

A Review of Univariate Tail Estimation

Una rassegna sulla stima di code di distribuzioni univariate

Beirlant Jan
Department of Mathematics
Katholieke Universiteit Leuven
jan.beirlant@wis.kuleuven.ac.be

Riassunto: Un argomento centrale nell'analisi dei valori estremi, che é stato oggetto di grande attenzione, é la stima adattiva del parametro di forma γ . Il caso di γ positivo, i.e. quando $\bar{F}$ é una funzione a variazione regolare, é stato trattato estesamente. Per questo caso particolare é noto lo stimatore di Hill (1975). Tuttavia esistono molti altri stimatori, soprattutto stimatori di tipo kernel e stimatori dei minimi quadrati ponderati della pendenza basati sul QQ plot di Pareto o sul plot di Zipf (Csorgo and Viharos (1998)). Inoltre, utilizzando un modello di regressione esponenziale (ERM) per gli intervalli tra successive statistiche d'ordine estreme, Beirlant *et al.* (1999) e Feuerverger and Hall (1999) hanno proposto stimatori con distorsione ridotta. Stimare i parametri nell'ERM permette anche di selezionare il numero appropriato di dati estremi quando si utilizzano gli stimatori citati in precedenza, in particolare lo stimatore di Hill.

Nel caso di γ a valori reali lo stimatore di Hill é stato generalizzato a uno stimatore di tipo dei momenti (Dekkers *et al.* (1989)), e a uno stimatore tipo Hill basato sul plot dei quantili generalizzati introdotto da Beirlant *et al.* (1996a). Un metodo di stima consegue dalla stima di massima verosimiglianza basata sulla distribuzione generalizzata di Pareto con vincoli per parametri di forma negativi (Smith (1987)). In questo articolo si precisa come gli stimatori già menzionati per $\gamma > 0$ possono essere generalizzati al caso di valori reali di γ , per esempio sostituendo il plot di Zipf con il plot dei quantili generalizzati.

Infine, é discussa un'applicazione alla stima dei quantili estremi e della coda di una distribuzione.

Keywords: extreme value index, least squares, bias, mean squared error, quantile plots.

1. Introduction

Let $X_1, X_2, \dots, X_n$ be a sequence of independent and identically distributed random variables with distribution function F and tail quantile function U defined by $U(x) = \inf\{y; F(y) \geq 1 - 1/x\}$. We denote the order statistics by $X_{1,n} \leq \dots \leq X_{n,n}$. The statistical model considered in this paper is given by the basic maximal domain of attraction condition which governs extreme value theory :

Suppose that there exist sequences of constants $(a_n; a_n > 0)$ and (b_n) , and some $\gamma \in \mathbb{R}$,

such that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{X_{n,n} - b_n}{a_n} \leq x \right) = G_\gamma(x) \text{ for all } x, \quad (1)$$

with $G_\gamma(x) = \exp \left(-(1 + \gamma x)^{-\frac{1}{\gamma}} \right)$.

The main aim of this paper is to discuss the estimation problem of the extreme value index γ under this model, and of extreme quantiles and tail probabilities. The extreme

value index γ is known to be the main indicator about the decay of the distribution tail, ranging from distributions with finite endpoint ($\gamma < 0$), over exponentially fast decreasing tails ($\gamma = 0$), to polynomially decreasing Pareto-type tails ($\gamma > 0$).

2. Pareto type tails

Most research in this area concentrates on the heavy tailed distributions with $\gamma > 0$. An excellent recent overview of this literature can be found in Csorgo and Viharos (1998). When γ is strictly positive, it follows from (1) that X is of Pareto-type, i.e.

$$\bar{F}(tx)/\bar{F}(x) \rightarrow t^{-1/\alpha} \text{ as } x \rightarrow \infty, \text{ for all } t > 0.$$

with $\alpha = \gamma$. This is equivalent to the regular variation of the tail function $U(x) = Q(1 - \frac{1}{x})$ with Q the quantile function of F :

$$U(x) = x^\alpha L(x), \quad (2)$$

where L is a slowly varying function, i.e. satisfying $L(tx)/L(x) \rightarrow 1$ as $x \rightarrow \infty$, for all $t > 0$. For the estimation of α under this regular variation model, Hill proposed the following estimator:

