Statistica Sinica **13**(2003), 853-880

ESTIMATING THE EXTREME VALUE INDEX AND HIGH QUANTILES WITH EXPONENTIAL REGRESSION MODELS

G. Matthys and J. Beirlant

Katholieke Universiteit Leuven

Abstract: In this paper we present exponential regression models for spacings, or ordered excesses over a given threshold, and for log-ratios of such spacings under maximum domain of attraction conditions. From these we derive estimators for the extreme value index (EVI) and for high quantiles, which share many attractive properties of the maximum likelihood estimators from the peaks-over-thresholds method, but offer the extra advantage of being generally applicable without restriction on the value of the EVI. Further, the exponential regression models can be refined with parameters of second order regular variation, which reduces the bias of the resulting estimators. The refined models also give rise to insightful and practical techniques to select the threshold in the estimation of the EVI and of high quantiles. We demonstrate asymptotic normality of the newly proposed estimators and compare their small sample behaviour to some classical methods in a simulation study.

Key words and phrases: Bias correction, peaks-over-thresholds method, second order regular variation, threshold selection.

1. Introduction

The second half of the 20th century was characterised by a boom in economic activity, which led to an unprecedented increase in the total wealth of the world population and, consequently, the risks it is exposed to. Many of these hazards stem from political, economic, natural or accidental events, the precise occurrence of which is hard to predict. In order to prevent or prepare for their potential adverse effects a very diverse and evergrowing range of insurance instruments is being developed and, especially in recent years, risk management has entered all aspects of economic processes.

An important topic in managing risk is the analysis, modelling and prediction of rare but dangerous extreme events, so-called ‘worst-case events’. Indeed, the most dramatic impacts on a system are typically inflicted under extraordinary circumstances, when common experience and safety measures break down and the logic of avalanching causes and devastating effects take over. This may lead to the total failure of the system, as with some stock market crashes and spectacular bankruptcies in the recent past.

Clearly, such unfavourable scenarios need analysis, preferably before they actually become reality, but by their very nature historical data on extreme events tend to be scarce. Since the mid-seventies, however, some innovative techniques based on stochastic extreme value theory (EVT) have been devised to describe and predict extreme events more or less accurately while using only a limited amount of data.

Put in statistical terms, the core problem in risk management is how to model the tails and estimate extreme quantiles of the distribution of the process at risk. Let us denote with $X_1, \ldots, X_n, \ldots$ data on such a process (e.g., daily loss returns on a particular stock, or measurements of the wind speed at a particular spot) and suppose that these data are independent and identically distributed (i.i.d., quite a crude simplification for stock returns) according to some probability distribution F. For the worst-case scenario analysis we are then interested in levels x_p that will only be exceeded with a probability $p \in (0,1)$ close to 0, i.e., $F(x_p) = 1-p$, close to 1. Defining the quantile function Q as the generalised inverse of F, $Q(r) := \inf\{x : F(x) \geq r\}$, one sees that the required levels x_p correspond to the quantiles $Q(1-p)$ with small exceedance probability p. Estimating such high quantiles is directly linked to the accurate modelling of the tail of the distribution $\bar{F}(x) := 1 - F(x) = P(X_i > x)$ for large thresholds x.

It is well-known from EVT that one specific parameter, namely the extreme value index (EVI), dominates the tail behaviour of a distribution. This real-valued parameter indicates the heaviness of the tail, i.e., how extreme and frequent extreme events can be under the given probability distribution. There is a substantial number of publications on estimators for this EVI and we are still learning and understanding more about it. The different estimators are all inspired by various (equivalent) conditions that assure convergence of the distribution of the sample maximum $X_{n,n} = \max\{X_1, \ldots, X_n\}$ to a limiting distribution of the extreme value type. It makes sense to require this convergence, otherwise no hope exists that there is something meaningful to say about extreme quantiles at the border of, or even beyond, the sample range.

However a good quantile estimator calls for a more than good EVI estimator since some renowned EVI estimators produce poor quantile estimators. Though it is true that the EVI determines the asymptotic behaviour of the quantiles and tails of a distribution, it should be stressed that additional parameters (e.g., of scale and location) are no less important for accurate quantile estimation. The quality of the latter depend to a large extent on the model the EVI estimator is derived from, and on the corresponding estimators for the other parameters in this model.

For the heavy-tailed case, i.e., for distributions with EVI greater than 0, Feuerverger and Hall (1999) and Beirlant, Dierckx, Goegebeur and Matthys

(1999) proposed an exponential regression model for log-spacings of order statistics based on the theory of slow variation with remainder. This model proved succesful in EVI estimation and optimal sample fraction selection (see Matthys and Beirlant (2000a)) as well as in extreme quantile estimation (see Matthys and Beirlant (2000b)).

The present paper extends the slow variation with remainder approach to the general case of a real-valued EVI. Section 2 gives a brief overview of the classical estimation methods for extreme quantiles. In Section 3 we present two non-linear exponential regression models for spacings and log-ratios of spacings of order statistics, respectively. These allow us to construct maximum likelihood estimators for the real-valued EVI and for extreme quantiles. Technical details and asymptotic properties are deferred to the Appendix. Section 4 then refines the regression models by including parameters of slow variation with remainder. This leads to bias-corrected estimators and techniques for optimal threshold selection.

2. Extreme Value Methods for High Quantile Estimation

In this section we review some classical methods to construct estimators for the tail $\bar{F}(x)$ of a continuous distribution F, and for high quantiles $Q(1-p)$ with $p \in (0,1)$ close to 0.

We suppose that a sequence $X_1, \ldots, X_n, \ldots$ of i.i.d. observations from F is given and we require that the properly centred and normed sample maxima $X_{n,n} = \max\{X_1, \ldots, X_n\}$ converge in distribution to a non-degenerate limit. Gnedenko (1943) showed that this limit distribution is necessarily of extreme value type, i.e., for some $\gamma \in I\!R$ there exist sequences of constants $a_n > 0$, $b_n \in I\!R$ such that

$$\lim_{n\to\infty} P\left(\frac{X_{n,n} - b_n}{a_n} \le x\right) \to H_\gamma(x) \tag{1}$$

for all continuity points of the extreme value distribution $H_\gamma(x)$, defined as

$$H_\gamma(x) = \begin{cases} \exp(-(1+\gamma x)^{-1/\gamma}), & \text{for} \quad \gamma \neq 0,\ 1+\gamma x > 0, \\ \exp(-e^{-x}), & \text{for} \quad \gamma = 0. \end{cases}$$

The distribution function F is then said to belong to the maximum domain of attraction of H_γ, denoted as $F \in MDA(H_\gamma)$. Most common continuous distribution functions satisfy this weak condition, which arises quite naturally when studying the behaviour of extreme quantiles.

The real-valued parameter γ is referred to as the extreme value index (EVI) of F. Distributions with $\gamma > 0$ are called heavy-tailed, as their tail $\bar{F}$ typically decays slowly as a power function. Examples in this Fréchet class are the Pareto, Burr, Student's t, α-stable ($\alpha < 2$) and loggamma distributions. The

Gumbel class of distributions with $\gamma = 0$ encompasses the exponential, normal, lognormal, gamma and classical Weibull distributions, the tail of which diminishes exponentially fast. Finally, the Weibull class consists of distributions with $\gamma < 0$, which all have a finite right endpoint $x_+ := Q(1)$. Examples in this class are the uniform, beta, reversed Pareto and reversed Burr distributions.

