Change Point Analysis of Extreme Values

Goedele Dierckx

Economische Hogeschool Sint Aloysius, Brussels, Belgium

e-mail: goedele.dierckx@hubrussel.be

Jef L. Teugels

Katholieke Universiteit Leuven, Belgium &

EURANDOM, Technical University Eindhoven, the Netherlands

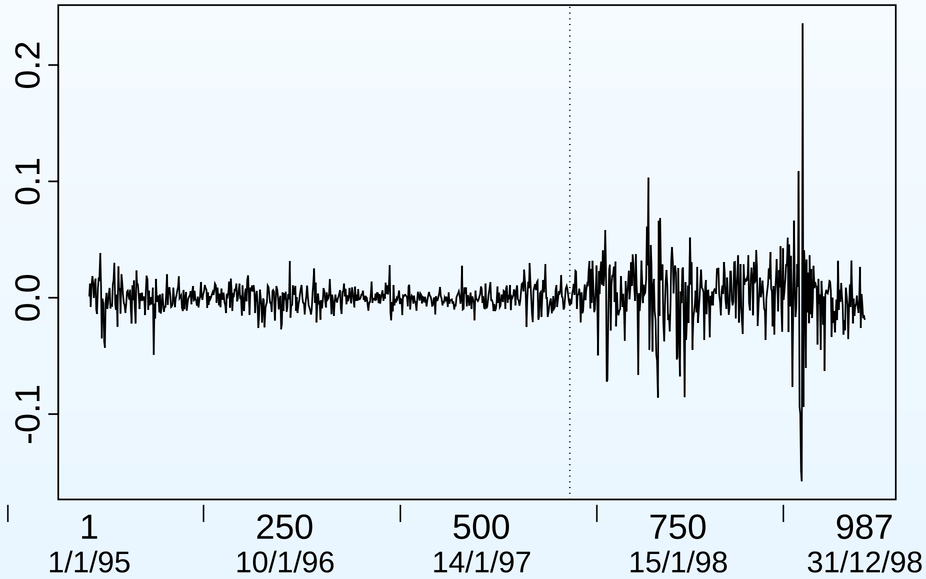
e-mail: jef.teugels@wis.kuleuven.be

Overview

1. Introduction
2. Test statistic
   (a) Construction
   (b) Extreme value situation
   (c) Asymptotics
   (d) Practical procedure
3. Examples
   (a) Simulation
   (b) Malaysian Stock Index.
       - Classical Approach
       - Improved Approach
   (c) Earthquake Data
   (d) Nile Data
   (e) Swiss-Re Catastrophic Data
4. Conclusions
5. References

1. INTRODUCTION

We start with an example where a change point has occurred. 987 measurements of the **Daily Stock Market Returns** of the **Malaysian Stock Index**. Jan. 1995 – Dec. 1998, covering the Asian financial crisis, July 1997.



Changes in

- distribution?
- in parameters of a distribution?
    - central behavior?
    - tail behavior?

2.a. Construction of Test Statistic

Start with a sample $X_1, \ldots, X_{m^*}, X_{m^*+1}, \ldots X_n$, from a density function $f(x; \theta_i, \eta)$. Csörgő and Horváth (1997) test whether θ_i changes at some point m^*

$$\begin{aligned} H_0 &: \quad \theta_1 = \theta_2 = \ldots = \theta_n \quad \text{versus} \\ H_1 &: \quad \theta_1 = \ldots = \theta_{m^*} \neq \theta_{m^*+1} = \ldots = \theta_n \text{ for some } m^*. \end{aligned}$$

using the **test statistic**

$$Z_n = \sqrt{\max_{1 \leqslant m < n} \left(-2 \log \Lambda_m\right)} \, ,$$

where

$$\Lambda_m = \frac{\sup_{\theta,\eta} \prod_{i=1}^{n} f(X_i; \theta, \eta)}{\sup_{\theta,\tau,\eta} \prod_{i=1}^{m} f(X_i; \theta, \eta) \prod_{i=m+1}^{n} f(X_i; \tau, \eta)} \, .$$

2.TEST STATISTIC

Example

For the **exponential distribution** where X_i has mean θ_i

$$-2\log\Lambda_m = 2\left[-m\log\frac{1}{m}\sum_{i=1}^{m}X_i-(n-m)\log\frac{1}{n-m}\sum_{i=m+1}^{n}X_i+n\log\frac{1}{n}\sum_{i=1}^{n}X_i\right]$$

For large n, m and $n-m$ one can expect 'normal' behaviour expressed in terms of Brownian motions.

2.TEST STATISTIC

2.b. Extreme Value Situation

Assume that $X_{n,n}$ is the maximum in a sample of independent random variables with a common distribution. Maximum domain of attraction condition

$$\lim_{n\to\infty} P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) = G_\gamma(x) \ .$$

Under very weak conditions we get the approximation

$$P\left(X_{n,n} \leq y\right) \approx G_\gamma(b_n + a_n\, x)$$

where γ is a real-valued **extreme value index** and

$$G_\gamma(x) = \exp -\{1 + \gamma x\}_+^{-1/\gamma}$$

an **extremal law.**

When $\gamma > 0$ we end up with heavy right-tailed distributions, the **Pareto-Fréchet Case.**

2.TEST STATISTIC

We concentrate on changes of parameters that describe the tail of distributions appearing in extreme value analysis.

- X has a **Pareto-type distribution** with parameter $\theta = \gamma$, when the relative excesses of X over a high threshold u, given that X exceeds u satisfy the condition

$$P\left(\frac{X}{u} > x | X > u\right) \rightarrow x^{-\frac{1}{\gamma}} \; . \; u \rightarrow \infty,$$

- More generally X follows a **Generalized Pareto distribution (GPD)** with parameter $\theta = (\gamma, \sigma)$ if the behavior of the absolute excesses over a high threshold u satisfies the condition

$$P\left(X - u > x | X > u\right) \rightarrow \left(1 + \frac{\gamma x}{\sigma}\right)^{-\frac{1}{\gamma}}, \quad u \rightarrow \infty.$$

2.TEST STATISTIC

For large values, $\log$ of Pareto-type with extreme value index γ_i is close to be exponential with mean γ_i.

- The most classical approach for the **estimation** of the extreme value index $\gamma > 0$ is to use the **Hill estimator**:

$$H_{k,n} = \frac{1}{k}\sum_{i=1}^{k} \log X_{n-i+1,n} - \log X_{n-k,n} \ .$$

  Hence, only a segment of the available data is used.

