Change Point Analysis of Extreme Values

Goedele Dierckx

Economische Hogeschool Sint Aloysius, Brussels, Belgium

Jef L. Teugels

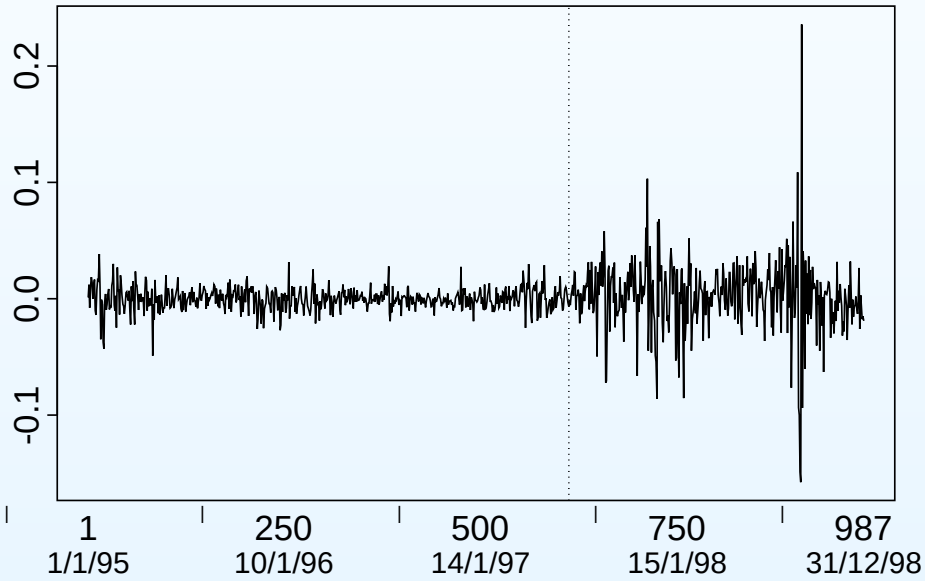
Katholieke Universiteit Leuven, Belgium

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
   (c) Malaysian Stock Index. Improved Approach
   (d) Nile Data
   (e) Swiss-Re Catastrophic Data
   (f) Further Examples
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

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

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.b. Extreme Value Situation

Put the data in increasing order: $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$. Hence $X_{n,n}$ is the maximum in a sample of independent random variables with a common distribution. The maximum domain of attraction condition is expressed as

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

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**.

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

- 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. 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.

We concentrate on two specific examples that regularly appear 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

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

- 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, given that X exceeds u, satisfies the condition

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

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

$$H_0 \quad : \quad \gamma_1 = \gamma_2 = \ldots = \gamma_n = \gamma \text{ versus}$$

$$H_1 \quad : \quad \gamma_1 = \gamma_2 = \ldots = \gamma_{m^*} \neq \gamma_{m^*+1} = \ldots = \gamma_{n-1} = \gamma_n \text{ for some } m^* .$$

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

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

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

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 proved the following result.

Theorem Suppose $X_1, \ldots, X_m, X_{m+1}, \ldots X_n$ are independent and identically distributed. 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.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^*	γ = 1		γ_1 = 1 γ_2 = 2	γ_1 = 2 γ_2 = 1	γ_1 = 1 γ_2= .5	γ_1 = .5 γ_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^*	γ_1 = 1 γ_2 = 2	γ_1 = 2 γ_2 = 1	γ_1 = 1 γ_2= .5	γ_1 = .5 γ_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

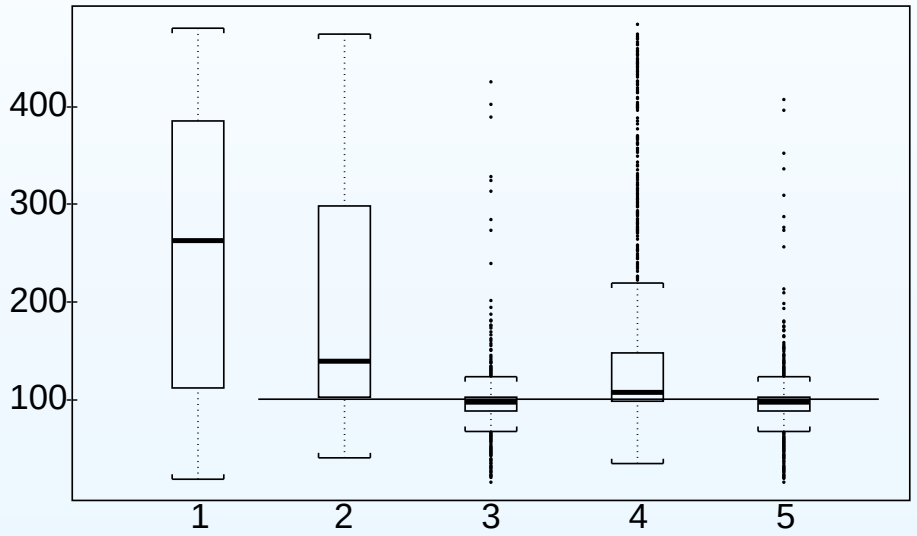
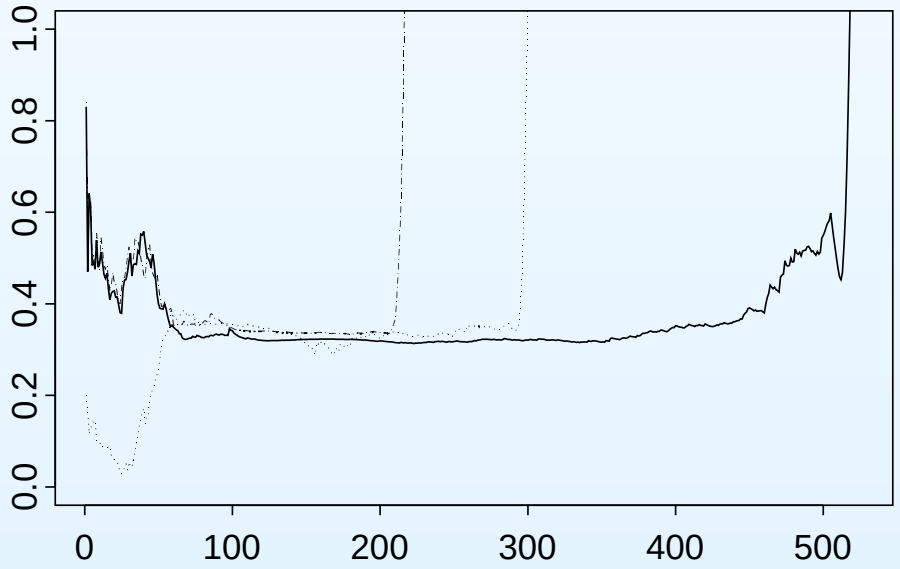