Writing

$$P\left(\frac{X_{n,n} - b_n}{a_n} \le x\right) = (1 - \bar{F}(a_n x + b_n))^n$$

one sees that (1) is equivalent to

$$\lim_{n\to\infty} n\bar{F}(a_n x + b_n) \to -\log H_\gamma(x) = (1 + \gamma x)^{-1/\gamma}$$

(to be read as e^{-x} for $\gamma = 0$) for all continuity points of H_γ. This suggests to approximate the tail $\bar{F}$ for large thresholds y $(= a_n x + b_n)$ by

$$\bar{F}(y) \approx \frac{1}{n}\left(1 + \gamma \frac{y - b_n}{a_n}\right)^{-1/\gamma}, \tag{2}$$

which, after inversion, yields the following approximation for high quantiles with exceedance probability p close to 0:

$$x_p = Q(1 - p) \approx a_n \frac{(pn)^{-\gamma} - 1}{\gamma} + b_n.$$

Estimators for x_p will typically be of this form, with appropriate estimators for the EVI γ and the normalising constants a_n and b_n.

Weissman (1978) discussed the estimation of extreme quantiles for the three extreme value classes separately, assuming *prior* knowledge on the EVI γ. In practice, however, this knowledge is often not available as it is mostly not immediately obvious to which class the underlying distribution of a random phenomenon belongs. Therefore, extreme value methods that treat all EVI classes at once and on equal basis are particularly useful. Although the number of publications on EVI estimation is growing rapidly, the literature on fully elaborate tail and high quantile modelling is still rather restricted. In this section we outline some classical methods, starting with the generalised Pareto model for excesses over a high threshold.

2.1. The POT method

The peaks-over-thresholds (POT) method embeds approximation (2) in a more precise theoretical foundation. Denote with $F_u(x) := P(X - u \le x | X > u)$ the distribution of the excess of X over u, given that u is exceeded, and with

$G_{\gamma,\mu,\sigma}$ the generalised Pareto distribution (GPD) defined by

$$\bar{G}_{\gamma,\mu,\sigma}(x) = \begin{cases} \left(1+\gamma\frac{x-\mu}{\sigma}\right)^{-1/\gamma}, & \text{for} \quad \gamma \neq 0, \\ \exp\left(-\frac{x-\mu}{\sigma}\right), & \text{for} \quad \gamma = 0. \end{cases} \tag{3}$$

Pickands' (1975) result on the limiting distribution of excesses over a high threshold then states that (1) holds if and only if

$$\lim_{u\to x_+} \sup_{0<x<x_+-u} |F_u(x) - G_{\gamma,0,\sigma(u)}(x)| = 0 \tag{4}$$

for some positive scaling function $\sigma(u)$ depending on u.

Thus, if one fixes a high threshold u and selects from a sample $X_1, \ldots, X_n$ only those observations $X_{i_1}, \ldots, X_{i_{N_u}}$ that exceed u, a GPD with parameters γ, $\mu = 0$ and $\sigma = \sigma(u)$ is likely to be a good approximation for the distribution F_u of the N_u excesses $Y_j := X_{i_j} - u$.

Smith (1987) describes how a GPD can be fitted to the excesses $Y_1, \ldots, Y_{N_u}$ with maximum likelihood (ML) techniques and shows that the resulting estimators for γ and σ are asymptotically normal if $\gamma > -1/2$. The asymptotic variance of the POT ML estimator $\hat{\gamma}_u^{MLP}$ for the EVI is then $(1+\gamma)^2/N_u$. For $\gamma \in (-1, -1/2)$ the ML estimator converges with rate of consistency $n^{-\gamma}$ to a non-normal limit distribution.

Hosking and Wallis (1987) derive a simple method of moments to estimate γ and σ, but this only works if $\gamma < 1/2$. They also apply a variant with probability weighted moments (PWM) and find that the corresponding EVI estimator is a good alternative to the ML estimator for $\gamma < 1$. Castillo and Hadi (1997) propose an elemental percentile method (EPM) that does not impose any restrictions on the EVI γ, whereas Coles and Powell (1996) use Bayesian methods.

After obtaining estimates $\hat{\gamma}_u$ and $\hat{\sigma}_u$ for γ and σ by one of the above methods the conditional tail $\bar{F}_u$ of F can be estimated by

$$\widehat{\bar{F}_u}(x) = \left(1 + \hat{\gamma}_u \frac{x}{\hat{\sigma}_u}\right)^{-1/\hat{\gamma}_u}, \quad 0 < x < x_+ - u,$$

and the (unconditional) tail $\bar{F}(x) = \bar{F}(u) \cdot \bar{F}_u(x-u)$ by

$$\widehat{\bar{F}}(x) = \frac{N_u}{n}\left(1 + \hat{\gamma}_u \frac{x-u}{\hat{\sigma}_u}\right)^{-1/\hat{\gamma}_u}, \quad u < x < x_+, \tag{5}$$

where $\bar{F}(u)$ is estimated with the empirical exceedance probability N_u/n.

Inverting (5) then yields the following POT estimator for high quantiles above the threshold u:

$$\widehat{Q}_u(1-p) = u + \hat{\sigma}_u \frac{\left(\frac{N_u}{np}\right)^{\hat{\gamma}_u} - 1}{\hat{\gamma}_u} \quad \text{for} \quad p < \frac{N_u}{n}.$$

Often u is chosen equal to one of the order statistics $X_{1,n} \leq \ldots \leq X_{n,n}$, which are the data points $X_1, \ldots, X_n$ rearranged in ascending order. Taking $u = X_{n-k,n}$, the $(k+1)$th largest observation, gives $N_u = k$ and defines the quantile estimator

$$\hat{x}_{p,k+1}^{POT} := X_{n-k,n} + \hat{\sigma}_{k+1} \frac{\left(\frac{k}{np}\right)^{\hat{\gamma}_{k+1}} - 1}{\hat{\gamma}_{k+1}} \quad \text{for} \quad p < \frac{k}{n},$$

with self-evident indexing by $k+1$ instead of u.

Concerning the three POT estimation methods (EPM, ML and PWM), it is our personal experience from extensive simulations that the ML method mostly provides the best estimators if γ is estimated to be positive, whereas the EPM is to be preferred if γ is estimated to be less than 0.

Up to now no adaptive procedure to select the optimal threshold $u = X_{n-k,n}$ for any of the POT EVI and high quantile estimators has been described.

2.2. The Pickands estimator

Pickands (1975) defined the following well-known estimator for a general EVI $\gamma \in \mathbb{R}$:

$$\hat{\gamma}_{4k}^{Pi} := \frac{1}{\log 2} \cdot \log \frac{X_{n-k+1,n} - X_{n-2k+1,n}}{X_{n-2k+1,n} - X_{n-4k+1,n}} \quad \text{for} \quad k \leq n/4.$$

The associated quantile estimator for $Q(1-p)$ is

$$\hat{x}_{p,4k}^{Pi} := X_{n-k+1,n} + \frac{\left(\frac{k}{(n+1)p}\right)^{\hat{\gamma}_{4k}^{Pi}} - 1}{1 - 2^{-\hat{\gamma}_{4k}^{Pi}}} \cdot (X_{n-k+1,n} - X_{n-2k+1,n}).$$

The asymptotic properties of $\hat{\gamma}_k^{Pi}$ and $\hat{x}_{p,k}^{Pi}$ are discussed by Dekkers and de Haan (1989).