- The determination of the quantity k is important. Alternatively, we look at extremes above a **threshold** $u = X_{n-k,n}$. The Hill estimator has
  - small bias but large variance for small k
  - large bias but small variance for large k.

  As a compromise we select k such that the **empirical mean squared error** is minimal.

2.TEST STATISTIC

1. Pareto-type density

Suppose $X_1, \ldots, X_m, X_{m+1}, \ldots X_n$ are independent and Pareto-type distributed. We denote the extreme value index for X_i by γ_i. In order to determine whether the index γ changes or not, we perform the following test

$$\begin{aligned} H_0 &: \quad \gamma_1 = \gamma_2 = \ldots = \gamma_n = \gamma \ \text{ versus} \\ H_1 &: \quad \gamma_1 = \gamma_{m^*} \neq \gamma_{m^*+1} = \gamma_n \text{ for some } m^* \end{aligned}$$

Hence $Z_n = \sqrt{\max_{1 \leqslant m < n} \left(-2 \log \Lambda_m\right)}$

where in turn

$$\log \Lambda_m = \left[k_1 \log H_{k_1,m} + (k - k_1) \log H_{k-k_1,n-m} - k \log H_{k,n}\right]$$

$$+ \left[\frac{1}{H_{k,n}} \left(k_1 H_{k_1,m} + (k - k_1) H_{k-k_1,n-m} - k H_{k,n}\right)\right] .$$

2.TEST STATISTIC

2. GPD. Suppose now that X_i is GPD with parameters $\theta_i = (\gamma_i, \sigma_i)$.To perform the test

$$\begin{aligned} H_0 &: \quad \theta_1 = \theta_2 = \ldots = \theta_n \quad \text{versus} \\ H_1 &: \quad \theta_1 = \ldots = \theta_{m^*} \neq \theta_{m^*+1} = \ldots = \theta_n \text{ for some } m^* \end{aligned}$$

we use as test statistic $Z_n = \sqrt{\max_{1 \leqslant m < n} (-2 \log \Lambda_m)}$, where

$$-2 \log \Lambda_m = 2 \left[L_{k_1}(\hat{\theta}_{k_1}) + L^+_{k_1}(\hat{\theta}^+_{k_1}) - L_k(\hat{\theta}_k) \right]$$

$$\begin{aligned} L_m(\hat{\theta}_m) &= -m \log \hat{\sigma}_m - \left(\frac{1}{\hat{\gamma}_m} + 1 \right) \sum_{i=1}^{m} \log \left(1 + \hat{\gamma}_m \frac{x}{\hat{\sigma}_m} \right) \\ L^+_m(\hat{\theta}^+_m) &= -(n-m) \log \hat{\sigma}^+_m - \left(\frac{1}{\hat{\gamma}^+_m} + 1 \right) \sum_{i=m+1}^{n} \log \left(1 + \hat{\gamma}^+_m \frac{x}{\hat{\sigma}^+_m} \right) \end{aligned}$$

and likelihood estimators $(\hat{\gamma}_m, \hat{\sigma}_m)$ resp. $(\hat{\gamma}^+_m, \hat{\sigma}^+_m)$ based on $X_1, X_2, \ldots, X_m$ and $X_{m+1}, \ldots X_n$ are obtained by numerical procedures.

2.c. Asymptotics

Using the procedure suggested by Csörgő and Horváth we have

Theorem Suppose $X_1, \ldots, X_m, X_{m+1}, \ldots X_n$ are independent and identically distributed. We set the threshold at $u = X_{n-k,n}$. Define

$$Z_n = \sqrt{\max_{c_n \leqslant m < n-d_n} (-2 \log \Lambda_m)},$$

with $-2 \log \Lambda_m$ as before. Let $n, k \to \infty$ such that $k/n \to 0$. Let further c_n and d_n be intermediate sequences for which $c_n/n \to 0$ and $d_n/n \to 0$. Then, under H_0 of our test,

$$Z_n \to_d \begin{cases} \sqrt{\sup_{0 \leqslant t < 1} \frac{B^2(t)}{t(1-t)}} & \text{if Pareto-type ,} \\ \sqrt{\sup_{0 \leqslant t < 1} \frac{B_2^2(t)}{t(1-t)}} & \text{if GPD .} \end{cases}$$

$B(t)$ is a Brownian bridge, $B_2(t)$ is a sum of two independent Brownian bridges.

2.d. Practical Procedure

Consecutive steps

1. Check on Pareto-type behavior of the data by $Q-Q-$plots.
2. Select a threshold u or the value of $k = k_{opt,n}$ that minimizes the asymptotic mean square error of the Hill estimator. We choose the **optimal** threshold $u = X_{n-k_{opt,n}}$.
3. (a) Define c_n as the smallest number such that at least $k_{min} = (\log k_{opt,n})^{3/2}$ of the data points $X_1, \cdots, X_{c_n}$ are larger than u.

   (b) Define d_n as the smallest number such that at least k_{min} of the data points $X_{n-d_n+1}, \ldots, X_n$ are larger than u.
4. Repeat the next step for all m from c_n up to $n - d_n$.

   (a) Split the data up in two groups $X_1, X_2, \ldots, X_m$ and $X_{m+1}, \ldots, X_n$.

   (b) Calculate $-2\log\Lambda_m$.
5. Calculate $Z_n = \sqrt{\max_{c_n \leqslant m < n-d_n} (-2\log\Lambda_m)}$ and compare Z_n with the critical values for sample size k.

3. EXAMPLES

3.a. Simulation

We simulate 1000 data sets of size n (with $n = 100$, $n = 500$) from the ***Burr distribution*** $Burr(\beta, \tau, \lambda)$ with parameters as given by

$$P(X > x) = \left(\frac{\beta}{\beta + x^\tau}\right)^\lambda ,$$

an example of a GPD with $\gamma = (\lambda\tau)^{-1}$. The rejection probabilities are given below.

		H_0 **true**			H_0 **false**		
n	m^*	$\gamma = 1$		$\gamma_1 = 1$, $\gamma_2 = 2$	$\gamma_1 = 2$, $\gamma_2 = 1$	$\gamma_1 = 1$, $\gamma_2 = .5$	$\gamma_1 = .5$, $\gamma_2 = 2$
100	20	.096		.191	.460	.486	.182
	50	.075		.517	.512	.519	.559
500	50	.029		.181	.782	.799	.144
	100	.044		.378	.955	.951	.645
	250	.019		.894	.951	.966	.909

3. EXAMPLES

The corresponding median of $\hat{m}$ is given in the table below.