Figure shows Boxplot of $\hat{m}$ for the Burr cases for $n = 500$ and $m^* = 100$.

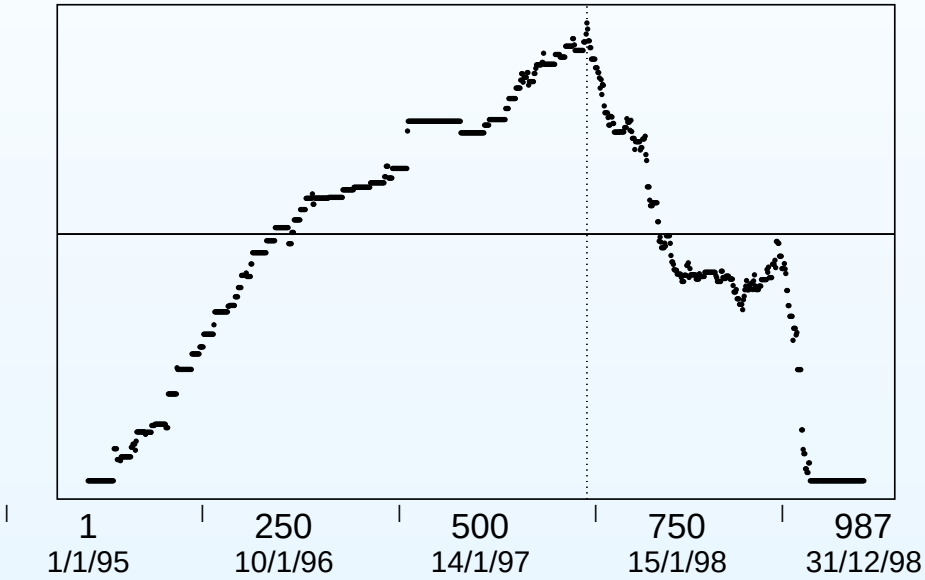
3.b. Malaysian Stock Index: Classical approach

The figure below indicates that the data are Pareto-type distributed. The mean squared error of the Hill estimator based on the whole data set attains a local minimum at the value $k = k_{opt} = 224$ which results in threshold $u = X_{987-224,987} = 0.0099$.



1. Pareto-type distribution

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

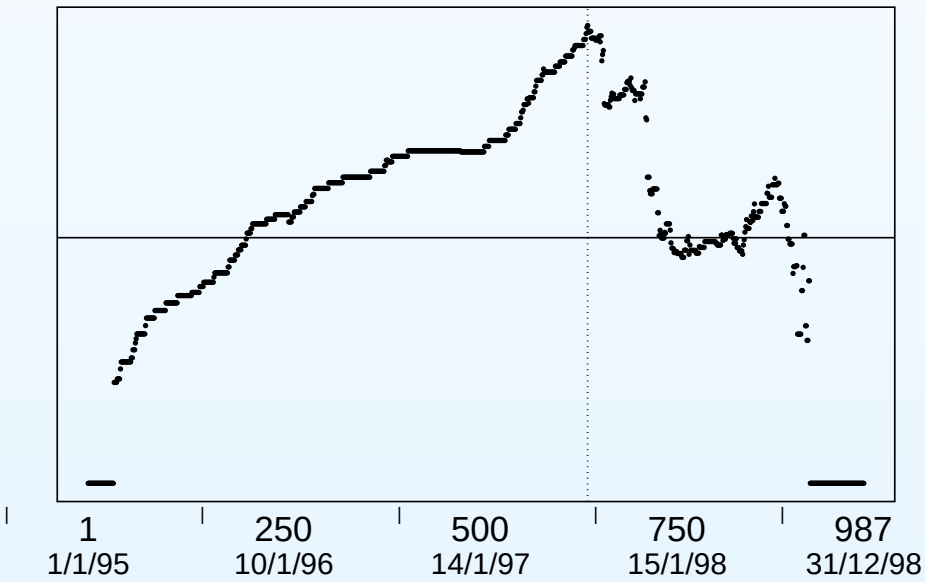


Graph of $(m,\sqrt{-2\log\Lambda_m})$ with critical value indicated with a horizontal line.