Usually, the Pickands estimators $\hat{\gamma}_k^{Pi}$ and $\hat{x}_{p,k}^{Pi}$ depend heavily on the number k of order statistics used, and their path as a function of k is quite jagged. Therefore the estimators are rather unworkable in practice for small and moderate sample sizes. Drees (1996) introduces refined Pickands estimators that suffer less from instability. He also derives the sample fractions k that are theoretically optimal for this refined estimation of the EVI and of high quantiles. On refined Pickands estimators see also Segers (2001).

2.3. The moment estimator

Another celebrated estimator for the EVI of a distribution was introduced by Dekkers, Einmahl and de Haan (1989). It is defined for $k \in \{2, \ldots, n-1\}$ by

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

where $M_{k+1}^{(l)} = (1/k)\sum_{i=1}^{k}(\log X_{n-i+1,n} - \log X_{n-k,n})^l$, $l = 1,2$. The moment estimator generalises the estimator $M_{k+1}^{(1)}$, proposed by Hill (1975) for the case $\gamma > 0$, to all EVI classes.

An estimator for high quantiles on the basis of the moment estimator is

$$\hat{x}_{p,k+1}^{M} := X_{n-k,n} + \hat{a}_{n,k+1}^{M} \frac{\left(\frac{k}{np}\right)^{\hat{\gamma}_{k+1}^{M}} - 1}{\hat{\gamma}_{k+1}^{M}} \quad \text{for} \quad k < n,$$

with the choices

$$\hat{a}_{n,k+1}^{M} = \frac{X_{n-k,n} M_{k+1}^{(1)}}{\rho_1(\hat{\gamma}_{k+1}^{M})},$$

$$\rho_1(\gamma) = \begin{cases} 1 & \text{for} \quad \gamma \geq 0, \\ \frac{1}{1-\gamma} & \text{for} \quad \gamma < 0. \end{cases}$$

The asymptotic normality of this quantile estimator under various conditions on the tail of the distribution and on the limiting order of $p = p_n$ for $n \to \infty$ is proved by Dekkers, Einmahl and de Haan (1989), and by de Haan and Rootzén (1993). Indeed, for high quantiles one is typically interested in the case $n \to \infty$ and $p = p_n \to 0$, but the asymptotic properties may differ as $np_n \to \infty$, $np_n \to c \in (0, \infty)$, or $np_n \to 0$. In practice, when dealing with finite samples, the distinction between these three cases is less clear-cut, of course. Ferreira, de Haan and Peng (2003) propose an adaptive bootstrap procedure to estimate the number of order statistics k that is asymptotically optimal for the quantile estimator $\hat{x}_{p,k}^{M}$.

3. Exponential Regression Models

In this section two non-linear exponential regression methods are presented that yield simple ML estimators for the EVI and for high quantiles. We make frequent use of the following equalities (in distribution, denoted by $\stackrel{d}{=}$), where $U_{j,n}$ $(1 \leq j \leq n)$ and $V_{j,k}$ $(1 \leq j \leq k)$ are order statistics from i.i.d. uniform(0,1) samples of size n, respectively size k, and $E_{j,k}$ $(1 \leq j \leq k)$ are order statistics from an i.i.d. standard exponential sample of size k:
with $U(r) := Q(1 - 1/r) = x_{1/r}$,

$$\begin{aligned} X_{n-j+1,n} &\stackrel{d}{=} U\left(U_{j,n}^{-1}\right) && \text{for} \quad j \leq n, \\ \frac{U_{j,n}}{U_{k+1,n}} &\stackrel{d}{=} V_{j,k} && \text{for} \quad j \leq k < n, \quad \text{and} \\ -\log(V_{j,k}) &\stackrel{d}{=} E_{k-j+1,k} && \text{for} \quad j \leq k. \end{aligned}$$

The $V_{j,k}$ in the second equality are independent of $U_{k+1,n}$. A key result in the argumentation is the Rényi representation of standard exponential order statistics, which states that

$$E_{k-j+1,k} \stackrel{d}{=} \sum_{i=j}^{k} \frac{f_{k-i+1}}{i} \quad \text{for} \quad j \leq k, \tag{6}$$

where the f_i are i.i.d. standard exponential random variables.

3.1. A maximum likelihood estimator for the EVI

Next to (1) and (4) an equivalent condition can be formulated in terms of the tail quantile function U. de Haan (1970) states that $F \in MDA(H_\gamma)$ if and only if there exists a positive measurable function a_U such that $\forall t > 0$

$$\lim_{r\to\infty} \frac{U(tr) - U(r)}{a_U(r)} = \begin{cases} \frac{t^\gamma - 1}{\gamma}, & \text{for} \quad \gamma \neq 0, \\ \log t, & \text{for} \quad \gamma = 0. \end{cases} \tag{7}$$

The norming function $a_U(r)$ is then equivalent to $\sigma(U(r))$ for $r \to \infty$, with $\sigma(.)$ as in (4). Note that for a GPD (3) we have $\sigma(u) = \sigma + \gamma(u - \mu)$, $U(r) = \mu + \sigma(r^\gamma - 1)/\gamma$, and the equality in (7) holds for all $t > 0$ and for all $r > 0$ (i.e., not only in the limit for $r \to \infty$) with $a_U(r) = \sigma r^\gamma = \sigma(U(r))$.

For a fixed $k < n$ and for $1 \leq j \leq k$, condition (7) on the tail quantile function inspires the following approximation :

$$\begin{aligned} X_{n-j+1,n} - X_{n-k,n} &\stackrel{d}{=} U(U_{j,n}^{-1}) - U(U_{k+1,n}^{-1}) \\ &\stackrel{d}{=} U(V_{j,k}^{-1} U_{k+1,n}^{-1}) - U(U_{k+1,n}^{-1}) \\ &\stackrel{d}{\approx} a_{n,k+1} \frac{V_{j,k}^{-\gamma} - 1}{\gamma}, \end{aligned} \tag{A0}$$

where $a_{n,k+1}$ stands for $a_U(U_{k+1,n}^{-1})$.

As $a_{n,k+1} = a_U(U_{k+1,n}^{-1}) \sim \sigma(X_{n-k,n})$, (A0) is in fact the counterpart, expressed in terms of the order statistics, for the GPD tail approximation (4) with $u = X_{n-k,n}$.

