Trends and change points in the tail behaviour of a heavy tailed distribution

Dierckx, Goedele
Hogeschool-Universiteit Brussel, Department of Mathematics and Statistics
Stormstraat 2
BE-1000 Brussel, Belgium
E-mail: Goedele.Dierckx@hubrussel.be

1 Introduction

We examine the extreme values of a phenomenon Y with a heavy tailed distribution function F through the sample $Y_1, \ldots, Y_n$. For a heavy tailed distribution, also called a Pareto type distribution, $1 - F(y) = y^{-1/\gamma}\ell(y)$, with $\ell(y)$ a slowly varying function. The parameter γ is called the extreme value index and is positive for heavy tailed distributions, where the larger γ, the heavier the tail. An extensive treatment of Pareto type distributions can e.g. be found in [1]. The Hall class, with $P(Y > y) \sim Cz^{-1/\gamma}(1 + Dz^{\rho})$, with C, D constants and $\rho < 0$ constitutes a wide subclass of the Pareto-type distributions.

In this text we study a variable $Y|x$ that can vary with some covariate information x (e.g. x=time) in which the main interest lies in the extreme data points. The extreme value index might depend on the covariate information and is denoted as $\gamma(x)$ as such. In Dierckx et al. (2010) we concentrated on testing whether an instantaneous change occurs in the value of the extreme value index γ. Now we illustrate with an explicit example that in some cases the extreme value index seems to change gradually rather than instantaneously according to some trend. More generally there might be a change point in the trend parameters as well.

Example Swiss Re. Worldwide major catastrophes often have a grave humanitarian impact with regard to losses and victims. Therefor, Swiss Re, one of the leading global reinsurers, listed the biggest disasters of different types (hurricanes, ...). We analyse the losses Y (adapted for inflation) of these disasters from January 1, 1970 until January 1, 2009 (n=62). These losses are plotted in time in Figure 1. This plot seems to indicate that the losses become more heavy tailed in time, suggesting a time varying extreme value index.

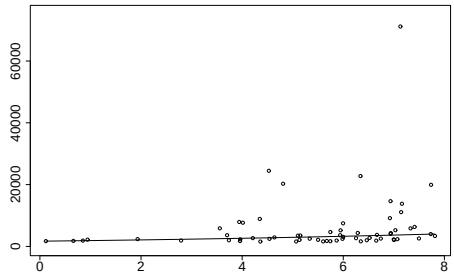

Figure 1: Swiss Re Loss data: (x, Y) where x denotes the number of 5 years since 1970

In Section 2 we will focus on studying a trend in the extreme behaviour. Some comments on change points of the trend parameters are made in Section 3..

2 Trend behaviour

2.1 Trend models

The problem of non-stationary extremes has already been addressed in the literature, especially with applications in climatology. Note however that in that context the authors often suggest to use a constant extreme value index γ, whereas other parameters account for the non-stationarity.

Several authors focus on fitting a Generalized Extreme Value (GEV) distribution

$$F(x) = exp\left(-\left(1+\gamma\frac{x-\mu}{\sigma}\right)^{-1/\gamma}\right),$$

which, in extreme value theory, is known to be the limiting distribution for maxima and containing the extreme value index. The non-stationarity can be introduced via the parameters in the GEV distribution as was done e.g. in katz et al. (2002), Kharin et al. (2005) and Cooley (2009). These authors introduced a parametric model for the parameters in de GEV distribution. Other authors use a non-parametric approach, e.g. Davison et al. (2005) and Chavez et al. (2005).

Although commonly used in climatology, analyzing e.g. annual maxima of temperatures, the GEV approach is not very efficient. Therefor, in this article, we prefer to use, what is called, a peaks over threshold method. Instead of only working with maxima, we will model all data exceeding a predefined threshold. For this threshold we adopt the following definition

$$1 - F_Y(u) = \frac{k}{n}, \tag{1}$$

where k/n is a small number, indicating what percentage of data will be called extreme. The choice of the threshold u then corresponds to the choice of k. Typically, different k values are taken in extreme value statistics to obtain estimators to get a clearer view.
From Extreme Value Theory it is known that the distribution of absolute excesses $Y - u$, over a high threshold u can be approximated well by a Generalized Pareto Distribution

$$P(Y - u \leqslant x|Y > u) \sim 1 - \left(1+\gamma\frac{x}{\sigma}\right)^{-1/\gamma}.$$

Again several authors proposed to introduce non-stationarity via the parameters of this model. Although some authors use a fixed threshold, the need for a varying threshold $u(x)$ depending on x has recently been recognized by several authors, e.g. Smith (1989), Nogaj et al. (2007), Coelho et al. (2008), Jagger et al. (2009), Eastoe et al. (2009) and Kyselý at al.(2010).