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.

2. GPD

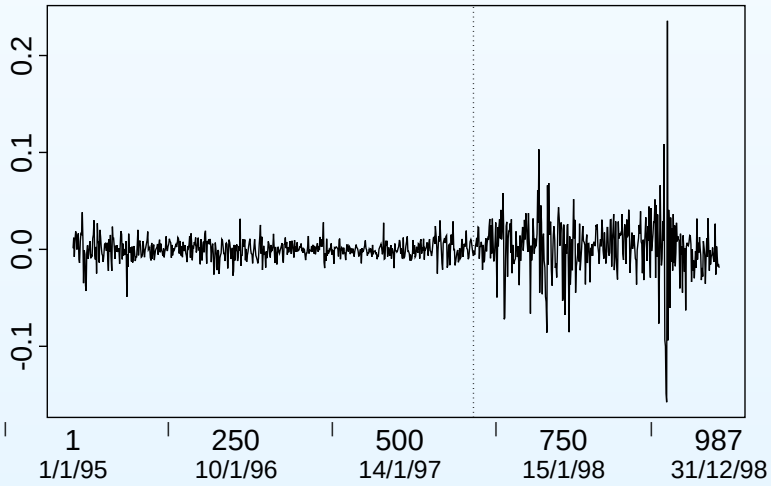
Now $\sqrt{-2\log\Lambda_m}, 1 \leqslant m \leqslant n-1$ is plotted below.

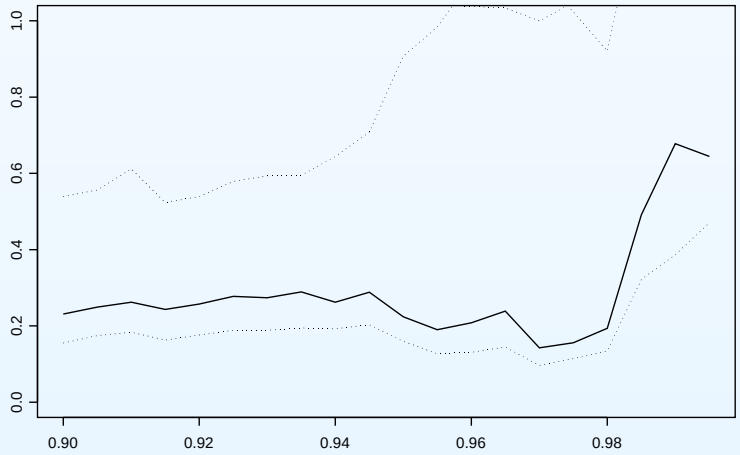


Since $Z_n = \sqrt{\max(-2\log\Lambda_m)} = 5.93$ is above the critical value 3.18 we again reject H_0. The instant of change is for $\hat{m} = 636$ and yields 2/08/1997, almost as before.

3.c. 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. Below we repeat the time series and an estimate of the extremal index that should be close to 1 if the data are independent.

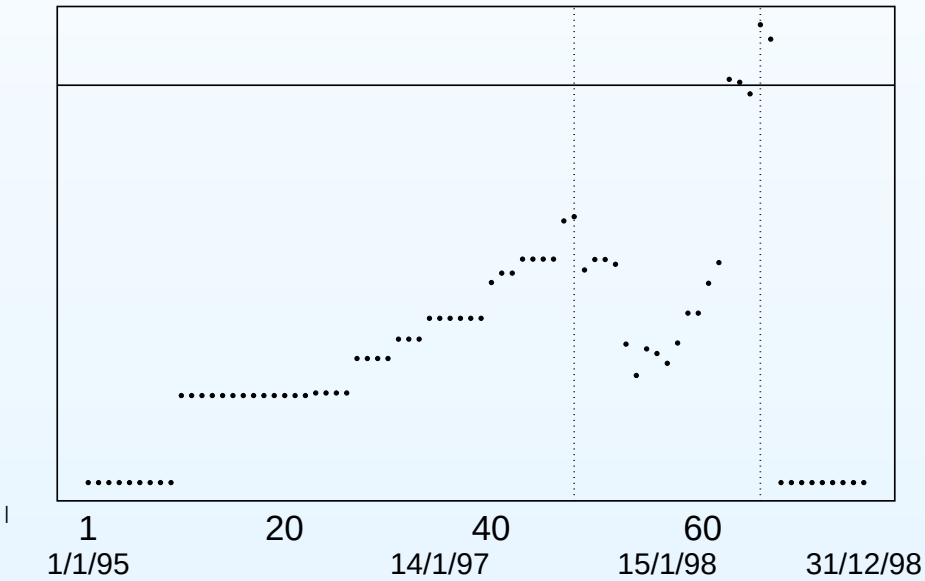




A **declusturing scheme** cuts the data into clusters that can safely be taken as independent.