For the log-ratio of spacings of order statistics we then obtain

$$\log \frac{X_{n-j+1,n} - X_{n-k,n}}{X_{n-j,n} - X_{n-k,n}} \stackrel{d}{\approx} \log \frac{V_{j,k}^{-\gamma} - 1}{V_{j+1,k}^{-\gamma} - 1} \quad \text{for} \quad 1 \leq j < k.$$

The Mean Value Theorem applied to the right-hand side gives, with $E_{j,k}^* \in (E_{k-j,k}, E_{k-j+1,k})$ and $V_{j,k}^* = \exp(-E_{j,k}^*)$,

$$\log \frac{V_{j,k}^{-\gamma} - 1}{V_{j+1,k}^{-\gamma} - 1} \stackrel{d}{=} \log(e^{\gamma E_{k-j+1,k}} - 1) - \log(e^{\gamma E_{k-j,k}} - 1)$$

$$\stackrel{d}{=} (E_{k-j+1,k} - E_{k-j,k}) \cdot \frac{\gamma e^{\gamma E^*_{j,k}}}{e^{\gamma E^*_{j,k}} - 1} \stackrel{d}{=} \frac{f_{k-j+1}}{j} \cdot \frac{\gamma}{1 - (V^*_{j,k})^\gamma}.$$

The last equality follows from (6). Estimating $V^*_{j,k}$ by $j/(k+1)$ we find the following non-linear exponential regression model for log-ratios of spacings:

$$j \log \frac{X_{n-j+1,n} - X_{n-k,n}}{X_{n-j,n} - X_{n-k,n}} \stackrel{d}{\approx} \frac{\gamma}{1 - (\frac{j}{k+1})^\gamma} f_{k-j+1} \quad \text{for} \quad 1 \le j < k, \tag{A1}$$

with f_j $(1 \le j \le k)$ i.i.d. standard exponential random variables.

For a fixed value $k < n$, a ML estimator $\hat{\gamma}^A_{k+1}$ for the EVI can easily be implemented on the basis of (A1) by calculating the scaled log-ratios of spacings $Y_j := j \log[(X_{n-j+1,n} - X_{n-k,n})/(X_{n-j,n} - X_{n-k,n})]$ for $1 \le j < k$ and maximising numerically the loglikelihood

$$\mathcal{L}^{A1}_{(\gamma)}(\mathbf{Y}) := \sum_{j=1}^{k-1} \left\{ \log\left(\frac{1 - (\frac{j}{k+1})^\gamma}{\gamma} \right) - \frac{1 - (\frac{j}{k+1})^\gamma}{\gamma} Y_j \right\}.$$

The resulting EVI estimator $\hat{\gamma}^A$ is, by construction, invariant under shifts and rescaling of the data. Also, under a few technical conditions on the tail of the distribution, $\hat{\gamma}^A_{k+1}$ is consistent and asymptotically normally distributed. A precise statement and proof of this property are given in Appendix A.1.

As for the POT ML estimator and the moment estimator, the asymptotic variance of $\sqrt{k}(\hat{\gamma}^A_k - \gamma)$ is 1 for $\gamma = 0$. For other γ-values it can be calculated numerically. In Figure 1 we compare the asymptotic variance of $\hat{\gamma}^A$ with the one of the moment estimator $\hat{\gamma}^M$ and the POT ML estimator $\hat{\gamma}^{MLP}$ for the range $-3 \le \gamma \le 3$ ($-1/2 < \gamma \le 3$ for the POT ML estimator). Note that for $\gamma > 0$ the asymptotic variance of $\hat{\gamma}^A$ almost equals the one of the POT ML estimator. For negative γ-values the asymptotic variance of $\hat{\gamma}^A$ is substantially lower than for the moment estimator.

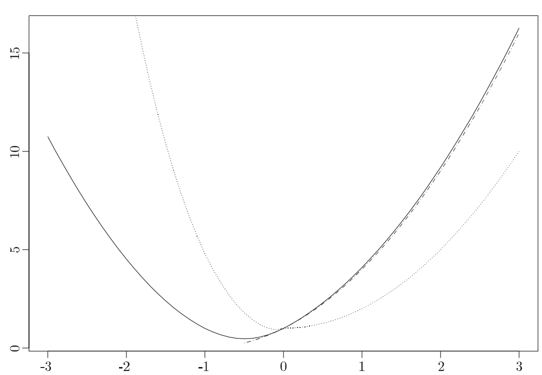

Figure 1. Asymptotic variances of $\sqrt{k}(\hat{\gamma}_k - \gamma)$ for $\hat{\gamma}^A$ (solid line), the moment estimator $\hat{\gamma}^M$ (dots) and the POT ML estimator $\hat{\gamma}^{MLP}$ (dots-dashes).

In Appendix A.1 we give an integral expression for the asymptotic bias of $\hat{\gamma}_k^A$ which can be calculated numerically and which is again strikingly similar to the asymptotic bias of the POT ML estimator for $\gamma > 0$. This bias is caused by deviations of the spacings from model (A0) when the threshold $X_{n-k,n}$ is not far enough in the tail for the asymptotic limit in (7) to hold approximately. For examples such as the loggamma and the lognormal distribution the convergence rate to the limit is rather slow, which results in an important bias even at the lower k-values.

As a rule, it is always necessary for any EVI estimator to make a trade-off between variance, which diminishes when more order statistics are used in the estimation, and bias, which increases when the threshold is put at a lower value. The point where the mean squared error (MSE, equal to bias squared plus variance) of the estimator is minimal can therefore be considered as an optimal choice for the threshold.

As an illustration Figure 2 shows the medians and empirical MSE's of the estimator $\hat{\gamma}^A$, together with the POT ML estimator $\hat{\gamma}^{MLP}$ and the moment estimator $\hat{\gamma}^M$, applied to simulated data from a Burr(1,0.5,2) and a reversed Burr(1,0.5,3) distribution.

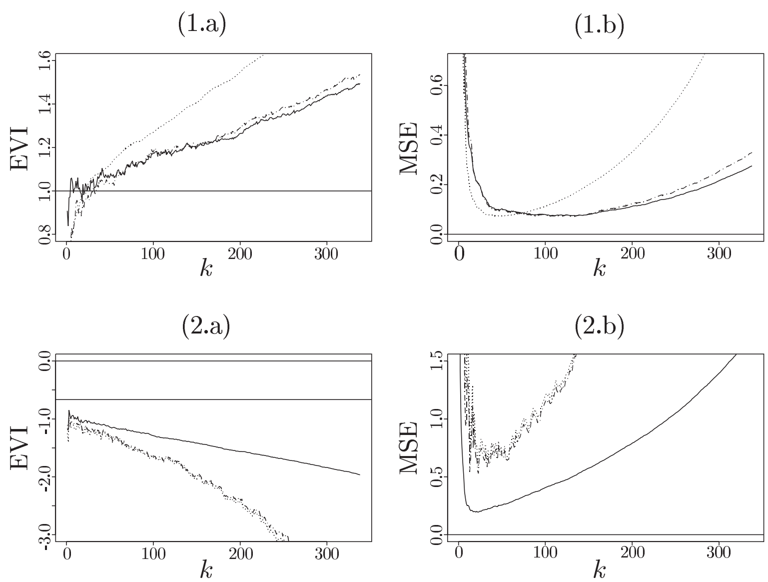

Figure 2. (a) Medians and (b) empirical MSE's of $\hat{\gamma}_k^A$ (solid line), $\hat{\gamma}_k^{MLP}$ (dots-dashes) and $\hat{\gamma}_k^M$ (dots), $k = 3, \ldots, 340$, for 100 simulated samples of size $n = 500$ with (1) a Burr(1,0.5,2) distribution ($\gamma = 1$), and (2) a reversed Burr(1,0.5,3) distribution ($\gamma = -2/3$). Horizontal lines indicate the true value of γ.