In this text, since we are working with Pareto type distribution, we will not make use of the general Generalized Pareto Distribution as the limiting distribution of the absolute excess over a high threshold u. Instead we will focus on the the limiting distribution of the relative excesses over the high threshold $u(x)$ to estimate the trend $\gamma(x)$ in a positive extreme value index. Let us denote the relative excesses of $Y|x$ over a high threshold u(x), given that $Y|x$ exceeds $u(x)$ by $Z|x$. For a Pareto-type distribution with parameter $\gamma(x)$, Z satisfies the condition

$$P\left(\frac{Y}{u} > z|Y > u\right) = P\left(Z > z|X = x\right) \to z^{-\frac{1}{\gamma(x)}}, \;\; u(x) \to \infty. \tag{2}$$

Although one can expect similarities with the GPD approach from above, some clear difference can be remarked.

First, note that the scale parameter σ does not occur in the limit for the relative excesses for Pareto type distributions. In case of numerical estimation procedures, this might enhance the convergence to stable solutions of such procedures.

Next, as in the GPD approach, the threshold u(x) has to be determined. We will work with a varying threshold and explain in Section 2.2 how we can determine this threshold in our setting.

Further, although second order properties of the GPD have been studied in a univariate context (see e.g. Drees et al. (2002), Beirlant et al. (2009)), and although Chavez-demoulin et al. (2005) remark, based on some simulations, that the overall biases in the trend setting seem to be sufficiently small, the second order properties are not systematically studied when trends appear. In this text we assume the following second order condition

$$P(Z > z) = z^{-1/\gamma}\left(1 + b(u)\frac{z^{\rho}-1}{\rho}\right), \;\; b(u) \sim B\, u^{\rho} \to 0, \; B\text{constant}; \rho < 0, \qquad u \to \infty \tag{3}$$

which is mild, as one can e.g. easily check that distributions of the Hall class satisfy the condition in (3). Note that Y, u, γ and B, ρ and $b(u)$ may depend on the covariate x.
Using a second order condition as in (3) allows for a study of the asymptotic bias of the resulting estimates which, together with the asympotic variance, could be usefull when determining an appropriate k-value.

Finally, we will concentrate on a linear trend function for $\gamma(x, \alpha) = \alpha_1 + \alpha_2 x$. This choice can be argumented by the fact that in practice one does not expect a huge change in γ, such that (at least locally) a linear trend might describe most practical data sets well, at least locally. In case this is not true, the proposed methods can be easily adapted for any kind of trend function.

2.2 Threshold choice

As stated before, it is already recognized by several authors that a varying threshold is needed. In this text, we focus on studying how the threshold alters with the covariate information due to a change in the heaviness of the tail. This is now illustrated in Figure 2 for a simulated data set.

Simulated data.
We illustrate the trend functions of the different parameters for a simulated data set of size $n = 500$ from the heavy tailed Fréchet distribution. Recall that the distribution function of a Fréchet(α) distribution is given by

$$1 - F(x) = 1 - exp(-x^{-\alpha}) \; .$$

In Figure 2, a simulated data set from a Fréchet($1/(0.5+0.1x)$), with $x = 0.01, 0.03, \ldots, 10$ as covariate information, is shown. For a Fréchet distribution, the extreme value index equals $\gamma = 1/\alpha$, leading to $\gamma(x, \alpha) = 0.5 + 0.1x$. From Figure 2 it is clear that the data become more heavy tailed as x grows.
From the definition of the threshold in (1) it follows that $u(x, \psi) = \left(-\log\left(1 - \frac{k}{n}\right)\right)^{-1/\alpha}$. For $k = 50$, $x = 1$, resp. $x = 9$ this becomes $u = 6.36$, resp. $u = 75.05$. Since the data are becoming more heavy tailed as x grows, a value of 15 is out of ordinary for small x, where it is not so exceptional for larger x.
We propose to apply the parametric quantile regression methodology of Koenker and Basset (1987) to derive an expression for the threshold $u(x)$. The authors determine the p^{th} regression quantile using the following optimization problem

$$min_{\psi} \sum \left(p\,(Y_i - u(x_i, \psi))^{+} + (1-p)\,(Y_i - u(x_i, \psi))^{-}\right), \tag{4}$$

