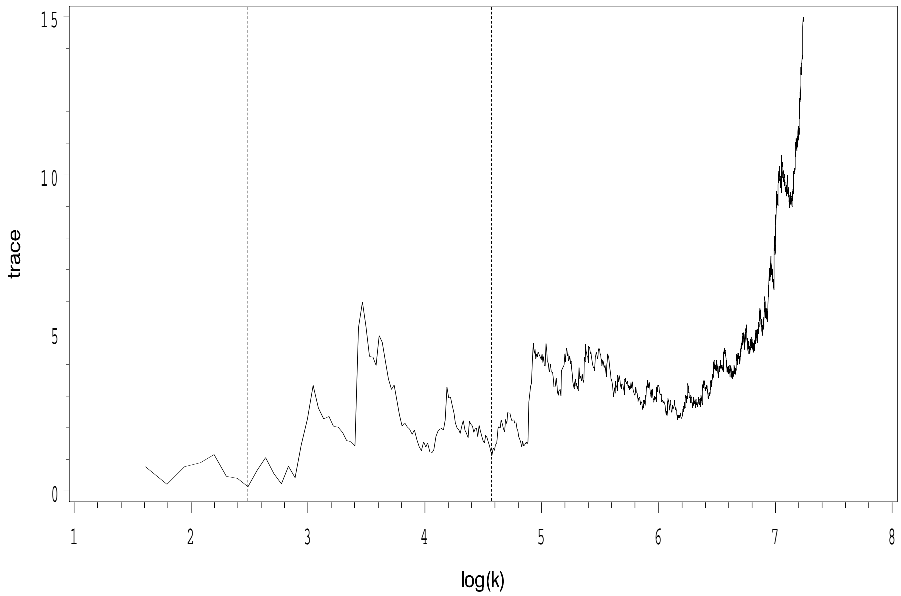


Figure 5: Earthquake data: $\mathrm{tr}\,\hat{\Omega}(k)$ vs $\log(k)$.

3. LINEAR MODEL FORMULATION, $\gamma \in \mathbb{R}$

3.1. Description of the model

In this section we discuss the simultaneous estimation of several extreme-value indices in the general case $\gamma \in \mathbb{R}$. Consider again a sample of independent and identically distributed positive random variables $X_1^{(j)}, ..., X_{n_j}^{(j)}$ according to some distribution function $F_{X^{(j)}}$, $j=1, ..., G$, with G denoting the number of groups. Further, assume that the G groups are independent of each other and that $F_{X^{(j)}} \in \mathcal{D}(H_{\gamma_j})$ for some $\gamma_j \in \mathbb{R}$. In Theorem 1 of Beirlant *et al.* (1996) it is shown that $F_{X^{(j)}} \in \mathcal{D}(H_{\gamma_j})$ implies that

$$UH_{X^{(j)}}(x) \;=\; x^{\gamma_j}\, \breve{l}_j(x)$$

with $UH_{X^{(j)}}$ and $\breve{l}_j$ denoting the UH function respectively the slowly varying function of group j.

Under the slow variation with remainder condition imposed on $\breve{l}_j$, the following relation holds (see Beirlant *et al.*, 2000)

$$(i{+}1)\log\frac{UH^{(j)}_{i,n_j}}{UH^{(j)}_{i+1,n_j}} - (i{+}1)\log\frac{i{+}1}{i} + \frac{i{+}1}{i} \;=\; \gamma_j + g_j\left(\frac{n_j{+}1}{k{+}1}\right)\left(\frac{i{+}1}{k{+}1}\right)^{-\breve{\rho}_j} + \varepsilon^{(j)}_i \tag{3.1}$$

$$i{=}1,...,k, \quad j{=}1,...,G\ ,$$

where g_j is some generic notation for a function decreasing to zero for increasing values of the argument and

$$UH^{(j)}_{i,n_j} \;=\; X^{(j)}_{n_j-i,n_j}\left(\frac{1}{i}\sum_{m=1}^{i}\log X^{(j)}_{n_j-m+1,n_j} - \log X^{(j)}_{n_j-i,n_j}\right) .$$

The residuals of model (3.1) have a mean approximately equal to zero and a covariance matrix given by

$$\boldsymbol{\Sigma}_{\varepsilon^{(j)}} \;=\; \left[\mathrm{Cov}\big(\varepsilon^{(j)}_s,\varepsilon^{(j)}_t\big)\right]_{s,t} \sim \begin{cases} \dfrac{\gamma_j}{t} & s<t \\ (\gamma_j-1)^2+\dfrac{1+2s}{s^2} & s=t \end{cases} \qquad j{=}1,...,G\ .$$

After introduction of the classical one-way ANOVA parametrization described above and deletion of the terms $g_j\big(\frac{n_j+1}{k+1}\big)\big(\frac{i+1}{k+1}\big)^{-\breve{\rho}_j}$ in model (3.1) (these terms tend to 0 as $n_j\to\infty$ and $k/n_j\to 0$), the following estimators are obtained:

$$\tilde{\beta}_0 = \frac{1}{G}\sum_{j=1}^{G}\tilde{\gamma}_j \quad\text{and}\quad \tilde{\beta}_j = \tilde{\gamma}_j - \tilde{\beta}_0\ , \qquad j=1,...,G\ , \tag{3.2}$$

with

$$\tilde{\gamma}_j \;=\; \frac{1}{k}\sum_{i=1}^{k}\left((i{+}1)\,\log\frac{UH^{(j)}_{i,n_j}}{UH^{(j)}_{i+1,n_j}} \;-\; (i{+}1)\,\log\frac{i{+}1}{i} + \frac{i{+}1}{i}\right) . \tag{3.3}$$

Using least squares computations, the asymptotic covariance matrix of $\tilde{\boldsymbol{\beta}}' = (\tilde{\beta}_0,\tilde{\beta}_1,...,\tilde{\beta}_{G-1})$ is given by:

$$\mathrm{Acov}(\tilde{\boldsymbol{\beta}}) \;=\; (\boldsymbol{L}'\boldsymbol{L})^{-1}\boldsymbol{L}'\,\boldsymbol{\Sigma}\,\boldsymbol{L}(\boldsymbol{L}'\boldsymbol{L})^{-1} \tag{3.4}$$

with $\boldsymbol{\Sigma} = \text{Block-diag}(\boldsymbol{\Sigma}_{\varepsilon^{(j)}};\, j=1,...,G)$. For the main term of the bias of the estimators we have

$$\mathrm{Abias}(\tilde{\beta}_0) \;=\; \frac{1}{G}\sum_{j=1}^{G}\frac{g_j\big(\frac{n_j+1}{k+1}\big)}{1-\breve{\rho}_j}\ ,$$

$$\mathrm{Abias}(\tilde{\beta}_j) \;=\; \frac{g_j\big(\frac{n_j+1}{k+1}\big)}{1-\breve{\rho}_j} - \frac{1}{G}\sum_{l=1}^{G}\frac{g_l\big(\frac{n_l+1}{k+1}\big)}{1-\breve{\rho}_l}\ , \qquad j=1,...,G{-}1\ .$$

Application of the estimators defined by (3.2) and (3.3) requires the selection of the number of *UH* statistics k to be used in the estimation. Again the asymptotic variance and bias are combined in an AMSE criterion. Hence, the optimal k-value is defined as

$$k_{\rm opt} \;=\; \arg\min\, {\rm tr}\Big[(\boldsymbol{L}'\boldsymbol{L})^{-1}\boldsymbol{L}'\,\boldsymbol{\Sigma}\,\boldsymbol{L}(\boldsymbol{L}'\boldsymbol{L})^{-1} + \breve{\boldsymbol{\kappa}}\,\breve{\boldsymbol{\kappa}}'\Big]$$

with $\breve{\boldsymbol{\kappa}}$ the vector containing the asymptotic bias expressions given above.

For k/n_j sufficiently small, hypothesis tests about the regression coefficients can be based on the reduced model

$$(i{+}1)\,\log\frac{UH^{(j)}_{i,n_j}}{UH^{(j)}_{i+1,n_j}} \;-\; (i{+}1)\,\log\frac{i{+}1}{i} + \frac{i{+}1}{i} \;\approx\; \beta_0+\beta_j+\varepsilon^{(j)}_i \;, \tag{3.5}$$
$$i=1,...,k, \quad j=1,...,G \;.$$

After transformation of model (3.5) by a matrix $\boldsymbol{C}$ such that ${\rm Cov}(\boldsymbol{C}\boldsymbol{\varepsilon})=\boldsymbol{I}$ (see Beirlant *et al.*, 2000), where $\boldsymbol{\varepsilon}'=(\varepsilon^{(j)}_i;\; i{=}1,...,k,\; j{=}1,...,G)$, hypothesis about $\boldsymbol{\beta}$ can be tested using a classical F-test statistic.

3.2. Simulation results and a practical example

3.2.1. Simulation results

First, we apply the above proposed estimation procedure for $\gamma_j\in\mathbb{R}$, $j=1,...,G$, to the simulated Burr datasets described in the previous section. Tables 4 and 5 contain the results for samples with 200 respectively 500 observations per group. Unlike the algorithm for $\gamma_j>0$, $j{=}1,...,G$, where the empirical MSE gains obtained from using a common k tend to increase with ρ, here the gains are quite stable with respect to the parameter $\breve{\rho}$.

Next we examine the small sample properties of the proposed procedure using datasets with $\gamma_j<0$, $j=1,...,G$. Datasets containing observations on 2, 3 and 4 groups were generated from reversed Burr distributions. The reversed Burr distribution function, given by

$$1-F_X(x) \;=\; \left(\frac{2}{1+(1{-}x)^{-\tau}}\right)^{\delta} \qquad 0<x<1;\quad \delta,\tau>0\;,$$

belongs to the maximum domain of attraction of the GEV for some $\gamma<0$, and hence the UH_X function can be written as in (1.9) with $\gamma=-1/(\delta\tau)$.

Table 4: Burr data, 200 observations/group, 500 simulation runs.

		$G=2$		$G=3$			$G=4$			
		β_0	β_1	β_0	β_1	β_2	β_0	β_1	β_2	β_3
value		1.5	−0.5	2	−1	0	2.5	−1.5	−0.5	0.5
$\lambda=0.5$	mean	1.5256	−0.5099	2.0626	−1.0214	−0.0032	2.5822	−1.5482	−0.5062	0.5205
	sd	0.1892	0.1640	0.2035	0.2177	0.2399	0.2199	0.2427	0.2820	0.3430
	RMSE	0.1907	0.1641	0.2128	0.2186	0.2397	0.2345	0.2472	0.2818	0.3433
	ratio	1.0344	0.8736	1.0406	0.9479	0.8351	1.0570	0.9699	0.8559	0.8567
$\lambda=1$	mean	1.6101	−0.5434	2.1938	−1.0937	−0.0098	2.7627	−1.6551	−0.5598	0.5514
	sd	0.2761	0.2220	0.2981	0.2765	0.3298	0.3168	0.3290	0.3654	0.4256
	RMSE	0.2970	0.2260	0.3553	0.2917	0.3297	0.4113	0.3634	0.3699	0.4283
	ratio	1.0450	0.8703	1.0929	0.9609	0.8416	1.1019	0.9746	0.8044	0.8383
$\lambda=2$	mean	1.9493	−0.6621	2.6153	−1.3241	0.0035	3.3430	−2.0366	−0.6866	0.6582
	sd	0.3754	0.3112	0.4224	0.3845	0.4239	0.5087	0.4792	0.4817	0.6217
	RMSE	0.5853	0.3506	0.7461	0.5026	0.4235	0.9843	0.7191	0.5161	0.6409
	ratio	1.0126	0.8442	1.0240	0.9237	0.7982	1.0679	0.9670	0.8137	0.7989
$\lambda_j=1/j$	mean	1.5676	−0.4476	2.0855	−0.9109	0.0056	2.5848	−1.3936	−0.4717	0.4442
	sd	0.2193	0.1847	0.1951	0.1872	0.2328	0.1942	0.2151	0.2328	0.3021
	RMSE	0.2293	0.1918	0.2128	0.2072	0.2326	0.2118	0.2398	0.2343	0.3069
	ratio	1.0134	0.8453	1.0516	0.8570	0.8256	1.0663	0.8528	0.7954	0.8536

Table 5: Burr data, 500 observations/group, 500 simulation runs.

		$G=2$		$G=3$			$G=4$			
		β_0	β_1	β_0	β_1	β_2	β_0	β_1	β_2	β_3
value		1.5	−0.5	2	−1	0	2.5	−1.5	−0.5	0.5
$\lambda=0.5$	mean	1.5418	−0.5127	2.0358	−1.0206	0.0090	2.5558	−1.5364	−0.4929	0.4996
	sd	0.1224	0.1080	0.1477	0.1434	0.1580	0.1559	0.1647	0.1766	0.2327
	RMSE	0.1293	0.1087	0.1519	0.1447	0.1581	0.1655	0.1685	0.1765	0.2325
	ratio	0.9691	0.8252	1.0148	0.8856	0.7395	1.0485	0.8854	0.7377	0.8045
$\lambda=1$	mean	1.5714	−0.5267	2.1576	−1.1004	0.0046	2.7033	−1.6413	−0.5303	0.5206
	sd	0.2143	0.1610	0.2171	0.2016	0.2072	0.2236	0.2291	0.2584	0.3106
	RMSE	0.2257	0.1630	0.2681	0.2250	0.2070	0.3020	0.2689	0.2599	0.3110
	ratio	1.0499	0.8362	1.0968	0.9123	0.7304	1.1378	0.9745	0.7719	0.7635
$\lambda=2$	mean	1.8209	−0.6052	2.4664	−1.2380	0.0095	3.1143	−1.8831	−0.6011	0.5939
	sd	0.2977	0.2396	0.3398	0.3064	0.2972	0.3628	0.3374	0.3519	0.4398
	RMSE	0.4375	0.2615	0.5769	0.3878	0.2971	0.7132	0.5103	0.3658	0.4493
	ratio	1.0465	0.8265	1.0787	0.8909	0.7386	1.1111	0.9947	0.7491	0.7358
$\lambda_j=1/j$	mean	1.5557	−0.4606	2.0663	−0.9269	−0.0044	2.5786	−1.4014	−0.4880	0.4672
	sd	0.1504	0.1255	0.1336	0.1346	0.1623	0.1226	0.1346	0.1659	0.2057
	RMSE	0.1603	0.1314	0.1490	0.1531	0.1622	0.1455	0.1668	0.1662	0.2081
	ratio	1.0041	0.7875	1.0173	0.8254	0.7941	1.1045	0.8256	0.7483	0.8470

Also, the $\breve{l}$ function associated with the UH_X function satisfies the slow variation with remainder condition with $\breve{\rho} = -\min\left(\frac{1}{\delta}, \frac{1}{\delta\tau}\right)$. In this simulation a common δ value was used for all groups. Further, values for the τ parameters were selected such that $\gamma_j = -j$, $j=1, ..., G$. Two cases were considered: $\delta = 1$ and $\delta = 2$ giving $\breve{\rho} = -1$ and $\breve{\rho} = -0.5$. Table 6 summarizes the simulation results. As expected, the estimators are more biased as the $\breve{\rho}$-parameter increases. The gains obtained by using a common k-value compared to a separate analysis of each group are quite stable with respect to $\breve{\rho}$.

Table 6: Reversed Burr data, 500 observations/group, 500 simulation runs.

		$G=2$		$G=3$			$G=4$			
		β_0	β_1	β_0	β_1	β_2	β_0	β_1	β_2	β_3
value		−1.5	0.5	−2	1	0	−2.5	1.5	0.5	−0.5
$\delta=1$	mean	−1.6382	0.5689	−2.1603	1.1005	−0.0094	−2.6820	1.6190	0.5077	−0.5401
	sd	0.1310	0.1124	0.1696	0.1510	0.1445	0.1861	0.1718	0.1832	0.2433
	RMSE	0.1903	0.1317	0.2332	0.1813	0.1447	0.2602	0.2089	0.1832	0.2464
	ratio	0.9650	0.9516	1.0028	1.0213	0.7320	1.0149	1.0585	0.8410	0.8030
$\delta=2$	mean	−1.8888	0.6507	−2.4375	1.2155	−0.0080	−2.9633	1.7470	0.5565	−0.5651
	sd	0.1646	0.1301	0.2234	0.1969	0.1785	0.2716	0.2466	0.2353	0.2821
	RMSE	0.4221	0.1990	0.4911	0.2918	0.1785	0.5369	0.3488	0.2418	0.2892
	ratio	0.9628	0.9593	0.9294	0.9865	0.7342	0.9667	1.0708	0.7785	0.7152

Finally, the procedure was applied to datasets containing groups for which the γ_j, $j=1,...,G$, can have a different sign and/or be equal to zero. Here, datasets containing observations on 2 and 3 groups were generated from the generalized Pareto distribution with distribution function given by (1.3). The slowly varying function $\breve{l}$ of the GPD satisfies the slow variation with remainder condition with $\breve{\rho} = -|\gamma|$. In this simulation, we took $\sigma = 1$. Table 7 contains the results for these problem sets. Also here we see that using a common optimal k can yield important efficiency gains compared to a separate analysis of each group, except in case β_0, β_1 are both negative.

Table 7: GPD data, 500 observations/group, 500 simulation runs.

	$G=2$		$G=2$		$G=2$		$G=3$		
	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1	β_2
value	−0.25	−0.25	0	−0.5	0.25	−0.25	0	−0.5	0
mean	−0.1582	−0.2591	0.0970	−0.5046	0.3442	−0.2518	0.0977	−0.5063	−0.0024
sd	0.0749	0.0734	0.0867	0.0800	0.1031	0.0819	0.0725	0.0789	0.0865
RMSE	0.1184	0.0739	0.1300	0.0800	0.1396	0.0819	0.1216	0.0791	0.0865
ratio	1.0551	0.9719	1.0204	0.8498	1.0317	0.8839	1.1251	0.8635	0.8507

3.2.2. Practical example 3: US wind speed data

The wind speed database, provided by the National Institute of Standards and Technology (NIST), contains information about 49 weather stations in the U.S.. The data have been filed for a period of 15 to 26 years. They are the daily fastest-mile wind speeds, measured by anemometers 10 m above ground. For more information about these data we refer to Simiu *et al.* (1979) and Simiu and Heckert (1995). We restrict our attention to three cities: Des Moines (Iowa), Grand Rapids (Michigan) and Albuquerque (New Mexico). Boxplots of the daily fastest wind speeds (in miles per hour) are given in Figure 6. The generalized quantile plot for each city is given in Figure 7. These plots allow to distinguish between the wind speed tail behavior of the different cities: the Des Moines data (Figure 7 (a)) are heavy tailed ($\gamma > 0$), the Grand Rapids data (Figure 7 (b)) seem to be moderately tailed with $\gamma \approx 0$ and the Albuquerque data (Figure 7 (c)) are weakly tailed ($\gamma < 0$). The line structures in these plots are the result of an inherent grouping of the data due to a loss of accuracy during the data collection process. Consequently many wind speed levels are registered more than once. Application of the above described procedure resulted in $\hat{k}_{\text{opt}} = 357$ with $\tilde{\gamma}_1 = 0.144$, $\tilde{\gamma}_2 = 0.053$ and $\tilde{\gamma}_3 = -0.088$. On each generalized quantile plot we superimposed the line passing through the anchor point $\left(\log(\frac{n_j}{k+1}), \log UH^{(j)}_{k+1,n_j}\right)$ with slope $\tilde{\gamma}_j$.

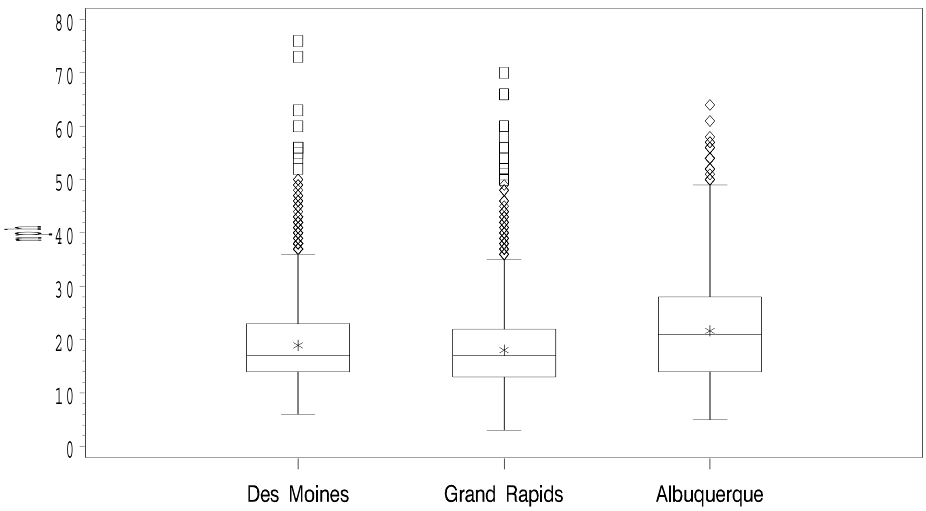
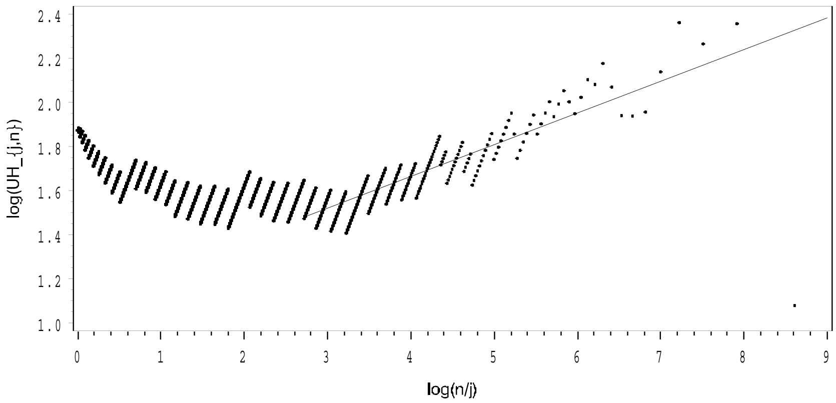
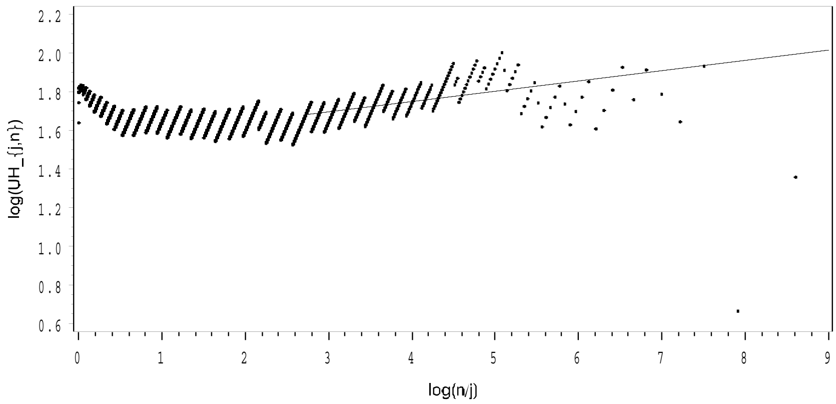


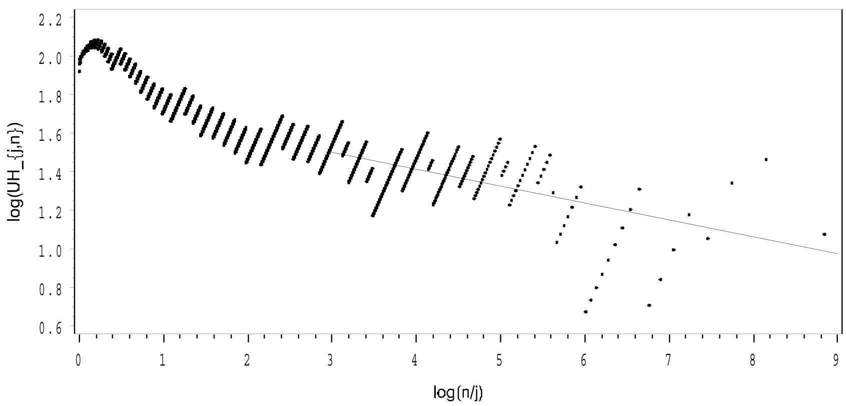
Figure 6: Wind speed data: Boxplots of the daily fastest wind speeds in miles per hour.