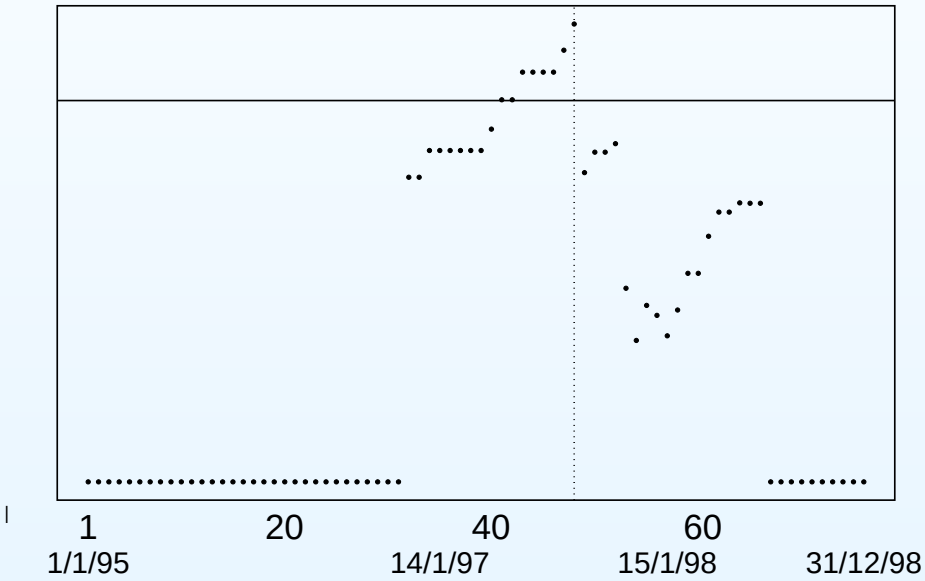
Apply the previous procedure to the 76 cluster maxima.

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 $m = 854$ on 22/6/98). We cannot reject the hypothesis.

2. GPD



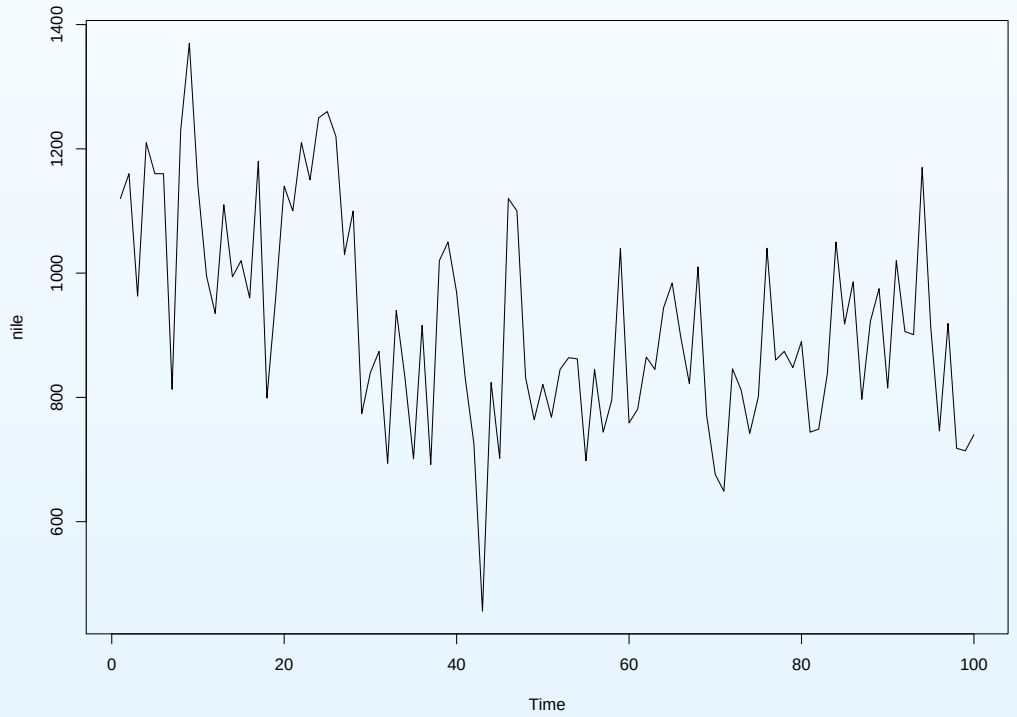
Now $\sqrt{-2\log\Lambda_m}, 1 \leqslant m \leqslant n-1$ is plotted in the figure. The maximum Z_n is attained for cluster maximum 48 which corresponds to $m = 631$ on 28/07/1997. The critical value 2.95 for the test is indicated with a horizontal line. On the basis of this test, we reject the hypothesis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