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

Here $k = k_n$ is a sequence of positive integers ($1 \leq k < n$) which, in theoretical asymptotic considerations, satisfies the conditions

$$k \rightarrow \infty \text{ and } \frac{k}{n} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

It has been mentioned in literature that this and many other estimators can be viewed as estimators of the slope in a Pareto quantile plot: if L is constant (i.e. when X follows a strict Pareto distribution) the Pareto quantile plot or Zipf plot

$$\left(\log \left(\frac{n+1}{j} \right), \log X_{n-j+1,n} \right), \quad j = 1, \dots, n \quad (3)$$

is overall linear with slope approximately equal to α . If L is not constant, the Pareto quantile plot eventually exhibits this feature for smaller values of j . Hence a large group of estimators of α evolves from different possible regression fits on Pareto quantile plots. Beirlant *et al.* (1996b) pointed out that fitting a constrained weighted least squares line to the k upper points in (3) leads to the class of kernel estimators

$$\hat{\alpha}_{k,n}^K = \frac{\sum_{j=1}^k \frac{j}{k} K\left(\frac{j}{k}\right) [\log X_{n-j+1,n} - \log X_{n-j,n}]}{\frac{1}{k} \sum_{j=1}^k K\left(\frac{j}{k}\right)}$$

with a kernel K integrating to 1. The Hill estimator is obtained for $K = I_{(0,1]}$. Furthermore, Kratz and Resnick (1996) and Schultze and Steinebach (1996) introduced the Zipf estimator which is an unconstrained least squares estimator based on (3):

$$\hat{\alpha}_{k,n}^Z = \frac{\sum_{j=1}^k \log \frac{k+1}{j} \log X_{n-j+1,n} - \frac{1}{k} \sum_{j=1}^k \log \frac{k+1}{j} \sum_{j=1}^k \log X_{n-j+1,n}}{\sum_{j=1}^k \log^2 \frac{k+1}{j} - \frac{1}{k} \left(\sum_{j=1}^k \log \frac{k+1}{j} \right)^2}.$$

This estimator was generalized in Csorgo and Viharos (1998) to the class of weight estimators

$$\hat{\alpha}_{k,n}^{WZ} = \frac{\sum_{j=1}^k [\int_{(j-1)/k}^{j/k} J(s) ds] \log X_{n-j+1,n}}{\sum_{j=1}^k [\int_{(j-1)/k}^{j/k} J(s) ds] \log(k/j)}$$

with a weight function J integrating to 0.

The asymptotic nature of the definition of the Pareto-type model implies that any estimator will contain quantities, the selection of which plays a crucial role for successful application of such a technique: especially the choice of the number of extremes k has received a lot of interest; see for example Beirlant *et al.* (1996b), Resnick and Stărică (1997), Drees and Kaufmann (1998), Drees *et al.* (2000) and Danielsson *et al.* (2001). The issue is important: the extreme volatility of the Hill plot $\{(k, H_{k,n}) : 1 \leq k < n\}$ makes it difficult to use the estimator in practice if no guideline is given for the choice of k . Minimizing the mean squared error of the estimation technique has been a constant guideline throughout almost all publications on this topic: due to the asymptotic nature of the nuisance part of the model the bias is smallest for small k , while the variance is of course reduced with increasing k .

However, next to the choice of k , at instances the appearance of a substantial bias is considered to be a serious problem. This typically happens when $|\rho|$ is small in the Hall model given by

$$L(x) = M_1(1 + M_2 x^\rho \{1 + o(1)\}) \quad (4)$$

where $\rho < 0$, $M_1 > 0$ and $M_2 \in \mathbb{R}$. Recently, in Beirlant *et al.* (1999) and Feuerverger and Hall (1999), the regression problem defined by the upper subsets of a Pareto quantile plot was further specified taking into account a second order slow variation condition, essentially induced by the Hall model (4). In this way, the bias can be reduced or, the bias of the Hill estimator can be estimated. From this, also an adaptive estimation procedure to choose k can be given, as discussed in Beirlant *et al.* (2001).

3. The case $\gamma \in \mathbb{R}$

The estimation of $\gamma \in \mathbb{R}$ has been studied less extensively. There are mainly two sets of solutions which originated from two different formulations of the model, which are equivalent to (1).