			H_0 **false**		
n	m^*	$\gamma_1 = 1$ $\gamma_2 = 2$	$\gamma_1 = 2$ $\gamma_2 = 1$	$\gamma_1 = 1$ $\gamma_2 = .5$	$\gamma_1 = .5$ $\gamma_2 = 2$
100	20	48	21	45	21
	50	55	44	56	45
500	50	175	47	92	48
	100	139	97	107	97
	250	252	247	252	248

3. EXAMPLES

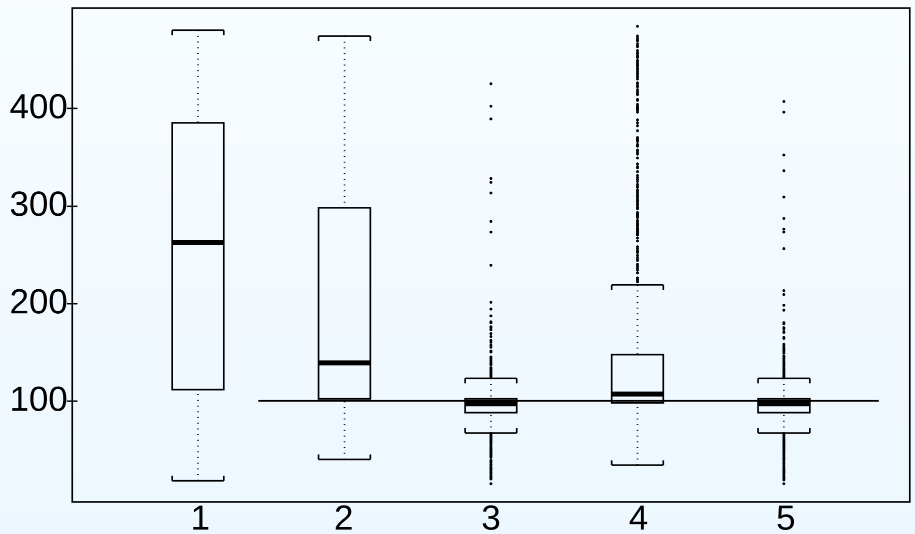
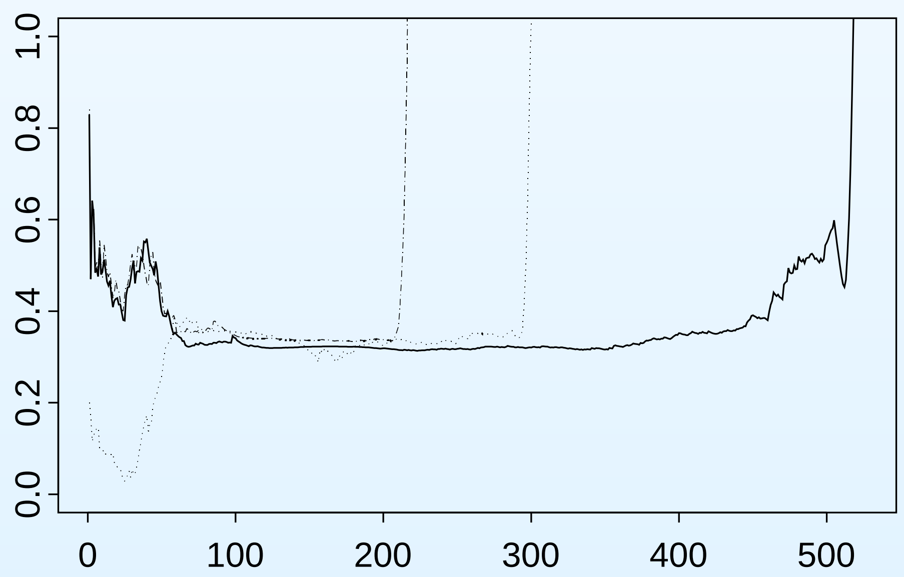


Figure shows Boxplot of $\hat{m}$ for the Burr cases for $n = 500$ and $m^* = 100$.

3. EXAMPLES

3.b. Malaysian Stock Index: Classical approach

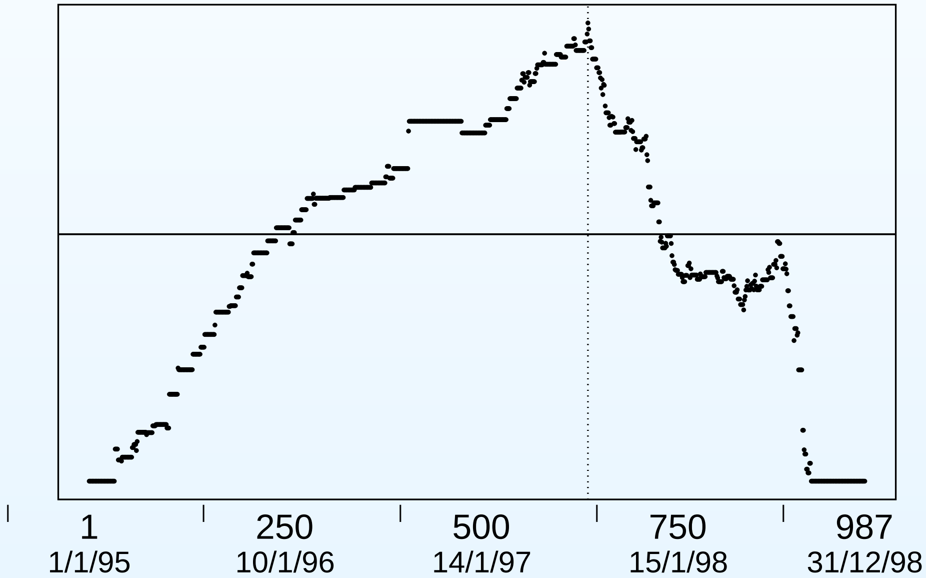
Figure below indicates that the data are Pareto-type distributed. If we accept that July 1997 was a change point, then the data before that date give an extreme value index γ_1 between 0.1 and 0.2 while those after that date give γ_2 around 0.5. The mean squared error of the Hill estimator based on the whole data set attains a local minimum for the threshold u given by $X_{987-224,987} = 0.0099$ so that $k = k_{opt} = 224$.



3. EXAMPLES

1. Pareto-type distribution

First $\sqrt{-2\log\Lambda_m}, 1 \leqslant m \leqslant n-1$ is plotted below.