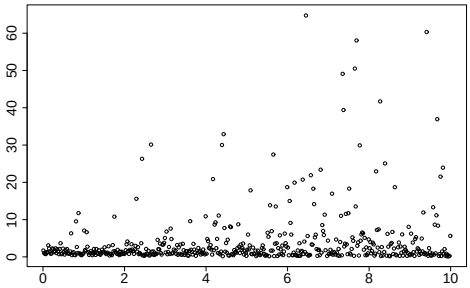

Figure 2: $(x, Y|x)$ for a simulation of size 500 of a Fréchet$(1/(0.5 + 0.1x))$. Note that the y-axis is limited to 65, although 7 Y-values exceed this value. The maximum Y-value of 2290 is attained for $x = 8.69$

with $x+ = max(0, x)$ and $x- = max(0, -x)$.
Since $u(x)$ was defined such that $P(Y|x > u(x)) = k/n$ for k/n small, $u(x)$ is actually the $p = (n-k)/(n+1)^{th}$ regression quantile.

In our setting, the threshold u(x) might indeed be assumed to change according to some known law $u(x_i, \psi)$. As mentioned and argumented before, we will concentrate on a linear trend function for $\gamma(x, \alpha) = \alpha_1 + \alpha_2 x$ in this text. Therefore, a suitable choice for u might be $u(x, \psi) = exp(\psi_1 + \psi_2 x)$, since for Pareto-type distributions belonging to the Hall subclass $\frac{k}{n} = \bar{F}(u(x, \psi)) \sim Du^{-1/\gamma(x)}$, $u \to \infty$. This shape of the function u can be found in the literature as well (see e.g. Eastoe et al. (2009)).

Simulated data set. The Koenker and Basset technique, is illustrated on the simulated data set. It will be argumented later that k=179 might be a good choice. The estimators of the threshold $u(x)$ for this k-value is shown on top of the scatterplot in Figure 3. The true $u(x)$ is also added and practically coincides with the estimated threshold.

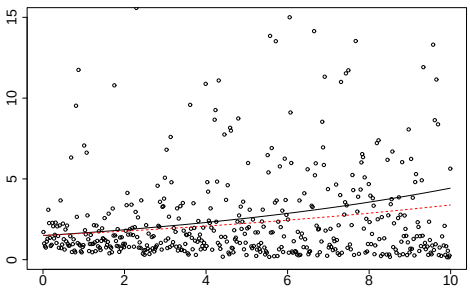

Figure 3: $(x, Y|x)$ as in Figure 2 together with the estimated $u(x)$ for $k = 179$. The dashed line is showing the true $u(x)$.

Note that it might be interesting to study the underlying mechanism of the variation of the threshold as a function of x. This variation indeed might be twofolded. As illustrated before, the threshold might alter with the covariate information due to a change in the heaviness of the tail. However, a change in location also results in a threshold evolving with the covariate information x. We assume that such a change in location has been removed. The technique of local polynomial quantile regression can be now e.g. be used to remove a trend due to a change in location.

From now on, we will retain a data point (x_j, Y_j) as extreme only if Y_j exceeds the corresponding

threshold $u(x_j)$. As such we are left with k data combinations $(x_i, Z_i), i = 1 \dots k$, where $Z_1, \dots, Z_k$ are notations for the relative excesses of Y_j as before. Based on these data, estimators concerning the extreme value index are obtained.

2.3 Estimators concerning $\gamma(x, \alpha)$

Maximum likelihood estimators for $\gamma(x, \alpha)$ will be provided. The asymptotical variance and bias of the corresponding estimators will be studied and lead to a hypothesis test to detect whether a trend is showing in the extreme value index.

The parameters α_1 and α_2 occuring in the extreme value index $\gamma = \alpha_1 + \alpha_2 x$ can be calculated using the maximum likelihood method based on the model in (2). This leads to the following score functions

$$(5) \qquad \begin{aligned} \frac{\delta \log L}{\delta \alpha_1} &= -\sum_{i=1}^{k} \frac{1}{\gamma(x_i)} + \sum_{i=1}^{k} \frac{\log(Z_i)}{\gamma^2(x_i)} = 0 \\ \frac{\delta \log L}{\delta \alpha_2} &= -\sum_{i=1}^{k} \frac{x_i}{\gamma(x_i)} + \sum_{i=1}^{k} \frac{x_i \, \log(Z_i)}{\gamma^2(x_i)} = 0 \end{aligned}$$

The resulting maximum likelihood estimators are denoted by $\hat{\alpha}_1$ and $\hat{\alpha}_2$. Note that also estimators for small probabilities and large quantiles can be proposed based on $\hat{u}(x)$ and $\hat{\gamma}(x)$.