First, the POT (Peaks over Threshold) method (see for instance Smith (1987), Davison and Smith (1990)) is based on results given by Balkema and de Haan (1974) and Pickands (1975), stating that the limit distribution of the exceedances over a threshold u when $u \rightarrow \infty$ is given by a generalized Pareto distribution (GPD) with distribution function

$$H_{\gamma,\sigma}(x) = 1 - \left(1 + \gamma \frac{x}{\sigma}\right)^{-1/\gamma}.$$

The fit of the generalized Pareto distribution over a high threshold can be performed by maximum likelihood (Smith (1987)), probability-weighted moments (Hosking *et al.* (1985)), Bayesian analysis methods (Coles and Powell (1996)), or by a percentile method given by Castillo and Hadi (1997). Here most estimating methods are only valid with some restrictions on the value of γ .

Secondly, several other estimators based on k upper order statistics have been proposed motivated by the following asymptotic relations which are equivalent to (1) (see Dekkers *et al.* (1989)):

there exists a positive function a such that for all $t > 0$

$$\lim_{x \rightarrow \infty} \frac{U(tx) - U(x)}{a(x)} = \begin{cases} \log t & \gamma = 0 \\ \frac{t^\gamma - 1}{\gamma} & \gamma \neq 0 \end{cases} \quad (5)$$

or, equivalently,

$$\lim_{x \rightarrow \infty} \frac{\log U(tx) - \log U(x)}{a(x)/U(x)} = \begin{cases} \log t & \gamma \geq 0 \\ \frac{t^\gamma - 1}{\gamma} & \gamma < 0. \end{cases} \quad (6)$$

Pickands' (1975) estimator

$$\hat{\gamma}_{k,n}^P = \frac{1}{\log 2} \log \left(\frac{X_{n-\lfloor k/4 \rfloor, n} - X_{n-\lfloor k/2 \rfloor, n}}{X_{n-\lfloor k/2 \rfloor, n} - X_{n-k, n}} \right)$$

can be seen to be consistent on the basis of (5); for more details see Dekkers and de Haan (1989). Refined estimators of this type and more general classes of estimators in this spirit were discussed in Drees (1995), Drees (1996), Drees (1998) and Segers (2001).

Based on (6), Dekkers *et al.* (1989) proposed the following estimator, termed *the moment estimator*, which brings about an adaptation of the Hill estimator to estimate $\gamma \in \mathbb{R}$:

$$\hat{\gamma}_k^M := \hat{\gamma}_{k,n}^M = H_{k,n} + 1 - \frac{1}{2} \left(1 - \frac{H_{k,n}^2}{S_{k,n}} \right)^{-1},$$

where

$$S_{k,n} = \frac{1}{k} \sum_{j=1}^k (\log X_{n-j+1, n} - \log X_{n-k, n})^2.$$

Next, based on (6), Beirlant *et al.* (1996a) proposed an estimator of $\gamma \in \mathbb{R}$ based on a generalized quantile plot, which takes over the role of the Pareto quantile plot in this more general setting. To construct this plot one observed, under (6), that UH is regularly varying at infinity with index γ where $H(x) = \mathbb{E}[\log X - \log U(x) | X > U(x)]$, i.e.

$$UH(x) = U(x)H(x) = x^\gamma \ell(x), \quad (7)$$

with ℓ again a slowly varying function at infinity. From this it can then be seen that the generalized quantile plot

$$\left(\log \left(\frac{n+1}{j} \right), \log UH_{j,n} \right), \quad j = 1, \dots, n, \quad (8)$$

becomes ultimately linear for small j , where

$$UH_{j,n} = X_{n-j, n} \left(j^{-1} \sum_{i=1}^j \log X_{n-i+1, n} - \log X_{n-j, n} \right)$$

is an empirical substitute for $UH((n+1)/j)$.

As in the $\gamma > 0$ case, one can now construct several regression based estimators for $\gamma \in \mathbb{R}$. The simplest estimator of this kind is given by the generalized Hill estimator

$$\hat{\gamma}_{k,n}^H = \frac{1}{k} \sum_{j=1}^k \log UH_{j,n} - \log UH_{k+1,n},$$

where k denotes again the number of extremes that is used in the estimation. Of course a class of kernel estimators $\hat{\gamma}_{k,n}^K$ generalizing the Pareto index estimators $\hat{\alpha}_{k,n}^K$ can also be introduced by formal replacement of $X_{n-j+1,n}$ by $UH_{j,n}$, as introduced in Beirlant *et al.* (1996a).