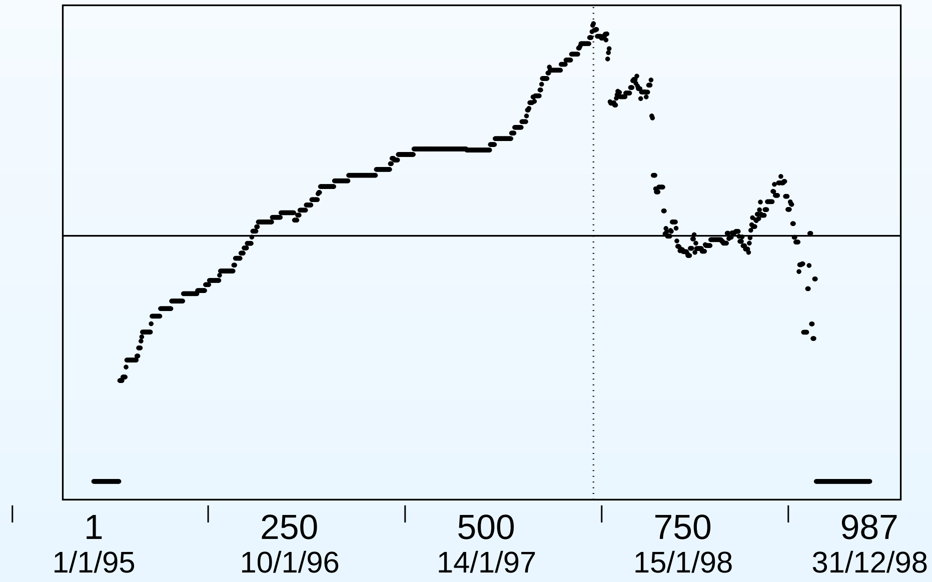
Graph of $(m, \sqrt{-2\log\Lambda_m})$ with critical value indicated with a horizontal line.

We see that $Z_n = \sqrt{\max(-2\log\Lambda_m)} = 5.8$ falls above the critical value 3.14 and we reject H_0. The maximum is attained at $m = 635$, which corresponds to 1/08/1997, shortly after the beginning of the Asian crisis.

3. EXAMPLES

2. GPD

Now $\sqrt{-2\log\Lambda_m}, 1 \leqslant m \leqslant n-1$ is plotted below.

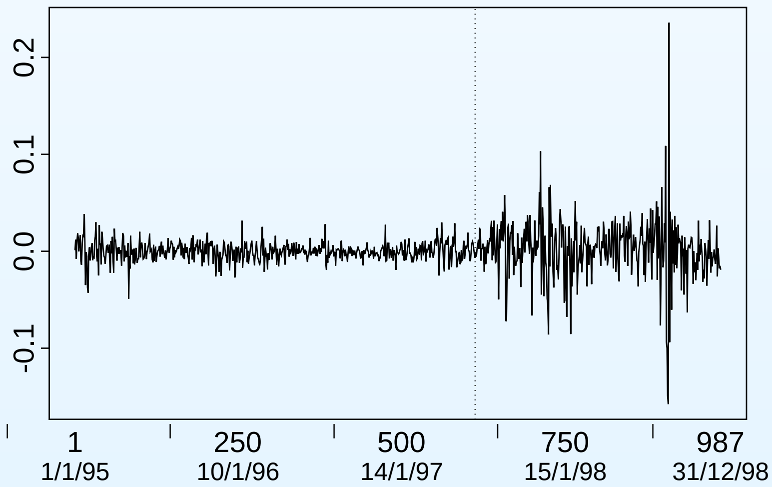


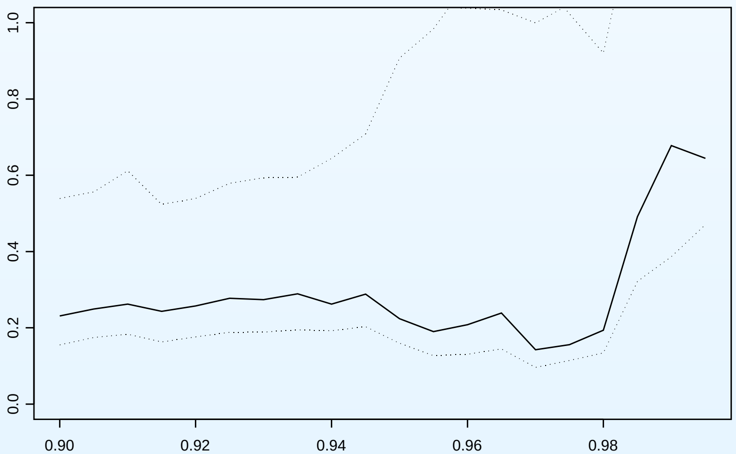
Since $Z_n = \sqrt{max(-2\log\Lambda_m)} = 5.93$ is above the critical value 3.18 we again reject H_0. Also the instant of change $\hat{m} = 636$ is again very close to the value before.

3. EXAMPLES

3.b. Malaysian Stock Index: Improved approach

In the above analysis, we assumed that the data were independent. But market data are hardly ever independent. However, it is known that the Hill estimator withstands many forms of dependence. Alternatively, one can proceed as follows. The time series and an estimate of the extremal index are given below.

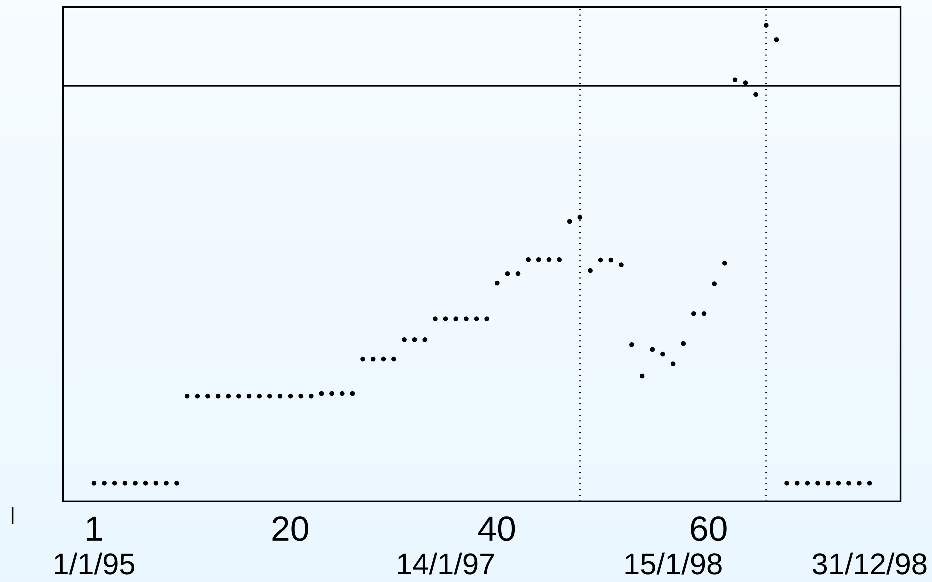




A **declustering scheme** cuts the data into clusters that can safely be taken as independent. Apply the previous procedure to the 76 cluster maxima.