3.d. Nile Data

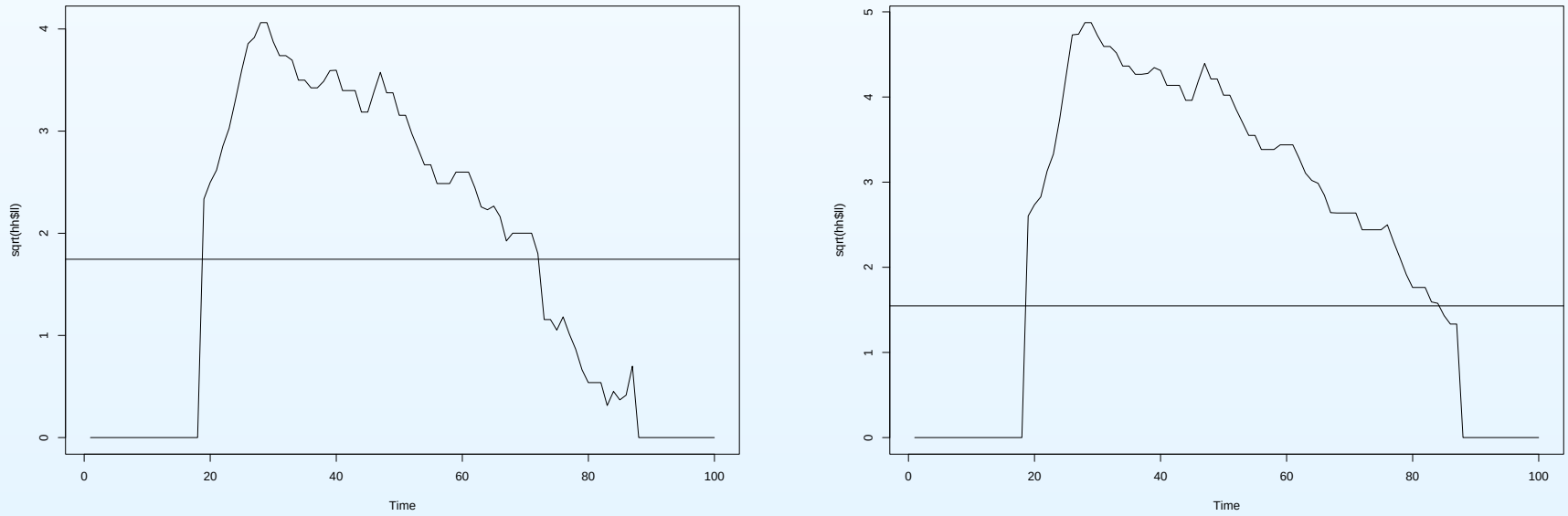
Annual flow volume of the Nile River at Aswan from 1871 to 1970.

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

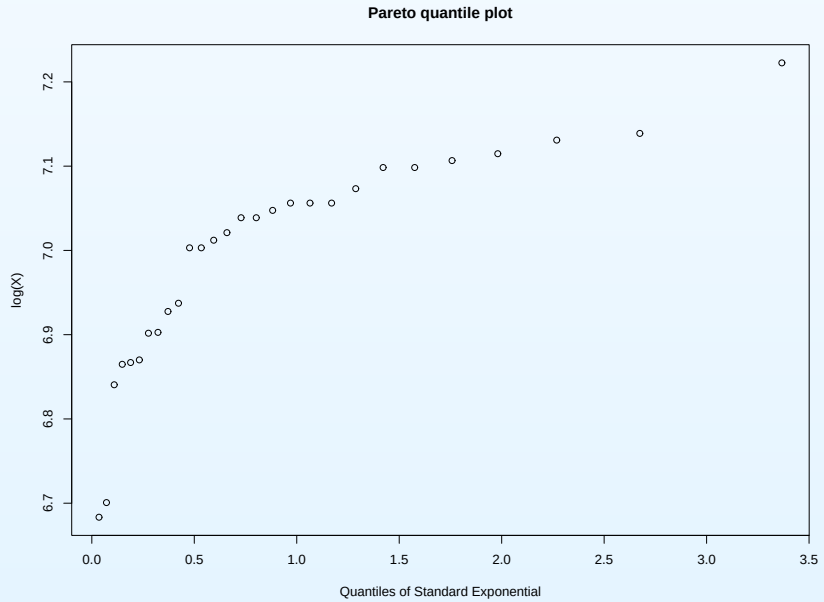
The graph of the series is given below.

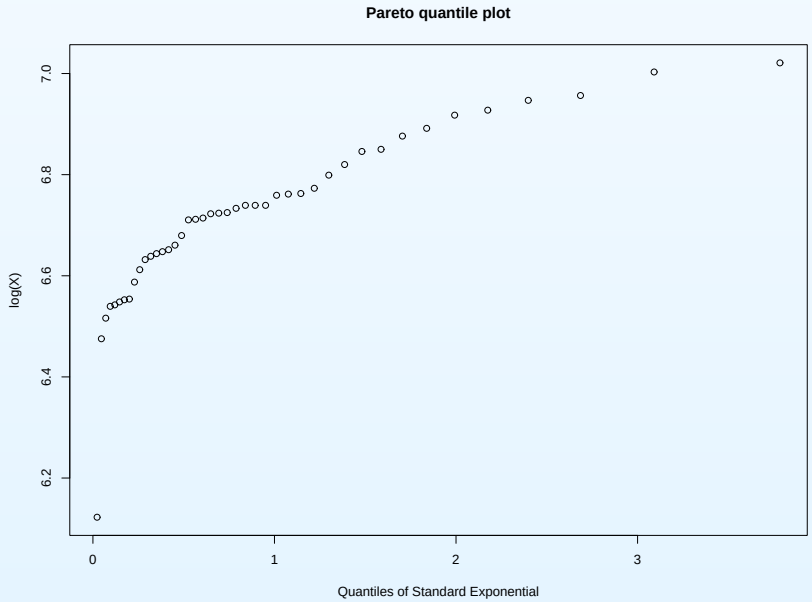


The change point detection based on the Pareto and the GPD model are given in the figure, both leading to a significant change point at $\hat{m} = 28$ at the beginning of construction of the Aswan dam.



To illustrate the extremal behaviour of both segments, take as group 1 the first 28 points and as group 2 the remaining 71 points. The two Pareto QQ plots are given below. They are based on the optimal values $k = 17$, resp. $k = 13$ leading to the estimates 0.07 and 0.13.