In Section 4.1 we discuss the origin of the bias problem in more detail and try to remedy it by introducing the concept of slow variation with remainder in the exponential regression model. Alternatively, this approach also provides an adaptive estimation procedure to select the optimal threshold for $\hat{\gamma}_k^A$ in cases where its bias is not excessive (Section 4.2).

3.2. Estimating high quantiles

In this section we present a second exponential regression model, which allows a simple estimator for the scaling factor $a_{n,k+1} = a_U(U_{k+1,n}^{-1})$ that is needed in the subsequent construction of an estimator for high quantiles. The derivation is similar to the one for model (A1), but instead of considering log-ratios of spacings we now employ (A0) to approximate the spacings of order statistics themselves :

$$X_{n-j+1,n} - X_{n-j,n} \overset{d}{\approx} a_{n,k+1} \frac{V_{j,k}^{-\gamma} - V_{j+1,k}^{-\gamma}}{\gamma} \quad \text{for} \quad 1 \leq j < k.$$

With the same notation for $E_{j,k}^*$ and $V_{j,k}^*$ as above, an application of the Mean Value Theorem to the second factor on the right gives

$$\begin{aligned}\frac{V_{j,k}^{-\gamma} - V_{j+1,k}^{-\gamma}}{\gamma} &\overset{d}{=} \frac{e^{\gamma E_{k-j+1,k}} - e^{\gamma E_{k-j,k}}}{\gamma} \\ &\overset{d}{=} (E_{k-j+1,k} - E_{k-j,k}) e^{\gamma E_{j,k}^*} \overset{d}{=} \frac{f_{k-j+1}}{j} \cdot (V_{j,k}^*)^{-\gamma}.\end{aligned}$$

Estimating $V_{j,k}^*$ by $j/(k+1)$ we arrive at the non-linear regression model with exponential responses

$$j(X_{n-j+1,n} - X_{n-j,n}) \overset{d}{\approx} a_{n,k+1} \left(\frac{j}{k+1}\right)^{-\gamma} f_{k-j+1} \quad \text{for} \quad 1 \leq j \leq k, \tag{A2}$$

with f_j $(1 \leq j \leq k)$ i.i.d. standard exponential random variables, as before.

Model (A2) makes it possible to calculate an estimator for $a_{n,k+1}$ by maximising the corresponding loglikelihood $\mathcal{L}^{A2}_{(a_{n,k+1},\gamma)}$ of the scaled spacings $Z_j := j\,(X_{n-j+1,n} - X_{n-j,n})$ for $1 \leq j \leq k$:

$$\hat{a}_{n,k+1} = \frac{1}{k} \sum_{j=1}^{k} Z_j \left(\frac{j}{k+1}\right)^{\gamma}. \tag{8}$$

Condition (7) now leads to the following approximation for high quantiles:

$$\begin{aligned}x_p - X_{n-k,n} &\overset{d}{=} U(p^{-1}) - U(U_{k+1,n}^{-1}) \\ &\overset{d}{=} U((U_{k+1,n}/p)U_{k+1,n}^{-1}) - U(U_{k+1,n}^{-1}) \\ &\overset{d}{\approx} \hat{a}_{n,k+1} \frac{(U_{k+1,n}/p)^{\gamma} - 1}{\gamma}.\end{aligned}$$

We estimate $U_{k+1,n}$ by its expected value $(k+1)/(n+1)$, γ by $\hat{\gamma}^A_{k+1}$, and $a_{n,k+1}$ by replacing $\hat{\gamma}^A_{k+1}$ for γ in formula (8):

$$\hat{a}^A_{n,k+1} := \frac{1}{k}\sum_{j=1}^{k} j(X_{n-j+1,n} - X_{n-j,n})\left(\frac{j}{k+1}\right)^{\hat{\gamma}^A_{k+1}}.$$

Doing so, we arrive at the quantile estimator

$$\hat{x}^A_{p,k+1} := X_{n-k,n} + \hat{a}^A_{n,k+1}\frac{\left(\frac{k+1}{p(n+1)}\right)^{\hat{\gamma}^A_{k+1}} - 1}{\hat{\gamma}^A_{k+1}} \quad \text{for} \quad k < n.$$

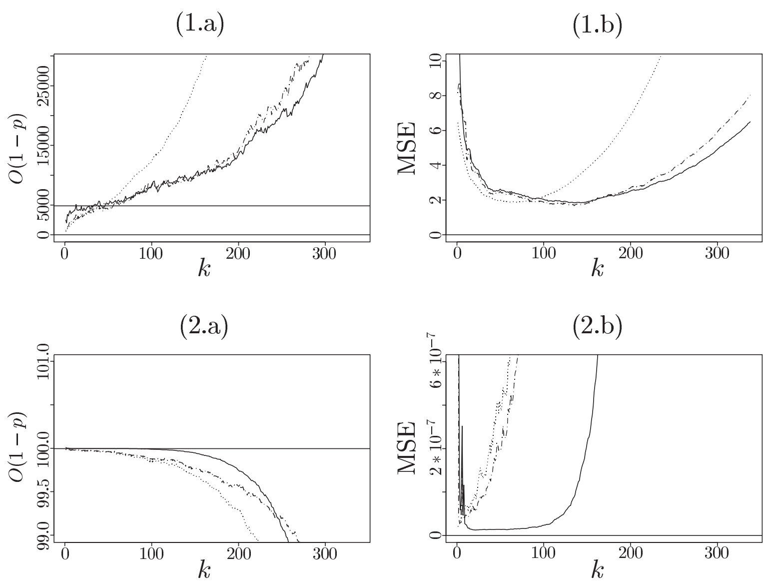

Figure 3. (a) Medians and (b) empirical MSE's of $\hat{x}^A_{p,k}$ (solid line), $\hat{x}^{MLP}_{p,k}$ (dots-dashes) and $\hat{x}^M_{p,k}$ (dots), with $p = 1/5000$, $k = 3,\ldots,340$, for 100 simulated samples of size $n = 500$ with (1) a Burr(1,0.5,2) distribution ($\gamma = 1$), and (2) a reversed Burr(1,0.5,3) distribution ($\gamma = -2/3$). Horizontal lines indicate the true value of x_p.

In Figure 3 the estimator $\hat{x}^A_p$ is compared to the POT ML estimator $\hat{x}^{MLP}_p$ and the moment estimator $\hat{x}^M_p$ for the simulated data from Figure 2, with $p = (1/10)(1/n) = 1/5000$. As the distributions of these estimators are heavily skewed to the right, we use a slightly adapted version of the normed MSE, with squared log-ratios instead of squared differences. If N is the number of simulation runs ($N = 100$ for this example) and $\hat{x}^{(i)}_{p,k}$ is the value of the quantile estimator

$\hat{x}_{p,k}$ in the i-th simulation, then the empirical MSE of $\hat{x}_{p,k}$ is calculated as

$$\widehat{MSE}_k = \frac{1}{N}\sum_{i=1}^{N}\left(\log\frac{\hat{x}_{p,k}^{(i)}}{x_p}\right)^2 \tag{9}$$

instead of the usual average $(1/N)\sum_{i=1}^{N}(\hat{x}_{p,k}^{(i)}/x_p - 1)^2$. This definition is used in all further MSE-plots for the quantile estimators.