Simulated data set 1. The estimators $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are given in Figure 4. E.g. for $k = 179$, this leads to the estimators $\hat{\alpha}_1$=0.525 and $\hat{\alpha}_2$=0.106. Together they give rise to an estimate of $\gamma(x)$. This estimate is compared to the true $\gamma(x)$ in Figure 5 for $x = 0.1, 1, 3, 5, 7, 9$.

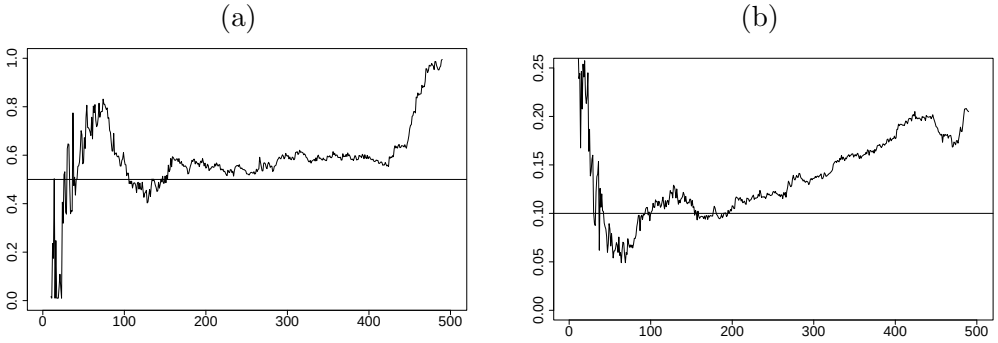

Figure 4: (a) For the same data as in Figure 2: (k, $\hat{\alpha}_1$); (b) (k, $\hat{\alpha}_2$). The horizontal line is indicating the true values $\alpha_1 = 0.5$ and $\alpha_2 = 0.1$.

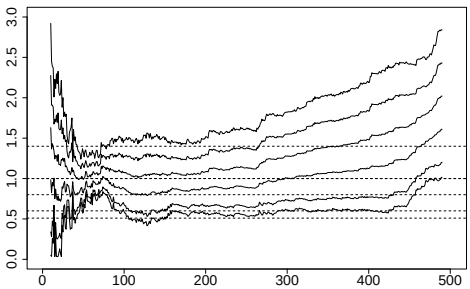

Figure 5: For the same data as in Figure 2: $(k, \hat{\gamma}(x))$ together with the true values (dashed lines) for $x = 0.1, 1, 3, 5, 7, 9$.

The asymptotic variance of the estimators $\hat{\alpha}_1$ and $\hat{\alpha}_2$ can be found from the inverse information matrix $I(\alpha) = -E\left[\frac{\delta^2 \log L}{\delta\alpha_i \delta\alpha_j}\right]$. Exploiting the fact that $E[\log Z_i] \sim \gamma(x_i)$, it follows immediately that asymptotically

$$\text{(6)} \qquad \text{Var}\,(\hat{\alpha}_1) = \frac{1}{D}\sum_{i=1}^{k}\frac{x_i^2}{\gamma^2(x_i)};\ \text{Var}\,(\hat{\alpha}_2) = \frac{1}{D}\sum_{i=1}^{k}\frac{1}{\gamma^2(x_i)};\ Cov(\hat{\alpha}_1,\hat{\alpha}_2) = -\frac{1}{D}\sum_{i=1}^{k}\frac{x_i}{\gamma^2(x_i)}$$

where the denominator D equals $\sum_{i=1}^{k}\frac{1}{\gamma^2(x_i)}\sum_{i=1}^{k}\frac{x_i^2}{\gamma^2(x_i)} - \left(\sum_{i=1}^{k}\frac{x_i}{\gamma^2(x_i)}\right)^2$.