3. EXAMPLES

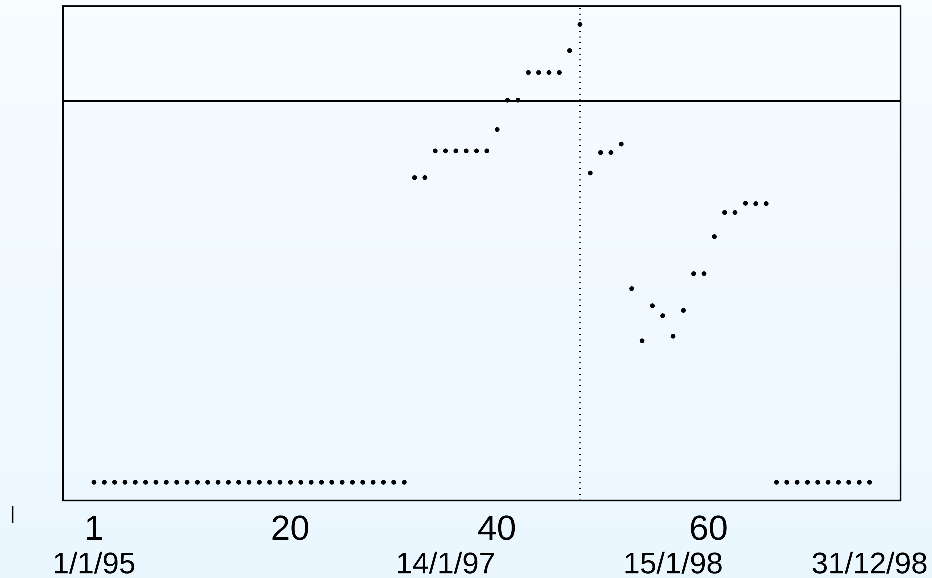
1. Pareto-type distribution



There is a local maximum for cluster maximum 48 which corresponds to $m = 631$ on (28/07/1997). However this local maximum is not larger than the critical. The actual maximum Z_n is attained for cluster maximum 66 which corresponds to $m = 854$ (22/6/98). We cannot reject the hypothesis.

3. EXAMPLES

2. GPD



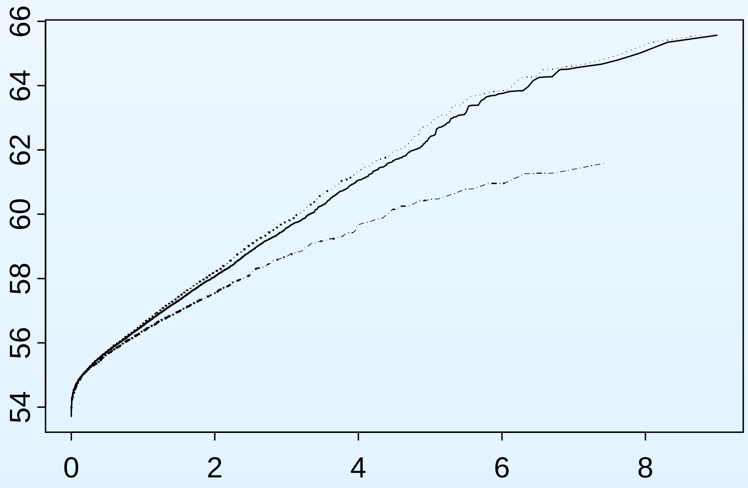
Now $\sqrt{-2\log\Lambda_m}, 1 \leqslant m \leqslant n-1$ is plotted in the figure. The maximum Z_n is attained for cluster maximum 48 which corresponds to $m = 631(28/07/1997)$. The critical value 2.95 for the test is indicated with a horizontal line. On the basis of this test, we reject the hypothesis of no change.

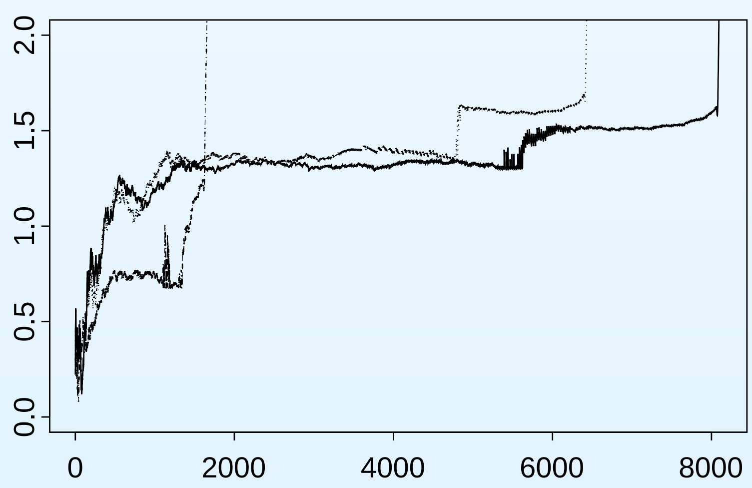
3.c. Earthquake Data

Data give seismic moment measurements of shallow earthquakes over the period from 1977 to 2000 for subduction zones and midocean ridge zones.

- First 6458 points for subduction zones
- further 1664 points for the midocean ridge zone

Figure shows Pareto-type distribution. Extreme value indices are 1.4 and 0.9. Optimal k is 1333 and the corresponding threshold $u = X_{8123-1333,8123} = 1.28\ 10^{25}$.





3. EXAMPLES

1. Pareto-type distribution

We test for a change in γ. $\sqrt{-2\log\Lambda_m}, 1 \leqslant m \leqslant n-1$ is plotted in the figure with the critical value 3.27 added on. As the maximum of $\sqrt{-2\log\Lambda_m}$ is larger, we reject H_0. The time of change can be estimated by $\hat{m} = 6957$.