(**a**)



(**b**)



(**c**)

Figure 7: Wind speed data: generalized quantile plots for (**a**) Des Moines, (**b**) Grand Rapids and (**c**) Albuquerque.

4. CONCLUSION

In this paper we discussed the simultaneous estimation of tail indices when data on several independent groups are available. The proposed methods are based on regression models linking statistics related to the tail of the underlying distribution function to the extreme-value index and parameters describing the second order tail behaviour. The optimal number of extremes (in case $\gamma > 0$) or *UH* statistics (in case $\gamma \in \mathbb{R}$) was derived from the trace of the AMSE matrix. It appears from the simulation results that combining data from several groups can lead to significant improvements in the estimation of the extreme-value index. A drawback of using a common k-value is that the procedure can run into difficulties when the design is severely unbalanced. However, this problem is easily remedied by using a common relative tail sample fraction. Future work will concentrate on the further extension of the proposed methods towards the estimation of other tail characteristics such as extreme quantiles or small exceedance probabilities.

ACKNOWLEDGMENTS

The authors would like to thank R. Verlaak and P. Vynckier (Aon Re Belgium) for providing the claim data and V.F. Pisarenko and D. Sornette for providing the earthquake data. The authors also take pleasure in thanking the referee for providing useful and interesting comments.

REFERENCES

[1] Beirlant, J.; Dierckx, G.; Goegebeur, Y. and Matthys, G. (1999). Tail index estimation and an exponential regression model, *Extremes*, **2**, 177–200.

[2] Beirlant, J.; Dierckx, G. and Guillou, A. (2000). *Estimation of the extreme value index and regression on generalized quantile plots*, Technical Report, Department of Mathematics, K.U. Leuven.

[3] Beirlant, J.; Dierckx, G.; Guillou, A. and Stărică, C. (2002). On exponential representations of log-spacings of extreme order statistics, *Extremes*, **5**, 157–180.

[4] Beirlant, J.; Vynckier, P. and Teugels, J.L. (1996). Excess functions and estimation of the extreme-value index, *Bernoulli*, **2**, 293–318.

[5] Bingham, N.H.; Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*, Cambridge University Press, Cambridge.

[6] Burr, I.W. (1942). Cumulative frequency functions, *Annals of Mathematical Statistics*, **13**, 215–232.

[7] Coles, S. and Tawn, J. (1998). *Statistical Methods for Extreme Values*, A course presented at the 1998 RSS conference, Strathclyde, September, 1998.

[8] Csörgő, S.; Deheuvels, P. and Mason, D.M. (1985). Kernel estimates of the tail index of a distribution, *Annals of Statistics*, **13**, 1050–1077.

[9] Davison, A.C. and Smith, R.L. (1990). Models for exceedances over high thresholds, *Journal of the Royal Statistical Society B*, **52**, 393–442.

[10] Dekkers, A.L.M.; Einmahl, J.H.J. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution, *Annals of Statistics*, **17**, 1833–1855.

[11] Feuerverger, A. and Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution, *Annals of Statistics*, **27**, 760–781.

[12] Fisher, R.A. and Tippett, L.H.C. (1928). Limiting forms of the frequency distribution in the largest particle size and smallest number of a sample, *Proceedings of the Cambridge Philosophical Society*, **24**, 180–190.

[13] Gumbel, E.J. (1967). *Statistics of Extremes*, Columbia University Press.

[14] Hall, P. (1982). On some simple estimates of an exponent of regular variation, *Journal of the Royal Statistical Society B*, **44**, 37–42.

[15] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *Annals of Statistics*, **3**, 1163–1174.

[16] Kratz, M. and Resnick, S. (1996). The qq-estimator of the index of regular variation, *Communications in Statistics: Stochastic Models*, **12**, 699–724.

[17] Pickands III, J. (1975). Statistical inference using extreme order statistics, *Annals of Statistics*, **3**, 119–131.

[18] Pisarenko, V.F. and Sornette, D. (2003). Characterization of the frequency of extreme earthquake events by the generalized Pareto distribution, *Pure and Applied Geophysics*, **160**, 2343–2364.

[19] Schultze, J. and Steinebach, J. (1996). On least squares estimates of an exponential tail coefficient, *Statistics and Decisions*, **14**, 353-372.

[20] Simiu, E.; Changery, M.J. and Filliben, J.J. (1979). *Extreme Wind Speeds at 129 Stations in the Contiguous United States*, NBS Building Science Series 118, National Bureau of Standards (U.S.), Washington, DC.

[21] Simiu, E. and Heckert, N.A. (1995). *Extreme wind distribution tails: a peak over threshold approach*, NIST, Gaitersburg.

[22] Smith, R.L. (1985). Maximum likelihood estimation in a class of nonregular cases, *Biometrika*, **72**, 67–90.

[23] Smith, R.L. (1987). Estimating tails of probability distributions, *Annals of Statistics*, **15**, 1174–1207.

REVSTAT – Statistical Journal
Volume 2, Number 1, June 2004

CHANGES OF STRUCTURE IN FINANCIAL TIME SERIES AND THE GARCH MODEL

Authors: Thomas Mikosch*
– Laboratory of Actuarial Mathematics, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, and MaPhySto, the Danish Network for Mathematical Physics and Stochastics, Denmark (mikosch@math.ku.dk)

Cătălin Stărică
– Department of Mathematical Statistics, Chalmers University of Technology, S–412 96 Gothenburg, Sweden (starica@math.chalmers.se)

Received: November 2003 Accepted: December 2003

Abstract:

- In this paper we propose a goodness of fit test that checks the resemblance of the spectral density of a GARCH process to that of the log-returns. The asymptotic behavior of the test statistics are given by a functional central limit theorem for the integrated periodogram of the data. A simulation study investigates the small sample behavior, the size and the power of our test. We apply our results to the S&P500 returns and detect changes in the structure of the data related to shifts of the unconditional variance. We show how a long range dependence type behavior in the sample ACF of absolute returns might be induced by these changes.

Key-Words:

- *integrated periodogram; spectral distribution; functional central limit theorem; Kiefer–Müller process; Brownian bridge; sample autocorrelation; change point; GARCH process; long range dependence; IGARCH; non-stationarity.*

AMS Subject Classification:

- Primary: 62P20; Secondary: 60G10 60F17 60B12 60G15 62M15.

*This research supported in part by a grant of the Dutch Science Foundation (NWO), by DYNSTOCH, a research training network under the programme Improving Human Potential financed by The 5$^{\text{th}}$ Framework Programme of the European Commission, the Danish Research Council Grant No 21-01-0546, and MaPhySto, the Danish Network for Mathematical Physics and Stochastics, funded by The Danish National Research Foundation.

1. INTRODUCTION

In this paper we introduce a goodness of fit test for the GARCH process. In its simplest form this model is given by

$$\begin{cases} X_t = \sigma_t\, Z_t\,, \\ \sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + \alpha_1 X_{t-1}^2\,, \end{cases} \qquad t \in \mathbb{Z}\,, \tag{1.1}$$

where $(Z_t)_{t\in\mathbb{Z}}$ is a sequence of iid random variables with $EZ_1 = 0$, $EZ_1^2 = 1$. The parameters α_1 and β_1 are non-negative and α_0 is necessarily positive.

Our test decides if the data at hand is a white noise whose squares have a covariance structure which is in agreement with the second order structure of the hypothesized squared GARCH process. The test is related to the classical Grenander–Rosenblatt or Bartlett goodness of fit tests for the spectral distribution of a time series; see for example Priestley [38]. Such tests are analogues to the Kolmogorov–Smirnov test for the distribution of a sample. Other testing procedures exist in the literature. Among them we mention the approach that uses the sequential empirical process for the residuals of an ARCH process; see Horváth et al. [26]. Besides being restricted to the ARCH case, these asymptotic tests present another drawback. The limit distribution of the test statistic depends in general on the distribution of the noise Z_t and the parameters of the model. An advantage of the test proposed in this paper is that the limit distribution of the test statistic is distribution free and, as in the Kolmogorov-Smirnov test, is a function of the Brownian bridge. Moreover, we prove that the limit distribution of our test statistics are insensitive to the replacement of the parameters by their estimators under the null hypothesis.

Although attractive as a model, there is copious empirical evidence in the econometrics literature, coming especially from the analysis of long series of log-returns, that argues against the GARCH$(1,1)$ model. For example, although the squares of a GARCH$(1,1)$ process follow the dynamics of an ARMA process (in particular the ACF goes to zero exponentially fast), the sample ACFs of the absolute values and their squares tend to stabilize around a positive value for larger lags (the so-called long range dependence in absolute returns or in volatility). For longer samples the estimated parameters α_1 and β_1 sum up to values close to 1 (Engle and Bollerslev [19], Mikosch and Stărică [35]). This fact, known as the integrated GARCH finding, implies infinite variance (see Bollerslev [8]) for the returns, a conclusion in strong disagreement with the accepted results of semi-parametric tail analysis that find at least a finite third moment (Embrechts et al. [17]).

The second contribution of the paper is the analysis based on our goodness of fit procedure of a long portion of the S&P500 log-return series (January 1953 to December 1990) which could provide an answer to these critics. For the data under investigation we detect structural changes related to movements of the unconditional variance and show how a long range dependence type behavior in the sample ACF of absolute returns might be induced by these shifts. Our procedure identifies most of the recessions of this period as being structurally different. The major structural change is detected between 1973 and 1975 and seems to correspond to the oil crises. Our analysis seems to indicate that one simple GARCH(1,1) process (which models the first ten years of the data quite well) cannot describe the complicated dynamics of longer, possibly non-stationary log-return time series.

Our paper is organized as follows. In Section 2 we formulate our main theoretical result, a functional central limit theorem for the integrated periodogram of the GARCH process. Then we indicate how this result can be used to build an asymptotic goodness of fit test for the spectral distribution of the GARCH process. We also discuss the behavior of the test statistics under the alternative hypothesis of a different GARCH process. The proofs are rather technical and therefore postponed to Appendix A1. In Section 3 we investigate by means of simulations the small sample properties, the size and the power of our test statistic while in Section 4 we apply our method to the study of a long portion of the S&P log-return series. Some concluding remarks are given in Section 5.

2. LIMIT THEORY FOR THE TWO-PARAMETER INTEGRATED PERIODOGRAM

In fields as diverse as time series analysis and extreme value theory it is generally assumed that the observations or a suitable transformation of them constitute a stationary sequence of random variables. In the context of this section, stationarity is always understood as strict stationarity. One of the aims of this paper is to provide a procedure for testing how good the fit of a stationary GARCH(p,q) model to data is. This section provides the limit theory for a certain two-parameter process which is the basis for the statistical procedure we propose in Section 2.2. This theory is slightly more general than needed for the purposes of this paper. However, the theory for the corresponding one-parameter process (which will be used intensively in the rest of the paper) is essentially the same as for the case of two parameters. The latter case can be used for change point detection in the spectral domain while the former one yields goodness of fit tests. As already mentioned, in the context of this paper, we are mainly interested in test statistics for the goodness of fit of GARCH processes. The statistical procedure will allow us to single out the parts of the data which are not well described by the hypothesized model.

To be precise, we assume that the data come from a stationary generalized autoregressive conditionally heteroscedastic process of order (p,q), for short GARCH(p,q):

$$X_t = \sigma_t Z_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-i}^2 , \qquad t \in \mathbb{Z} , \tag{2.1}$$

where (Z_t) is an iid symmetric sequence with $EZ^2 = 1$, non-negative parameters α_i and β_j, and the stochastic volatility σ_t is independent of Z_t for every fixed t. We also assume that Z_1 has a Lebesgue density on the real line. This ensures that (X_t) is α-mixing with geometric rate; see Boussama [11]. In what follows, we write σ for a generic random variable with the distribution of σ_1, X for a generic random variable with the distribution of X_1, etc.

This kind of model is most popular in the econometrics literature for modeling the log-returns of stock indices, share prices, exchange rates, etc., and has found its way into the practice of forecasting financial time series. See for example Engle [18] for a collection of papers on ARCH. We assume that, for a particular choice of parameters α_i and β_i, the sequence $((X_t, \sigma_t))$ is stationary. Assumptions for stationarity of a GARCH process can be found in Bougerol and Picard [10] for the general GARCH(p,q) case and in Nelson [36] for the GARCH$(1,1)$ case. For a recent overview on the mathematics of GARCH processes, we refer to Mikosch [32].

Our analysis is based on the spectral properties of the underlying time series. Consider the classical estimator of the spectral density, the periodogram, given by

$$I_{n,X}(\lambda) = \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathrm{e}^{-i\lambda t} X_t \right|^2 , \qquad \lambda \in [0,\pi] .$$

Under general conditions, the integrated periodogram or empirical spectral distribution function

$$\frac{1}{2\pi} J_{n,X}(\lambda) = \frac{1}{2\pi} \int_0^\lambda I_{n,X}(x)\, dx , \qquad \lambda \in [0,\pi] , \tag{2.2}$$

is a consistent estimator of the spectral distribution function given by

$$F_X(\lambda) = \int_0^\lambda f_X(x)\, dx , \qquad \lambda \in [0,\pi] ,$$

provided the spectral density f_X is well defined.

2.1. Main results

As a motivation for our main result, we start by considering the two-parameter process $J_{n,X}(x,\lambda)$ related to (2.2) (see also Appendix A1):

$$(2.3)\quad \begin{aligned} J_{n,X}(x,\lambda) &= \int_0^\lambda \left(\gamma_{n,[nx],X}(0) + 2\sum_{h=1}^{[nx]-1} \gamma_{n,[nx],X}(h)\cos(yh)\right) dy \\ &= \lambda\,\gamma_{n,[nx],X}(0) + 2\sum_{h=1}^{[nx]-1} \gamma_{n,[nx],X}(h)\,\frac{\sin(\lambda h)}{h}\ , \end{aligned}$$

where

$$\gamma_{n,[nx],X}(h) = \frac{1}{n}\sum_{t=1}^{[nx]-h} X_t\,X_{t+h}\ , \qquad h=0,1,2,\ldots,[nx]-1, \quad x\in[0,1]\ .$$

Clearly,

$$\gamma_{n,X}(h) := \gamma_{n,n,X}(h)$$

denotes a version of the sample autocovariance at lag h; the standard version of the sample autocovariance is defined for the centered random variables $X_t - \overline{X}_n$, where $\overline{X}_n$ is the sample mean. We also write

$$\gamma_X(h) = \mathrm{cov}(X_0,X_h) \quad \text{and} \quad v_X(h) = \mathrm{var}(X_0X_h) = E(X_0X_h)^2\ , \qquad h\in\mathbb{Z}\ .$$

The processes $\gamma_{n,[n\cdot],X}(h)$ satisfy a fairly general functional central limit theorem (FCLT). Recall that $\mathbb{D}([0,1],\mathbb{R}^m)$ is the Skorokhod space of $\mathbb{R}^m$-valued cadlag functions on $[0,1]$ (continuous from the right in $[0,1)$, limits exist from the left in $(0,1]$) endowed with the J_1-topology and the corresponding Borel σ-field; see for example Jacod and Shiryaev [27] or Bickel and Wichura [6].

Lemma 2.1. *Consider the* GARCH(p,q) *process* (X_t) *given by* (2.1). *Assume that*

$$(2.4)\qquad E|X|^{4+\delta} < \infty \qquad \text{for some}\quad \delta > 0\ .$$

Then for every $m\ge 1$, *as* $n\to\infty$

$$(2.5)\qquad \sqrt{n}\Big(\gamma_{n,[nx],X}(h),\ h=1,\ldots,m\Big)_{x\in[0,1]} \xrightarrow{d} \Big(v_X^{1/2}(h)\,W_h(x),\ h=1,\ldots,m\Big)_{x\in[0,1]}\ ,$$

in $\mathbb{D}([0,1],\mathbb{R}^m)$, *where* $W_h(\cdot)$, $h=1,\ldots,m$, *are iid standard Brownian motions on* $[0,1]$.

The proof of the lemma is given in Appendix A1. A naive argument, based on Lemma 2.1 and the decomposition (2.3), suggests that

$$\sqrt{n}\Big(J_{n,X}(x,\lambda) - \lambda\,\gamma_{n,[nx],X}(0)\Big)_{x\in[0,1],\ \lambda\in[0,\pi]} \xrightarrow{d}$$

$$\xrightarrow{d} 2\left(\sum_{h=1}^{\infty} v_X^{1/2}(h)\, W_h(x)\, \frac{\sin(\lambda h)}{h}\right)_{x\in[0,1],\ \lambda\in[0,\pi]},$$

in $\mathbb{D}\big([0,1]\times[0,\pi]\big)$. This result can be shown to be true; one can follow the lines of the proof of Theorem 2.1 below. However, the two-parameter Gaussian limit field has a distribution that explicitly depends on the covariance structure of (X_t^2), which is not a very desirable property. Indeed, since we are interested in using functionals of the limit process for a goodness of fit procedure, we would like that the asymptotic distribution of those functionals is independent of the null hypothesis we test. In other words, we want a "standard" Gaussian process in the limit since otherwise we would have to evaluate the distributions of its functionals by Monte–Carlo simulations for every choice of parameters of the GARCH(p,q) we consider in the null hypothesis.

A glance at the right-hand side of (2.3) suggests another approach. The dependence of the limiting Gaussian field on the covariance structure of (X_t^2) comes in through the FCLT of Lemma 2.1. However, it is intuitively clear that, if we replaced in (2.3) the processes $\gamma_{n,[n\cdot],X}(h)$ by

$$\widetilde{\gamma}_{n,[n\cdot],X}(h) = \frac{\gamma_{n,[n\cdot],X}(h)}{v_X^{1/2}(h)},$$

the limit process would become independent of the covariance structure of (X_t^2).

Therefore we introduce the following two-parameter process which is a straightforward modification of $J_{n,X}(x,\lambda)$:

$$C_{n,X}(x,\lambda) = \sum_{h=1}^{[nx]-1} \widetilde{\gamma}_{n,[nx],X}(h)\, \frac{\sin(\lambda h)}{h}, \qquad x\in[0,1], \quad \lambda\in[0,\pi].$$

Our main result is a FCLT for $C_{n,X}$.

Theorem 2.1. *Let* (X_t) *be a stationary* GARCH(p,q) *process given by* (2.1). *Assume that* (2.4) *holds. Then*

$$\sqrt{n}\big(C_{n,X}(x,\lambda)\big)_{x\in[0,1],\ \lambda\in[0,\pi]} \xrightarrow{d}$$

$$\xrightarrow{d} \big(K(x,\lambda)\big)_{x\in[0,1],\ \lambda\in[0,\pi]} = \left(\sum_{h=1}^{\infty} W_h(x)\, \frac{\sin(\lambda h)}{h}\right)_{x\in[0,1],\ \lambda\in[0,\pi]}, \tag{2.6}$$

in $\mathbb{D}\big([0,1]\times[0,\pi]\big)$ *where* $(W_h(\cdot))_{h=1,\ldots}$ *is a sequence of iid standard Brownian motions on* $[0,1]$*. The infinite series on the right-hand side converges with probability* 1 *and represents a Kiefer–Müller process, i.e., a two-parameter Gaussian field with covariance structure*

$$
\begin{aligned}
E\Big(K(x_1,\lambda_1)\,K(x_2,\lambda_2)\Big) &= \min(x_1,x_2)\sum_{t=1}^{\infty}\frac{\sin(\lambda_1 t)\,\sin(\lambda_2 t)}{t^2}\\
&= 2^{-1}\,\pi^2\min(x_1,x_2)\left(\min\Big(\frac{\lambda_1}{\pi},\frac{\lambda_2}{\pi}\Big)-\frac{\lambda_1}{\pi}\,\frac{\lambda_2}{\pi}\right)\,.
\end{aligned}
\tag{2.7}
$$

The proof of the theorem is given in Appendix A1.

The series representation of the Kiefer–Müller process can be found in Klüppelberg and Mikosch [29]. This process is known in empirical process theory as the limiting Gaussian field for the sequential empirical process; see Shorack and Wellner [41].

Remark 2.1. The statement of Theorem 2.1 remains valid for wider classes of stationary sequences. In particular the result holds if the conditions in Remark 1.1 are satisfied and in addition, (X_t) is symmetric and $(|X_t|)$ and $(\mathrm{sign}(X_t))$ are independent. The latter conditions are satisfied by any stochastic volatility model of the form $X_t=\sigma_t\,Z_t$, where (Z_t) is a sequence of iid symmetric random variables and the random variables σ_t are adapted to the filtration $\sigma(Z_{t-1},Z_{t-2},...)$, or alternatively, (σ_t) and (Z_t) are independent.

Immediate consequences of Theorem 2.1 and the continuous mapping theorem are limit theorems for continuous functionals of the process $C_{n,X}$ which can be used for the construction of goodness of fit tests and tests for detecting changes in the spectrum of the time series.

Corollary 2.1. *Under the assumptions of Theorem 2.1,*

$$
\begin{aligned}
\sqrt{n}\sup_{x\in[0,1],\ \lambda\in[0,\pi]}\big|C_{n,X}(x,\lambda)\big| &\stackrel{d}{\longrightarrow} \sup_{x\in[0,1],\lambda\in[0,\pi]}\big|K(x,\lambda)\big|\,,\\
n\int_0^1\!\!\int_0^{\pi} C_{n,X}^2(x,\lambda)\,dx\,d\lambda &\stackrel{d}{\longrightarrow} \int_0^1\!\!\int_0^{\pi} K^2(x,\lambda)\,dx\,d\lambda\,.
\end{aligned}
$$

For $x=1$, convergence in (2.6) yields

$$
\sqrt{n}\,\widetilde{C}_{n,X}(\cdot) := \sqrt{n}\sum_{h=1}^{n-1}\frac{\gamma_{n,X}(h)}{v_X^{1/2}(h)}\,\frac{\sin(\cdot\,h)}{h}\;\stackrel{d}{\longrightarrow}\; B(\cdot) := \sum_{h=1}^{\infty}W_h(1)\,\frac{\sin(\cdot h)}{h}\,,
\tag{2.8}
$$

in $\mathbb{C}\,[0,\pi]$. The series on the right-hand side is the so-called Paley–Wiener representation of a Brownian bridge on $[0,\pi]$; see (2.7) with $x=1$ (see for example Hida [25]).

The one-parameter process $\widetilde{C}_{n,X}$ will be our basic process for testing the goodness of fit of the sample $X_1, ..., X_n$ to a GARCH process. The convergence of the following functionals can be used for constructing Kolmogorov–Smirnov and Cramér–von Mises type goodness of fit tests for a GARCH(p, q) process.

Corollary 2.2. *Under the assumptions of Theorem 2.1,*

$$\widetilde{S}_n := \sqrt{n} \sup_{\lambda\in[0,\pi]} \left|\widetilde{C}_{n,X}(\lambda)\right| \xrightarrow{d} \sup_{\lambda\in[0,\pi]} |B(\lambda)| \,, \tag{2.9}$$

$$n \int_0^\pi \widetilde{C}^2_{n,X}(\lambda)\, d\lambda \xrightarrow{d} \int_0^\pi B^2(\lambda)\, d\lambda \,.$$

2.2. The goodness of fit test

In what follows, we focus on the GARCH$(1, 1)$ case but a similar theory can be developed for the general GARCH(p, q) case. The quantities $v_X(h)$ are continuous functions of the GARCH parameters and the fourth moments of the iid noise Z_t. We refer to Appendix A2 where v_X is explicitly given for the GARCH$(1, 1)$ case. For an application of the results above it is natural to replace the unknown quantities $v_X(h)$ in the definition of $\widetilde{\gamma}_{n,k,X}(h)$ by their sample versions $\widehat{v}_X(h)$, i.e., the parameters α_i and β_1 are replaced by some estimators $\widehat{\alpha}_i$ and $\widehat{\beta}_1$ and EZ^4 is replaced by the sample mean of the 4^{th} powers of the residuals $\widehat{EZ^4} = n^{-1} \sum_{i=1}^n \widehat{Z}_t^4$, where $\widehat{Z}_t = X_t/\widehat{\sigma}_t$ and $\widehat{\sigma}_t^2 = \widehat{\alpha}_0 + \widehat{\alpha}_1 X_{t-1}^2 + \widehat{\beta}_1 \widehat{\sigma}_{t-1}^2$ and $\widehat{\sigma}_0^2$ and X_0^2 are arbitrarily chosen, but fixed. Denoting by

$$\widehat{\gamma}_{n,[n\cdot],X}(h) \;=\; \frac{\gamma_{n,[n\cdot],X}(h)}{\widehat{v}_X^{1/2}(h)} \,,$$

we produce the straightforward modification of $C_{n,X}(x, \lambda)$:

$$\widehat{C}_{n,X}(x, \lambda) \;=\; \sum_{h=1}^{[nx]-1} \widehat{\gamma}_{n,[nx],X}(h)\, \frac{\sin(\lambda h)}{h} \,, \qquad x \in [0, 1], \quad \lambda \in [0, \pi] \,,$$

and that of $\widetilde{S}_n$:

$$S_n \;:=\; \sqrt{n} \sup_{\lambda\in[0,\pi]} \left|\widehat{C}_{n,X}(\lambda)\right| \,. \tag{2.10}$$

The following result states that the theory developed in this section remains valid if v_X is replaced by its sample analogue.