By neglecting the terms of order $b(u(x_i))$ a bias will be introduced in the estimators $\hat{\alpha}_1$ and $\hat{\alpha}_2$. Formulas for these biases can derived and are given by

$$\begin{aligned} Bias(\hat{\alpha}_1) &= E\left[-\frac{1}{F}\left(\sum_{i=1}^{k}\frac{x_i^2}{\gamma^2(x_i)}\sum_{i=1}^{k}b_{x_i}(u(x_i))Z_i^{\hat{\rho}} + \sum_{i=1}^{k}\frac{x_i}{\gamma^2(x_i)}\sum_{i=1}^{k}x_i b_{x_i}(u(x_i))Z_i^{\hat{\rho}}\right)\right] \\ \text{(7)} \qquad Bias(\hat{\alpha}_2) &= E\left[-\frac{1}{F}\left(\sum_{i=1}^{k}\frac{1}{\gamma^2(x_i)}\sum_{i=1}^{k}x_i b_{x_i}(u(x_i))Z_i^{\hat{\rho}} + \sum_{i=1}^{k}\frac{x_i}{\gamma^2(x_i)}\sum_{i=1}^{k}b_{x_i}(u(x_i))Z_i^{\hat{\rho}}\right)\right] \end{aligned}$$

where the denominator F equals $\sum_{i=1}^{k}\frac{1}{\gamma^2(x_i)}\sum_{i=1}^{k}\frac{x_i^2}{\gamma^2(x_i)} - \left(\sum_{i=1}^{k}\frac{x_i}{\gamma^2(x_i)}\right)^2$.
The order of this bias is determined by the order of $b(u(x_i))$, which tends to zero as $u \to \infty$. The bias could be estimated using the second order parameters. However, using maximumlikelihood to estimate a parametrized version of (3), leads to unstable estimators. Using a bootstrap method to estimate the bias seems to lead to better results and so this method is used in this article.

The bias and variance of $\hat{\gamma}(x_i)$ follow from the bias, variance and covariance of the estimators $\hat{\alpha_1}$ and $\hat{\alpha_2}$. Indeed if, $\gamma(x_i) = \alpha_1 + \alpha_2 x_i$ then

$$\begin{aligned} \text{(8)} \qquad \text{Var}\,(\hat{\gamma}(x_i)) &= \text{Var}\,(\hat{\alpha}_1) + x_i^2\,\text{Var}\,(\hat{\alpha}_2) + 2x_i Cov(\hat{\alpha}_1,\hat{\alpha}_2) \\ Bias(\hat{\gamma}(x_i)) &= Bias(\hat{\alpha}_1) + x_i Bias(\hat{\alpha}_2) \end{aligned}$$

Note that as such the AMSE could be calculated as well. One might adopt the k-value minimizing the AMSE as optimal.

Simulated data set 1. The variance and the covariance of the estimators can be estimated based on the data when plugging in the estimator $\hat{\gamma}(x_i)$ in (6), as is illustrated for the variance in Figure 6. Estimators of the squared bias, using the bootstrap methodology, are added in the same graph. Based on the AMSE for $\gamma(x)$, the 'optimal' k-value of 179.

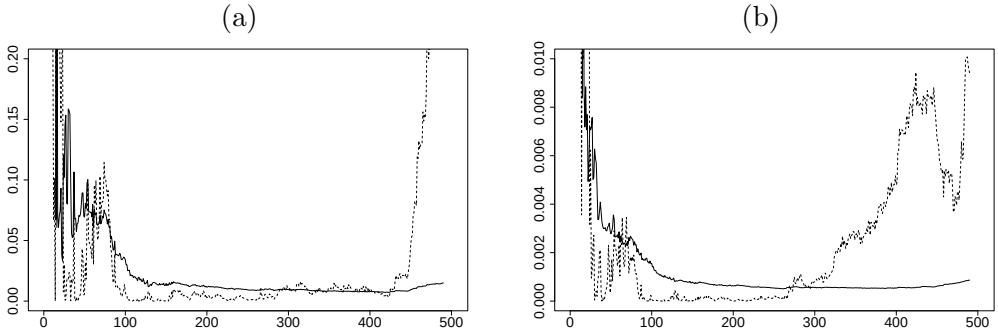

Figure 6: For the same data as in Figure 2: (k, Var $(\hat{\alpha})$)(full line) and $(k, Bias^2(\hat{\alpha}))$ (dotted line) for (a); $\hat{\alpha}_1$; (b) $\hat{\alpha}_2$.

Finally one could also look at the following diagnostic plot. According to (3), $\log(Z)$ divided by $\gamma(x)$ is asymptotically distributed as a standard exponential distribution. Thus when a good estimate of $\gamma(x)$ is used, we expect the exponential QQ plot for $\log(Z)/\hat{\gamma}(x)$ to show approximatly a linear pattern with slope 1 for a suitable k-value.

Simulated data set 1. In Figure 7 an exponential QQ plot is given for $\log(Z)/\hat{\gamma}(x)$ for $k = 179$.