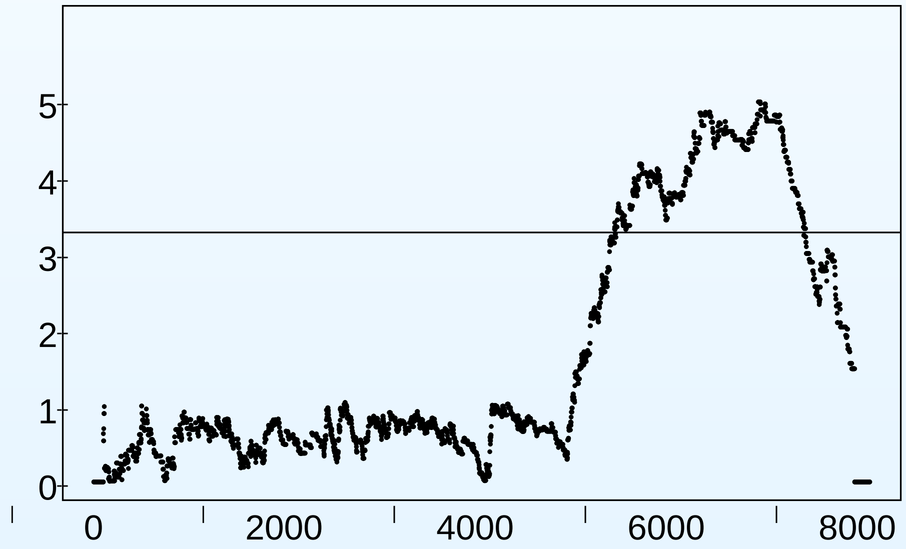
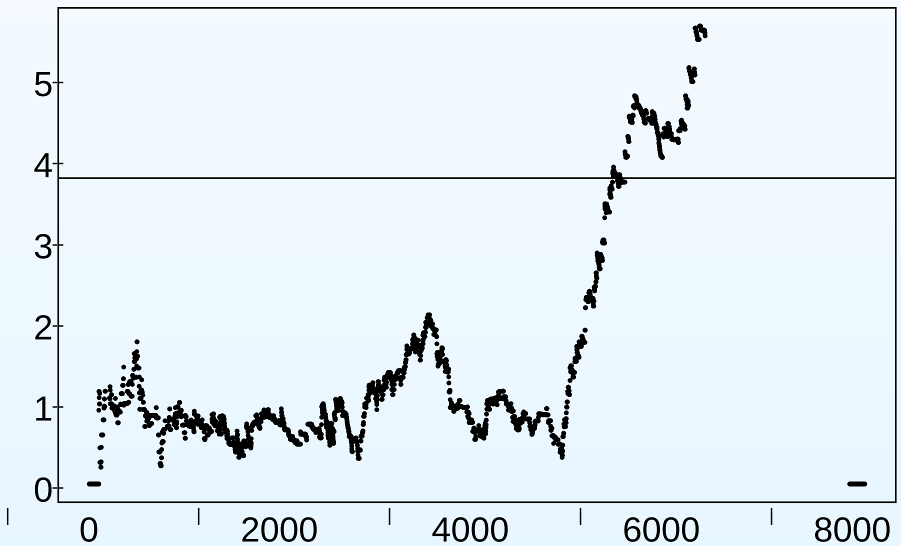


Figure of $(m, \sqrt{-2\log\Lambda_m})$ with critical value.

3. EXAMPLES

2. GPD

We test for a change point in $\theta = (\gamma, \sigma)$. $\sqrt{-2\log\Lambda_m}, 1 \leqslant m \leqslant n-1$ is plotted in the figure.



Picture gives (m,$\sqrt{-2\log\Lambda_m}$) with the critical value 5.64 indicated with a horizontal line. Since the maximum of $\sqrt{-2\log\Lambda_m}$ is larger than the critical value we reject H_0. The estimated time point of the change is $\hat{m} = 6391$.

3.d. Nile Data

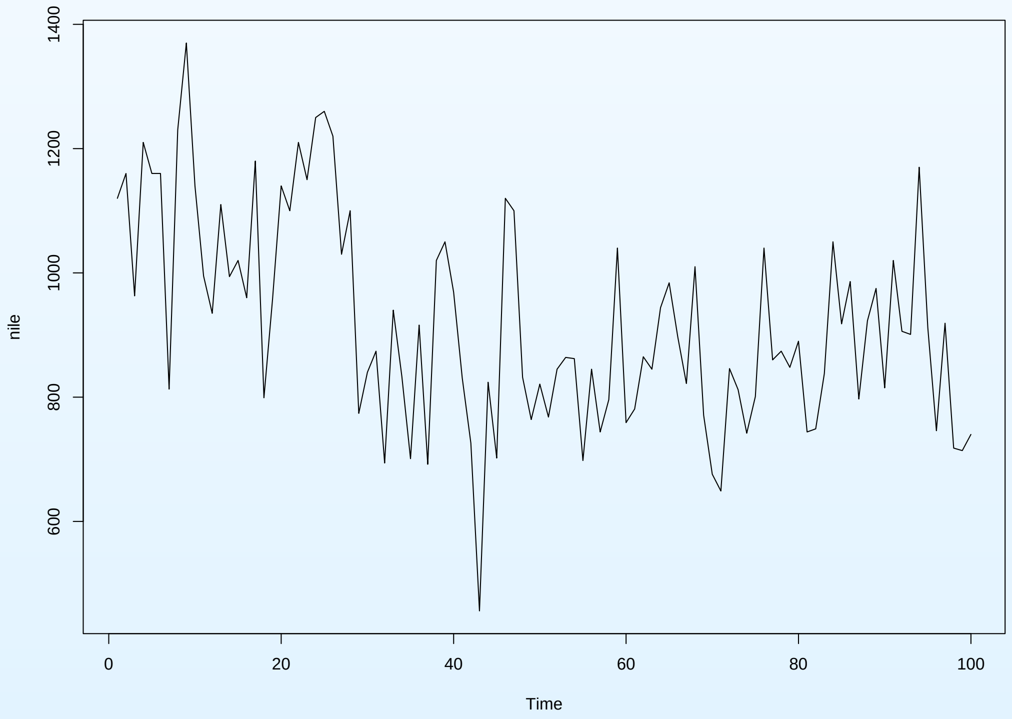
Annual flow volume of the Nile River at Aswan from 1871 to 1970.