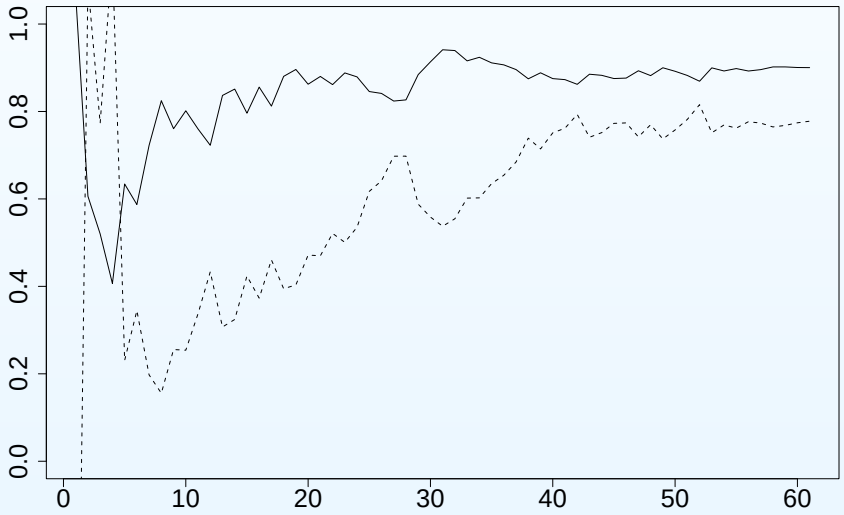
3.e. Catastrophic losses (2010)

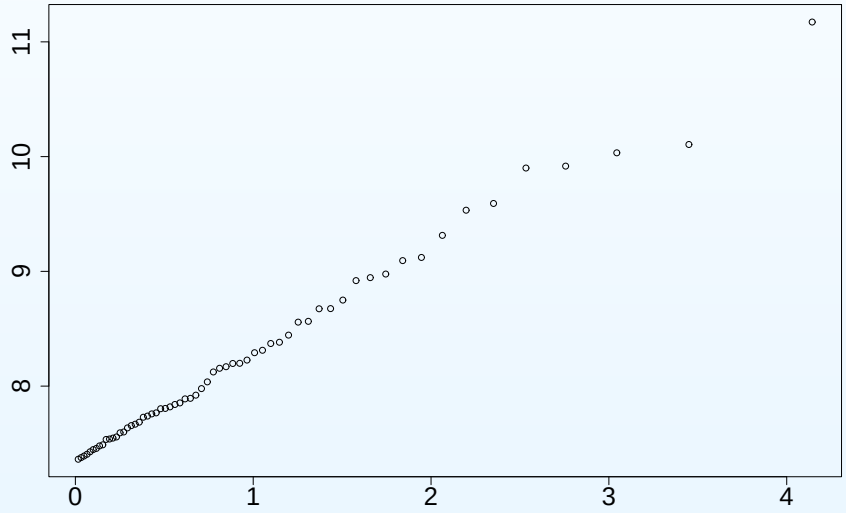
EVENT	T	DATE	LOSS	EVENT	T	DATE	LOSS
Katrina	H	24.08.05	71.163	Frances	H	26.08.04	5.848
Andrew	H	23.08.92	24.479	Vivian	ES	25.02.90	5.242
WTC-attack	M	11.09.01	22.767	Bart	T	22.09.99	5.206
Northridge	E	17.01.94	20.276	Georges	H	20.09.98	4.649
Ike	H	06.09.08	19.940	Allison	S	05.06.01	4.369
Ivan	H	02.09.04	14.642	Jeanne	H	13.09.04	4.321
Wilma	H	16.10.05	13.807	Songda	T	06.09.04	4.074
Rita	H	20.09.05	11.089	Gustav	H	26.08.08	3.988
Charley	H	11.08.04	9.148	X_{106}	US	02.05.03	3.740
Mireille	T	27.09.91	8.899	Floyd	H	10.09.99	3.637
Hugo	H	15.09.89	7.916	Piper Alpha	M	06.07.88	3.631
Daria	ES	25.01.90	7.672	Opal	H	01.10.95	3.530
Lothar	ES	25.12.99	7.475	Kobe, Japan	E	17.01.95	3.482
Kyrill	ES	18.01.07	6.309	Klaus	ES	24.01.09	3.372
X_{100}	ES	15.10.87	5.857	Martin	S	27.12.99	3.093

X_{101}	ES	06.08.02	2.755	X_{107}	US	27.04.02	1.999
X_{114}	US	20.10.91	2.680	Gilbert	H	10.09.88	1.984
X_{102}	US	06.04.01	2.667	X_{108}	US	03.05.99	1.914
X_{103}	ES	25.06.07	2.575	X_{109}	US	17.12.83	1.885
Isabel	H	18.09.03	2.540	X_{110}	US	04.04.03	1.880
Fran	H	05.09.96	2.488	X_{1}	US	02.04.74	1.873
Anatol	ES	03.13.99	2.454	Mississippi	F	25.04.73	1.787
Iniki	H	11.09.92	2.448	X_{111}	US	15.05.98	1.770
Frederic	H	12.09.79	2.361	Loma Pieta	E	17.10.89	1.732
X_{104}	ES	19.08.05	2.340	Celia	H	04.08.70	1.714
Petro US	M	23.10.89	2.296	Vicki	T	19.09.98	1.682
Tsunami	E	26.12.04	2.273	Fertilizer	M	21.09.01	1.646
Fifi	S	18.09.74	2.177	X_{112}	US	05.01.98	1.621
X_{105}	ES	04.07.97	2.139	X_{113}	US	05.05.95	1.599
Luis	H	03.09.95	2.113	Grace	H	20.10.91	1.576
Erwin	ES	08.01.05	2.071				

3.e.1. Pareto-type?