From a systematic simulation study we may conclude the following concerning the finite sample behaviour of the new estimators $\hat{\gamma}^A$ and $\hat{x}_p^A$. First, as in Figures 2(1) and 3(1), the POT ML estimators and the new $\hat{\gamma}^A$ and $\hat{x}_p^A$ usually show a particularly similar behaviour if γ is estimated to be positive. This results from the fact that the exponential regression models (A1) and (A2) from which $\hat{\gamma}^A$ and $\hat{x}_p^A$ are derived are close approximations to the POT GPD model. However, by using log-ratios of spacings for the EVI estimator $\hat{\gamma}^A$, its regularity is now extended to the whole real line $\mathbb{R}$.

Further one observes that the biases of $\hat{\gamma}_{k+1}^A$ and $\hat{x}_{p,k+1}^A$ typically grow at a slower rate (with k) than these of the moment estimators, especially for distributions in the class $\gamma < 0$. Therefore, $\hat{\gamma}_{k+1}^A$ and $\hat{x}_{p,k+1}^A$ reach their optimal points of minimal MSE deeper in the sample, i.e., at lower threshold values $X_{n-k,n}$. Mostly, these minimal empirical MSE's of $\hat{\gamma}_{k+1}^A$ and $\hat{x}_{p,k+1}^A$ are comparable to, or even lower than, the minimal MSE's of the moment estimators.

4. Bias Correction and Threshold Selection

In this section we refine the exponential regression models (A1) and (A2) with concepts from EVT on slow variation with remainder. A positive function ℓ is slowly varying at infinity if

$$\lim_{r\to\infty}\frac{\ell(tr)}{\ell(r)} = 1 \quad \text{for all} \quad t > 0. \tag{10}$$

A function f is called regularly varying with index γ, denoted by $f \in R_\gamma$, if $f(x) = x^\gamma \ell(x)$ with ℓ slowly varying. R_0 thus symbolizes the class of slowly varying functions.

From EVT it is well-known that, if the tail quantile function U satisfies (7) with EVI γ, then the norming function a_U varies regularly at infinity with the same index γ. Moreover,

- if $\gamma > 0$: $U \in R_\gamma$, i.e., $U(r) = r^\gamma \ell(r)$ with $\ell \in R_0$, and $a_U(r)/U(r) \to \gamma$ for $r \to \infty$, such that $a_U/U \in R_0$. Thus for all $t, r > 0$,

$$\frac{U(tr)-U(r)}{a_U(r)} = \frac{U(r)}{a_U(r)}\left(\frac{U(tr)}{U(r)} - 1\right) = \frac{1}{d(r)}\left(t^\gamma\frac{\ell(tr)}{\ell(r)} - 1\right),$$

with $d(r) := a_U(r)/U(r) \to \gamma$ for $r \to \infty$, $d \in R_0$;

- if $\gamma < 0$: $x_+ < \infty$, $x_+ - U \in R_\gamma$, i.e., $x_+ - U(r) = r^\gamma \ell(r)$ with $\ell \in R_0$, and $a_U(r)/(x_+ - U(r)) \to -\gamma$ for $r \to \infty$, such that $a_U/(x_+ - U) \in R_0$. Thus for all $t, r > 0$,
$$\frac{U(tr) - U(r)}{a_U(r)} = -\frac{x_+ - U(r)}{a_U(r)} \left(\frac{x_+ - U(tr)}{x_+ - U(r)} - 1 \right) = \frac{1}{d(r)} \left(t^\gamma \frac{\ell(tr)}{\ell(r)} - 1 \right),$$
with $d(r) := -a_U(r)/(x_+ - U(r)) \to \gamma$ for $r \to \infty$, $-d \in R_0$;

- if $\gamma = 0$: $U \in R_0$, i.e., $U(r) = \ell(r)$ with $\ell \in R_0$, and $a_U(r)/U(r) \to 0$ for $r \to \infty$, with $a_U/U \in R_0$. Thus for all $t, r > 0$,
$$\frac{U(tr) - U(r)}{a_U(r)} = \frac{U(r)}{a_U(r)} \left(\frac{U(tr)}{U(r)} - 1 \right) = \frac{1}{d(r)} \left(\frac{\ell(tr)}{\ell(r)} - 1 \right),$$
with $d(r) := a_U(r)/U(r) \to 0$ for $r \to \infty$, $d \in R_0$.

Hence we obtain for all $\gamma \in I\!R$ that, if U satisfies (7), there exist a slowly varying function ℓ and a function d with $\pm d \in R_0$ and $d(r) \to \gamma$ for $r \to \infty$, such that for all $t, r > 0$,
$$\frac{U(tr) - U(r)}{a_U(r)} = \frac{1}{d(r)} \left(t^\gamma \frac{\ell(tr)}{\ell(r)} - 1 \right). \tag{11}$$

Conversely, for $\gamma \neq 0$ the limit result in (7) follows immediately from (11) by the definition (10) of slow variation. For $\gamma = 0$ the tail quantile function $U = \ell$ is not only required to belong to R_0, but also to the special subclass of so-called Π-varying functions (see Geluk and de Haan (1987)). These are characterised by the existence of a measurable function a, such that for all $t > 0$, $\lim_{r\to\infty}(\ell(tr) - \ell(r))/a(r) = \log t$, which is precisely condition (7) for $\gamma = 0$.

The bias of classical EVI estimators and of the newly introduced $\hat{\gamma}^A$ arises from taking the limit results (7) or (4) as equalities, while these actually only hold when the threshold $U(r)$ grows to x_+. When $U(r)$ moves further away from x_+, or when the ratio $\ell(tr)/\ell(r)$ in (11) tends to 1 at a slow rate, the approximation will be poorer, with a more or less biased estimate for the EVI as a consequence.

We will show, however, how the exponential regression models (A1) and (A2) can be refined on the basis of (11) by parametrising the convergence rate of $\ell(tr)/\ell(r)$ in order to obtain bias-corrected estimates for the EVI and for high quantiles. To this purpose consider the following condition on ℓ, which describes slow variation with remainder (see Section 3.12.1 of Bingham, Goldie and Teugels (1987)).

Assumption (R_ℓ)**.** There exists a real constant $\rho \leq 0$ and a rate function b satisfying $b(r) \to 0$ as $r \to \infty$, such that for all $t \geq 1$, as $r \to \infty$,
$$\log \frac{\ell(tr)}{\ell(r)} \sim \begin{cases} b(r) \dfrac{t^\rho - 1}{\rho}, & \text{for} \quad \rho \neq 0, \\ b(r) \log t, & \text{for} \quad \rho = 0. \end{cases}$$

Apart from a possible change of sign the function b is then regularly varying with index ρ.

The slowly varying parts ℓ in (11) of most common distributions in statistics satisfy this rather weak assumption. Note that for the Gumbel class of distributions with $\gamma = 0$, $U = \ell$ is Π-varying so that necessarily $\rho = 0$ and $b(r) \sim a_U(r)/U(r)$.