Theorem 2.2. *Assume that the parameter estimators $\widehat{\alpha}_1$ and $\widehat{\beta}_1$ based on $X_1, ..., X_n$ are independent of* $(\mathrm{sign}(X_t))$ *and consistent, i.e., $\widehat{\alpha}_1 \stackrel{P}{\to} \alpha_1$ and $\widehat{\beta}_1 \stackrel{P}{\to} \beta_1$. Then Theorem 2.1 and its corollaries remain valid for a* GARCH(1,1) *process if v_X is replaced by its sample analogue $\widehat{v}_X$. In particular*

$$S_n \stackrel{d}{\longrightarrow} \sup_{\lambda\in[0,\pi]} |B(\lambda)| \,. \tag{2.11}$$

Remark 2.2. The Whittle parameter estimators of a GARCH process are consistent if $EX^4 < \infty$, and so are the Gaussian quasi maximum likelihood estimators; see Giraitis and Robinson [22] and Mikosch and Straumann [33] for the former case and Berkes et al. [5] for the latter case. Moreover, by their definitions they are calculated from the X_t^2's and σ_t^2's only and therefore they are independent of $(\mathrm{sign}(X_t))$.

The results we presented so far are sufficient for providing a theoretical understanding of the behavior of tests based on functionals of $\widehat{C}_{n,X}$ (for example S_n). These are tests of the null hypothesis that the sample $X_1, ..., X_n$ comes from a GARCH(1,1) model with given parameters α_i and β_i against the alternative of another GARCH(1,1) model with parameters α_i^a, $i = 0, 1$, and β_1^a. They reject the null hypothesis if the functional is in a certain region. The rejection region giving the test the right size is constructed based on the quantiles of the appropriate functional of the limit process in Theorem 2.2 (i.e. the supremum of a Brownian bridge in the case of the statistic S_n). As for the power of the test, similar arguments as for the proof of Theorem 2.2 yield under the alternative the following result.

Theorem 2.3. *Assume that (X_t) and (Y_t) are two stationary* GARCH(1,1) *processes* (2.1) *with coefficients α_i, $i=0,1$, β_1, and α_i^a, $i=0,1$, β_1^a respectively. Assume that the parameter estimators $\widehat{\alpha}_1$ and $\widehat{\beta}_1$ (based on the sample $X_1, ..., X_n$) are independent of* $(\mathrm{sign}(X_t))$ *and consistent, i.e., $\widehat{\alpha}_1 \stackrel{P}{\to} \alpha_1$ and $\widehat{\beta}_1 \stackrel{P}{\to} \beta_1$. Define*

$$\widehat{\gamma}^a_{n,[n\cdot]}(h) = \frac{\gamma_{n,[n\cdot],Y}(h)}{\widehat{v}_X^{1/2}(h)}\,,$$

$$\widehat{C}^a_{n,X,Y}(x,\lambda) = \sum_{h=1}^{[nx]-1} \widehat{\gamma}^a_{n,[nx]}(h)\,\frac{\sin(\lambda h)}{h}\,, \qquad x\in[0,1], \quad \lambda\in[0,\pi]\,.$$

Then

$$\sqrt{n}\left(\widehat{C}^a_{n,X,Y}(x,\lambda)\right)_{x\in[0,1],\,\lambda\in[0,\pi]} \stackrel{d}{\longrightarrow} \left(\sum_{h=1}^{\infty} \frac{v_X^{a\;1/2}(h)}{v_X^{1/2}(h)}\, W_h(x)\,\frac{\sin(\lambda h)}{h}\right)_{x\in[0,1],\,\lambda\in[0,\pi]}, \tag{2.12}$$

in $\mathbb{D}\left([0,1]\times[0,\pi]\right)$ where $(W_h(\cdot))$ is a sequence of iid standard Brownian motions on $[0,1]$ while $v_X^a(h) = E(Y_0Y_h)^2$.

This result yields a theoretical description of the power of tests based on functionals of $\widehat{C}_{n,X}$. It individuates the functional of the Gaussian process on the right hand side of equation (2.12) as an asymptotic equivalent of the desired functional of $\widehat{C}^a_{n,X,Y}$. Note that the distributions of the functionals of the limit processes depend on the parameters of the alternative hypothesis in a rather complicated way. This makes the direct use of the Theorem 2.3 in applications rather difficult. For this reason we will rely on Monte–Carlo based calculations of the distribution of $\widehat{C}^a_{n,X,Y}$. See Section 3 for a simulation study on the size and the power of a test based on the statistics S_n.

Results similar to Theorem 2.3 can be derived for the alternative hypothesis that the sample $X_1, ..., X_n$ consists of subsamples from different GARCH(p, q) processes. Clearly, the asymptotic distribution will be even more complex and the Monte–Carlo approach again inevitable.

3. A SIZE AND POWER MONTE CARLO STUDY

The aim of this section is to investigate the size and the power of a test based on the statistic S_n in (2.10). The set up is relevant to the real data analysis performed in Section 4. There we check the goodness of fit of a GARCH$(1, 1)$ process with parameters estimated on the first 3 years of data (750 observations)

$$\alpha_0 = 8.58 \times 10^{-6}, \quad \alpha_1 = 0.072, \quad \beta_1 = 0.759, \quad \nu = 5.24\ , \tag{3.1}$$

to various segments of the data set. Here ν is the number of degrees of freedom of the t-distributed noise sequence (Z_t). The corresponding value of the fourth moment of the estimated residuals of the model 3.1 is $EZ^4 = 7.82$.

The first choice we need to make when applying our test on data is precisely that of the size of the window that guarantees a correct behavior of the statistic S_n. Theoretically, the correct size of the test will be guaranteed by a choice of the rejection region based on the asymptotic behavior of S_n described by Theorem 2.2. These results are only asymptotic and provide the right size if the data window used to calculate the S_n statistic is large. For reasons that are explained in Section 4, we want to keep the length of the window as small as possible. It is by means of simulations that we find the right balance between these opposing requirements on the window size. As a byproduct of the simulation study, we will understand how to adjust the interval provided by Theorem 2.2 in order to maintain the correct size.

The top graph in Figure 1 displays the QQ-plot of 1000 simulated values of $\widetilde{S}_{125}$ (the quantiles on the x-axis), calculated on samples of 125 observations from a GARCH process with Student-t innovations and parameters (3.1) against

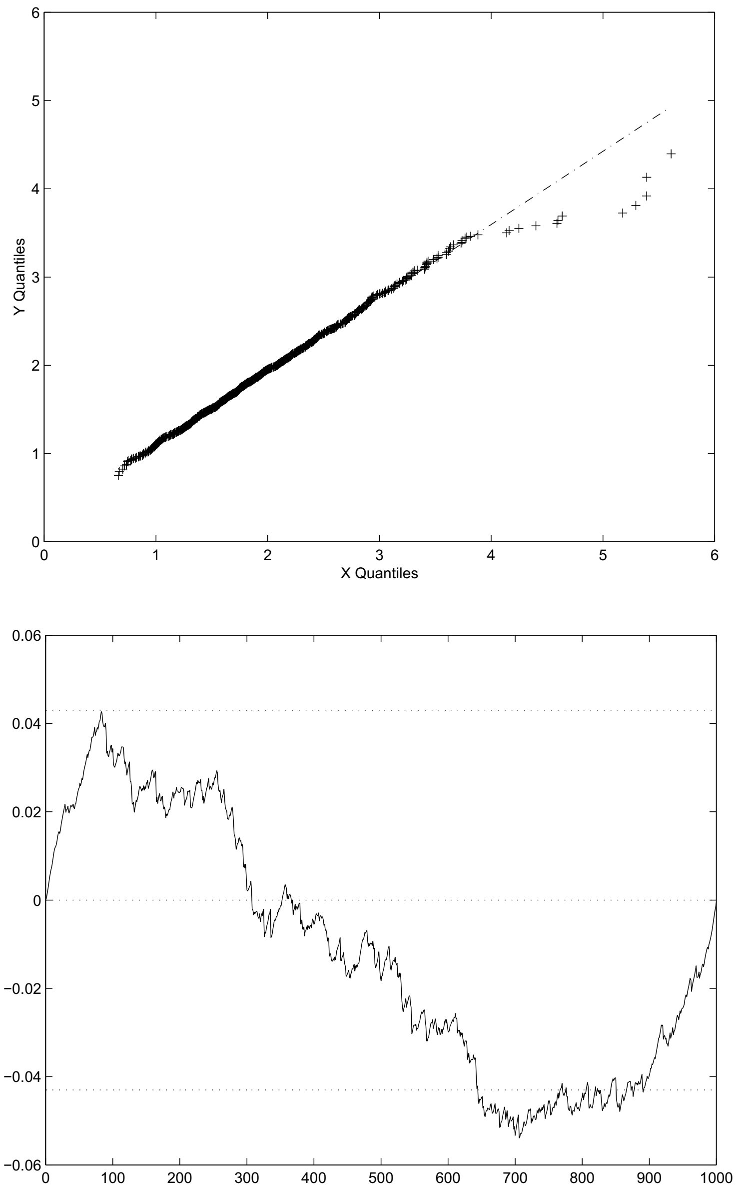


Figure 1: *Top:* QQ-plot of 1000 values of $\widetilde{S}_{125}$ (x-axis) against the quantiles of the supremum of a Brownian bridge (y-axis).
Bottom: The difference between the sample cdf of 1000 simulated values of $\widetilde{S}_{125}$ and the theoretical cdf of the supremum of a Brownian bridge with the Kolmogorov-Smirnov 95% confidence bands.

the quantiles of the supremum of a Brownian bridge (on the y-axis). The bottom graph in the same figure together with the graphs in Figure 2 shows the goodness of fit of the distribution of the supremum of a Brownian bridge to samples of 1000 simulations of $\widetilde{S}_{125}$ (Figure 1), $\widetilde{S}_{500}$ and $\widetilde{S}_{1000}$ (Figure 2) respectively. The statistic

Figure 2: The difference between the sample cdf of the 1000 simulated values of $\widetilde{S}_{500}$ *(Top)* and of $\widetilde{S}_{1000}$ *(Bottom)* and the theoretical cdf of the supremum of a Brownian bridge with the Kolmogorov-Smirnov 95% confidence bands.

$\widetilde{S}$ is calculated using the parameters (3.1). The goodness of fit is based on the Kolmogorov–Smirnov test. The solid line in these graphs represents the difference between the sample cdf and the theoretical cdf of a Brownian bridge while the dotted lines are the 95% confidence intervals stipulated by the Kolmogorov–Smirnov test. This test seems to indicate that the asymptotic behavior fully works for sample sizes of the order 1000 while the qualitative differences between sample sizes of order 125 and 500 are not too big. This observation together with the good fit showed by the QQ-plot in Figure 1 motivate our choice of a window size of 125 data points (or half a business year).

A next issue that we need to clarify is the behavior of S_{125}. Recall that Theorem 2.2 stipulates that the asymptotic behaviors of S_n and $\widetilde{S}_n$ are the same. A verification of this statement is provided in Figure 3 which displays the QQ-plot of 2500 simulated values of S_{125} against 2500 simulated values of $\widetilde{S}_{125}$. In all cases the data generating process is a GARCH model with Student-t innovations and parameters (3.1). To obtain one value of $\widetilde{S}_{125}$, 125 simulated data and the true parameters (3.1) are used, while in the case of S_{125}, 875 data points are simulated, the parameters are estimated on the first 750 data points and the last 125 observations together with the estimated parameters are used to produce the statistic. The two distributions seem indeed very close to each other.

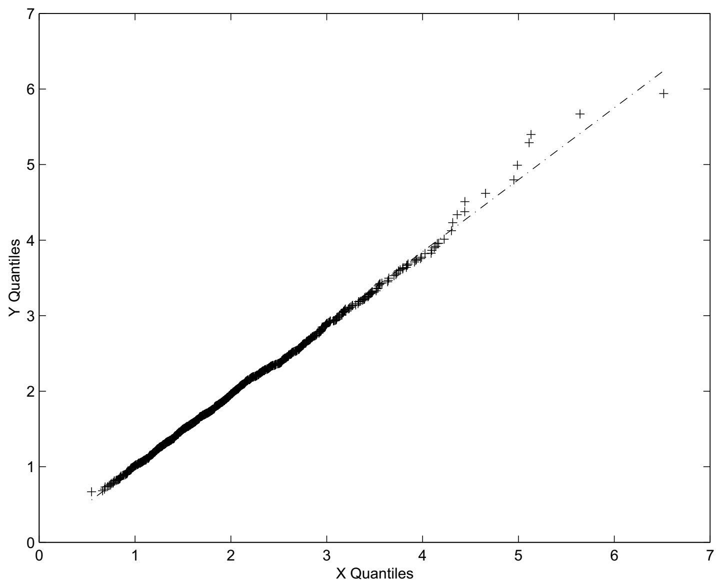


Figure 3: QQ-plot of 2500 values of S_{125} (x-axis) against 2500 values of $\widetilde{S}_{125}$ (y-axis).

This part of the simulation study serves also to define the rejection regions of the test based on the statistic S_{125}. The rejection intervals for a 95% size one-sided, respectively two-sided test for the S_{125} statistic are $(0, 1.01)$ and $(3.32, \infty)$, and $(0, 0.92) \cup (3.8, \infty)$ respectively. The interval that gives a size of 99% to our two-sided test (and that will be used for the data analysis in the next section) is $(0, 0.785) \cup (4.9, \infty)$.

The alternatives we consider in our study are those of GARCH processes with parameters different from those in (3.1). Although Theorem 2.3 gives the theoretical power of our test against various GARCH alternatives, its complicated form renders it of little practical help. For understanding the behavior of the S_n statistic under various GARCH alternatives we again have to turn to a Monte Carlo analysis.

The results are displayed in Figures 4–7. The tests have size 95% and are based on the choice of intervals given by our simulation study as discussed above. In Figures 4 and 5 the parameters α_0 and EZ^4 are kept constant while the other two parameters α_1 and β_1 are made to vary between 0.04 and 0.14 and 0.6 and 0.92 respectively. In Figures 6 and 7 the parameters α_1 and β_1 are kept constant while the other two parameters α_0 and EZ^4 are made to vary between 1.5×10^{-6} and 4.05×10^{-5} and 3 and 9 respectively. For every alternative, 500 simulations were produced. The top and center graphs in Figures 4 and 6 display the power of the one-sided tests (for the top graph, the rejection interval is $(0, 1.01)$, for the center graph $(3.32, \infty)$) while the graphs on the bottom row display the difference between the standard deviation of the alternative models and that of the model with parameters (3.1). The top graphs in Figures 5 and 7 show the power of the two-sided test (rejection region $(0, 0.785) \cup (4.9, \infty)$) while the bottom graphs in the two pictures display the absolute value of the difference between the log of the standard deviation of the alternative models and that of the model with parameters (3.1).

The graphs in Figures 4 and 6 shed light on the relationship between the difference of the unconditional variances and the distribution of the S_{125} statistic under the alternative. They show that the sampling distribution of the statistic S_{125} (calculated with the parameters of the null hypothesis) for GARCH models with lower (higher) unconditional variance is dominated (dominates) the sampling distribution for the null model. Hence rejecting for small (high) values of the statistic S_{125} gives power against alternative models with smaller (larger) unconditional variance. The graphs in Figures 5 and 7 show a strong connection between the power of the test and the absolute value of the difference of the log unconditional variances of the two models. The higher the size of the difference, the higher the power. Even more, Figures 6 and 7 show that the test has equal power against alternatives of equal variance. Note that the variance of the alternative GARCH$(1, 1)$ processes does not depend on the EZ^4 parameter.

As a conclusion, the study motivates the interpretation of the rejection of the null hypothesis not only as signaling the need for another GARCH model but also as a clear sign of a change in the unconditional variance of the time series. More concretely, a rejection on the upper (lower) end of the rejection region also signals an increase (a decrease, respectively) in the unconditional variance of the time series.

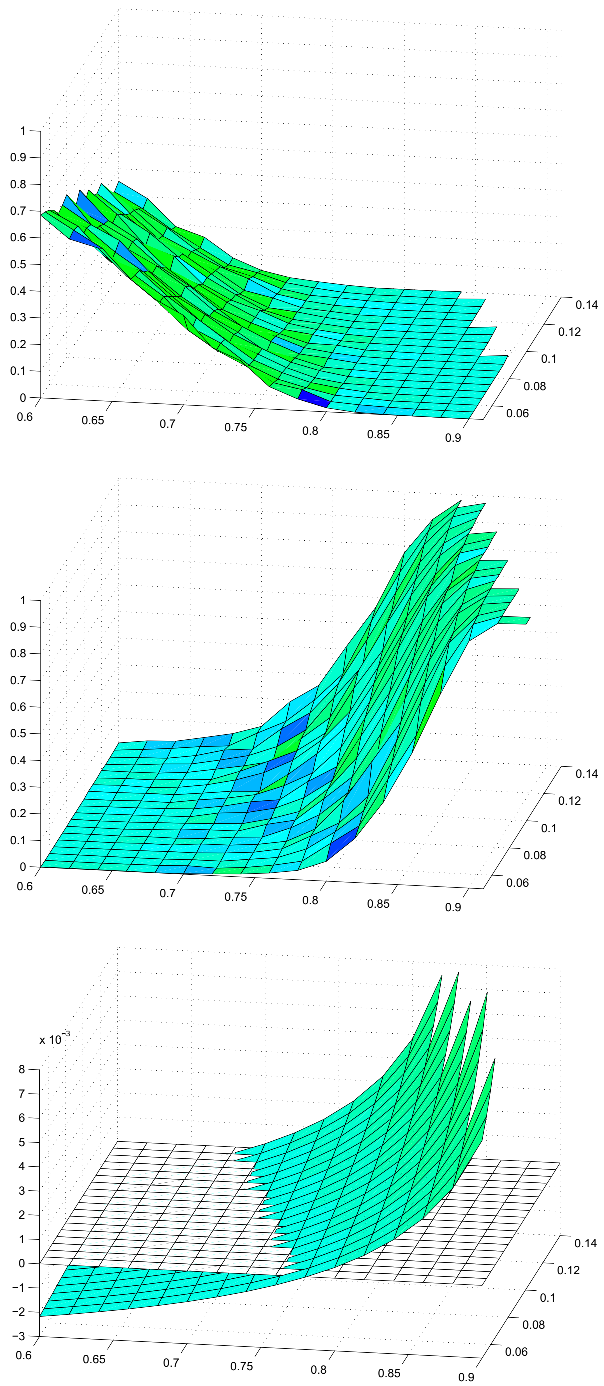
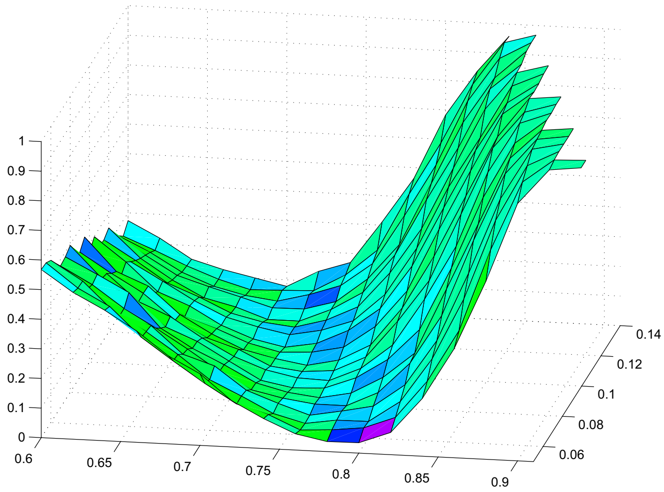


Figure 4: *Top and Center:* Power against GARCH alternatives of a test based on S_{125} and the rejection interval $(0, 1.01)$ *(Top)* and $(3.32, \infty)$ *(Center)*. The parameters α_0^a and ν^a are kept constant and equal to the values in (3.1), i.e., 8.58×10^{-6} and 5.24 respectively. The x- and y-axes show the β_1^a- and α_1^a-values of the alternatives.
Bottom: The difference between the standard deviations of the alternative models and that of the model with parameters (3.1).



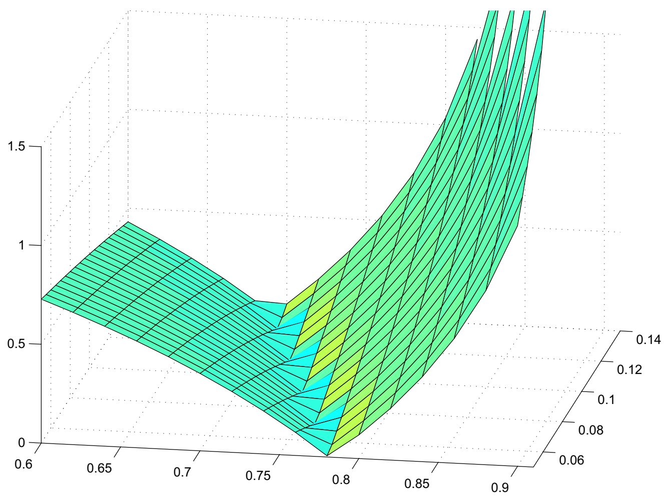


Figure 5: *Top:* Power against GARCH alternatives of a test based on S_{125} and the two-sided rejection region $(0, 0.785) \cup (4.9, \infty)$. The parameters α_0^a and ν^a are kept constant and equal to the values in (3.1), i.e., 8.58×10^{-6} and 5.24, respectively. The x- and y-axes show the β_1^a- and α_1^a-values of the alternatives.
Bottom: The absolute value of the differences between the log of the standard deviations of the alternative models and that of the model with parameters (3.1).

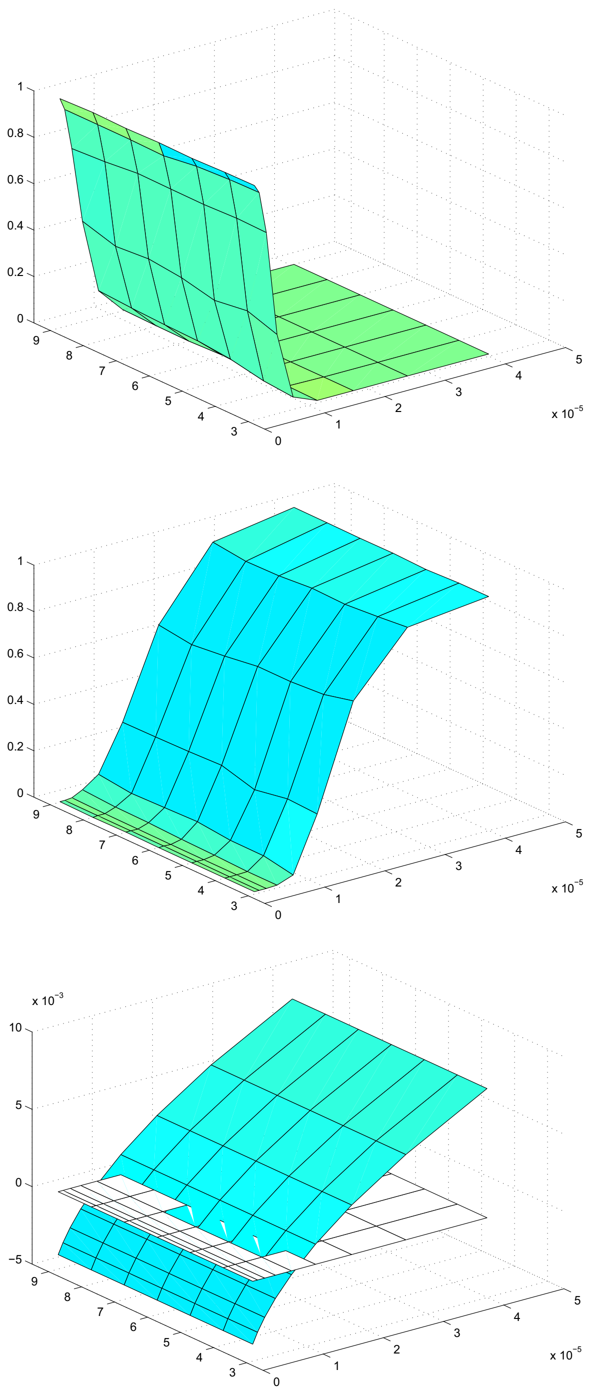


Figure 6: *Top and Center:* Power against GARCH alternatives of a test based on S_{125} and the one-sided rejection intervals $(0, 1.01)$ *(Top)* and $(3.32, \infty)$ *(Center)*. The parameters α_1^a and β_1^a are kept constant to the values in (3.1), i.e., 0.072 and 0.759, respectively. The x- and y-axes show the α_0^a- and ν^a-values of the alternatives.
Bottom: The difference between the standard deviations of the alternative models and that of the model with parameters (3.1).