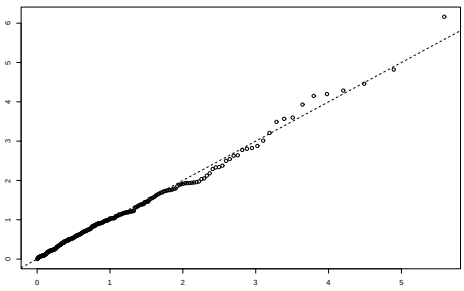

Figure 7: For the same data as in Figure 2: Exponential QQ plot ($k = 179$) (first bissector is added in dotted line); .

2.4 Real life data

Let us now analyse the losses Y (adapted for inflation) from the Swiss Re example. These losses are plotted in time in Figure 8(a). The extreme value index of these losses seems to increase in time as can be concluded from the Moving Hill plot in Figure 8(b).

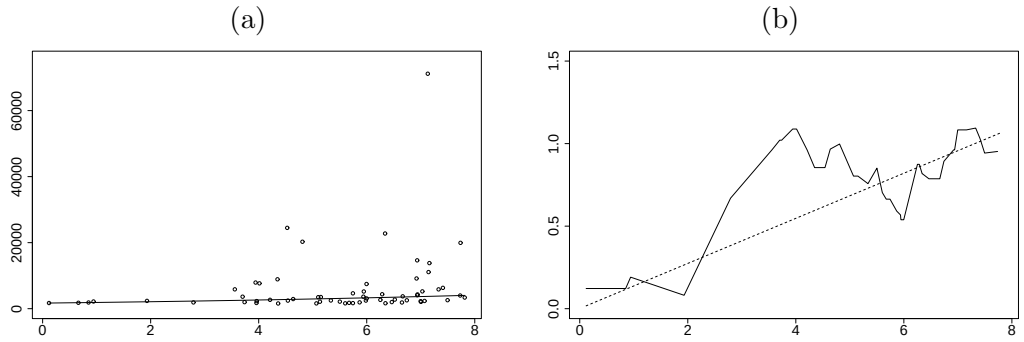

Figure 8: Swiss Re Loss data: (a) (x, Y) where x denotes the number of 5 years since 1970 with the estimation for the threshold u superimposed for $k = 31$; (b) Moving Hill plot with $\hat{\gamma}(x)$ for $k = 31$ added in dashed line.

Let us assume that the extreme value index changes according to some linear trend $\gamma(x) = \alpha_1 + \alpha_2 x$, where x denotes the number of 5-years since January 1, 1970. The corresponding maximum likelihood estimators are shown in Figure 9. In this example the choice of k is not crucial as the estimators are quite stable. E.g. for $k = 31$, minimizing $AMSE(x = 7)$, $\hat{\alpha}_1 = 0.001$, whereas $\hat{\alpha}_2 = 0.1366$.
It seems that γ can be estimated as $0.001 + 0.1366x$. Note that α_2 is significantly different from 0. This linear trend model summarizes well the moving Hill plot as be seen in Figure 8(b). Another diagnostic that can be used to assess the linear trend funcion is the exponential quantile plot. In Figure 10(b), such an exponential quantile plot is provided, together with the first bissector which is linked to the expected standard exponential. There seem to be no clear indications that a linear trend function is unappropriate.

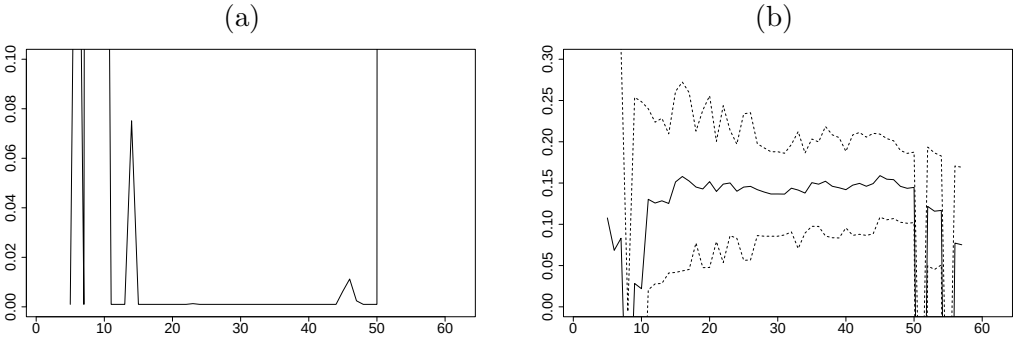

Figure 9: Swiss Re Loss data: (a) $(k, \hat{\alpha}_1)$; (b) $(k, \hat{\alpha}_2)$ with 95% confidence intervals.