4.1. A bias-corrected estimator for the EVI

For spacings of order statistics with $1 \le j \le k < n$, (11) gives

$$\begin{aligned} X_{n-j+1,n} - X_{n-k,n} &\stackrel{d}{=} U(V_{j,k}^{-1} U_{k+1,n}^{-1}) - U(U_{k+1,n}^{-1}) \\ &\stackrel{d}{=} c_{n,k+1} \left[V_{j,k}^{-\gamma} \frac{\ell(V_{j,k}^{-1} U_{k+1,n}^{-1})}{\ell(U_{k+1,n}^{-1})} - 1 \right], \end{aligned}$$

where $c_{n,k+1}$ denotes $a_U(U_{k+1,n}^{-1})/d(U_{k+1,n}^{-1})$. With assumption (R_ℓ) we then obtain the approximation

$$X_{n-j+1,n} - X_{n-k,n} \stackrel{d}{\approx} c_{n,k+1} \left[V_{j,k}^{-\gamma} \exp\Big(b_{n,k+1} \frac{V_{j,k}^{-\rho} - 1}{\rho} \Big) - 1 \right], \tag{B0}$$

where $b_{n,k+1}$ stands for $b(U_{k+1,n}^{-1})$.

We now proceed in a similar way as for model (A1) with the log-ratio of spacings of order statistics:

$$\log \frac{X_{n-j+1,n} - X_{n-k,n}}{X_{n-j,n} - X_{n-k,n}} \stackrel{d}{\approx} \log \frac{V_{j,k}^{-\gamma} \exp\Big(b_{n,k+1} \frac{V_{j,k}^{-\rho} - 1}{\rho} \Big) - 1}{V_{j+1,k}^{-\gamma} \exp\Big(b_{n,k+1} \frac{V_{j+1,k}^{-\rho} - 1}{\rho} \Big) - 1} \quad \text{for} \quad 1 \le j < k.$$

For the right-hand side, the Mean Value Theorem, with the same notations for $E_{j,k}^*$ and $V_{j,k}^*$ as in Section 3, and (6) yield

$$\begin{aligned} &\log \frac{V_{j,k}^{-\gamma} \exp\Big(b_{n,k+1} \frac{V_{j,k}^{-\rho} - 1}{\rho} \Big) - 1}{V_{j+1,k}^{-\gamma} \exp\Big(b_{n,k+1} \frac{V_{j+1,k}^{-\rho} - 1}{\rho} \Big) - 1} \\ &\stackrel{d}{=} \log \frac{\exp\Big(\gamma E_{k-j+1,k} + b_{n,k+1} \frac{e^{\rho E_{k-j+1,k}} - 1}{\rho} \Big) - 1}{\exp\Big(\gamma E_{k-j,k} + b_{n,k+1} \frac{e^{\rho E_{k-j,k}} - 1}{\rho} \Big) - 1} \\ &\stackrel{d}{=} (E_{k-j+1,k} - E_{k-j,k}) \cdot \frac{\gamma + b_{n,k+1} e^{\rho E_{j,k}^*}}{1 - \exp\Big(-\gamma E_{j,k}^* + b_{n,k+1} \frac{e^{\rho E_{j,k}^*} - 1}{-\rho} \Big)} \end{aligned}$$

$$\stackrel{d}{=} \frac{f_{k-j+1}}{j} \cdot \frac{\gamma + b_{n,k+1}(V_{j,k}^*)^{-\rho}}{1-(V_{j,k}^*)^{\gamma}\exp\Big(b_{n,k+1}\frac{(V_{j,k}^*)^{-\rho}-1}{-\rho}\Big)}.$$

Hence the following non-linear exponential regression model for log-ratios of spacings with $1 \leq j < k$:

$$j \log \frac{X_{n-j+1,n} - X_{n-k,n}}{X_{n-j,n} - X_{n-k,n}} \stackrel{d}{\approx} \frac{\gamma + b_{n,k+1}\left(\frac{j}{k+1}\right)^{-\rho}}{1-\left(\frac{j}{k+1}\right)^{\gamma}\exp\Big(b_{n,k+1}\frac{(\frac{j}{k+1})^{-\rho}-1}{-\rho}\Big)} f_{k-j+1}, \qquad \text{(B1)}$$

where the f_j are i.i.d. standard exponential random variables. Note that (B1) simplifies to model (A1) if $b_{n,k+1} = 0$.

Model (B1) makes it possible to calculate joint estimates for the parameters γ, $b_{n,k+1}$ and ρ by maximising numerically the corresponding loglikelihood of the log-ratios of spacings for $1 \leq j < k < n$. We denote these ML estimators by $\hat{\gamma}_{k+1}^B$, $\hat{b}_{n,k+1}^B$ and $\hat{\rho}_{k+1}^B$, respectively. In most cases the new EVI estimator reduces the bias of $\hat{\gamma}^A$ to a large extent. As an example Figure 4 compares $\hat{\gamma}^B$ with $\hat{\gamma}^A$, the POT ML estimator and the moment estimator, for the same simulated data as in Figures 2 and 3.

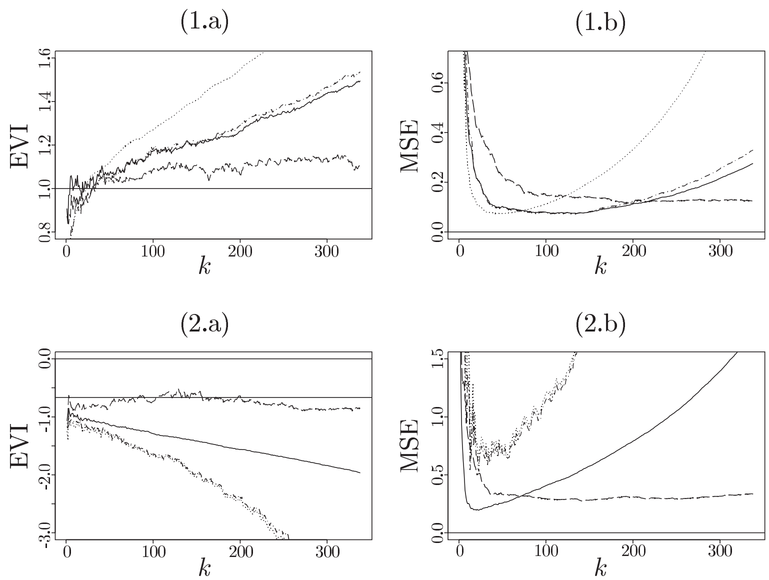

Figure 4. (a) Medians and (b) empirical MSE's of $\hat{\gamma}_k^A$ (solid line), $\hat{\gamma}_k^B$ (dashes), $\hat{\gamma}_k^{MLP}$ (dots-dashes) and $\hat{\gamma}_k^M$ (dots), $k = 3, \ldots, 340$, for 100 simulated samples of size $n = 500$ with (1) a Burr(1,0.5,2) distribution ($\gamma = 1$), and (2) a reversed Burr(1,0.5,3) distribution ($\gamma = -2/3$). Horizontal lines indicate the true value of γ.

We note that other exponential regression models can be derived, using properties of generalised second order regular variation as described by de Haan and Stadtmüller (1996) rather than the condition (R_ℓ) on slow variation with remainder. However, model (B1) and the alternative models all have equivalent linear expansions in $b_{n,k+1}$, and in simulations we found that the functional form of (B1) gave the best bias-reduced estimator for the EVI.