1120	1160	963	1210	1160	1160	**813**	1230	1370	1140
995	935	1110	994	1020	960	1180	799	958	1140
1100	1210	1150	1250	1260	1220	1030	1100	**774**	840
874	694	940	833	701	916	692	1020	1050	969
831	726	**456**	824	702	1120	1100	832	764	821
768	845	864	862	698	845	744	796	1040	759
781	865	845	944	984	897	822	1010	771	676
649	846	812	742	801	1040	860	874	848	890
744	749	838	1050	918	986	797	923	975	815
1020	906	901	1170	912	746	919	718	714	740

3. EXAMPLES

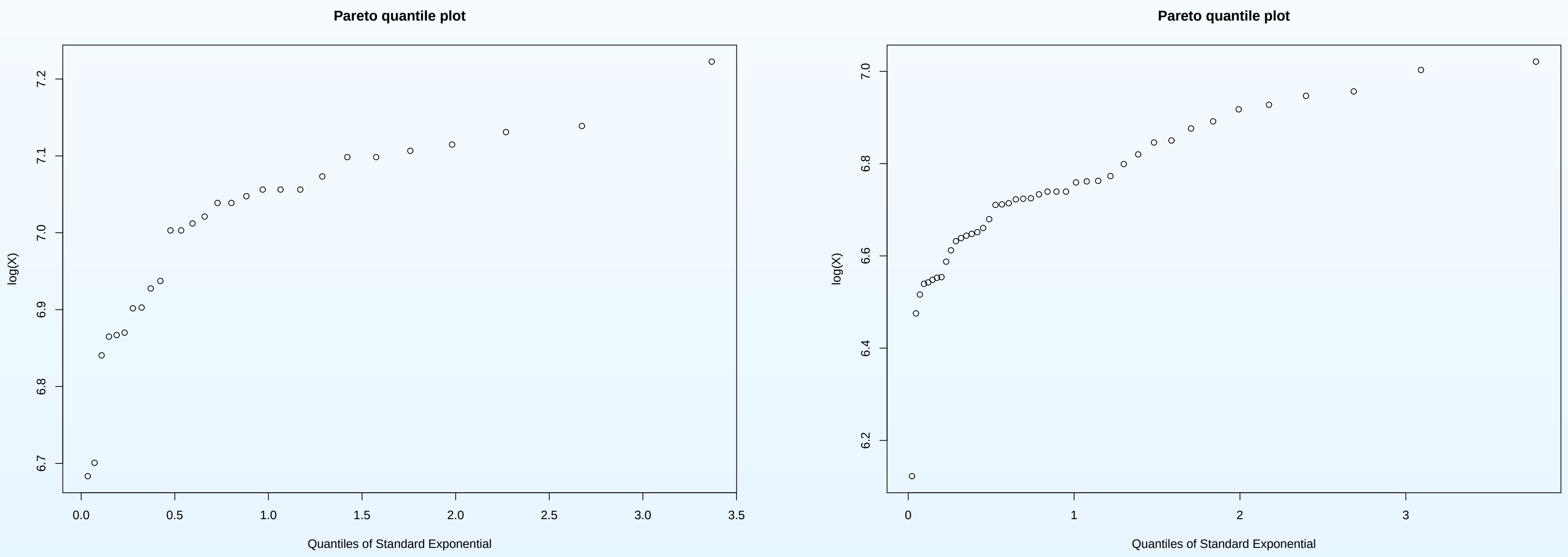
Prior studies indicate

- 1877 (measurement 813) candidate for additive outlier,
- 1913 (measurements 456) candidate for additive outlier,
- 1899 (measurement 774) indicates start of construction of Aswan dam.



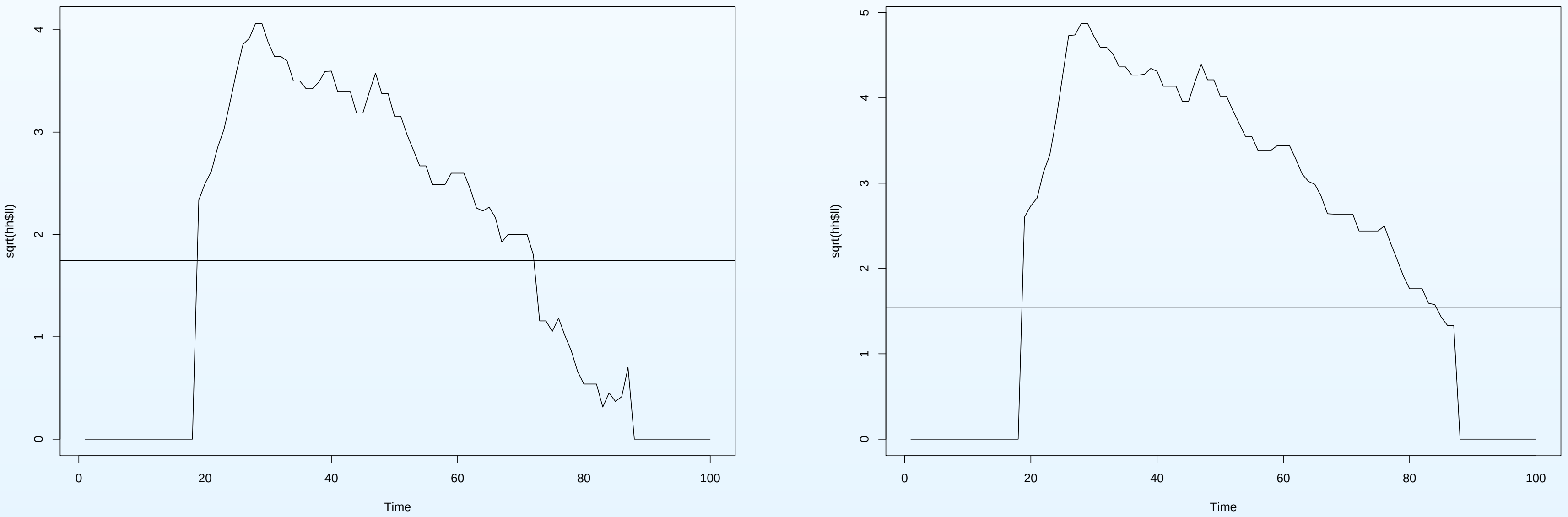
3. EXAMPLES

Group 1: first 28 points – Group 2: remaining 71 points with Pareto QQ plots for both groups. Optimal values $k = 17$, resp. $k = 13$ lead to the estimators 0.07 and 0.13.



3. EXAMPLES

The change point detection based on the Pareto and the GPD model are given in the figure, both leading to a significant change point at $\hat{m} = 28$ at the beginning of construction of the Aswan dam.