One can also consider the generalized unconstrained least squares estimator $\hat{\alpha}_{k,n}^Z$ to the case where $\gamma \in \mathbb{R}$:

$$\hat{\gamma}_{k,n}^Z = \frac{\sum_{j=1}^k \log \frac{k+1}{j} \log UH_{j,n} - \frac{1}{k} \sum_{j=1}^k \log \frac{k+1}{j} \sum_{j=1}^k \log UH_{j,n}}{\sum_{j=1}^k \log^2 \frac{k+1}{j} - \frac{1}{k} \left(\sum_{j=1}^k \log \frac{k+1}{j} \right)^2}$$

which can be approximated by $\frac{1}{k} \sum_{j=1}^k \left(\log \frac{k+1}{j} - \frac{1}{k} \sum_{i=1}^k \log \frac{k+1}{i} \right) \log UH_{j,n}$ using the fact that $\frac{1}{k} \sum_{j=1}^k \log^2 \frac{k+1}{j} - \left(\frac{1}{k} \sum_{j=1}^k \log \frac{k+1}{j} \right)^2 \sim 1$, as $k \rightarrow \infty$ and $\frac{k}{n} \rightarrow 0$.

As in Csorgo and Viharos (1998) one can extend this idea to a class of weight estimators

$$\hat{\gamma}_{k,n}^{WZ} = \frac{\sum_{j=1}^k \left[\int_{(j-1)/k}^{j/k} J(s) ds \right] \log UH_{j,n}}{\sum_{j=1}^k \left[\int_{(j-1)/k}^{j/k} J(s) ds \right] \log(k/j)}$$

with a weight function J integrating to 0. More specifically the weight functions

$$J_\theta(s) = \frac{1+\theta}{\theta} \left(1 - (1+\theta)s^\theta \right), \quad s \in [0, 1]$$

with $\theta \geq 0$, which contains the Zipf weight function $-\log(s) - 1$ as a special case taking $\theta = 0$. This leads to a natural class of estimators $\hat{\gamma}_{\theta,k,n}^{WZ}$.

In case of the Hall model

$$\begin{cases} U(x) = M_1 x^\gamma (1 + M_2 x^\rho (1 + o(1))) & (x \rightarrow \infty) & \text{if } \gamma > 0 \\ U(x) = U(\infty) - M_1 x^\gamma (1 + M_2 x^\rho (1 + o(1))) & (x \rightarrow \infty) & \text{if } \gamma < 0 \end{cases} \quad (9)$$

with $M_1 > 0$ and $M_2 \in \mathbb{R}$, the following minimal AMSE values can be found with

$$b^\#(x) = \begin{cases} b(x) \left(1 + \frac{\rho}{\gamma(1-\rho)} \right) & \text{if } \gamma > 0 \\ \frac{a(x)}{U(x)} & \text{if } \rho < \gamma < 0 \\ b(x) \frac{(\rho+\gamma)(1-\gamma)}{\gamma(1-\gamma-\rho)} & \text{if } \gamma < \rho, \end{cases}$$

where $b(x) = M_2 \rho x^\rho$ and, in case $\gamma < 0$, $a/U(x) = \frac{M_1 |\gamma|}{U(\infty)} x^\gamma$:

$$AMSE(\hat{\gamma}_{k_{opt}}^H) = \begin{cases} \left(\frac{(-2\rho)^\rho b^\#(n)}{(1-\rho)(1+\gamma^2)^\rho} \right)^{\frac{2}{1-2\rho}} (-2\rho + 1) & \text{if } \gamma > 0 \\ \left(\frac{[-2\rho(1-2\gamma)]^\rho b^\#(n)}{(1-\rho)[(1-\gamma)(1+\gamma+2\gamma^2)]^\rho} \right)^{\frac{2}{1-2\rho}} (-2\rho + 1) & \text{if } \gamma < 0, \end{cases}$$