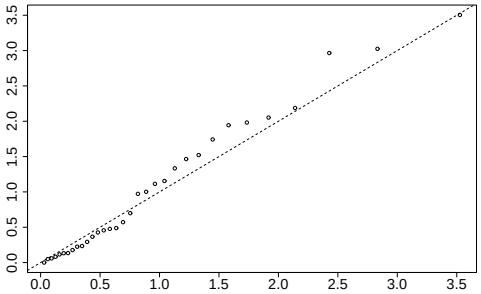

Figure 10: Swiss Re Loss data: Exponential QQ plot for $logY/\gamma(x)$, with first bissector added in dashed line.

Finally, the estimator $\hat{\gamma}(x)$ can be used to determine small probabilities and large quantiles as in Figure 11. For comparison, also the estimators based on the classical Weismann type estimators using the Hill estimator under the assumption that no time component is to be taken into account, are added in the same Figure. Clearly these last estimators do not describe well the behavior of the extreme values, indicating that one can not defend that the extreme value index stays unaltered in time.

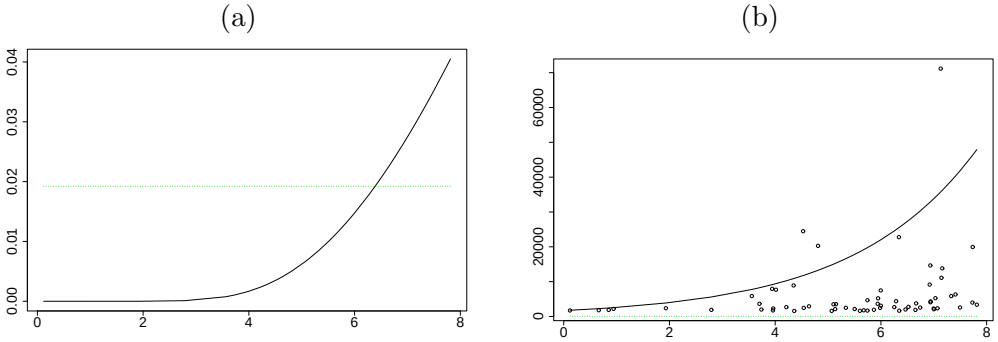

Figure 11: Swiss Re Loss data: (a) $(x, \hat{p}(x; 60000)$; (b) $(x, \hat{q}(x, 0.01))$; based on the trend function $\gamma(x) = 0.001 + 0.159x$. The estimation is also performed assuming that the extremal behavior of the data does not change in time and is added with a dotted line.

3 Change point in trend parameters

To study whether a change point in the trend parameters occurs, we denote the sample as $Y_1, \ldots Y_m$, $Y_{m+1}, \ldots Y_n$. The following hypothesis test for α e.g. might be of interest.

$$
\begin{aligned}
&H_0 \;:\; \alpha_1 = \alpha_2 = \ldots = \alpha_n = \alpha \\
&\qquad\qquad\qquad \text{versus} \qquad\qquad\qquad (9)\\
&H_1 \;:\; \alpha_1 = \alpha_{m^*} \neq \alpha_{m^*+1} = \alpha_n \text{ for some } m^*,
\end{aligned}
$$

To detect a change point in $Y_1, \ldots Y_{m^*}, Y_{m^*+1}, \ldots Y_n$ at m^*, for the parameter θ, Csörgő and Horváth (1997) introduce a likelihood based test statistic. First we will shortly repeat the general procedure as introduced by Csörgő and Horváth (1997). Then we will adapt the general procedure to perform the hypothesis test in (9).

3.1 General procedure

Csörgő and Horváth (1997) basically compare a likelihood of the whole data set to the corresponding likelihoods of the two groups $Y_1, \ldots Y_{m^*}$ and $Y_{m^*+1}, \ldots Y_n$.
In general, denoting the loglikelihood based on the whole data set by $\log L_n$ and the loglikelihoods of the two subgroups by $\log L_m$ resp. $\log L_m^*$, the following test statistic is suggested

$$
Z_n \;=\; \sqrt{\max_{1 \leqslant m < n} \; (-2 \log \Lambda_m)}\,, \qquad (10)
$$

with

$$
-2 \log \Lambda_m = 2 \left[L_m(\hat{\theta}_m) + L_m^*(\hat{\theta}_m^*) - L_n(\hat{\theta}_n) \right] \qquad (11)
$$

where $\hat{\theta}_m$ is the likelihood estimator based on $(Y_1, \ldots, Y_m)$, $\hat{\theta}_m^+$ on $(Y_{m+1}, \ldots, Y_n)$ and $\hat{\theta}_n$ on $(Y_1, \ldots, Y_n)$.