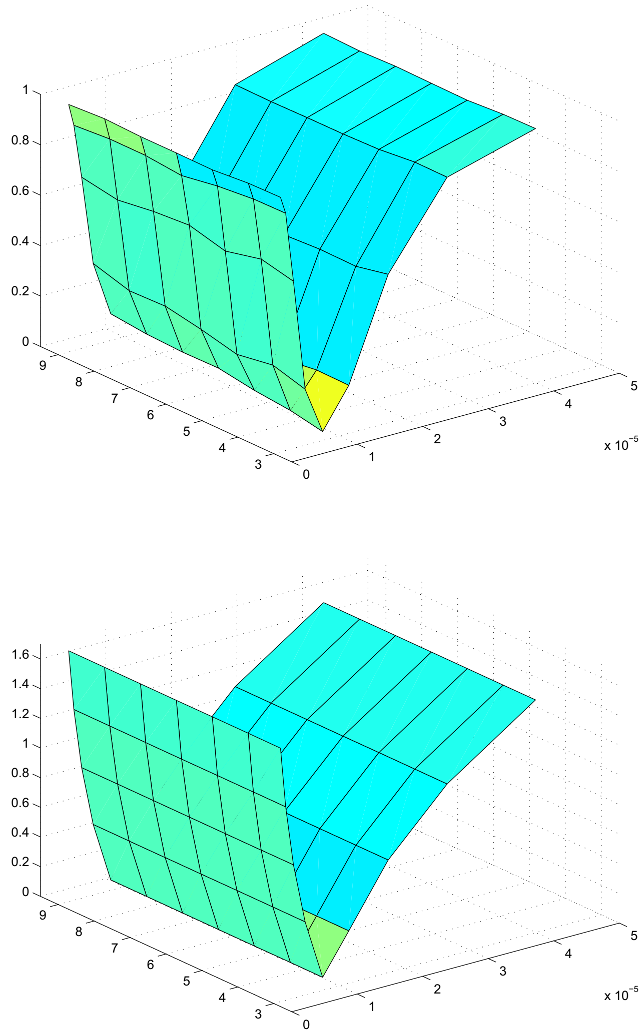


Figure 7: *Top:* Power against GARCH alternatives of a test based on S_{125} and the two-sided rejection interval $(0, 0.785) \cup (4.9, \infty)$ and $(3.32, \infty)$. The parameters α_1^a and β_1^a are kept constant to the values in (3.1), i.e., 0.072 and 0.759, respectively. The x- and y-axes show the α_0^a- and ν^a-values of the alternatives.
Bottom: The absolute value of the difference between the log of the standard deviations of the alternative models and that of the model with parameters (3.1).

4. A STUDY OF THE STANDARD & POOR'S 500 SERIES

We now proceed to analyze a time series that has been previously used to exemplify the presence of LRD in financial log-return series: the Standard 90 and Standard and Poor's 500 composite stock index. This series, covering the period between January 3, 1928, to August 30, 1991, was used in Ding et al. [15], Granger et al. [23], Ding and Granger [14] for an analysis of its autocorrelation structure. It led the authors to the conclusion that the powers of the absolute values of the log-returns are positively correlated over more than 2500 lags, i.e., 10 years. It is hard to believe that this time series is likely to be stationary. It covers the Great Depression, a world war together with the most recent period, marked by major structural changes in the world's economy. In addition, there was a compositional change in the S&P composite index that happened in January 1953 when the Standard 90 was replaced by the broader Standard and Poor's 500 index. Despite all these, Ding et al. [15] conclude the section which describes the data as follows (page 85): "During the Great Depression of 1929 and early 1930s, volatilities are much higher than any other period. There is a sudden drop in prices on Black Monday's stock market crash of 1987, but unlike the Great Depression, the high market volatility did not last very long. *Otherwise, the market is relatively stable.*" Bollerslev and Mikkelsen [9] used the daily returns on the Standard and Poor's 500 composite stock index from January 2, 1953, to December 31, 1990 (a total of 9559 observations) to fit a FIGARCH model under the assumptions of stationarity and LRD. (It is unknown whether the FIGARCH has a stationary version, and if it existed, it had infinite variance marginals, thus the definition of LRD via the ACF would break down. See Giraitis et al. [21] and Mikosch and Stărică [34] for some discussions.)

In the sequel we perform a detailed analysis of the same data set covering the time span from January 2, 1953, to December 31, 1990; see Figure 8. The first goal of the analysis is to check the goodness of fit of a GARCH process with parameters estimated on the first 3 years of data, the period from the beginning in 1953 until the beginning of 1956 (750 observations)

$$\alpha_0 = 8.58 \times 10^{-6}, \quad \alpha_1 = 0.072, \quad \beta_1 = 0.759, \quad \nu = 5.24\ , \tag{4.1}$$

to various segments of the data set. Here ν is the number of degrees of freedom of the t-distributed noise sequence (Z_t). The corresponding value of the fourth moment of the estimated residuals of the model (4.1) is $EZ^4 = 7.82$. The analysis verifies if this GARCH$(1,1)$ model which provides a good description to the beginning of the sample can be used to model later periods. In the case of a negative answer we are interested in understanding the type of changes that occurred and, if possible, to pin them to new economic conditions. In other words, the second goal of the analysis is the timing of possible changes in the structure of the data. We try to achieve this goal by evaluating the statistic S_n on a window that moves sequentially through the data. We will chose the window as small as possible to make sure the statistic reacts promptly to possible structural changes. In the end of the section we document the effect which changes in the variance have on the sample ACF. We find that the shape of the sample ACF changes drastically after episodes of increased variance that cannot be properly described by the estimated model.

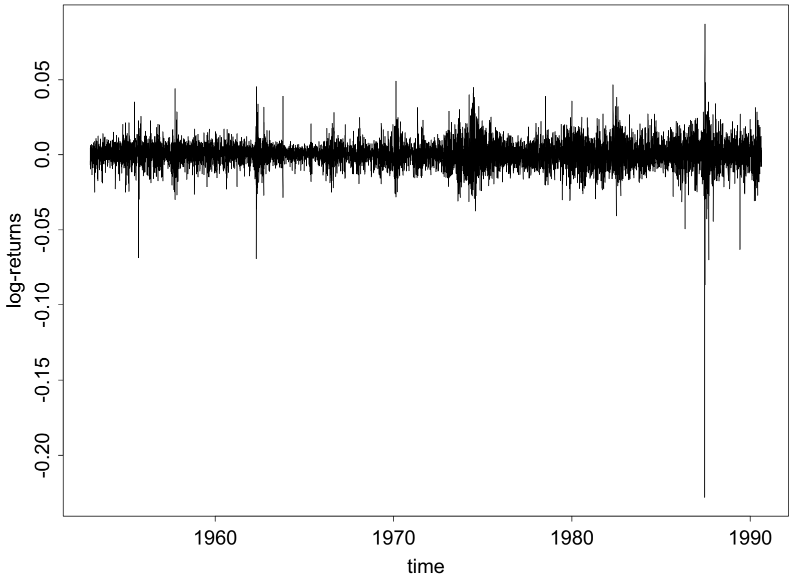


Figure 8: Plot of 9558 *S&P500* daily log-returns.
The year marks indicate the beginning of the calendar year.

The top graph in Figure 9 shows the results of calculating the statistic S_n (see (2.10)) on a weekly basis (i.e., every 5^{th} instant of time) with blocks of $n = 125$ past observations, corresponding to approximately 6 months of previous observations. The horizontal lines correspond to the ends of the rejection region of a goodness of fit test of size 99% based on S_{125} statistic as obtained from the simulation study in Section 3. The dotted vertical lines mark the start and the end of economic recessions as determined by the US National Bureau of Economic Research. This graph shows that one simple GARCH$(1,1)$ process (which, according to the S_{125} statistic, models the first ten years of data or so quite well cannot describe the complicated dynamics of longer, possibly non-stationary log-return time series. More precisely, the graph shows that most of the more pronounced violations of the confidence interval are on the upper side. It also shows that most of the recessions of the period under study (apart the one in the beginning of the 60s) are associated with larger than acceptable values of the S_{125} statistic. Recalling the simulation results of Section 3, these two findings also seem to imply that the unconditional variance of the log-returns changes through time and that most of the recessions of the period under study are characterized by higher unconditional variance than the periods of normal economic activity.

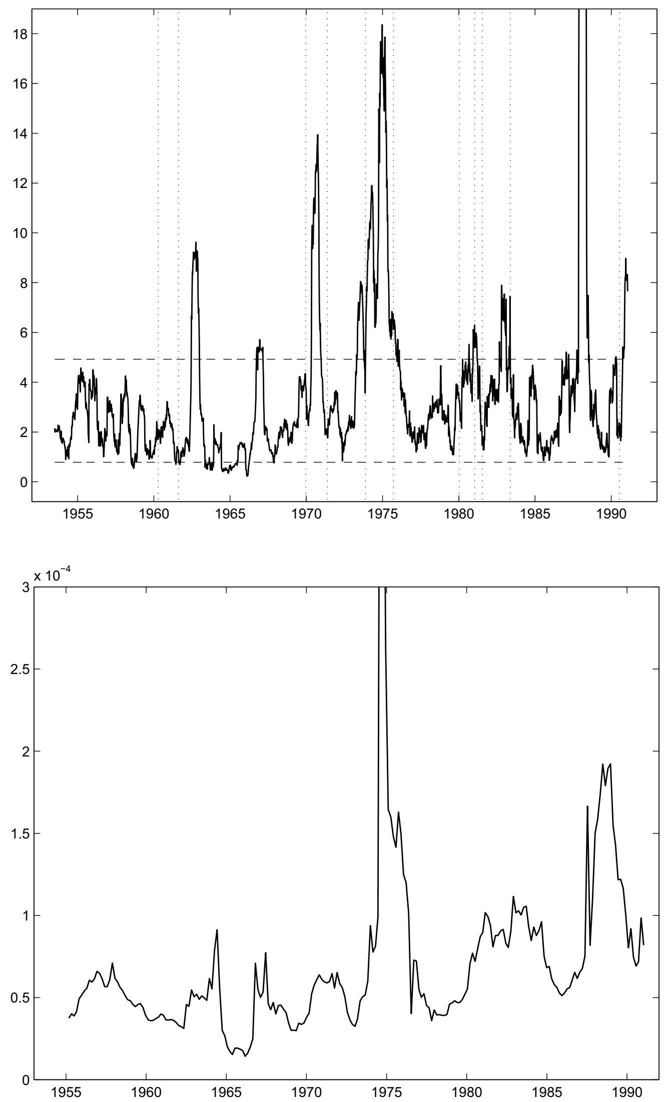


Figure 9: *Top:* The goodness of fit test statistic S_{125} for the *S&P500* data. The horizontal lines are the limits of the 99% confidence interval of S_{125} as obtained from the simulation study in Section 3. The dotted vertical vertical lines mark the start and the end of economic recessions as determined by the National Bureau of Economic Research.
Bottom: The implied GARCH(1, 1) unconditional variance of the *S&P500* data. A GARCH(1, 1) model is estimated every 2 months using the previous 2 years of data (i.e., 508 observations). The graph displays the variances $\sigma_X^2 = \alpha_0/(1-\alpha_1-\beta_1)$; see (A2.1).

A closer look at the S&P 500 plot in Figure 8 together with the top graph in Figure 9 reveals an almost one-to-one correspondence between the periods of larger absolute log-returns (larger volatility) and the periods when the goodness of fit test statistic S_{125} falls outside and above the confidence region.

If the unconditional variance changes through time, as our analysis seems to indicate, no GARCH$(1,1)$ model could be a good model for the whole period. It is then interesting to verify whether a periodically updated GARCH$(1,1)$ model could account for the more pronounced volatility periods that cannot be explained by the GARCH$(1,1)$ model (4.1). One way to do this is to calculate the implied unconditional GARCH$(1,1)$ variance of a periodically re-estimated GARCH$(1,1)$ model, i.e., one calculates the variance

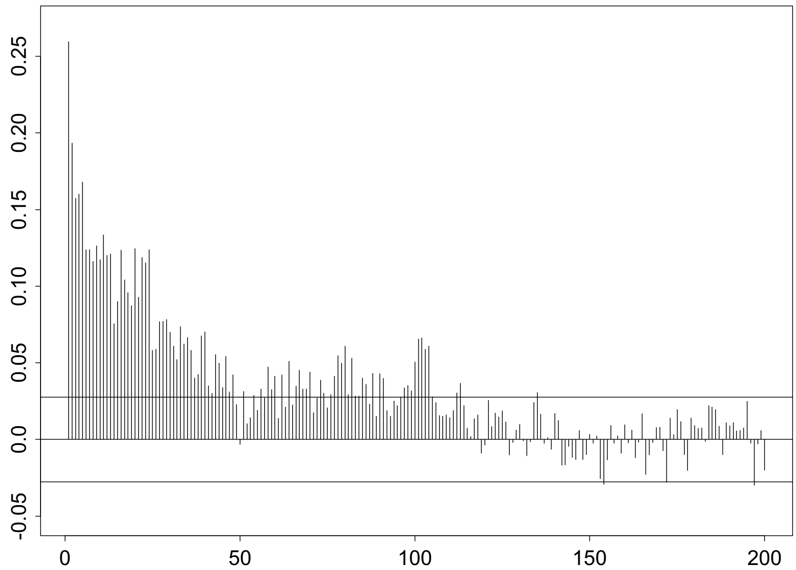
$$\sigma_X^2 = \alpha_0 / \big(1 - (\alpha_1 + \beta_1)\big)$$

based on the periodically re-estimated parameters α_1 and β_1; see (A2.1).

More concretely, we re-estimated a GARCH$(1,1)$ model every 2 months, i.e., every 42 days, on a moving window of 508 past observations, equivalent to roughly two business years of daily log-returns. We then plotted the implied variance σ_X^2. The results of this procedure are displayed in the bottom graph of Figure 9. One notices that the pattern of increased implied unconditional variance is quite similar to the pattern of the excursions of the statistic S_{125} above the 99% quantile threshold. This similarity seems to imply that one can capture the changing patterns of volatility present in the data by periodically updating the GARCH$(1,1)$ model. However a more in-depth study would be needed to substantiate such a statement.

Let us now analyze the impact which these periods of different structural behavior detected by the goodness of fit test statistic S_{125} have on the sample ACF of the time series. The top graph in Figure 9 identifies the period beginning in 1973 and lasting for almost 4 years as the longest and most significant deviation from the hypothesized model. This period is centered around the longest economic recession in the analyzed data. Figure 10 displays the sample ACF of the absolute values $|X_t|$ up to the moment when the change is detected, i.e., beginning of 1973, next to the sample ACF including the 4-year period that followed. The impact of the change in the structure of the time series between 1973 and 1977 on the sample ACF is extremely severe as one sees from the second graph of Figure 10. The graph clearly displays the LRD effect as explained in [34, 35]: exponential decay at small lags followed by almost constant plateau for larger lags together with strictly positive correlations.

Contrary to the belief that the LRD characteristic carries meaningful information about the price generating process, these graphs show that the LRD behavior could be just an artifact due to very plausible structural changes in the log-return data: variations of the unconditional variance due to the business cycle.



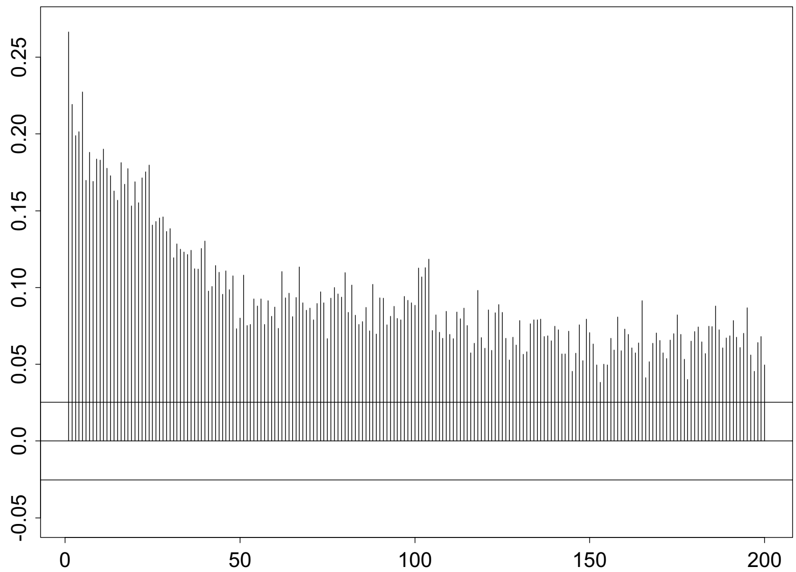


Figure 10: The sample ACF for the absolute log-returns of the first 20 and 24 years *(top and bottom)* of the *S&P500* data.

5. CONCLUDING REMARKS

In this paper we have argued that long financial time series display complicated volatility structures for which the simplifying assumption of constant unconditional variance and constant other moments is too rigid. Modeling the changing unconditional variance (possibly together or instead of the changing conditional one) is an important component of the modelization of long log-returns time series.

A1. APPENDIX

Proof of Lemma 2.1: We have to show the convergence of the finite-dimensional distributions and the tightness in $\mathbb{D}\,([0,1],\mathbb{R}^m)$. Notice first that for every fixed h,

$$\text{(A1.1)} \qquad \sqrt{n}\Big(\gamma_{n,[nx],X}(h)\Big)_{x\in[0,1]} \stackrel{d}{\longrightarrow} \Big(v_X^{1/2}(h)\; W_h(x)\Big)_{x\in[0,1]} .$$

in $\mathbb{D}\,[0,1]$; see Oodaira and Yoshihara [37]; cf. Doukhan [16], Theorem 1 on p. 46. In the latter theorem one has to ensure that $E|X_0X_h|^{2+\epsilon}<\infty$ for some $\epsilon>0$ (this follows from (2.4)) and that the sequence (X_tX_{t+h}) is α-mixing with a sufficiently fast rate for the mixing coefficients; see (A1.2). However, the GARCH(p,q) is strongly mixing with geometric rate since we assume that Z has a Lebesgue density on $\mathbb{R}$ (see Boussama [11]), and so the mixing coefficients converge to zero at an exponential rate, which implies the conditions in the aforementioned theorem.

Thus each of the processes $\sqrt{n}\,\gamma_{n,[n\cdot],X}(h)$ is tight in $\mathbb{D}\,[0,1]$. Using a generalization of the argument for Lemma 4.4 in Resnick [39], one obtains that the map from $(\mathbb{D}\,[0,1])^m$ into $\mathbb{D}\,([0,1],\mathbb{R}^m)$ defined by

$$\Big(x_1,...,x_m\Big) \longrightarrow \Big(x_1(t),...,x_m(t)\Big)_{t\in[0,1]}$$

is continuous at $(x_1,...,x_m)$ in $(\mathbb{C}\,[0,1])^m$. This and the sample path continuity of the limit process ensure that the processes on the left-hand side of (2.5) are tight in $\mathbb{D}\,([0,1],\mathbb{R}^m)$.

Notice that the multivariate CLT

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[nx]}\Big(X_t\,X_{t+1},\;...,\;X_t\,X_{t+h}\Big) \stackrel{d}{\longrightarrow} \Big(v_X^{1/2}(1)\,W_1(x),\;...,\;v_X^{1/2}(h)\,W_h(x)\Big)$$

holds for every fixed x. This is again a consequence of the aforementioned CLT for α-mixing sequences in combination with the Cramér–Wold device. A similar argument for a finite number of x-values yields the convergence of the finite-dimensional distributions. This proves the lemma. □

Remark 1.1. It follows from the argument in the proof of Lemma 2.1 that (2.5) remains valid for stationary strongly mixing sequences (X_t) with $EX=0$, $E|X|^{4+\delta}<\infty$ for some $\delta>0$ and such that $EX_0X_h=0$ for $h\geq 1$, $\operatorname{cov}(X_0X_h, X_0X_l)=0$ for all $h\neq l\geq 1$, and with α-mixing coefficients $\widetilde{\alpha}_i$ satisfying

$$\sum_{i=1}^{\infty} \widetilde{\alpha}_i^{\delta/(2+\delta)} < \infty \; . \tag{A1.2}$$

The latter conditions are needed for the validity of the FCLT in (A1.1); see Oodaira and Yoshihara [37].

Proof of Theorem 2.1: We proceed analogously to Klüppelberg and Mikosch [29]. It follows from Lemma 2.1 and the continuous mapping theorem that, for every fixed $m\geq 1$, in $\mathbb{D}\left([0,1]\times[0,\pi]\right)$

$$\sum_{h=1}^{m} \sqrt{n}\, \widetilde{\gamma}_{n,[nx],X}(h)\, \frac{\sin(\lambda h)}{h} \;\stackrel{d}{\longrightarrow}\; \sum_{h=1}^{m} W_h(x)\, \frac{\sin(\lambda h)}{h} \; . \tag{A1.3}$$

According to Theorem 4.2 in Billingsley [7], it remains to show that for every $\epsilon>0$,

$$\lim_{m\to\infty} \limsup_{n\to\infty} P\left(\sup_{0\leq x\leq 1} \sup_{0\leq\lambda\leq\pi} \left| \sum_{h=m+1}^{[nx]-1} \sqrt{n}\, \widetilde{\gamma}_{n,[nx],X}(h)\, \frac{\sin(\lambda h)}{h} \right| > \epsilon \right) = 0 \; . \tag{A1.4}$$

Since Z is symmetric the sequences $(r_t)=(\operatorname{sign}(X_t))$ and $(|X_t|)$ are independent. Conditionally on $(|X_t|)$,

$$\sum_{h=m+1}^{k} \sqrt{n}\, \widetilde{\gamma}_{n,k,X}(h)\, \frac{\sin(\cdot h)}{h} \; , \qquad k=m+1,...,n-1 \; ,$$

is a sequence of quadratic forms in the iid Rademacher random variables r_t and with values in the Banach space $\mathbb{C}\,[0,\pi]$ endowed with the sup-norm. Now condition on $(|X_t|)$. Use first a decoupling inequality for Rademacher quadratic forms (e.g. de la Peña and Montgomery–Smith [13], Theorem 1) then the Lévy maximal inequality for sums of iid symmetric random variables, then again the decoupling inequality in reverse order, and finally take expectations with respect to $(|X_t|)$. Then we obtain the inequality

$$P\left(\sup_{0\leq x\leq 1} \sup_{0\leq\lambda\leq\pi} \left| \sum_{h=m+1}^{[nx]} \sqrt{n}\, \widetilde{\gamma}_{n,[nx],X}(h)\, \frac{\sin(\lambda h)}{h} \right| > \epsilon \right) \leq$$

$$\leq \; c_1\, P\left(c_2 \sup_{0\leq\lambda\leq\pi} \left| \sum_{h=m+1}^{n-1} \sqrt{n}\, \widetilde{\gamma}_{nX}(h)\, \frac{\sin(\lambda h)}{h} \right| > \epsilon \right)$$

for certain positive constants c_1, c_2. The right-hand probability can be treated in the same way as the derivation of (6.3) in [28], pp. 1873–1876. Instead of Theorem 3.1 in Rosiński and Woyczyński [40] one can simply use the Cauchy–Schwarz inequality in the first display on p. 1876 in [28] with $\mu = 2$. Then all the calculations for (6.3) remain valid, implying that (A1.4) holds. This concludes the proof of the theorem. □

Remark 1.2. The condition of symmetry of Z is needed only for the application of the Lévy maximal inequality for sums of independent random variables. Alternatively, one can proceed as in the proof of Theorem 3.1 in Klüppelberg and Mikosch [29], p. 980, last display, where instead of the Lévy maximal inequality Doob's 2^{nd} moment maximal inequality for submartingales was used. Then one can follow the lines of the proof of Theorem 1 in Grenander and Rosenblatt [24], Chapter 6.4.