We plot the pairs $(x_i, Y_i), i = 1, \ldots, 62$. The variable Y consists of the losses of the catastrophes in million US-dollars, corrected for inflation. To avoid empty places and erratic behavior, we re-scale the time axis so that one unit represents 5 years. The extreme value index γ is clearly positive and its estimation is done with the Hill estimator and the Peak Over Threshold estimator. The value $k = 52$ minimizes the empirical mean square error. The corresponding estimate for $\hat{\gamma} = 0.87$, hence the underlying distribution has finite mean but infinite variance.

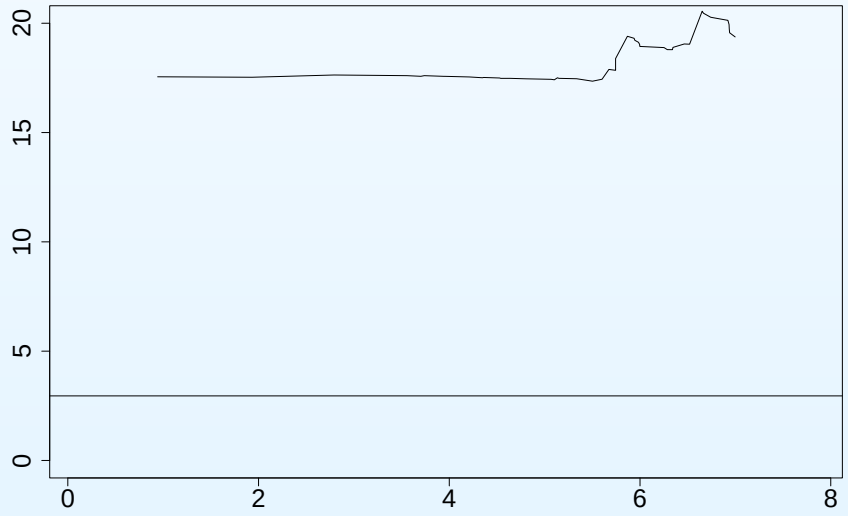




The Pareto QQ plot of the data is very close to linearity, suggesting that almost all the losses can be used in the estimation.

3.e.2. Change point?

In the figure below, the test statistic is well over the critical value 2.96 for all points, suggesting that there is not one single change point, but that the extreme value index keeps changing over time. The maximum hints at a value of 1982.



3.e.3. Trend Analysis

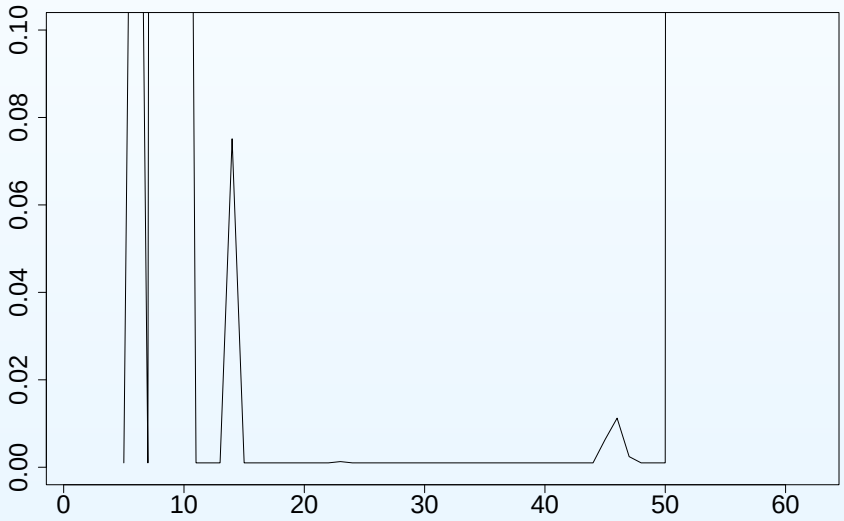
What about a trend model? We assume now that the relative excesses over some threshold $u(x)$ follow a Pareto-type distribution, this means

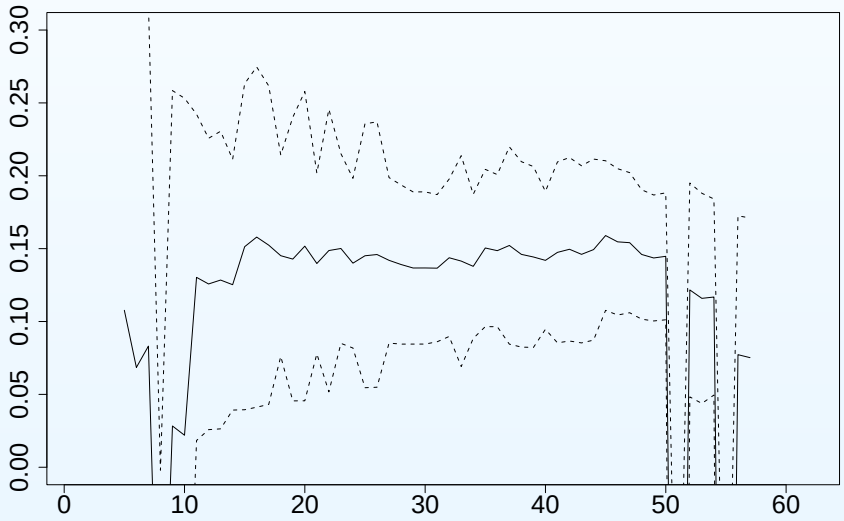
$$P(Y/u(x) > y|Y > u(x)) \sim y^{-1/\gamma(x)}$$

when $u(x) \to \infty$ where we further assume that the extreme value index follows a linear trend