For a change in 2 parameters Csörgő and Horváth (1997) suggest to use the critical values based on

$$
\sqrt{\sup_{\epsilon \leqslant t < 1-\epsilon} \; \frac{B_2^2(t)}{t(1-t)}},
$$

with $B_2(t)$ the sum of two Brownian bridges.

3.2 Specific procedure

First find an expression for the threshold $u(x)$ as in Section 2.2 for the whole data set, using the quantile regression methodology of Koenker and Basset (1978). Next, again only retain the data points Y overshooting $u(x)$, denoted by Z. The Z-values of the first group are denoted by $Z_1, \ldots Z_{k_1}$. In the same way $Z_{k_1+1}, \ldots Z_k$ denote the $Z - values$ of the second group. Further put $k = k_1 + k_2$.

Under H_0, Z_1, …, Z_k asymptotically follow a Pareto distribution with parameter $\gamma(x, \alpha)$ as in (2). The loglikelihood function $L_n(\alpha)$ is given by

$$
\log L(z) = \sum_{i=1}^{k} \log \gamma(x_i) - \sum_{i=1}^{k} \left(\frac{1}{\gamma(x_i)} + 1 \right) \log Z_i
$$

The loglikelihood based on the whole data set is denoted by $\log L_n$. The loglikelihoods of the two subgroups are denoted by $\log L_m$ resp. $\log L_m^*$. Test statistic (10) can then be used, with

$$
-2 \log \Lambda_m = 2 \left[L_m(\hat{\alpha}_m) + L_m^*(\hat{\alpha}_m^*) - L_n(\hat{\alpha}_n) \right] \qquad (12)
$$

where the likelihood estimator $\hat{\alpha}_m$ is based on $(Y_1, \ldots, Y_m)$, $\hat{\alpha}_m^+$ on $(Y_{m+1}, \ldots, Y_n)$ and $\hat{\alpha}_n$ on $(Y_1, \ldots, Y_n)$.

References

[1] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J.L. (2004). Statistics of extremes. *Wiley.* 490p.

[2] Beirlant, J., Joossens, E. and Segers, J. (2009). Second-order refined peaks-over-threshold modelling for heavy-tailed distributions. *Journal of Statistical Planning and Inference* **139**, 2800-2815.

[3] Chavez-Demoulin, V. and Davison, A. (2005). Generalized additive modelling of sample extremes. *J R Stat Soc Ser C Appl Stat* **54**, 207-222.

[4] Coelho, C., Ferro, C., Stephenson, D. and Steinskog, D. (2008). Methods for Exploring Spatial and Temporal Variability of Extreme Events in Climate Data. *Journal of Climate* **21**, 2072-2092.

[5] Cooley, D. (2009). Extreme value analysis and the study of climate change. *Climate Change* **97**, 77-83.

[6] Csörgő, M., Horváth, L. (1997). Limit Theorems in Change Point Analysis. *Wiley, Chichester*

[7] Dierckx, G., Teugels, J. (2010). Changepoint Analysis of extreme values. *Environmetrics.* **21** 7-8, 659-867.

[8] Drees, H., Ferreira, A. and De Haan, L. (2004). On maximum likelihood estimation of the extreme value index. *The Annals of Applied Probability* **14** 3, 1179-1201.

[9] Eastoe, E. and Tawn, J. (2009). Modelling non-stationary extremes with application to surface level ozone. *Applied Statistics* **58** 1, 25-45.

[10] Gomes, I., Oliveira, O. (2002). The Bootstrap Methodology in Statistics of Extremes - Choice of the Optimal Sample Fraction. *Extremes* **4**, 331-358.

[11] Jagger, T. and Elsner, J. (2009). Modeling tropical cyclone intensity with quantile regression. *Int. J. Climatol* **29**, 1351-61.

[12] Katz, R., Parlange, M. and Naveau, P. (2002). Extremes in hydrology. *Adv Water Resour* **25**, 1287-1304.

[13] Kharin, V. and Zwiers, F. (2005). Estimating extremes in transient climate change simulations. *Journal of Climate* **18**, 1156-1173.

[14] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46**, 33-50.

[15] Kyselý, J., Picek, J. and Beranov'a, R. (2010). Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold. *Global and Planetary Change* **72**, 55-68.

[16] Smith, R. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Stat Sci* **4**, 367-393.