Proof of Theorem 2.2: We start by showing that $\widehat{EZ^4} \xrightarrow{P} EZ^4$. Indeed, consistency of the estimators $\widehat{\alpha}_i$ and $\widehat{\beta}_1$ implies consistency of $\widehat{EZ^4}$. We have by induction, using the definitions of σ_t^2 and $\widehat{\sigma}_t^2$,

$$\begin{aligned}
\widehat{EZ^4} - n^{-1}\sum_{i=1}^{n} Z_t^4 &= \\
&= n^{-1}\sum_{i=1}^{n}\frac{X_t^4}{\widehat{\sigma}_t^4} - n^{-1}\sum_{i=1}^{n}\frac{X_t^4}{\sigma_t^4} \\
&= n^{-1}\sum_{i=1}^{n} X_t^4\,\frac{\sigma_t^4-\widehat{\sigma}_t^4}{\sigma_t^4\,\widehat{\sigma}_t^4} \\
&= n^{-1}\sum_{i=1}^{n} X_t^4\,\frac{(\sigma_t^2-\widehat{\sigma}_t^2)\,(\sigma_t^2+\widehat{\sigma}_t^2)}{\sigma_t^4\,\widehat{\sigma}_t^4} \\
&= n^{-1}\sum_{i=1}^{n} X_t^4\Big[(\alpha_0-\widehat{\alpha}_0)+(\alpha_1-\widehat{\alpha}_1)X_{t-1}^2+(\beta_1-\widehat{\beta}_1)\sigma_{t-1}^2+\widehat{\beta}_1(\sigma_{t-1}^2-\widehat{\sigma}_{t-1}^2)\Big]\times \\
&\qquad\times\frac{\sigma_t^2+\widehat{\sigma}_t^2}{\sigma_t^4\,\widehat{\sigma}_t^4} \\
&= (\alpha_0-\widehat{\alpha}_0)\,n^{-1}\sum_{i=1}^{n} X_t^4\left(1+\widehat{\beta}_1+\widehat{\beta}_1^2+\cdots+\widehat{\beta}_1^t\right)\frac{\sigma_t^2+\widehat{\sigma}_t^2}{\sigma_t^4\,\widehat{\sigma}_t^4} \\
&\quad+(\alpha_1-\widehat{\alpha}_1)\,n^{-1}\sum_{i=1}^{n} X_t^4\left(X_{t-1}^2+\widehat{\beta}_1X_{t-2}^2+\cdots+\widehat{\beta}_1^tX_0^2\right)\frac{\sigma_t^2+\widehat{\sigma}_t^2}{\sigma_t^4\,\widehat{\sigma}_t^4} \\
&\quad+(\beta_1-\widehat{\beta}_1)\,n^{-1}\sum_{i=1}^{n} X_t^4\left(\sigma_{t-1}^2+\widehat{\beta}_1\sigma_{t-2}^2+\cdots+\widehat{\beta}_1^t\sigma_0^2\right)\frac{\sigma_t^2+\widehat{\sigma}_t^2}{\sigma_t^4\,\widehat{\sigma}_t^4} \\
&\quad+(\sigma_0^2-\widehat{\sigma}_0^2)\,n^{-1}\sum_{i=1}^{n} X_t^4\,\widehat{\beta}_1^t\,\frac{\sigma_t^2+\widehat{\sigma}_t^2}{\sigma_t^4\,\widehat{\sigma}_t^4} \\
&= (\alpha_0-\widehat{\alpha}_0)\,I_1+(\alpha_1-\widehat{\alpha}_1)\,I_2+(\beta_1-\widehat{\beta}_1)\,I_3+(\sigma_0^2-\widehat{\sigma}_0^2)\,I_4\ .
\end{aligned}$$

Notice that, by consistency of the parameter estimators and since $(Z_t^4\sigma_t^2)$ is ergodic,

$$\begin{aligned} I_1 &\le (1-\widehat{\beta}_1)^{-1}\, n^{-1} \sum_{i=1}^n Z_t^4\, \frac{\sigma_t^2+\widehat{\sigma}_t^2}{\widehat{\sigma}_t^4} \\ &\le (1-\widehat{\beta}_1)^{-1}\, n^{-1} \sum_{i=1}^n Z_t^4 \left[\sigma_t^2\,\widehat{\alpha}_0^{-2}+\widehat{\alpha}_0^{-1}\right] \\ &\stackrel{\text{a.s.}}{\longrightarrow} (1-\beta_1)^{-1}\, EZ^4 \left[E\,\sigma^2\,\alpha_0^{-2}+\alpha_0^{-1}\right] . \end{aligned}$$

By similar arguments, for any $\delta>0$ and $\epsilon>0$ such that $\beta_1+\epsilon<1$,

$$\begin{aligned} &P\Big(I_2>\delta\Big) \le \\ &\le P\left(n^{-1}\sum_{i=1}^n X_t^4\Big(X_{t-1}^2+(\beta_1+\epsilon)X_{t-2}^2+\cdots+(\beta_1+\epsilon)^t X_0^2\Big)\,\frac{\sigma_t^2+\widehat{\sigma}_t^2}{\sigma_t^4\,\widehat{\sigma}_t^4}>\delta\right) \\ &\quad + P\Big(\widehat{\beta}_1>\beta_1+\epsilon\Big) \\ &\le P\left(n^{-1}\sum_{i=1}^n Z_t^4\Big(X_{t-1}^2+(\beta_1+\epsilon)X_{t-2}^2+\cdots+(\beta_1+\epsilon)^t X_0^2\Big)\left(\sigma_t^2\,\widehat{\alpha}_0^{-2}+\widehat{\alpha}_0^{-1}\right)>\delta\right) \\ &\quad + o(1)\ . \end{aligned}$$

It is not difficult to see, by an application of the Cauchy–Schwarz inequality, that the first moments of

$$n^{-1}\sum_{i=1}^n Z_t^4\left(X_{t-1}^2+(\beta_1+\epsilon)\,X_{t-2}^2+\cdots+(\beta_1+\epsilon)^t\,X_0^2\right)\sigma_t^2$$

and

$$n^{-1}\sum_{i=1}^n Z_t^4\left(X_{t-1}^2+(\beta_1+\epsilon)\,X_{t-2}^2+\cdots+(\beta_1+\epsilon)^t\,X_0^2\right),$$

are bounded, uniformly for n. Therefore I_2 is stochastically bounded, and a similar argument applies to I_3. Finally,

$$P\Big(n\,I_4>\delta\Big) \le P\left(\sum_{i=1}^n Z_t^4\,(\beta_1+\epsilon)^t\left(\sigma_t^2\,\widehat{\alpha}_0^2+\widehat{\alpha}_0^{-1}\right)>\delta\right)+P\Big(\widehat{\beta}_1>\beta_1+\epsilon\Big)\ .$$

The second probability vanishes by consistency of $\widehat{\beta}_1$. Moreover,

$$\sum_{i=1}^n Z_t^4\,(\beta_1+\epsilon)^t\,\sigma_t^2 \qquad \text{and} \qquad \sum_{i=1}^n Z_t^4\,(\beta_1+\epsilon)^t\ ,$$

have bounded first moments. This implies that I_4 is stochastically bounded, and $n^{-1}I_4\stackrel{P}{\to}0$. Collecting the bounds for all I_j, we conclude by the law of large numbers that $\widehat{EZ^4}\stackrel{P}{\to}EZ_4$.

For the remaining proof we follow the lines of the proof of Theorem 2.1. Write $\widehat{v}_X(h)$ for the sample version of $v_X(h)$. By consistency of $\widehat{EZ^4}$, $\widehat{\alpha}_i$, $\widehat{\beta}_1$ and the form of $v_X(h)$, see (A2.2), we have $\widehat{v}_X(h) \stackrel{P}{\to} v_X(h)$ for every h. This fact and the continuous mapping theorem immediately yield that (A1.3) remains valid with $v_X(h)$ replaced by $\widehat{v}_X(h)$. So it remains to show

$$\lim_{m\to\infty} \limsup_{n\to\infty} P\left(\sup_{0\le x\le 1} \sup_{0\le\lambda\le\pi} \left| \sum_{h=m+1}^{[nx]-1} \sqrt{n}\, \frac{\gamma_{n,[nx],X}(h)}{\widehat{v}_X^{1/2}}\, \frac{\sin(\lambda h)}{h} \right| > \epsilon \right) = 0 \,.$$

Notice that for every $h \ge 1$,

$$\widehat{v}_X(h)^{-1/2} \le \widehat{\sigma}_X^{-4} = \widehat{\alpha}_0^{-4}\big(1-\widehat{\varphi}_1\big)^{-4} \,.$$

See Appendix A2. By the assumptions, the estimators $\widehat{\alpha}_i$, $\widehat{\beta}_1$ are independent of $(\mathrm{sign}(X_t))$, and so is $\widehat{EZ^4}$ by construction of the residuals. Thus, conditionally on $(|X_t|)$,

$$\sum_{h=m+1}^{[nx]-1} \sqrt{n}\, \frac{\gamma_{n,[nx],X}(h)}{\widehat{v}_X^{1/2}}\, \frac{\sin(\lambda h)}{h}$$

is a random quadratic form in the variables $\mathrm{sign}(X_t)$, which, by symmetry of (Z_t) are independent of the coefficients of the quadratic form which only depend on the sequence $(|X_t|)$. An application of the contraction principle for Rademacher quadratic forms (cf. Kwapień and Woyczyński [31]) implies that for some constants $c_1, c_2 > 0$

$$\begin{aligned} & P\left(\sup_{0\le x\le 1} \sup_{0\le\lambda\le\pi} \left| \sum_{h=m+1}^{[nx]-1} \sqrt{n}\, \frac{\gamma_{n,[nx],X}(h)}{\widehat{v}_X^{1/2}}\, \frac{\sin(\lambda h)}{h} \right| > \epsilon \right) \le \\ & \quad \le c_1\, P\left(c_2 \max_h \widehat{v}_X^{-1/2} \sup_{0\le x\le 1} \sup_{0\le\lambda\le\pi} \left| \sum_{h=m+1}^{[nx]-1} \sqrt{n}\, \gamma_{n,[nx],X}(h)\, \frac{\sin(\lambda h)}{h} \right| > \epsilon \right) \\ & \quad \le c_1\, P\left(c_2\, \widehat{\alpha}_0^{-4} \big(1-\widehat{\varphi}_1\big)^{-4} \sup_{0\le x\le 1} \sup_{0\le\lambda\le\pi} \left| \sum_{h=m+1}^{[nx]-1} \sqrt{n}\, \gamma_{n,[nx],X}(h)\, \frac{\sin(\lambda h)}{h} \right| > \epsilon \right). \end{aligned}$$

Thus it remains to show that

$$\lim_{m\to\infty} \limsup_{n\to\infty} P\left(\sup_{0\le x\le 1} \sup_{0\le\lambda\le\pi} \left| \sum_{h=m+1}^{[nx]-1} \sqrt{n}\, \gamma_{n,[nx],X}(h)\, \frac{\sin(\lambda h)}{h} \right| > \epsilon \right) = 0 \,,$$

which follows along the lines of the proof of Theorem 2.1. □

A2. APPENDIX

Consider a GARCH$(1,1)$ process (X_t) with parameters $\alpha_0, \alpha_1, \beta_1$. We write $\varphi_1 = \alpha_1 + \beta_1$ and assume $EX^4 < \infty$. From the calculations below it follows that the condition

$$1 - \Big(\alpha_1^2\, EZ^4 + \beta_1^2 + 2\,\alpha_1\,\beta_1\Big) \;>\; 0$$

must be satisfied. The squared GARCH$(1,1)$ process can be rewritten as an ARMA$(1,1)$ process by using the defining equation (2.1):

$$X_t^2 - \varphi_1\, X_{t-1}^2 \;=\; \alpha_0 + \nu_t - \beta_1\, \nu_{t-1} \;,$$

where $(\nu_t) = (X_t^2 - \sigma_t^2)$ is a white noise sequence. Thus, the covariance structure of

$$U_t = X_t^2 - EX^2 \;, \qquad t \in \mathbb{Z} \;,$$

is that of a mean-zero ARMA$(1,1)$ process. The values of $\gamma_U(h)$ are given on p. 87 in Brockwell and Davis [12]:

$$\begin{aligned}
\gamma_U(0) \;&=\; \sigma_\nu^2 \left[1 + \frac{(\varphi_1 - \beta_1)^2}{1 - \varphi_1^2}\right] \;, \\
\gamma_U(1) \;&=\; \sigma_\nu^2 \left[\varphi_1 - \beta_1 + \frac{(\varphi_1 - \beta_1)^2\, \varphi_1}{1 - \varphi_1^2}\right] \;, \\
\gamma_U(h) \;&=\; \varphi_1^{h-1}\, \gamma_U(1) \;, \qquad h \geq 2 \;.
\end{aligned}$$

Straightforward calculation yields

$$\begin{aligned}
\sigma_\nu^2 \;&=\; (EZ^4 - 1)\; E\sigma_1^4 \;=\; \frac{1+\varphi_1}{1-\varphi_1}\, \frac{\alpha_0^2(EZ^4-1)}{1 - \Big(\varphi_1^2 + \alpha_1^2(EZ^4-1)\Big)} \;, \\
\sigma_X^2 \;&=\; \frac{\alpha_0}{1-\varphi_1} \;.
\end{aligned} \tag{A2.1}$$

Thus we can calculate the quantities

$$v_X(h) \;=\; E(X_0^2 X_h^2) \;=\; \gamma_U(h) + \sigma_X^4 \;, \qquad h \geq 1 \;,$$

which occur in the definition of the change point statistics and goodness of fit test statistics of Section 2. We obtain:

$$v_X(h) \;=\; \sigma_X^4 \left(\frac{(EZ^4 - 1)\; \alpha_1\big(1 - \varphi_1^2 + \alpha_1\,\varphi_1\big)}{1 - \Big(\varphi_1^2 + \alpha_1^2(EZ^4 - 1)\Big)}\; \varphi_1^{h-1} + 1 \right) \;, \qquad h \geq 1 \;. \tag{A2.2}$$

ACKNOWLEDGMENTS

We would like to thank the Newton Institute in Cambridge for its stimulating hospitality during the Managing Uncertainty period in June/July 2001.

REFERENCES

[1] Anderson, T.W. (1993). Goodness of fit tests for spectral distributions, *Ann. Statist.*, **21**, 830–847.

[2] Baillie, R.T.; Bollerslev, T. and Mikkelsen, H.O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity, *J. Econometrics*, **74**, 3–30.

[3] Bartlett, M.S. (1954). Problèmes de l'analyse spectrale des séries temporelles stationnaires, *Publ. Inst. Statist. Univ. Paris*, **III-3**, 119–134.

[4] Beran, J. (1994). *Statistics for Long–Memory Processes*, Monographs on Statistics and Applied Probability, No. 61, Chapman and Hall, New York.

[5] Berkes, I.; Horváth, L. and Kokoszka, P. (2003). GARCH processes: structure and estimation, *Bernoulli*, **9**, 201–228.

[6] Bickel, P.J. and Wichura, M.J. (1971). Convergence criteria for multiparameter stochastic processes and some applications, *Ann. Math. Statist.*, **42**, 1656–1670.

[7] Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

[8] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *J. Econometrics*, **31**, 307–327.

[9] Bollerslev, T. and Mikkelsen, H.O. (1996). Modeling and pricing long memory in stock market volatility, *J. Econometrics*, **73**, 151–184.

[10] Bougerol, P. and Picard, N. (1992). Stationarity of GARCH processes and of some nonnegative time series, *J. Econometrics*, **52**, 115–127.

[11] Boussama, F. (1998). *Ergodicité, mélange et estimation dans le modelès GARCH*, PhD Thesis, Université 7 Paris.

[12] Brockwell, P.J. and Davis, R.A. (1996). *Introduction to Time Series and Forecasting*, Springer, New York.

[13] de la Peña, V.H. and Montgomery-Smith, S.J. (1995). Decoupling inequalities for the tail probabilities of multivariate U-statistics, *Ann. Probab.*, **23**, 806–816.

[14] Ding, Z. and Granger, C.W.J. (1996). Modeling volatility persistence of speculative returns: A new approach, *J. Econometrics*, **73**, 185–215.

[15] Ding, Z.; Granger, C.W.J. and Engle, R. (1993). A long memory property of stock market returns and a new model, *J. Empirical Finance*, **1**, 83–106.

[16] Doukhan, P. (1994). *Mixing. Properties and Examples*, Lecture Notes in Statistics, **85**, Springer Verlag, New York.

[17] Embrechts, P.; Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin.

[18] Engle, R.F. (Ed.) (1995). *ARCH Selected Readings*, Oxford University Press, Oxford (U.K.).

[19] Engle, R.F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. With comments and a reply by the authors, *Econometric Rev.*, **5**, 1–87.

[20] Giraitis, L. and Leipus, R. (1992). Testing and estimating in the change-point problem of the spectral function, *Lith. Math. Trans. (Lit. Mat. Sb.)*, **32**, 20–38.

[21] Giraitis, L.; Kokoszka, P. and Leipus, R. (2000). Stationary ARCH models: dependence structure and central limit theorem, *J. Econom. Theory*, **16**, 3–22.

[22] Giraitis, L. and Robinson, P.M. (2001). Whittle estimation of ARCH models, *J. Econom. Theory*, **17**, 608–631.

[23] Granger, C.W.J. and Ding, Z. (1996). Varieties of long memory models, *J. Econometrics*, **73**, 61–77.

[24] Grenander, U. and Rosenblatt, M. (1984). *Statistical Analysis of Stationary Time Series*, 2nd edition, Chelsea Publishing Co., New York.

[25] Hida, T. (1980). *Brownian Motion*, Springer, New York.

[26] Horváth, L.; Kokoszka, P. and Teyssiere, G. (2001). Empirical process of squared residuals of an ARCH sequence, *Ann. Statist.*, **29**, 445–469.

[27] Jacod, J. and Shiryaev, A.N. (1987). *Limit Theorems for Stochastic Processes*, Springer, Berlin, New York.

[28] Klüppelberg, C. and Mikosch, T. (1996). The integrated periodogram for stable processes, *Ann. Statist.*, **24**, 1855–1879.

[29] Klüppelberg, C. and Mikosch, T. (1996). Gaussian limit fields for the integrated periodogram, *Ann. Appl. Probab.*, **6**, 969–991.

[30] Kokoszka, P. and Mikosch, T. (1997). The integrated periodogram for long-memory processes with finite or infinite variance, *Stoch. Proc. Appl.*, **66**, 55–78.

[31] Kwapień, S. and Woyczyński, W.A. (1992). *Random Series and Stochastic Integrals: Single and Multiple*, Birkhäuser, Basel.

[32] Mikosch, T. (2003). *Modelling dependence and tails of financial time series.* In: "Extreme Values in Finance, Telecommunications and the Environment" (B. Finkenstädt and H. Rootzén, Eds.), Chapman and Hall, Boca Raton, pp. 185–286.

[33] Mikosch, T. and Straumann, D. (2002). Whittle estimation in a heavy-tailed GARCH(1,1) model, *Stoch. Proc. Appl.*, **100**, 187–222.

[34] Mikosch, T. and Stărică, C. (2003). *Long range dependence effects and ARCH modeling.* In: "Long Range Dependence" (P. Doukhan, G. Oppenheim and M.S. Taqqu, Eds.), Birkhauser, Boston, pp. 439–460.

[35] Mikosch, T. and Stărică, C. (2004). Non-stationarities in financial time series, the long range dependence and the IGARCH effects, *Rev. Economics and Statistics*, **86**, 378–390.

[36] Nelson, D.B. (1990). Stationarity and persistence in the GARCH(1, 1) model, *J. Econom. Theory*, **6**, 318–334.

[37] Oodaira, H. and Yoshihara, K. (1972). Functional central limit theorems for strictly stationary processes satisfying the strong mixing condition, *Kōdai Math. Sem. Rep.*, **24**, 259–269.

[38] Priestley, M.B. (1981). *Spectral Analysis and Time Series*, vols. I and II, Academic Press, New York.

[39] Resnick, S.I. (1986). Point processes, regular variation and weak convergence, *Adv. Appl. Probab.*, **18**, 66–138.

[40] Rosiński, J. and Woyczyński, W.A. (1987). Multilinear forms in Pareto–like random variables and product random measures, *Coll. Math.*, **51**, 303–313.

[41] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics*, Wiley, New York.

[42] Whittle, P. (1951). *Hypothesis Testing in Time Series Analysis*, Almqvist och Wicksel, Uppsala.

REVSTAT – Statistical Journal
Volume 2, Number 1, June 2004

ESTIMATING PARETO TAIL INDEX BASED ON SAMPLE MEANS

Authors: Alena Fialová, Jana Jurečková
– Department of Statistics, Charles University in Prague, Czech Republic (fialova@karlin.mff.cuni.cz, jurecko@karlin.mff.cuni.cz)

Jan Picek
– Department of Applied Mathematics, Technical University in Liberec, Czech Republic (jan.picek@vslib.cz)

Received: October 2003 Revised: October 2003 Accepted: January 2004

Abstract:

- We propose an estimator of the Pareto tail index m of a distribution, that competes well with the Hill, Pickands and moment estimators. Unlike the above estimators, that are based only on the extreme observations, the proposed estimator uses all observations; its idea rests in the tail behavior of the sample mean $\bar{X}_n$, having a simple structure under heavy-tailed F. The observations, partitioned into N independent samples of sizes n, lead to N sample means whose empirical distribution function is the main estimation tool. The estimator is strongly consistent and asymptotically normal as $N \to \infty$, while n remains fixed. Its behavior is illustrated in a simulation study.

Key-Words:

- *domain of attraction; Pareto index; strong embedding of empirical process; tail behavior.*

AMS Subject Classification:

- 62G05, 62G30.

1. INTRODUCTION

Let $X_1, ..., X_n$ be independent nonnegative random variables, identically distributed with distribution function F. The exact shape of F is generally unknown, but we assume that F is absolutely continuous with density f and nondegenerate right tail of the Pareto type satisfying

$$\lim_{x\to\infty} \frac{-\log\bigl(1-F(x)\bigr)}{m\log x} \;=\; 1 \tag{1.1}$$

for some $m > 0$. Then, by the von Mises condition (see Embrechts *et al.* [4]),

$$1 - F(x) \;=\; x^{-m} L(x) \tag{1.2}$$

where $L(x)$ is a function, slowly varying at ∞, and hence F belongs to the domain of attraction of the Fréchet distribution with the distribution function $\Phi_m(x) = \exp\{-x^{-m}\}\,,\; x > 0$.

Among the estimators of the Pareto index m or its reciprocal $\gamma = \frac{1}{m}$, proposed in the literature, the Hill [9], Pickands [15] and moment estimators [3] are the most well-known. Either of these estimators is based only on the fraction of the observations, namely on k_n largest ones, where $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$. The consistency and asymptotic normality of these estimators was proved under various regularity conditions on k_n and on F, some of them not easy to verify. The problem of the estimating was considered by many other authors, e.g. Smith [16], Beirlant *et al.* [2], Feuerverger *et al.* [8], Gomes and Martins [7].

We propose another estimator of the Pareto index m, that competes well with the above estimators; the regularity conditions, required for its strong consistency and asymptotic normality, are apparently more transparent and less restrictive. The proposed estimator uses all observations, unlike the estimators mentioned above. The idea of the estimator is based on the tail behavior of the sample mean $\bar{X}_n$, that has a simple structure under heavy-tailed F, satisfying (1.1). The estimator is strongly consistent and asymptotically normal and it was also discussed by the same authors in [5, 6].

The tail behavior of the sample mean is described is Section 2. The estimator is defined in Section 3, along with the formulation of its consistency and asymptotic normality. Its behavior is illustrated in a simulation study in Section 4. The proofs of the main results are postponed to Section 5. In Section 6 we propose a test of a one-sided hypothesis on m, that can be used as a preliminary test before the estimation.