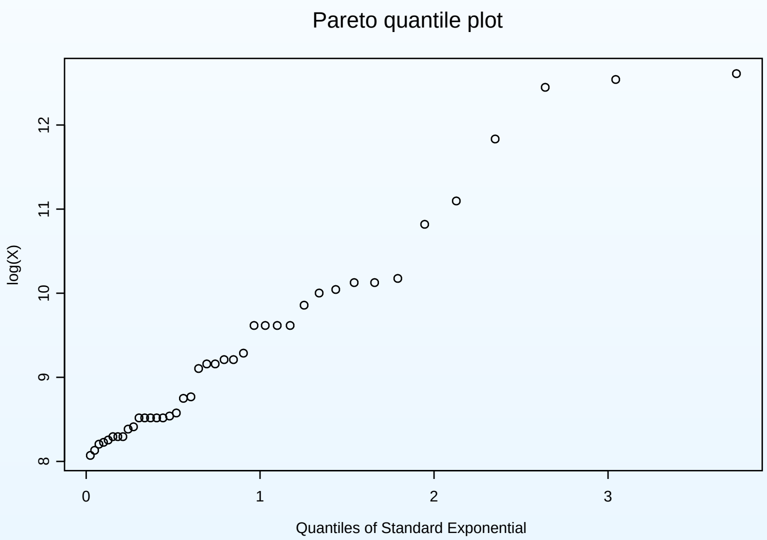
3. EXAMPLES

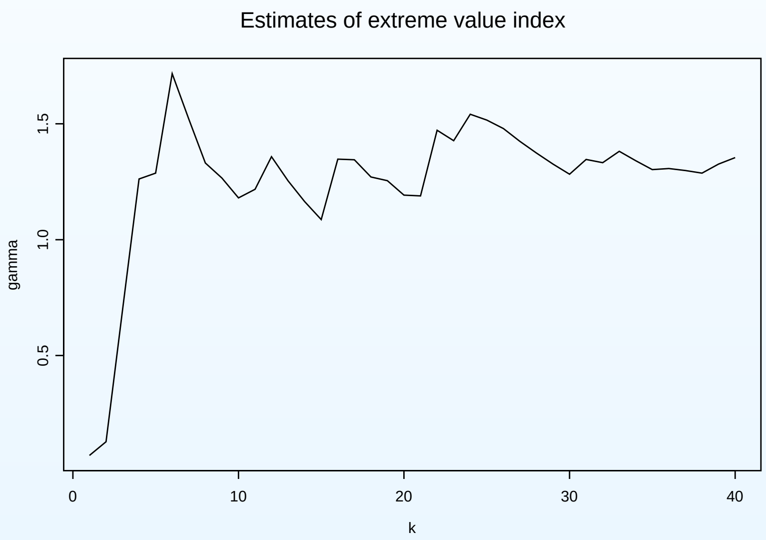
3.e. Swiss-Re Catastrophic Data

PLACE	DATE	VICTIMS	PLACE	DATE	VICTIMS
Bangladesh	0,87	300000	Indonesia	34,98	280000
China	6,57	255000	Bangladesh	21,33	138000
Peru	0,51	66000	Gilan (Iran)	20,47	50000
Bam (Iran)	33,98	26271	Tabas (Iran)	8,71	25000
Armenia	18,93	25000	Colombia	15,87	23000
Guatemala	6,10	22084	Izmit (Turkey)	29,63	19118
Gujarat (India)	31,07	15000	India	8,67	15000
India	29,83	15000	India	9,61	15000
India	1,83	10800	Venezuela	29,95	10000
Bangladesh	7,89	10000	Mexico	15,72	9500
India	23,75	9500	Honduras	28,81	9000
Kobe (Japan)	25,05	6425	Philippines	21,85	6304
Pakistan	4,99	5300	Brazil	31,87	5112

3. EXAMPLES

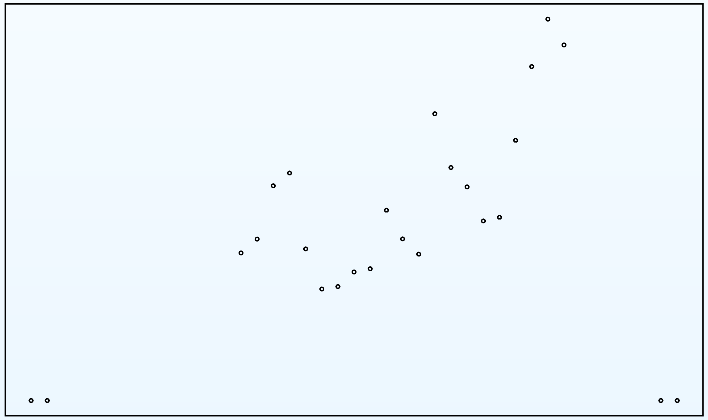
The Pareto QQ plots with the corresponding Hill estimators. The mean squared error is minimal at $k = 39$ for Pareto and $k = 22$ for GPD, both leading to $\hat{\gamma} = 1.3$.

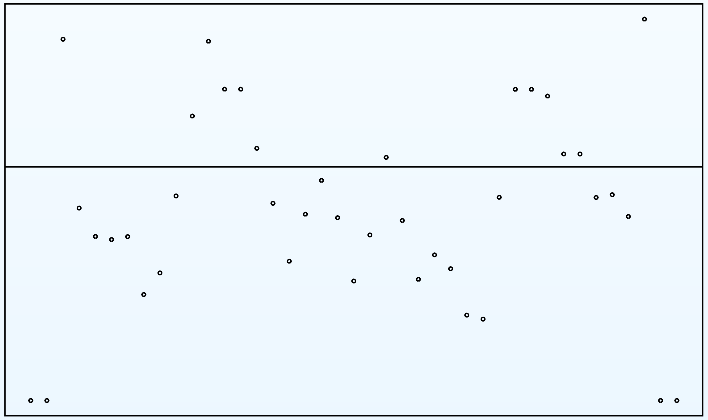




3. EXAMPLES

The likelihood expression $\sqrt{-2 \log \Lambda_m}$ based on the Pareto model and the GPD model as a function of m where m is indicating where the group is split up in two.





Pictures for Pareto model with critical value 1.6 and GPD model with critical value 1.4.

4. CONCLUSIONS

- What has been shown are just first attempts
- Rounded figures make accurate conclusions harder
- There is a need for sufficiently large data sets
- Need for studies under specific dependence structures
- Multivariate extensions should be possible
- Where to go from here?
- Good examples from environmetrics are very, very welcome

5. REFERENCES

- Beirlant, J., Goegebeur Y., Segers, J. and Teugels, J.L. (2004). Statistics of Extremes, Theory and Applications, *Wiley, Chichester*.
- Csörgő, M., Horváth, L. (1997) . Limit Theorems in Change Point Analysis. *Wiley, Chichester*.