$$\begin{aligned}
AMSE(\hat{\gamma}_{k_{opt}}^Z) &= \begin{cases} \left(\frac{(-2\rho)^\rho b^\#(n)}{(1-\rho)^2 [2(1+\gamma+\gamma^2)]^\rho} \right)^{\frac{2}{1-2\rho}} (-2\rho+1) & \text{if } \gamma > 0 \\ \left(\frac{[-2\rho(1-2\gamma)]^\rho b^\#(n)}{(1-\rho)^2 [2(1+2\gamma+\gamma^2-2\gamma^3)]^\rho} \right)^{\frac{2}{1-2\rho}} (-2\rho+1) & \text{if } \gamma < 0, \end{cases} \\
AMSE(\hat{\gamma}_{k_{opt}}^M) &= \begin{cases} \left(\frac{(-2\rho)^\rho b^\#(n)}{(1-\rho)(1+\gamma^2)^\rho} \right)^{\frac{2}{1-2\rho}} (-2\rho+1) & \text{if } \gamma > 0 \\ \left(\frac{[-2\rho(1-3\gamma)(1-4\gamma)]^\rho b^\#(n)}{(1-\rho)[(1-\gamma)^2(1-2\gamma)(6\gamma^2-\gamma+1)]^\rho} \right)^{\frac{2}{1-2\rho}} (-2\rho+1) & \text{if } \rho < \gamma < 0, \\ \left(\frac{(1-2\gamma)[-2\rho(1-3\gamma)(1-4\gamma)]^\rho b^\#(n)}{(1-2\gamma-\rho)[(1-\gamma)^2(1-2\gamma)(6\gamma^2-\gamma+1)]^\rho} \right)^{\frac{2}{1-2\rho}} (-2\rho+1) & \text{if } \gamma < \rho, \end{cases} \\
AMSE(\hat{\gamma}_{k_{opt}}^{ML}) &= \left(\frac{(-2\rho)^\rho b^\#(n)(1+\gamma)(\gamma+\rho)}{(1+\gamma)^{2\rho}(\gamma(1-\rho)+\rho)(1+\gamma-\rho)} \right)^{\frac{2}{1-2\rho}} (-2\rho+1) \quad \text{if } \gamma > 0.
\end{aligned}$$

It is seen that the generalized Zipf estimator is to be preferred especially when $|\rho^\#|$ is smaller than one. For negative values of γ the generalized Zipf estimator is systematically best. With respect to the maximum likelihood estimator one remarks that its bias disappears completely when $\gamma + \rho = 0$. When $\gamma + \rho > 0$ the maximum likelihood estimator performs better than the generalized Zipf estimator, and vice versa when $\gamma + \rho < 0$. Of course this analysis assumes that the threshold $X_{n-k,n}$ is chosen optimally. Adaptive methods to achieve this goal (in an asymptotic sense) can be constructed in similar ways as in the case of Pareto type distributions.

Another interesting feature of the Zipf estimator is the smoothness of the realizations as a function of k , which alleviates the problem of choosing k to some extent.

4. Applications to tail and quantile estimation

The abovementioned estimators can of course be used in estimation of extreme tail probabilities and extreme quantiles following the general technique presented for instance in Dekkers *et al.* (1989):

$$\hat{U}(1/p) = \hat{Q}(1-p) = X_{n-k,k} + \hat{a}(n/k) \frac{(k/np)^{\hat{\gamma}_{k,n}} - 1}{\hat{\gamma}_{k,n}} \quad \text{for some } p \in (0, 1/n],$$

and

$$\hat{F}(x) = \frac{k}{n} \max \left\{ 0, \left(1 + \hat{\gamma}_{k,n} \frac{x - X_{n-k,n}}{\hat{a}(n/k)} \right)^{-1/\hat{\gamma}_{k,n}} \right\} \quad \text{for some } x \geq X_{n-k,n}$$

where $\hat{\gamma}_{k,n}$ can be taken from the estimators given above, and where, for instance,

$$\hat{a}(n/k) = \frac{X_{n-k,n} H_{k,n}}{1 - \hat{\gamma}_{k,n} I_{(-\infty, 0)}(\hat{\gamma}_{k,n})}.$$

Asymptotic results for such estimators can be developed in analogy with the results given for instance in De Haan and Rootzen (1993) and Ferreira *et al.* (1999) and Ferreira (2002).

Among others Ferreira *et al.* (1999) have shown that for $\gamma > 0$ the analysis of the asymptotic mean squared error in the case of these corresponds to the analysis given above concerning the estimation of γ .