2. TAIL-BEHAVIOR OF THE SAMPLE MEAN

Let $X_1, ..., X_n$ be a random sample from a distribution with an absolutely continuous distribution function F and density f, positive on interval (K_f, ∞), $K_f \geq 0$. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

For a heavy-tailed F, symmetric around 0, Jurečková [10] showed that the tail behavior of $\bar{X}_n$ coincides with that of F. The following lemma demonstrates a similar behavior of the sample mean also for heavy-tailed F concentrated only on the positive half-axis.

Lemma 2.1. *Let $X_1, ..., X_n$ be a random sample from the distribution with absolutely continuous d.f. F and density f such that*

(i) $f(x) = 0$ for $x < 0$ and $0 < f(x) < \infty$ for $x \geq K_f \geq 0$.

(ii) F satisfies (1.1) for some m, $0 < m < \infty$.

Then, for any fixed n,

$$\lim_{a\to\infty} \frac{-\log \mathbb{P}_m(\bar{X}_n > a)}{-\log\big(1 - F(a)\big)} = \frac{-\log\big(1 - F_{\bar{X}_n}(a)\big)}{-\log\big(1 - F(a)\big)} = 1 \ . \tag{2.1}$$

Proof: Let $0 \leq X_{n:1} \leq ... \leq X_{n:n}$ be the order statistics corresponding to $X_1, ..., X_n$. Then

$$\mathbb{P}(\bar{X}_n > a) = \mathbb{P}\left(\sum_{i=1}^n X_i > na\right) \geq \mathbb{P}(X_{n:n} > na) \geq 1 - F(na)$$

and

$$\mathbb{P}(\bar{X}_n > a) \leq \mathbb{P}(X_{n:n} > a) = 1 - \big(F(a)\big)^n \leq n\big(1 - F(a)\big) ,$$

hence

$$\underline{\lim}_{a\to\infty} \frac{-\log \mathbb{P}(\bar{X}_n > a)}{-\log\big(1 - F(a)\big)} \geq \lim_{a\to\infty} \frac{-\log\big(n(1 - F(a)\big)}{-\log\big(1 - F(a)\big)} = 1$$

and

$$\overline{\lim}_{a\to\infty} \frac{-\log \mathbb{P}(\bar{X}_n > a)}{-\log\big(1 - F(a)\big)} \leq \lim_{a\to\infty} \frac{-\log\big(1 - F(na)\big)}{-\log\big(1 - F(a)\big)} = 1 ,$$

what implies (2.1). □

Notice that (2.1) and (1.1) imply

$$\lim_{a\to\infty} \frac{-\log \mathbb{P}_m(\bar{X}_n > a)}{m \log a} = 1 , \tag{2.2}$$

hence

$$m = \lim_{a\to\infty} m_n(a) , \tag{2.3}$$

where

$$m_n(a) = \frac{-\log \mathbb{P}_m(\bar{X}_n > a)}{\log a} = \frac{-\log\bigl(1 - F_{\bar{X}_n}(a)\bigr)}{\log a} \tag{2.4}$$

with $F_{\bar{X}_n}$ being the distribution function of $\bar{X}_n$. There are two possibilities how to estimate m with the aid of formula (2.4): First, we can estimate the unknown $F_{\bar{X}_n}$ in (2.4) by the empirical distribution function, based on N realizations of $\bar{X}_n$ (nonparametric approach). Second, the distribution function can be modelled by the by some parametric model whose parameters are then estimated. The "perturbed Pareto distribution", considered recently by Feuerverger and Hall [8], is a possible parametric model. Both approaches lead to the asymptotically normal estimators, that are generally biased, unless the distribution has exactly Pareto tails. The parametric model enables to reduce the bias, provided it is correct, e.g. using efficient estimators of its parameters. The bias in the nonparametric approach is expressed by means on the unknown slowly varying function; it can be still reduced if the slowly varying function can be further parametrized.

In the present paper, we shall develop the nonparametric approach, replacing $F_{\bar{X}_n}$ by the empirical distribution function. In this way we obtain a consistent estimator of m under $N \to \infty$, while n remains fixed. Because we need to estimate the limit of (2.4) as $a \to \infty$, the argument a_N of the empirical distribution function should be sufficiently large, but some observations should be still greater than a_N.

The estimator and its properties are described in the next section.

3. ESTIMATOR OF THE TAIL INDEX BASED ON SAMPLE MEANS

Let us partition the set of observations into N non-overlapping samples of the same sizes n (a modification to different sample sizes is possible), denoted as $(X_1^{(1)}, ..., X_n^{(1)})$, ..., $(X_1^{(N)}, ..., X_n^{(N)})$. Then the vector $(\bar{X}_n^{(1)}, ..., \bar{X}_n^{(N)})$ of the corresponding sample means is a random sample from a distribution with distribution function $F_{\bar{X}_n}(x) = \mathbb{P}(\bar{X}_n \le x)$ (unknown).

Denote $\widehat{F}^{(N)}_{\bar{X}_n}(x) = \frac{1}{N}\sum_{j=1}^{N} I[\bar{X}^{(j)}_n \le x]$ the empirical distribution function, based on $(\bar{X}^{(1)}_n, ..., \bar{X}^{(N)}_n)$.

The argument a_N of the empirical distribution function should be sufficiently large, but such that there are still some observations behind a_N. If we know that F is not lighter than the Pareto distribution with index m_0 for some fixed m_0, $0 < m_0 < \infty$, hence we know that $0 < m \le m_0$, then a possible choice of a_N is as in (3.1) below. This situation is considered in the present paper. We can either have such information from the experience or from the character of the experiment.

Remark 3.1. Another possibility would be a preliminary test estimation, when we first apply a preliminary test of the hypothesis **H**: $0 < m \le m_0$. In Section 5 we shall briefly describe one possible test of **H** based on the sample means. Other tests of **H** were recently proposed and numerically illustrated by Picek and Jurečková [14], Jurečková and Picek [12]; the test on the tail of errors in linear model was proposed by Jurečková [11]. A preliminary test estimator will be a subject of the next study.

Choose the sequence $\{a_N\}_{N=1}^{\infty}$, $a_N \to \infty$ as $N \to \infty$, in the following way:

$$a_N = N^{\frac{1-\delta}{m_0}}, \qquad \text{with a fixed } \ \delta \in (0,1) \tag{3.1}$$

and consider the sequence of random functions

$$\widehat{m}_N(a) \;=\; \tilde{m}_N(a)\, I\Big[0 < \widehat{F}^{(N)}_{\bar{X}_n}(a) < 1\Big] + m_0\, I\Big[\widehat{F}^{(N)}_{\bar{X}_n}(a) = 0 \text{ or } 1\Big], \quad a > 0\,, \tag{3.2}$$

where

$$\tilde{m}_N(a) \;=\; \frac{-\log\Big(1 - \widehat{F}^{(N)}_{\bar{X}_n}(a)\Big)}{\log a}\,, \qquad a > 0\,. \tag{3.3}$$

We propose $\widehat{m}_N = \widehat{m}_N(a_N)$ as an estimator of the parameter m; more precisely,

$$\widehat{m}_N \;=\; \tilde{m}_N(a_N)\, I\Big[0 < \widehat{F}^{(N)}_{\bar{X}_n}(a_N) < 1\Big] + m_0\, I\Big[\widehat{F}^{(N)}_{\bar{X}_n}(a_N) = 0 \text{ or } 1\Big] \tag{3.4}$$

with $\tilde{m}_N(a)$ defined in (3.3) and a_N defined in (3.1) with a fixed choice of δ, $0 < \delta < 1$.

We must first show that the estimator $\widehat{m}_N$ is well defined. It follows from the following lemma:

Lemma 3.1. *Let F satisfy the conditions of Lemma 2.1 with $0<m\le m_0$, $m_0>0$ fixed. Let $\{a_N\}$ be the sequence defined in (3.1). Then $a_N \to \infty$ and*

$$\mathbb{P}_m\Big(0 < \hat{F}^{(N)}_{\bar{X}_n}(a_m) < 1\Big) \to 1 \qquad \text{as} \quad N \to \infty \; . \tag{3.5}$$

Proof: If F is heavy-tailed with Pareto index m, satisfying (1.1), then, by Lemma 2.1,

$$\lim_{a\to\infty} \frac{1-F(a)}{m\log a} \;=\; \lim_{a\to\infty} \frac{1-F_{\bar{X}_n}(a)}{m\log a} \;=\; 1 \tag{3.6}$$

and both F and $F_{\bar{X}_n}$ belong to the domain of attraction of the Fréchet distribution Φ_m with the distribution function $\Phi_m(x) = \exp\{-x^{-m}\}$, $x > 0$. Let $\bar{\mathcal{X}}^{(N)}_n = \max_{1\le j\le N} \bar{X}^{(j)}_n$ denote the maximum of $\bar{X}^{(1)}_n, ..., \bar{X}^{(N)}_n$. Then

$$\mathbb{P}_m\bigg(\frac{\bar{\mathcal{X}}^{(N)}_n}{\xi_N} \le x\bigg) \to \Phi_m(x) \qquad \text{as} \quad N\to\infty \tag{3.7}$$

with ξ_N satisfying $N\big[1-F_{\bar{X}_n}(\xi_N)\big] = 1$, $N=1,2,...$; then we conclude from (3.10) that $\xi_N = N^{\frac{1}{m}} L^*_2(N)$ with some slowly varying function L^*_2 and, by (3.7),

$$\mathbb{P}_m\Big(\bar{\mathcal{X}}^{(N)}_n \le a_N\Big) = \mathbb{P}_m\bigg(\frac{\bar{\mathcal{X}}^{(N)}_n}{\xi_N} \le \frac{a_N}{\xi_N}\bigg) \to 0 \qquad \text{as} \quad N\to\infty \; . \tag{3.8}$$

It means that at least one $\bar{X}^{(j)}_n$ lies above a_N with probability tending to 1, and thus

$$\lim_{N\to\infty} \mathbb{P}_m\Big(\hat{F}^{(N)}_{\bar{X}_n}(a_N) < 1\Big) \;=\; 1 \; .$$

On the other hand, we obtain from (3.10),

$$\begin{aligned}
\mathbb{P}_m\bigg(\min_{1\le j\le N} \bar{X}^{(j)}_n \ge a_N\bigg) \;&=\; \Big(1 - F_{\bar{X}_n}(a_N)\Big)^N \\
&=\; a_N^{-mN}\big(L^*(a_N)\big)^N \\
&=\; N^{-\frac{m(1-\delta)}{m_0}N}\big(L^*(a_N)\big)^N \to 0 \qquad \text{as} \quad N\to\infty \; ,
\end{aligned}$$

and hence there is at least one $\bar{X}^{(j)}_n$ below a_N with probability tending to one. This completes the proof of (3.5). ☐

The first main property of estimator $\widehat{m}_N$ is its strong consistency with respect to the asymptotics $N \to \infty$:

Theorem 3.1. *Let $\{X_1, X_2, ...\}$ be a sequence of random variables, identically distributed according to distribution function F of the Pareto type (1.1), satisfying the conditions (i) and (ii) of Lemma 2.1 with $0 < m \leq m_0 < \infty$. Let $\widehat{m}_N$ be the estimator of m defined in (3.4). Then*

$$\widehat{m}_N \to m \quad \text{with probability } 1, \qquad \text{as } N \to \infty . \tag{3.9}$$

The second main result is the asymptotic normality of $\widehat{m}_N$. The problem of estimating m is semiparametric in its nature, involving an unknown slowly varying function. If distribution function F is of the type (1.1) with index m, then Lemma 2.1 implies that $F_{\bar{X}_n}$ also satisfies (1.1) with the same m; hence, by the von Mises condition, it has the form

$$1 - F_{\bar{X}_n}(x) \; = \; x^{-m} L^*(x) \; , \tag{3.10}$$

where $L^*(x)$ is a function, slowly varying at ∞. The presence of L^* can cause a bias in the asymptotic distribution of $\widehat{m}_N$, generally not asymptotically negligible, unless we impose some more restrictive condition on F. We shall see (Lemma 5.1) that $(\widehat{m}_N - m_n(a_N))$, with $m_n(\cdot)$ defined in (2.4), is asymptotically normal and unbiased, while the bias of $(\widehat{m}_N - m)$ is due to the term $(m_n(a_N) - m)$, that tends to 0, but generally not fast enough to eliminate function L^*.

Theorem 3.2. *Under the conditions of Theorem 3.1, the sequence*

$$N^{\frac{1}{2}} \log a_N \left(\frac{1 - F_{\bar{X}_n}(a_N)}{F_{\bar{X}_n}(a_N)} \right)^{\frac{1}{2}} \left(\widehat{m}_N - m + \frac{\log L^*(a_N)}{\log a_N} \right) \tag{3.11}$$

is asymptotically normally distributed as $N \to \infty$, where L^* is the function, defined in (3.10).

Remark 3.2. The order of the coefficient by $\left(\widehat{m}_N(a_N) - m + \frac{\log L^*(a_N)}{\log a_N} \right)$ in (3.11) can be alternatively expressed as

$$\begin{aligned} N^{\frac{1}{2}} \log a_N \left(\frac{1 - F_{\bar{X}_n}(a_N)}{F_{\bar{X}_n}(a_N)} \right)^{\frac{1}{2}} &\sim \tfrac{1-\delta}{m_0} \big(L^*(a_N) \big)^{\frac{1}{2}} \cdot N^{\frac{1}{2}\left(1 - \frac{m}{m_0}(1-\delta)\right)} \log N \\ &\geq \tfrac{1-\delta}{m_0} \big(L^*(a_N) \big)^{\frac{1}{2}} \cdot N^{\frac{\delta m}{2 m_0}} \quad \Big(\to \infty \quad \text{as} \quad N \to \infty \Big) \; , \end{aligned} \tag{3.12}$$

where $b_N \sim c_N$ means that $\lim_{N \to \infty} \frac{b_N}{c_N} \to 1$.

4. NUMERICAL ILLUSTRATION

The performance of the estimation procedure for different choices of m and δ is illustrated on the simulated random samples: The replications ($N = 200$ and $N = 2000$) of samples of sizes $n = 5$ were simulated 1000 times from the following distributions:

Pareto $$F(x) = 1 - \left(\frac{1}{1+x}\right)^m, \quad x \geq 0\ ;$$

Burr $$F(x) = 1 - \left(\frac{1}{1+x^m}\right)^\alpha, \quad x \geq 0\ ;$$

Generalized Pareto $$F(x) = \begin{cases} 1 - \left(1 + \dfrac{x}{m\beta}\right)^{-m} & \text{if } \ x \geq 0,\ \ 0 < m < \infty,\ \ \beta > 0, \\ 1 - \left(1 + \dfrac{x}{m\beta}\right)^{-m} & \text{if } \ 0 \leq x \leq -m\beta,\ \ m < 0,\ \ \beta > 0, \\ 1 - e^{-x/\beta} & \text{if } \ m = \infty,\ \ \beta > 0, \\ 0 & \text{otherwise}\ ; \end{cases}$$

Inverse normal $$F(x) = \begin{cases} 2\left(1 - \Phi\left(\dfrac{1}{\sqrt{x}}\right)\right) & x > 0, \\ 0 & x \leq 0\ . \end{cases}$$

For each distribution we proceeded as follows:

(1) we generated the independent observations $X_1,\ ...,\ X_n, X_{n+1},\ ...,\ X_{2n},\ ...,\ X_{Nn}$;

(2) computed sample means $\overline{X}_n^{(1)}, ..., \overline{X}_n^{(N)}$

(3) and found the empirical distribution function based on $\overline{X}_n^{(1)}, ..., \overline{X}_n^{(N)}$;

(4) for $a_N = N^{\frac{1-\delta}{m_0}}$ we calculated
$$\widehat{m}_N = \tilde{m}_N(a_N)\, I\left[0 < \widehat{F}^{(N)}_{\bar{X}_n}(a_N) < 1\right] + m_0\, I\left[\widehat{F}^{(N)}_{\bar{X}_n}(a_N) = 0 \text{ or } 1\right]\ ;$$

(5) Step (4) was repeated for various values m_0, δ;

(6) For a comparison, the Hill estimator

$$H(k) \;=\; \frac{1}{k}\sum_{i=1}^{k} \log X_{(Nn-i+1:Nn)} - \log X_{(Nn-k:Nn)} \ ,$$

the Pickands estimator

$$P(k) \;=\; \frac{1}{\log 2}\ \log\left(\frac{X_{Nn-k+1:Nn} - X_{Nn-2k+1:Nn}}{X_{Nn-2k+1:Nn} - X_{Nn-4k+1:Nn}}\right) \ ,$$

the moment estimator

$$M(k) \;=\; 1 + M(k)^{(1)} + \frac{1}{2}\left(\frac{\left(M(k)^{(1)}\right)^2}{M(k)^{(2)}} - 1\right)^{-1} ,$$

where

$$M(k)^{(j)} \;=\; \frac{1}{k}\sum_{i=1}^{k}\Big(\log X_{(Nn-i+1:Nn)} - \log X_{(Nn-k:Nn)}\Big)^{j} \ ,$$

and Gomes and Martins [7] estimator

$$GM(k) \;=\; \frac{1}{k}\sum_{i=1}^{k} U_i \;-\; \left(\frac{1}{k}\sum_{i=1}^{k} i\,U_i\right)\frac{\displaystyle\sum_{i=1}^{k}(2\,i-k-1)\,U_i}{\displaystyle\sum_{i=1}^{k} i\,(2\,i-k-1)\,U_i} \ ,$$

where

$$U_i \;=\; i\left[\log\frac{X_{Nn-i+1:Nn}}{X_{Nn-i:Nn}}\right] \ ,$$

were computed for $k = 1, ..., Nn-1$.

(7) steps (1)–(6) were repeated 1 000 times.

(8) Selected sample quantiles of estimates $\left(\widehat{m}_N^1, ..., \widehat{m}_N^{1000}\right)$ and selected sample statistics of pertaining estimates were computed and tabulated.

Selected sample quantiles for different distributions of the errors are summarized in Table 1 and Table 2. Fig. 1 and 2 show the behaviour of the tail index estimator with regard to δ and m_0 in 1000 simulated samples ($N = 2000$) of Pareto with $m = 1$.

For a comparison, the Hill, Pickands, moment and Gomes and Martins [7] estimators were computed. The question is the choice of k, respectively δ for our procedures. To compare we followed the standard approach of minimizing the mean squared error (MSE); the Table 3 give the selected sample statistics of estimators of m for various distribution shapes of errors.

Table 1: Sample quantiles of the estimation of Pareto index under different distributions for some values m_0 and δ $(N\!=\!200)$.

sample distr.	m_0	δ	min	5%	25%	50% quantile	75%	95%	max
Pareto	0.75	0.1	0.279	0.321	0.354	0.379	0.397	0.443	0.580
$m = 0.5$		0.5	0.205	0.242	0.263	0.278	0.297	0.327	0.355
	2	0.1	0.141	0.165	0.187	0.200	0.214	0.239	0.274
		0.5	0.043	0.063	0.080	0.092	0.105	0.123	0.164
Pareto	1.5	0.1	0.694	0.815	0.885	0.942	1.013	1.160	1.449
$m = 1$		0.5	0.548	0.663	0.721	0.763	0.808	0.884	1.038
	3	0.1	0.489	0.617	0.670	0.707	0.747	0.824	0.938
		0.5	0.253	0.311	0.356	0.388	0.428	0.479	0.551
Pareto	3.5	0.1	3.083	3.380	3.500	3.500	3.889	3.889	3.889
$m = 3$		0.5	3.042	3.422	3.717	3.958	4.253	4.874	6.084
	5	0.1	3.252	3.677	4.102	4.404	4.829	5.556	5.556
		0.5	2.272	2.774	2.991	3.185	3.344	3.711	4.254
Burr	0.75	0.1	0.283	0.327	0.354	0.379	0.407	0.443	0.615
$\alpha = 1$		0.5	0.220	0.267	0.293	0.310	0.327	0.360	0.416
$m = 0.5$	2	0.1	0.200	0.232	0.258	0.274	0.295	0.321	0.359
		0.5	0.164	0.202	0.238	0.259	0.280	0.314	0.373
Burr	1.5	0.1	0.694	0.837	0.912	0.975	1.055	1.164	1.667
$\alpha = 1$		0.5	0.654	0.752	0.820	0.870	0.926	1.020	1.250
$m = 1$	3	0.1	0.617	0.737	0.812	0.860	0.911	0.983	1.173
		0.5	0.541	0.667	0.741	0.796	0.843	0.930	1.066
Burr	3.5	0.1	3.083	3.500	3.500	3.500	3.889	3.889	3.889
$\alpha = 1$		0.5	3.500	3.958	4.429	4.633	5.168	6.084	7.000
$m = 3$	5	0.1	3.375	3.868	4.404	4.829	5.000	5.556	5.556
		0.5	3.290	3.847	4.166	4.443	4.767	5.310	6.327
General.	0.75	0.1	0.315	0.370	0.407	0.430	0.456	0.527	0.661
Pareto		0.5	0.270	0.323	0.355	0.376	0.398	0.435	0.528
$m = 0.5$	2	0.1	0.243	0.287	0.317	0.330	0.354	0.390	0.434
β=1		0.5	0.192	0.232	0.264	0.291	0.314	0.349	0.418
General.	1.5	0.1	0.694	0.837	0.912	0.975	1.055	1.164	1.667
Pareto		0.5	0.617	0.737	0.812	0.860	0.911	0.983	1.173
$m = 1$	3	0.1	0.654	0.752	0.820	0.870	0.926	1.020	1.250
$\beta = 1$		0.5	0.541	0.667	0.741	0.796	0.843	0.930	1.066
General.	3.5	0.1	1.901	2.129	2.363	2.574	2.708	3.083	3.889
Pareto		0.5	1.798	2.025	2.180	2.314	2.468	2.716	3.252
$m = 3$	5	0.1	1.635	1.971	2.126	2.266	2.421	2.646	3.042
$\beta = 1$		0.5	1.753	2.007	2.151	2.304	2.437	2.655	3.038
Inverse	0.75	0.1	0.304	0.354	0.388	0.407	0.443	0.488	0.661
normal		0.5	0.207	0.243	0.270	0.287	0.308	0.335	0.390
	2	0.1	0.242	0.297	0.323	0.341	0.365	0.398	0.456
		0.5	0.132	0.168	0.192	0.217	0.238	0.269	0.337

Table 2: Sample quantiles of the estimation of Pareto index under different distributions for some values m_0 and δ ($N = 2000$).

sample distr.	m_0	δ	min	5%	25%	50% quantile	75%	95%	max
Pareto	0.75	0.1	0.361	0.388	0.402	0.411	0.424	0.440	0.476
$m = 0.5$		0.5	0.313	0.327	0.336	0.342	0.349	0.358	0.374
	2	0.1	0.249	0.261	0.269	0.274	0.279	0.288	0.302
		0.5	0.137	0.146	0.151	0.156	0.161	0.168	0.179
Pareto	1.5	0.1	0.881	0.921	0.961	0.989	1.021	1.088	1.185
$m = 1$		0.5	0.825	0.857	0.880	0.899	0.915	0.942	0.978
	3	0.1	0.786	0.826	0.848	0.865	0.881	0.907	0.933
		0.5	0.540	0.562	0.577	0.590	0.604	0.623	0.649
Pareto	3.5	0.1	3.180	3.500	3.500	3.500	3.889	3.889	3.889
$m = 3$		0.5	3.899	4.112	4.288	4.447	4.570	4.879	5.350
	5	0.1	3.739	4.036	4.379	4.542	4.753	5.556	5.556
		0.5	3.648	3.780	3.877	3.968	4.051	4.202	4.488
Burr	0.75	0.1	0.361	0.390	0.402	0.411	0.424	0.440	0.480
$\alpha = 1$		0.5	0.322	0.336	0.345	0.351	0.358	0.368	0.383
$m = 0.5$	2	0.1	0.276	0.293	0.301	0.307	0.313	0.322	0.338
		0.5	0.237	0.248	0.257	0.263	0.270	0.279	0.294
Burr	1.5	0.1	0.881	0.921	0.961	0.989	1.021	1.088	1.185
$\alpha = 1$		0.5	0.843	0.891	0.915	0.933	0.953	0.980	1.028
$m = 1$	3	0.1	0.833	0.873	0.898	0.917	0.933	0.960	0.990
		0.5	0.735	0.783	0.805	0.820	0.838	0.863	0.897
Burr	3.5	0.1	3.180	3.500	3.500	3.500	3.889	3.889	3.889
$\alpha = 1$		0.5	4.112	4.391	4.570	4.712	4.976	5.208	6.362
$m = 3$	5	0.1	3.873	4.133	4.542	4.753	5.049	5.556	5.556
		0.5	4.171	4.391	4.528	4.635	4.752	4.907	5.285
General.	0.75	0.1	0.394	0.424	0.437	0.450	0.464	0.485	0.529
Pareto		0.5	0.373	0.392	0.402	0.410	0.418	0.429	0.444
$m = 0.5$	2	0.1	0.335	0.356	0.366	0.373	0.380	0.391	0.409
β=1		0.5	0.285	0.298	0.307	0.314	0.322	0.333	0.353
General.	1.5	0.1	0.881	0.921	0.961	0.989	1.021	1.088	1.185
Pareto		0.5	0.833	0.873	0.898	0.917	0.933	0.960	0.990
$m = 1$	3	0.1	0.843	0.891	0.915	0.933	0.953	0.980	1.028
$\beta = 1$		0.5	0.735	0.783	0.805	0.820	0.838	0.863	0.897
General.	3.5	0.1	2.410	2.539	2.662	2.765	2.893	3.065	3.889
Pareto		0.5	2.267	2.390	2.471	2.527	2.588	2.696	2.878
$m = 3$	5	0.1	2.178	2.270	2.326	2.379	2.430	2.510	2.671
$\beta = 1$		0.5	2.028	2.161	2.216	2.260	2.301	2.363	2.466
Inverse	0.75	0.1	0.377	0.411	0.426	0.437	0.450	0.468	0.523
normal		0.5	0.306	0.323	0.332	0.338	0.345	0.355	0.377
	2	0.1	0.354	0.369	0.380	0.386	0.394	0.406	0.420
		0.5	0.230	0.242	0.251	0.257	0.264	0.272	0.288

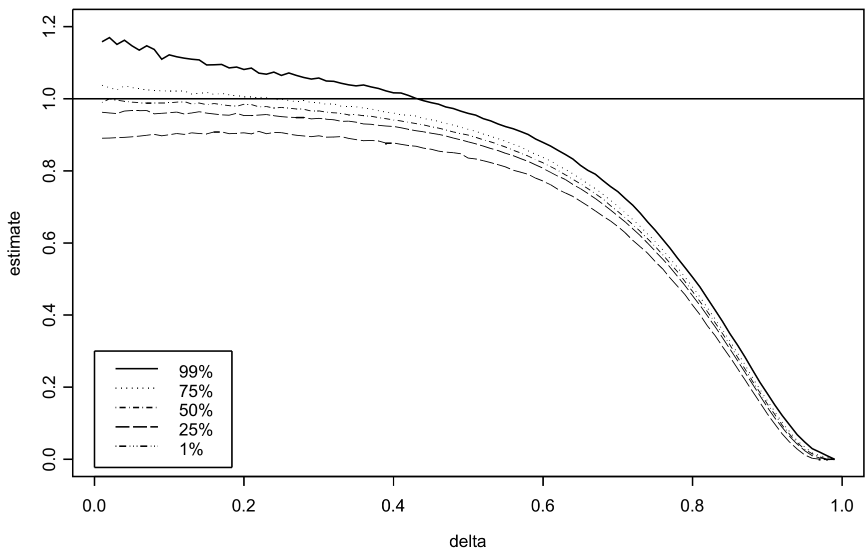


Figure 1: Dependence of tail index estimator in 1000 simulated samples of Pareto ($m = 1$) on the parameter δ for $m_0 = 1.5$.
Plotted are the median and the 1, 25, 75 and 99 percentiles.