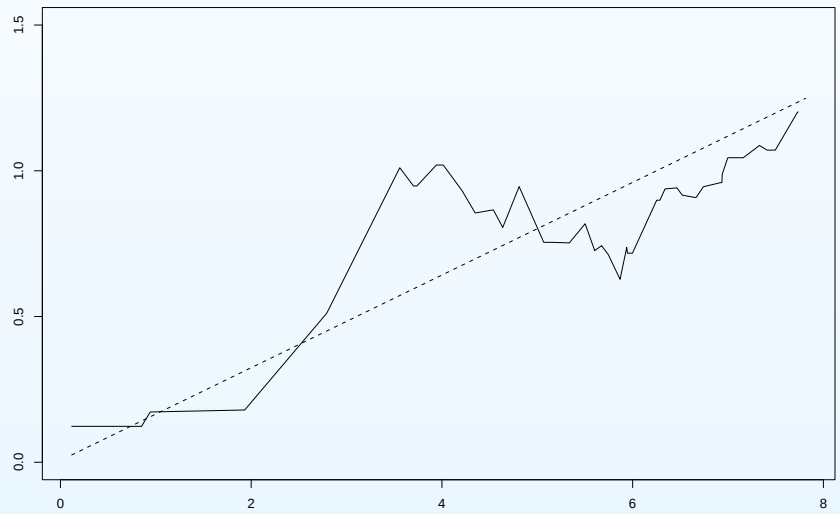
$$\gamma(x) = \alpha_1 + \alpha_2 x$$

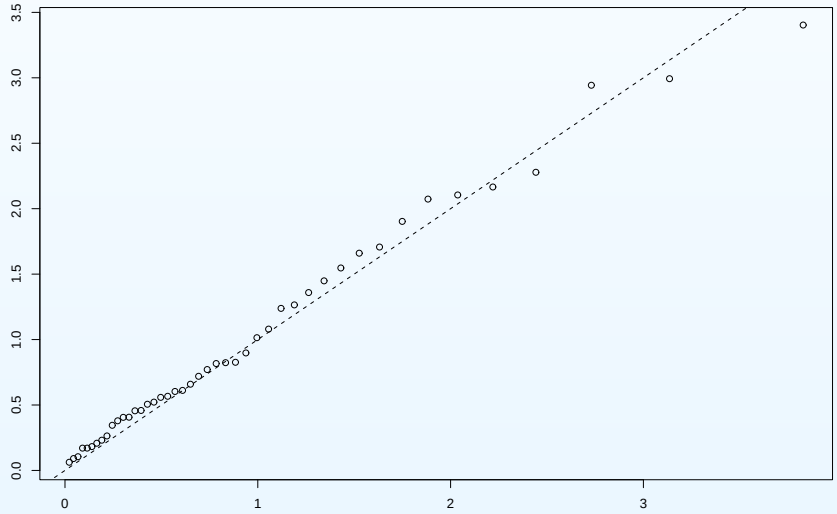
with $\alpha_i, i = 1, 2$ constants. The latter two parameters are estimated by maximum likelihood. This is illustrated in the figures below and is quite insensitive to the choice of k. For $k = 45$ $\hat{\alpha}_1 = 0.001$, whereas $\hat{\alpha}_2 = 0.159$.





On the left the estimation of α_1 is given, on the right that of α_2 with corresponding 95%-confidence bounds. Hence γ can be estimated as $\hat{\gamma}(x) = 0.001 + 0.159x$, where x denotes the number of 5-years since January 1, 1970. The estimated value for α_2 is significantly different from 0, as follows from the large sample behavior of maximum likelihood estimators.





The left figure graphs the fit of the data with the estimation of $\gamma(x)$. There is an alternative verification. When $Y|x$ follows a Pareto type distribution with extreme value index $\gamma(x)$, then the random variable $\log Y/\gamma(x)$ follows approximately a standard exponential distribution, at least for the largest data points. On the right, we drafted the exponential QQ-plot with the first bisector corresponding to the expected standard exponential.

3.f. Further Examples

Other examples from

1. Daily rain data from Quebec City. Period 1915 to 2007, giving 34596 data points.
   Change in 1943 caused by a move from city center to the airport.
   Same with daily minimal temperature data.
2. Pole vault data at Olympic Games.

4. CONCLUSIONS

- What has been shown are just first attempts
- Trend analysis seems to apply more often
- There is a need for sufficiently large data sets
- Need for studies under specific dependence structures
- Multivariate extensions should be possible

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*.
- Dierckx, G. and Teugels, J.L. (2010) Change point analysis of extreme values *Environmetrics*, 21, 661-686, 2010.
- Dierckx, G. and Teugels, J.L. (2011) Trend analysis of extreme values *Probability Approximations and Beyond: A Conference in honour of Louis Chen on his 70th Birthday*, A.D. Barbour, H.P. Chan & D. Siegmund, (eds.), *Lecture Notes in Statistics*, 205, 101–108, Springer-Verlag, Berlin, 2012.