References

- Balkema A. and de Haan L. (1974) Residual life at great age, *Ann. Probab.*, 2, 792–804.
- Beirlant J., Dierckx G., Goegebeur Y. and Matthys G. (1999) Tail index estimation and an exponential regression model, *Extremes*, 2, 177–200.
- Beirlant J., Dierckx G., Guillou A. and Stărică C. (2001) On exponential representations of log-spacings of extreme order statistics, *Technical report K.U. Leuven*.
- Beirlant J., Vynckier P. and Teugels J. (1996a) Excess functions and estimation of the extreme value index, *Bernoulli*, 2, 293–318.
- Beirlant J., Vynckier P. and Teugels J. (1996b) Tail index estimation, pareto quantile plots and regression diagnostics, *J. Amer. Statist. Assoc.*, 91, 1659–1667.
- Castillo E. and Hadi A. (1997) Fitting the generalized pareto distribution to data, *J. Amer. Statist. Assoc.*, 92, 1609–1620.
- Coles S. and Powell E. (1996) Bayesian methods in extreme value modelling: a review and new developments, *International Statist. Review*, 64, 119–136.
- Csorgo S. and Viharos L. (1998) Estimating the tail index, *Asymptotic Methods in Probability and Statistics*, North Holland, (B. Szyszkowicz, Ed.), 833–881.
- Danielsson J., de Haan L., Peng L. and de Vries C. (2001) Using a bootstrap based method to choose the sample fraction in tail index estimation, *J. Multivariate Analysis*, 76, 226–248.
- Davison A. and Smith R. (1990) Models for exceedances over high thresholds, *J. Roy. Statist. Soc. B*, 52, 393–442.
- De Haan L. and Rootzen H. (1993) On the estimation of high quantiles, *J. Statist. Plann. Inf.*, 35, 1–13.
- Dekkers A. and de Haan L. (1989) On the estimation of the extreme-value index and large quantile estimation, *Ann. Statist.*, 17, 1795–1832.
- Dekkers A., Einmahl J. and de Haan L. (1989) A moment estimator for the index of an extreme-value distribution, *Ann. Statist.*, 17, 1833–1855.
- Drees H. (1995) Refined pickands estimators of the extreme value index, *Ann. Statist.*, 23, 2059–2080.
- Drees H. (1996) Refined pickands estimators with bias correction, *Comm. Statist. Theory Methods*, 25, 837–851.
- Drees H. (1998) On smooth statistical tail functional, *Scand. J. Statist.*, 25, 187–210.
- Drees H., de Haan L. and Resnick S. (2000) How to make a hill plot, *Ann. Statist.*, 28, 254–274.
- Drees H. and Kaufmann E. (1998) Selecting the optimal sample fraction in univariate extreme value estimation, *Stoch. Proc. Applications*, 75, 149–172.
- Ferreira A. (2002) Optimal asymptotic estimation of small exceedance probabilities, *J. Statist. Plann. Inf.*, to appear.
- Ferreira A., de Haan L. and Peng L. (1999) Adaptive estimators for the endpoint and high quantiles of a probability distribution, *Eurandom Technical Report*.
- Feuerverger A. and Hall P. (1999) Estimating a tail exponent by modelling departure from a pareto distribution, *Ann. Statist.*, 27, 760–781.

- Hill B. (1975) A simple general approach to inference about the tail of a distribution, *Ann. Statist.*, 3, 1163–1174.
- Hosking J., Wallis J. and Wood E. (1985) Estimation of the generalized extreme-value distribution by the method of probability-weighted moments, *Technometrics*, 27, 251–261.
- Kratz M. and Resnick S. (1996) The qq-estimator and heavy tails, *Commun. Statist. Stochastic Models*, 12, 699–724.
- Pickands J. (1975) Statistical inference using extreme order statistics, *Ann. Statist.*, 3, 119–131.
- Resnick S. and Stărică C. (1997) Smoothing the hill estimator, *Adv. Appl. Probab.*, 29, 271–293.
- Schultze J. and Steinebach J. (1996) On least squares estimates of an exponential tail coefficient, *Statist. Decisions*, 14, 353–372.
- Segers J. (2001) *Extremes of a Random Sample: Limit Theorems and Statistical Applications*, PhD K.U. Leuven.
- Smith R. (1987) Estimating tails of probability distributions, *Ann. Statist.*, 15, 1174–1207.