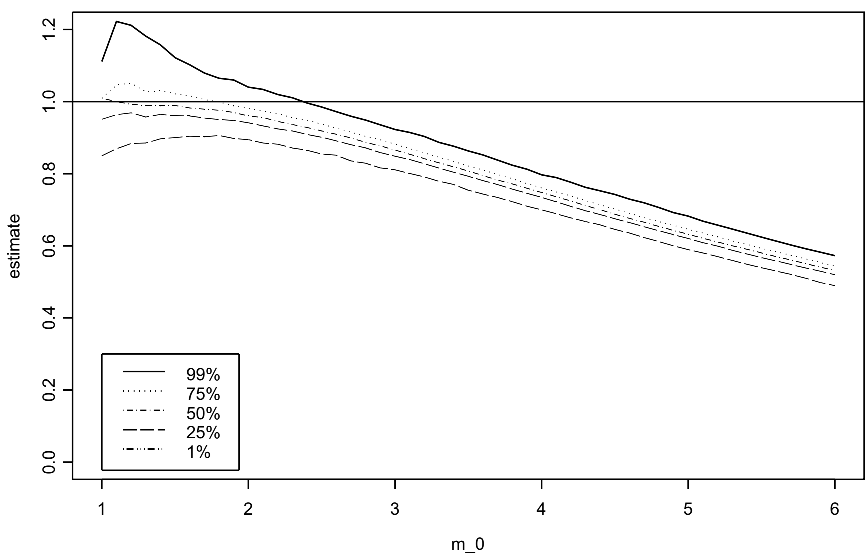


Figure 2: Dependance of tail index estimator in 1000 simulated samples of Pareto ($m = 1$) on the value m_0 for $\delta = 0.1$.
Plotted are the median and the 1, 25, 75 and 99 percentiles.

The Table 3 shows that the estimator based on the sample means (FJP) can be considered as comparable with the most popular estimators of the tail index. The regularity conditions, required for its strong consistency and asymptotic normality, are apparently more transparent and less restrictive.

Table 3: Sample statistics of the estimates of the Pareto index under different distributions for minimal MSE and $N = 200$ and $n = 5$

sample	method	fraction	MSE	mean	median	var	MAD
Pareto	Hill	$k = 998$	0.0010	1.0003	0.9984	0.0010	0.0321
$m = 1$	Moment	$k = 998$	0.0023	1.0053	1.0033	0.0022	0.0454
	Pickands	$k = 985$	0.0221	1.0177	0.9967	0.0218	0.1349
	Gomes	$k = 997$	0.0044	1.0016	0.9968	0.0044	0.0655
	FJP	$\delta = 0.15$	0.0123	0.9542	0.9371	0.0102	0.0900
Burr	Hill	$k = 112$	0.0098	0.9517	0.9489	0.0075	0.0885
$\alpha = 1$	Moment	$k = 257$	0.0101	0.9478	0.9383	0.0074	0.0802
$m = 1$	Pickands	$k = 985$	0.0221	1.0177	0.9967	0.0218	0.1349
	Gomes	$k = 998$	0.0012	1.0007	0.9989	0.0012	0.0345
	FJP	$\delta = 0.22$	0.0111	0.9574	0.9402	0.0093	0.0981
General.	Hill	$k = 310$	0.0010	0.4847	0.4841	0.0007	0.0261
Pareto	Moment	$k = 367$	0.0010	0.4880	0.4863	0.0009	0.0283
$m = 0.5$	Pickands	$k = 993$	0.0020	0.5030	0.4997	0.0020	0.0429
β=1	Gomes	$k = 482$	0.0025	0.5227	0.5210	0.0020	0.0440
	FJP	$\delta = 0.01$	0.0084	0.4177	0.4123	0.0016	0.0395
General.	Hill	$k = 112$	0.0098	0.9517	0.9489	0.0075	0.0885
Pareto	Moment	$k = 257$	0.0101	0.9478	0.9383	0.0074	0.0802
$m = 1$	Pickands	$k = 985$	0.0221	1.0177	0.9967	0.0218	0.1349
$\beta = 1$	Gomes	$k = 998$	0.0012	1.0007	0.9989	0.0012	0.0345
	FJP	$\delta = 0.22$	0.0111	0.9574	0.9402	0.0093	0.0981
General.	Hill	$k = 23$	0.5527	2.4329	2.3598	0.2314	0.4397
Pareto	Moment	$k = 257$	0.5037	2.5140	2.4248	0.2678	0.4368
$m = 3$	Pickands	$k = 890$	16.1112	3.6237	3.0364	15.7379	1.1255
$\beta = 1$	Gomes	$k = 102$	0.4795	2.4276	2.4020	0.1520	0.3966
	FJP	$\delta = 0.01$	0.2869	2.5618	2.5565	0.0949	0.2841
Inverse	Hill	$k = 360$	0.0008	0.4894	0.4888	0.0007	0.0250
normal	Moment	$k = 472$	0.0008	0.4889	0.4881	0.0007	0.0258
	Pickands	$k = 893$	0.0026	0.5142	0.5111	0.0024	0.0467
	Gomes	$k = 588$	0.0021	0.5202	0.5184	0.0017	0.0407
	FJP	$\delta = 0.01$	0.0127	0.3937	0.3890	0.0014	0.0347

5. PROOFS OF THEOREMS 3.1 AND 3.2

5.1. Asymptotic normality

We shall start with the asymptotic normality of $\widehat{m}_N$; and first prove that the sequence

$$N^{\frac{1}{2}} \log a_N \left(\frac{1 - F_{\bar{X}_n}(a_N)}{F_{\bar{X}_n}(a_N)}\right)^{\frac{1}{2}} \Big(\widehat{m}_N - m_n(a_N)\Big) ,$$

with $m_n(\cdot)$ given in (2.4), has asymptotically standard normal distribution:

Lemma 5.1. *Let $\{X_1, X_2, ...\}$ be a sequence of independent random variables, identically distributed with distribution function F of the Pareto type satisfying the conditions (i) and (ii) of Lemma 2.1 with $0 < m \leq m_0 < \infty$. Put $a_N = N^{\frac{1-\delta}{m_0}}$, $0 < \delta < 1$ and*

$$\begin{aligned} \widehat{m}_N &= \tilde{m}_N(a_N)\, I\Big[0 < \hat{F}^{(N)}_{\bar{X}_n}(a_N) < 1\Big] + m_0\, I\Big[\hat{F}^{(N)}_{\bar{X}_n}(a_N) = 0 \text{ or } 1\Big] , \\ \tilde{m}_N(a) &= \frac{-\log\Big(1 - \hat{F}^{(N)}_{\bar{X}_n}(a)\Big)}{\log a} , \qquad a > 0 . \end{aligned} \tag{5.1}$$

Then the sequence

$$N^{\frac{1}{2}} \log a_N \left(\frac{1 - F_{\bar{X}_n}(a_N)}{F_{\bar{X}_n}(a_N)}\right)^{\frac{1}{2}} \Big(\widehat{m}_N - m_n(a_N)\Big) \tag{5.2}$$

with $m_n(x)$ defined in (2.4), is asymptotically normally distributed $\mathcal{N}(0,1)$, as $N \to \infty$ and for any fixed n.

Proof: By the Hungarian embedding theorems (see, e.g., [13]), there exists a sequence of Brownian bridges $\{\mathcal{B}_N\}$, $\mathcal{B}_N$ dependent on $\bar{X}_n^{(1)}, ..., \bar{X}_n^{(N)}$, such that

$$\sup_{a \in \mathbb{R}} \left| \sqrt{N}\Big[1 - \hat{F}^{(N)}_{\bar{X}_n}(a) - \big(1 - F_{\bar{X}_n}(a)\big)\Big] + \mathcal{B}_N\big(F_{\bar{X}_n}(a)\big) \right| = \mathcal{O}\Big(N^{-\frac{1}{2}} \log N\Big) \ \text{a.s.} \tag{5.3}$$

as $N \to \infty$.

Because $B_N(F_{\bar{X}_n}(a))$ is normally distributed $\mathcal{N}\big(0, F_{\bar{X}_n}(a)(1 - F_{\bar{X}_n}(a))\big)$, then

$$\mathbb{P}_m\left(\mathcal{B}_N\big(F_{\bar{X}_n}(a)\big) > C\Big[F_{\bar{X}_n}(a)\Big(1 - F_{\bar{X}_n}(a)\Big)\Big]^{\frac{1}{2}}\right) = 1 - \Phi(C) ,$$

holds for all $a \in \mathbb{R}$ and all $C > 0$, where Φ is the standard normal distribution function; hence $\forall \varepsilon > 0\ \exists C > 0$ such that, for all $a \in \mathbb{R}$,

$$\mathbb{P}_m \left(\mathcal{B}_N\big(F_{\bar{X}_n}(a)\big) > C\Big[F_{\bar{X}_n}(a)\Big(1 - F_{\bar{X}_n}(a)\Big)\Big]^{\frac{1}{2}} \right) < \varepsilon \, . \tag{5.4}$$

Let us first consider the first term of $\widehat{m}_N$, i.e.

$$\tilde{m}_N(a_N)\, I\Big[0 < \hat{F}^{(N)}_{\bar{X}_n}(a_N) < 1\Big] \;=\; \frac{-\log\Big(1 - \hat{F}^{(N)}_{\bar{X}_n}(a_N)\Big)}{\log a_N}\, I\Big[0 < \hat{F}^{(N)}_{\bar{X}_n}(a_N) < 1\Big] \, .$$

We can write

$$\begin{aligned}
&\sqrt{N}\, \log a_N \left(\frac{1 - F_{\bar{X}_n}(a_N)}{F_{\bar{X}_n}(a_N)}\right)^{\frac{1}{2}} \Big(\tilde{m}_N(a_N) - m_n(a_N)\Big)\, I\Big[0 < \hat{F}^{(N)}_{\bar{X}_n}(a_N) < 1\Big] \;= \\
&\quad= \sqrt{N}\, \log a_N \left(\frac{1 - F_{\bar{X}_n}(a_N)}{F_{\bar{X}_n}(a_N)}\right)^{\frac{1}{2}} \\
&\qquad \cdot \left[\frac{-\log\Big(1 - \hat{F}^{(N)}_{\bar{X}_n}(a_N)\Big)}{\log a_N} - \frac{-\log\Big(1 - F_{\bar{X}_n}(a_N)\Big)}{\log a_N}\right] I\Big[0 < \hat{F}^{(N)}_{\bar{X}_n}(a_N) < 1\Big] \\
&\quad= \sqrt{N} \left(\frac{1 - F_{\bar{X}_n}(a_N)}{F_{\bar{X}_n}(a_N)}\right)^{\frac{1}{2}} \\
&\qquad \cdot \left(-\log \left[\frac{1 - \hat{F}^{(N)}_{\bar{X}_n}(a_N)}{1 - F_{\bar{X}_n}(a_N)} - 1 + 1\right]\right) I\Big[0 < \hat{F}^{(N)}_{\bar{X}_n}(a_N) < 1\Big] \, .
\end{aligned} \tag{5.5}$$

An expansion of $\log(1 + x)$ or $\log(1 - x)$, $x > 0$, gives

$$\begin{aligned}
&-\log \left[\frac{1 - \hat{F}^{(N)}_{\bar{X}_n}(a_N)}{1 - F_{\bar{X}_n}(a_N)} - 1 + 1\right] = \\
&\qquad = 1 - \frac{1 - \hat{F}^{(N)}_{\bar{X}_n}(a_N)}{1 - F_{\bar{X}_n}(a_N)} + \mathcal{O}\left(\left[1 - \frac{1 - \hat{F}^{(N)}_{\bar{X}_n}(a_N)}{1 - F_{\bar{X}_n}(a_N)}\right]^2\right);
\end{aligned} \tag{5.6}$$

further we obtain from (5.3)

$$\begin{aligned}
&\sqrt{N} \left(\frac{1 - F_{\bar{X}_n}(a_N)}{F_{\bar{X}_n}(a_N)}\right)^{\frac{1}{2}} \left[1 - \frac{1 - \hat{F}^{(N)}_{\bar{X}_n}(a_N)}{1 - F_{\bar{X}_n}(a_N)}\right] = \\
&\qquad = \frac{\mathcal{B}_N\big(F_{\bar{X}_n}(a_N)\big)}{\Big[F_{\bar{X}_n}(a_N)\Big(1 - F_{\bar{X}_n}(a_N)\Big)\Big]^{\frac{1}{2}}} \Big(1 + \mathbf{o}_p(1)\Big)
\end{aligned} \tag{5.7}$$

and

$$
\sqrt{N}\left(\frac{1-F_{\bar X_n}(a_N)}{F_{\bar X_n}(a_N)}\right)^{\frac12}\left[1-\frac{1-\hat F^{(N)}_{\bar X_n}(a_N)}{1-F_{\bar X_n}(a_N)}\right]^2 =
$$

$$
(5.8)\qquad = N^{-\frac12\left(1-\frac{m}{m_0}\right)-\frac{\delta m}{2m_0}}\,(L^*(a_N))^{-\frac12}\,\frac{\Big(\mathcal{B}_N\left(F_{\bar X_n}(a_N)\right)\Big)^2}{F_{\bar X_n}(a_N)\Big(1-F_{\bar X_n}(a_N)\Big)}
$$

$$
= \mathbf{o}_p\left(N^{-\delta/2}\right) .
$$

It follows from (5.6), (5.7), (5.8) that

$$
\sqrt{N}\,\log a_N\left(\frac{1-F_{\bar X_n}(a_N)}{F_{\bar X_n}(a_N)}\right)^{\frac12}\Big(\tilde m_N(a_N)-m_n(a_N)\Big)\,I\Big[0<\hat F^{(N)}_{\bar X_n}(a_N)<1\Big] =
$$

$$
= \left\{\frac{\mathcal{B}_N\big(F_{\bar X_n}(a_N)\big)}{\Big(\mathrm{var}\ \mathcal{B}_N\big(F_{\bar X_n}(a_N)\big)\Big)^{\frac12}}\Big(1+\mathbf{o}_p(1)\Big)+\mathcal{O}_p\big(N^{-\delta/2}\big)\right\} I\Big[0<\hat F^{(N)}_{\bar X_n}(a_N)<1\Big] ,
$$

hence

$$
\lim_{N\to\infty}\mathbb{P}_m\left(\sqrt{N}\,\log a_N\left(\frac{1-F_{\bar X_n}(a_N)}{F_{\bar X_n}(a_N)}\right)^{\frac12}\Big(\tilde m_N(a_N)-m_n(a_N)\Big)\le y\right) =
$$

$$
(5.9)\qquad = \lim_{N\to\infty}\mathbb{P}_m\left(\sqrt{N}\,\log a_N\left(\frac{1-F_{\bar X_n}(a_N)}{F_{\bar X_n}(a_N)}\right)^{\frac12}\Big(\tilde m_N(a_N)-m_n(a_N)\Big)\le y\,,\right.
$$

$$
\left. 0<\hat F^{(N)}_{\bar X_n}(a_N)<1\right) = \Phi(y) . \qquad \square
$$

Proof of Theorem 3.2: By (3.10), $1-F_{\bar X_n}(x)=x^{-m}L^*(x)$ where L^* is slowly varying at ∞. Moreover, $\widehat m_N-m=(\widehat m_N-m_n(a_N))+(m_n(a_N)-m)$, while

$$
\sqrt{N}\,\log a_N\left(\frac{1-F_{\bar X_n}(a_N)}{F_{\bar X_n}(a_N)}\right)^{\frac12}\Big(\widehat m_N-m_n(a_N)\Big)
$$

has asymptotically the standard normal distribution by Lemma 5.1. By (2.2), $m_n(a_N)-m\to 0$ as $N\to\infty$; more precisely,

$$
\lim_{N\to\infty}\Big(m_n(a_N)-m\Big) = \lim_{N\to\infty}\left[\frac{-\log\Big(1-F_{\bar X_n}(a_N)\Big)}{\log a_N}-m\right]
$$

$$
(5.10)\qquad = \lim_{N\to\infty}\left[\frac{m\log a_N-\log\big(L^*(a_N)\big)}{\log a_N}-m\right]
$$

$$
= \lim_{N\to\infty}\frac{-\log\big(L^*(a_N)\big)}{\log a_N} = 0 ,
$$

but the term

$$\sqrt{N}\left(\frac{1-F_{\bar{X}_n}(a_N)}{F_{\bar{X}_n}(a_N)}\right)^{\frac{1}{2}}\left|\log\bigl(L^*(a_N)\bigr)\right|$$

converges to 0 in only special cases and hence will create a bias; this, together with Lemma 5.1, implies the theorem. □

5.2. Strong consistency

We shall prove Theorem 3.1 with the aid of the following lemma.

Lemma 5.2. *Under the assumptions of Theorem 3.1,*

$$\Bigl(\tilde{m}_N(a_N)-m\Bigr)\,I\Bigl[0<\widehat{F}^{(N)}_{\bar{X}_n}(x)<1\Bigr]\;\to\;0 \tag{5.11}$$

with probability 1, as $N\to\infty$.

Proof: Because $\tilde{m}_N(a_N)-m=[\tilde{m}_N(a_N)-m_n(a_N)]+[m_n(a_N)-m]$ and because of (2.4), it suffices to prove that

$$\Bigl(\tilde{m}_N(a_N)-m_n(a_N)\Bigr)\,I\Bigl[0<\widehat{F}^{(N)}_{\bar{X}_n}(x)<1\Bigr]\;\to\;0 \tag{5.12}$$

with probability 1, as $N\to\infty$. Using (5.5) and (5.6), we obtain

$$\begin{aligned}
&\Bigl(\tilde{m}_N(a_N)-m_n(a_N)\Bigr)\,I\Bigl[0<\hat{F}^{(N)}_{\bar{X}_n}(a_N)<1\Bigr] \;=\\
&\quad=\;(\log a_N)^{-1}\left\{\left[1-\frac{1-\hat{F}^{(N)}_{\bar{X}_n}(a_N)}{1-F_{\bar{X}_n}(a_N)}\right]+\mathcal{O}\left(\left[1-\frac{1-\hat{F}^{(N)}_{\bar{X}_n}(a_N)}{1-F_{\bar{X}_n}(a_N)}\right]^2\right)\right\} \qquad (5.13)\\
&\quad=\;A^{(1)}_N+A^{(2)}_N
\end{aligned}$$

and, using again the strong embedding of empirical processes,

$$\begin{aligned}
A^{(1)}_N \;&=\;(\log a_N)^{-1}\Bigl(1-F_{\bar{X}_n}(a_N)\Bigr)^{-1}\\
&\quad\cdot\Bigl[N^{-\frac{1}{2}}\,\mathcal{B}_N\bigl(F_{\bar{X}_n}(a_N)\bigr)+\mathcal{O}_{\text{a.s.}}\bigl(N^{-1}\log N\bigr)\Bigr]\\
&=\;(\log a_N)^{-1}N^{\frac{1}{2}\left(\frac{m}{m_0}-1\right)-\frac{m\delta}{2m_0}}\bigl(L^*(a_N)\bigr)^{-\frac{1}{2}}\,\frac{\mathcal{B}_N\bigl(F_{\bar{X}_n}(a_N)\bigr)}{\Bigl(F_{\bar{X}_n}(a_N)\Bigl(1-F_{\bar{X}_n}(a_N)\Bigr)\Bigr)^{\frac{1}{2}}} \qquad (5.14)\\
&\quad+\;N^{\frac{m}{m_0}-1-\delta\frac{m}{m_0}}\bigl(L^*(a_N)\bigr)^{-1}\,\mathcal{O}_{\text{a.s.}}(1)\;.
\end{aligned}$$

The second term on the right-hand side of (5.14) converges to 0 almost surely as $N \to \infty$. The first term is normally distributed, hence, because $m \le m_0$, it holds for any $\varepsilon > 0$,

$$\sum_{N=1}^{\infty} \mathbb{P}_m \left((\log a_N)^{-1} N^{\frac{1}{2}\left(\frac{m}{m_0}-1\right)-\frac{m\delta}{2m_0}} \big(L^*(a_N)\big)^{-\frac{1}{2}} \frac{\left|\mathcal{B}_N\big(F_{\bar{X}_n}(a_N)\big)\right|}{\Big(F_{\bar{X}_n}(a_N)\Big(1-F_{\bar{X}_n}(a_N)\Big)\Big)^{\frac{1}{2}}} > \varepsilon \right) \le$$

$$\le\; 2 \sum_{N=1}^{\infty} \left[1 - \Phi\left(\varepsilon \, \frac{1-\delta}{m_0} \, N^{\frac{m\delta}{2m_0}} \, \log N \big(L_1^*(N)\big)^{\frac{1}{2}} \right) \right] . \tag{5.15}$$

Using the inequality $1-\Phi(x) \le \frac{1}{x\sqrt{2\pi}} \, \mathrm{e}^{-\frac{x^2}{2}}, \; x>0$ in (5.15), we obtain

$$\sum_{N=1}^{\infty} \frac{1}{N^{\frac{m\delta}{2m_0}} \log N} \exp\left\{ -\frac{\varepsilon^2}{2} \left(\frac{1-\delta}{m_0}\right)^2 (\log N)^2 \, N^{\frac{m\delta}{m_0}} L_1^*(N) \right\} \;\le$$

$$\le\; K_1 \sum_{N=1}^{\infty} \exp\big\{-K_2 \, N^{\kappa}\big\} \;<\; \infty$$

where $K_1, K_2, \kappa > 0$ are constants, and by the Borel–Cantelli lemma we conclude that the first term on the right-hand side of (5.14) also converges to 0 almost surely as $N \to \infty$. Similarly we prove that $A_N^{(2)} = \mathbf{o}(1)$ *a.s.* as $N \to \infty$. This proves (5.12) and, in turn, (5.11). □

Proof of Theorem 3.1: For any $\varepsilon > 0$, it holds

$$\sum_{N=1}^{\infty} \mathbb{P}_m\Big(\big|\widehat{m}_N(a_N) - m\big| > \varepsilon \Big) \;\le$$

$$\begin{aligned} &\le \sum_{N=1}^{\infty} \mathbb{P}_m\Big(\big|\tilde{m}_N(a_N) - m\big| \, I\Big[0 < \hat{F}_{\bar{X}_n}^{(N)}(a_N) < 1\Big] > \frac{\varepsilon}{2} \Big) \\ &\quad + \sum_{N=1}^{\infty} \mathbb{P}_m\Big((m_0 - m) \, I\Big[\hat{F}_{\bar{X}_n}^{(N)}(a_N) = 0 \text{ or } 1\Big] > \frac{\varepsilon}{2} \Big) . \end{aligned} \tag{5.16}$$

The convergence of the first series on the right-hand side of (5.16) follows from Lemma 5.2. The sum of the second series is bounded from above by

$$\begin{aligned} &\sum_{N=1}^{\infty} \mathbb{P}_m\Big((m_0 - m) \, I\Big[\hat{F}_{\bar{X}_n}^{(N)}(a_N) = 0\Big] > \frac{\varepsilon}{4} \Big) \;+ \\ &\qquad + \sum_{N=1}^{\infty} \mathbb{P}_m\Big((m_0 - m) \, I\Big[\hat{F}_{\bar{X}_n}^{(N)}(a_N) = 1\Big] > \frac{\varepsilon}{4} \Big) \;\le \\ &\le \sum_{N=1}^{\infty} \Big(1 - F_{\bar{X}_n}(a_N)\Big)^N + \sum_{N=1}^{\infty} \big(F_{\bar{X}_n}(a_N)\big)^N \\ &= \sum_{N=1}^{\infty} \big(a_N^{-m} L^*(a_N)\big)^N + \sum_{N=1}^{\infty} \Big(1 - a_N^{-m} L^*(a_N)\Big)^N . \end{aligned} \tag{5.17}$$

Because $a_N^{-\eta} < L^*(a_N) < a_N^{\eta}$ for $N > N_\eta$ and $\forall\, \eta > 0$, we conclude

$$(5.18) \qquad \big(a_N^{-m} L^*(a_N)\big)^N < \big(a_N^{-m+\eta}\big)^N \le \left(\frac{1}{N^{\frac{1-\delta}{m_0}(m-\eta)}}\right)^N$$

hence the first series on the right-hand side of (5.17) converges for sufficiently small η. Similarly,

$$(5.19) \qquad \begin{aligned} \Big(1 - a_N^{-m} L^*(a_N)\Big)^N &< \Big(1 - a_N^{-m-\eta}\Big)^N \\ &\le \left(1 - \frac{1}{N^{\frac{1-\delta}{m_0}(m+\eta)}}\right)^N \\ &\le \left[\exp\left\{-\big(N^{\frac{1-\delta}{m_0}(m+\eta)}\big)^{-1}\right\}\right]^N, \end{aligned}$$

what implies the convergence of the second series on the right-hand side of (5.17). □

6. TEST ON THE PARETO INDEX

We shall now briefly describe one possible test of the hypothesis on the Pareto index, based on the sample means. For other tests we refer to [12] and [14].

Because the problem is of semiparametric nature, we should first think over a proper formulation of the hypothesis. Following [12], we shall consider the hypothesis

$$(6.1) \qquad \mathbf{H}_{m_0} : \qquad x^{m_0}\Big(1 - F(x)\Big) \ge 1 \quad \forall\, x > x_0$$

with a hypothetical $m_0 > 0$ and with some $x_0 \ge 0$. Such hypothesis and hence the test are nonparametric; the test is based on splitting the set of observations into N subsamples of sizes n and on the empirical distribution function of the means of the subsamples; the asymptotics is for $N \to \infty$ and fixed n (eventually small), and the asymptotic null distribution of the test criterion is normal. The proposed test is consistent against exponentially tailed alternatives, as well as against heavy tailed alternatives with index $m > m_0$. The test is asymptotically unbiased for the broad family of distributions represented by $\mathbf{H}_{m_0}$ and its alternative. Such test may be used as a supplement to the usual tests of the Gumbel hypothesis $m = \infty$ against $m < \infty$, namely in the situation that the latter tests reject the hypothesis of exponentiality.

Similarly as in the estimation, we partition the set of observations into N non-overlapping samples of the same sizes n, denoted as

$$\big(X_1^{(1)},...,X_n^{(1)}\big), \quad ..., \quad \big(X_1^{(N)},...,X_n^{(N)}\big) \tag{6.2}$$

and denote $\bar{X}_n^{(1)},...,\bar{X}_n^{(N)}$ the respective sample means. Let $F_{\bar{X}_n}(x) = \mathbb{P}_m(\bar{X}_n \le x)$ be the common distribution function of the sample means and let $\widehat{F}_{\bar{X}_n}^{(N)}(x) = \frac{1}{N}\sum_{j=1}^{N} I[\bar{X}_n^{(j)} \le x]$ be the corresponding empirical distribution function. Let

$$a_N = N^{(1-\delta)/m_0}\;, \qquad 0 < \delta < 1\;. \tag{6.3}$$

We propose the test rejecting $\mathbf{H}_{m_0}$ if

$$\begin{aligned} &\text{either} \quad \hat{F}_{\bar{X}_n}^{(N)}(a_n) = 1 \\ &\text{or} \quad \hat{F}_{\bar{X}_n}^{(N)}(a_n) < 1 \quad \text{and simultaneously} \\ &N^{\delta/2}\Big[-\log\Big(1-\hat{F}_{\bar{X}_n}^{(N)}(a_n)\Big) - (1-\delta)\log N\Big] \ge \Phi^{-1}(1-\alpha) \end{aligned} \tag{6.4}$$

where Φ is the standard normal distribution function. If F satisfies (6.1) as an equality, as the Pareto distribution, then α is the asymptotic probability of the error of the first kind; for any other distribution satisfying (6.1), the asymptotic probability of the error of the first kind it is $\le \alpha$.

The asymptotic null distribution of the test is described in the following theorem:

Theorem 6.1. *Let $X_1, X_2, ...$ be independent observations, identically distributed according to absolutely continuous distribution function F satisfying (6.1). Let $\hat{F}_{\bar{X}_n}^{(N)}$ be the empirical distribution function of the means of samples (6.2). Then*

$$\lim_{N\to\infty} \mathbb{P}_{m_0}\Big(\hat{F}_{\bar{X}_n}^{(N)}(a_N) < 1\Big) = 1 \tag{6.5}$$

with a_N defined in (6.3), and

(6.6)
$$\lim_{N\to\infty} \mathbb{P}_{m_0}\Big\{N^{\delta/2}\Big[-\log\Big(1-\hat{F}_{\bar{X}_n}^{(N)}(a_N)\Big) - (1-\delta)\log N\Big] \ge \tau_\alpha\,,\; \hat{F}_{\bar{X}_n}^{(N)}(a_N) < 1\Big\} \le \alpha\,,$$

where $\tau_\alpha = \Phi^{-1}(1-\alpha)$, $0<\alpha<1$, and Φ is the standard normal distribution function.

Moreover, if there exists x_0 such that

$$x^{m_0}\Big(1 - F(x)\Big) = 1 \qquad \text{for} \quad x > x_0 \ , \tag{6.7}$$

then

(6.8)
$$\lim_{N\to\infty} \mathbb{P}_{m_0}\bigg\{N^{\delta/2}\Big[-\log\Big(1-\hat{F}^{(N)}_{\bar{X}_n}(a_N)\Big) - (1-\delta)\log N\Big] \ge \tau_\alpha \,, \ \hat{F}^{(N)}_{\bar{X}_n}(a_N) < 1\bigg\} = \alpha \,.$$

Proof: First, (6.5) follows from Lemma 3.1. Further, (6.8) follows from the proof of Lemma 5.1, namely from (5.9), where we insert the pertinent expressions for $m_n(a)$ and a_N according to (2.4), (5.1) and (6.3), respectively.

If F satisfies (6.1), then the right tail of $1-\hat{F}^{(N)}_{\bar{X}_n}$ is ultimately not smaller stochastically than that of F satisfying (6.7); this implies (6.6). □

The following Corollary shows the set of alternatives against which is the test asymptotically unbiased:

Corollary 6.1.

(i) Under the conditions of Theorem 6.1, the test with the critical region (6.4) is asymptotically unbiased for the hypothesis $\mathbf{H}_{m_0}$ against the alternative

$$x^{m_0}\Big(1 - F(x)\Big) < 1 \qquad \text{for} \quad x > x_0 \ . \tag{6.9}$$

(ii) The test attains the asymptotic power 1 against the alternative that F is of type (1.1) with index $m > m_0$, including $m = \infty$.

Proof: Under $\mathbf{H}_{m_0}$, (6.6) holds by Theorem 6.1, hence the asymptotic size of the test is equal to α for the whole hypothesis $\mathbf{H}_{m_0}$.

Under (6.9), $1-\hat{F}^{(N)}_{\bar{X}_n}$ is ultimately stochastically smaller than under Pareto with index m_0, hence

$$\lim_{N\to\infty} \mathbb{P}_{m_0}\bigg\{N^{\delta/2}\Big[-\log\Big(1-\hat{F}^{(N)}_{\bar{X}_n}(a_N)\Big) - (1-\delta)\log N\Big] \ge \tau_\alpha \,, \ \hat{F}^{(N)}_{\bar{X}_n}(a_N) < 1\bigg\} \ \ge \ \alpha \,.$$

This proves the asymptotic unbiasedness.

Let now F satisfy (1.1) with index $m > m_0$. Then $F_{\bar{X}_n}$ also satisfies (1.1) and it implies that, given $\varepsilon > 0$, there exists N_0 such that, for $N > N_0$,

$$N^{-\frac{m}{m_0}(1+\varepsilon)(1-\delta)} \ = \ a_N^{-m(1+\varepsilon)} \ \le \ 1 - F_{\bar{X}_n}(a_N) \ \le \ a_N^{-m(1-\varepsilon)} \ = \ N^{-\frac{m}{m_0}(1-\varepsilon)(1-\delta)} \ .$$

If $1-\hat{F}^{(N)}_{\bar{X}_n}(a_N)=0$, we reject $\mathbf{H}_{m_0}$. If $1-\hat{F}^{(N)}_{\bar{X}_n}(a_N)>0$, then

$$\mathbb{P}_{m_0}\Big\{N^{\delta/2}\Big[-\log\Big(1-\hat{F}^{(N)}_{\bar{X}_n}(a_N)\Big)-(1-\delta)\log N\Big]\geq\tau_\alpha\Big\} = \tag{6.10}$$

$$= \mathbb{P}_m\left\{N^{\delta/2}\left[-\log\left(\frac{1-\hat{F}^{(N)}_{\bar{X}_n}(a_N)}{1-F_{\bar{X}_n}(a_N)}\right)-\log\Big(1-F_{\bar{X}_n}(a_N)\Big)-(1-\delta)\log N\right]\geq\tau_\alpha\right\}$$

$$\to 1 \qquad \text{as} \quad N\to\infty\ ,$$

because the first term of the argument on the right-hand side of (6.10) is stochastically bounded under index m (cf. the proof of Lemma 3.1), while the second term tends to infinity for $m>m_0$. Hence, we reject $\mathbf{H}_{m_0}$ with probability tending to 1. The case $m=\infty$ corresponds to the exponential tail. □

The performance of the test procedure for different choices of m_0 is illustrated again on the simulated random samples. The replications ($N=200$) of samples of sizes $n=5$ were simulated 1000 times. Fig. 3–5 show the number of rejection of the null hypothesis $\mathbf{H}_{m_0}$ as a function of m_0 for Pareto ($m=1$), Burr ($m=2$) and generalized Pareto ($m=0.5$) distributions with $\delta=0.1,\ 0.5$ on the level $\alpha=0.01,\ 0.05$.

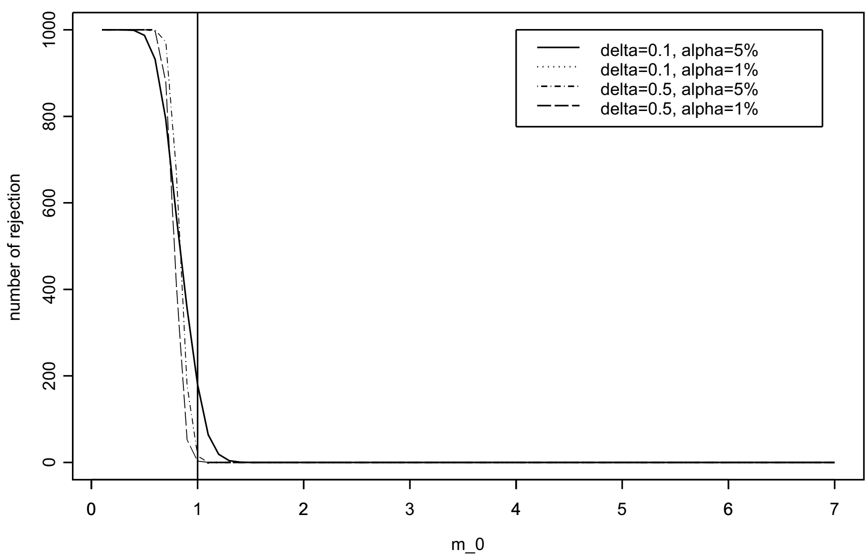


Figure 3: The number of rejection of $\mathbf{H}_{m_0}$ as a function of m_0 for Pareto distribution with $m=1$.

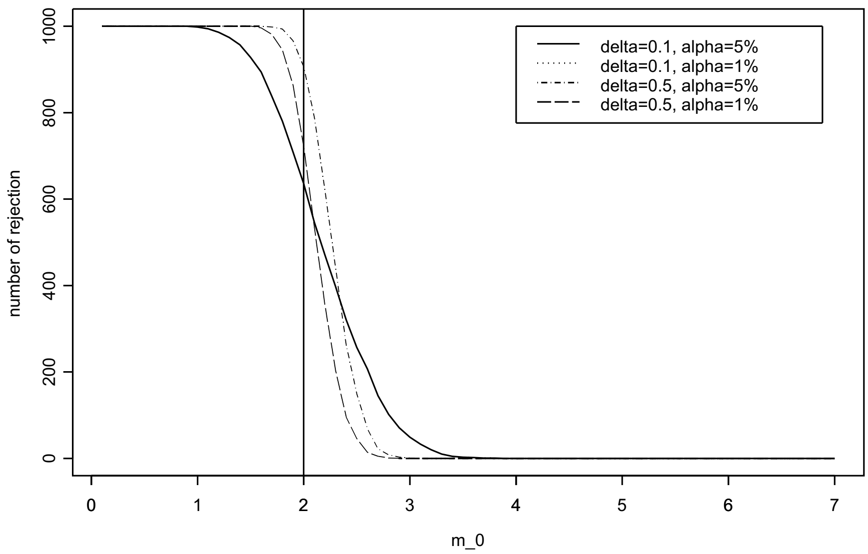


Figure 4: The number of rejection of $\mathbf{H}_{m_0}$ as a function of m_0 for Burr distribution with $m = 2$.

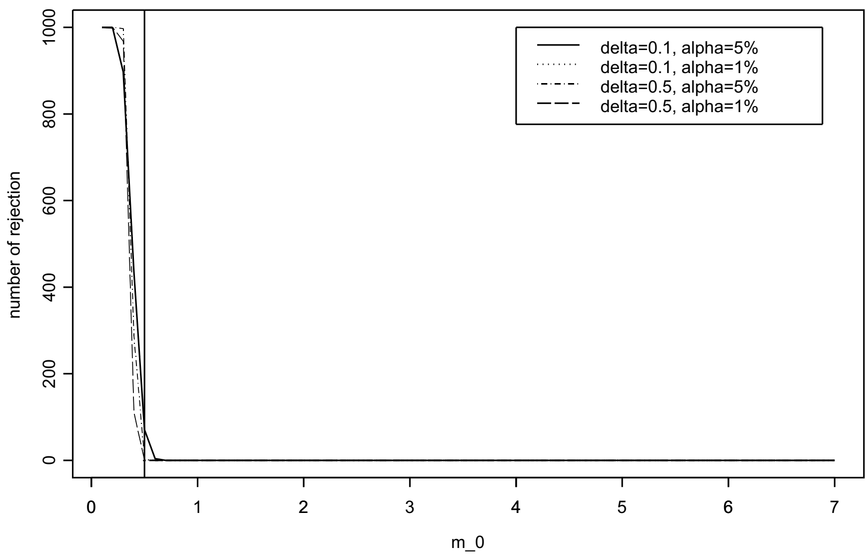


Figure 5: The number of rejection of $\mathbf{H}_{m_0}$ as a function of m_0 for Generalized Pareto distribution with $m = 0.5$.

ACKNOWLEDGMENTS

The authors thank two referees for their valuable comments and suggestions, which led to a better readability of the text and to further considerations.

The research of the first and the second authors was supported by the Research Project CEZ: MSM 113200008 J13/98:113200008 "Mathematical Methods in Stochastics", the third author was supported by postdoctoral grant from the Portuguese Fundação para a Ciência e a Tecnologia (FCT) at the University of Lisbon (the Centro de Estatística e Aplicações). The second author was also supported by the Czech Republic Grant 201/02/0621 and the third by GAČR 205/02/0871.

REFERENCES

[1] Beirlant, J.; Teugels, J.L. and Vynckier, P. (1996). *Practical Analysis of Extreme Values*, Leuven University Press.

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

[3] Dekkers, A.L.M.; Einmahl, J.H.J. and de Haan, L. (1989). A moment estimator for the index of the extreme-value distribution, *Ann. Statist.*, **17**, 1833–1855.

[4] Embrechts, P.; Klüppelberg, C. and Mikosch, T. (1997). *Modelling of Extremal Events in Insurance and Finance*, Springer Verlag.

[5] Fialová, A. (2001). *Estimators and tests of Pareto index* (in Czech), PhD Thesis, Charles University, Prague.

[6] Fialová, A.; Jurečková, J. and Picek, J. (2002). *New type of estimator of Pareto index*, Preprint #28, Dept. of Statistics, Charles University, Prague.

[7] Gomes, M.I. and Martins, M.J. (2002). Asymptotically unbiased estimators of the tail index based on external estimation of the second order parameter, *Extremes*, **5**(1), 5–31.

[8] Feuerverger, A. and Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution, *Ann. Statist.*, **27**, 760–781.

[9] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *Ann. Statist.*, **3**, 1163–1174.

[10] Jurečková, J. (1981). Tail-behavior of location estimators, *Ann. Statist.*, **9**, 578–585.

[11] Jurečková, J. (2000). Tests of tails based on extreme regression quantiles, *Statist. & Probab. Letters* **49**, 53–61.

[12] Jurečková, J. and Picek, J. (2001). A class of tests on the tail index, *Extremes*, **4**(2), 165–183.

[13] Komlós, J.; Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent R.V.'s and the sample DF, *Y. Wahrscheinlichkeitstheorie verw. Geb.*, **32**, 111–131.

[14] Picek, J. and Jurečková, J. (2001). *A class of tests on the tail index using the modified extreme regression quantiles.* In "ROBUST'2000" (J. Antoch, G. Dohnal, Eds.), Union of Czech Mathematicians and Physicists, Prague, 217–226.

[15] Pickands, J. (1975). Statistical inference using extreme order statistics, *Ann. Statist.*, **3**, 119–131.

[16] Smith, R.L. (1987). Estimating tails of probability distribution, *Ann. Statist.*, **15**(3), 1174–1207.

REVSTAT – STATISTICAL JOURNAL

Background

Statistical Institute of Portugal (INE), well aware of how vital a statistical culture is in understanding most phenomena in the present-day world, and of its responsibility in disseminating statistical knowledge, started the publication of the scientific statistical journal Revista de Estatística, in Portuguese, publishing three times a year papers containing original research results, and application studies, namely in the economic, social and demographic fields.

In 1998 it was decided to publish papers also in English. This step has been taken to achieve a larger diffusion, and to encourage foreign contributors to submit their work.

At the time, the Editorial Board was mainly composed by Portuguese university professors, being now composed by national and international university professors, and this has been the first step aimed at changing the character of Revista de Estatística from a national to an international scientific journal.

In 2001, the Revista de Estatística published three volumes special issue containing extended abstracts of the invited contributed papers presented at the 23^{rd} European Meeting of Statisticians.

The name of the Journal has been changed to REVSTAT – STATISTICAL JOURNAL, published in English, with a prestigious international editorial board, hoping to become one more place where scientists may feel proud of publishing their research results.

- The editorial policy will focus on publishing research articles at the highest level in the domains of Probability and Statistics with emphasis on the originality and importance of the research.
- All research articles will be refereed by at least two persons, one from the Editorial Board and another, external.
- The only working language allowed will be English.
- For 2004 two volumes are scheduled for publication.
- On average, four articles will be published per issue.

Aims and Scope

The aim of REVSTAT is to publish articles of high scientific content, in English, developing innovative statistical scientific methods and introducing original research, grounded in substantive problems.

REVSTAT covers all branches of Probability and Statistics. Surveys of important areas of research in the field are also welcome.

Abstract/indexed in

REVSTAT is expected to be abstracted/indexed at least in Current Index to Statistics, Mathematical Reviews, Statistical Theory and Method Abstracts, and Zentralblatt für Mathematic.

Instructions to Authors, special-issue editors and publishers

Papers may be submitted in two different ways:

- By sending a paper copy to the Executive Editor and one copy to one of the two Editors or Associate Editors whose opinion the author(s) would like to be taken into account, together with a postscript or a PDF file of the paper to the e-mail: revstat@fc.ul.pt.
- By sending a paper copy to the Executive Editor, together with a postscript or a PDF file of the paper to the e-mail: revstat@fc.ul.pt.

Submission of a paper means that it contains original work that has not been nor is about to be published elsewhere in any form.

Submitted manuscripts should be typed on one side, in double spacing, with a left margin of at least 3 cm and not have more than 30 pages.

The first page should include the name, affiliation and address of the author(s) and a short abstract with the maximum of 100 words, followed by the key words up to the limit of 6, and the AMS 2000 subject classification.

Authors are obliged to write the final version of accepted papers using LaTeX, in the REVSTAT style.

This style (REVSTAT.sty), and examples file (REVSTAT.tex), which may be download to PC Windows System (Zip format), Mackintosh, Linux and Solaris Systems (StuffIt format), and Mackintosh System (BinHex Format), are available in the REVSTAT link of the National Statistical Institute's Website: http://www.ine.pt/revstat.html

Additional information for the authors may be obtained in the above link.

Accepted papers

Authors of accepted papers are requested to provide the LaTeX files to the e-mail: liliana.martins@ine.pt.

Such e-mail message should include the author(s)'s name, mentioning that it has been accepted by REVSTAT. The authors should also mention if encapsulated postscript figure files were included, and submit electronics figures separately in .tiff, .gif, .eps or .ps format. Figures must be a minimum of 300 dpi.

Also send always the final paper version to:

Adrião Ferreira da Cunha
Executive Editor, REVSTAT – STATISTICAL JOURNAL
Instituto Nacional de Estatística
Av. António José de Almeida 5
1000-043 LISBOA
PORTUGAL

Copyright and Reprints

Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher, the INE, in order to ensure the widest possible dissemination of information, namely through the National Statistical Institute's Website (http://www.ine.pt).

After assigning the transfer copyright form, authors may use their own material in other publications provided that the REVSTAT is acknowledged as the original place of publication. The Executive Editor of the Journal must be notified in writing in advance.

Authors of articles published in the REVSTAT will be entitled to one free copy of the respective issue of the Journal and twenty-five reprints of the paper are provided free. Additional reprints may be ordered at expenses of the author(s), and prior to publication.