4.2. Selecting the threshold for $\hat{\gamma}^A$ and $\hat{x}_p^A$

The price to pay for the bias reduction of $\hat{\gamma}^B$ is a substantially higher variance than for the simple estimator $\hat{\gamma}^A$. Only in exceptional cases does $\hat{\gamma}^B$ outperform $\hat{\gamma}^A$ or the moment estimator in a MSE-sense. We therefore consider $\hat{\gamma}^B$ as an interesting data-analytical tool that can be used to complement $\hat{\gamma}^A$. When plotting both EVI estimators for a particular data set, $\hat{\gamma}^B$ will inform the analyst of the quality, and especially of the bias, of $\hat{\gamma}^A$. If the estimators show a similar pattern over a sizable range of k-values, then one can rely on $\hat{\gamma}^A$ with a proper choice for the position of the threshold $X_{n-k,n}$, ideally situated in or just beyond the region of congruence. If the patterns diverge rapidly one should be cautious concerning the resulting estimates.

Moreover, one can make these practical guidelines more explicit and use model (B1) to select the threshold $X_{n-k,n}$ for $\hat{\gamma}^A_{k+1}$ in an adaptive way. As in Beirlant et al. (1999) the estimators $\hat{\gamma}^B_{k+1}$, $\hat{b}^B_{n,k+1}$ and $\hat{\rho}^B_{k+1}$ are then used to estimate the asymptotic mean squared error (AMSE) of $\hat{\gamma}^A_{k+1}$. This is done by replacing γ, $b_{n,k+1}$ and ρ by $\hat{\gamma}^B_{k+1}$, $\hat{b}^B_{n,k+1}$ and $\hat{\rho}^B_{k+1}$ in the integral expressions for the asymptotic variance and asymptotic bias of $\hat{\gamma}^A$ (see Appendix A.1). The index k_0 which minimizes the obtained AMSE estimate is then an estimator for the optimal number of extremes to use with $\hat{\gamma}^A$. We note that the result of this procedure is immediately relevant to the POT method, as $\hat{\gamma}^A$ closely resembles the POT ML estimator for $\gamma \geq 0$.

In extensive simulations we found that this procedure can be simplified by putting $\rho \equiv -1$ in model (B1). The AMSE of $\hat{\gamma}^A_{k+1}$ is then estimated with the corresponding estimators for γ and $b_{n,k+1}$, and with $\rho = -1$. This mis-specification is in agreement with established adaptive threshold selection methods for the Hill estimator (see e.g., Hall (1990), Drees and Kaufmann (1998), Gomes and Oliveira (2001) and Matthys and Beirlant (2000a)) and often reduces the variability of the estimation results. Results of this simplified procedure applied to 100 simulated samples of size $n = 500$ from some common distributions (see Appendix A.3 for a precise description) are presented in Table 1 as root mean squared errors (RMSE) of $\hat{\gamma}^A_{k_0}$. In most cases these are of relatively small order, except for distributions where the bias dominates, e.g., for the Burr(1,0.25,4) and the reversed Burr(1,0.25,4). As mentioned above, the bias-corrected estimator $\hat{\gamma}^B$ is especially useful as a refinement of $\hat{\gamma}^A$ for such ill-behaved cases. To assess the quality of the adaptive threshold selection method we also report in Table 1 the ratio R of the RMSE of $\hat{\gamma}^A_{k_0}$ and the minimal empirical RMSE for $\hat{\gamma}^A_k$ found in the simulations:

$$R := \frac{\text{RMSE } \hat{\gamma}^A_{k_0}}{\min_k(\text{RMSE } \hat{\gamma}^A_k)}.$$

In our simulations with sample size $n = 500$ we limited the range from which k_0 could be selected to $\{1, \ldots, 350\}$. For the Gamma(2) and Uniform(1) distributions the ratio R is excessively large although their RMSE for $\hat{\gamma}^A_{k_0}$ is relatively small. This indicates that the optimal k_0 can even be beyond 350 out of 500 observations for extremely well-behaved distributions such as the Gamma(2) and Uniform(1).

Table 1. Simulation results of the adaptive threshold selection procedure for $\hat{\gamma}^A$.

Distribution	γ	ρ	RMSE $\hat{\gamma}^A_{k_0}$	R	Distribution	γ	ρ	RMSE $\hat{\gamma}^A_{k_0}$	R
Burr(1,0.25,4)	1	$-\frac{1}{4}$	0.78	1.27	Lognormal	0	0	0.30	1.12
Burr(1,0.5,2)	1	$-\frac{1}{2}$	0.34	1.25	Gamma(2)	0	0	0.11	1.94
Burr(1,1,1)	1	-1	0.11	1.03	Weibull(1,2)	0	0	0.23	1.25
Fréchet(1)	1	-1	0.13	1.06	rev. Burr(1,0.5,3)	$-\frac{2}{3}$	$-\frac{1}{3}$	0.47	1.07
Loggamma(1,2)	1	0	0.20	0.91	rev. Burr(1,0.25,4)	-1	$-\frac{1}{4}$	0.44	1.07
$\lvert t_2\rvert$	$\frac{1}{2}$	-1	0.14	1.04	rev. Burr(1,0.5,2)	-1	$-\frac{1}{2}$	0.27	1.11
					rev. Burr(1,1,1)	-1	-1	0.18	1.23
					Uniform(1)	-1	-1	0.11	2.07

Table 2. Simulation results of the adaptive threshold selection procedure for $\hat{x}^A_p$.

Distribution	γ	ρ	RMSE $\hat{x}^A_{p,k_0}$	R	Distribution	γ	ρ	RMSE $\hat{x}^A_{p,k_0}$	R
Burr(1,0.25,4)	1	$-\frac{1}{4}$	2.34	1.09	Lognormal	0	0	0.37	1.13
Burr(1,0.5,2)	1	$-\frac{1}{2}$	1.60	1.18	Gamma(2)	0	0	0.53	1.16
Burr(1,1,1)	1	-1	0.87	1.22	Weibull(1,2)	0	0	0.18	1.86
Fréchet(1)	1	-1	0.96	1.23	rev. Burr(1,2,2)	$-\frac{1}{4}$	$-\frac{1}{2}$	0.10	1.09
Loggamma(1,2)	1	0	0.91	1.14					
$\lvert t_2\rvert$	$\frac{1}{2}$	-1	0.55	1.02					

Analogously one can select the optimal threshold $X_{n-k,n}$ for the quantile estimator $\hat{x}^A_{p,k+1}$. One estimates the AMSE of $\hat{x}^A_{p,k+1}$ by replacing γ, $b_{n,k+1}$ and ρ by $\hat{\gamma}^B_{k+1}$, $\hat{b}^B_{n,k+1}$ and $\rho \equiv -1$ in the integral expression for AMSE($\hat{x}^A_{p,k+1}$) that is stated in Appendix A.2 and looks for the position k_0 where the AMSE estimate reaches its minimum value. Simulation results of this adaptive procedure are listed in Table 2 as the RMSE of $\hat{x}^A_{p,k_0}$ (according to (9)) for similar simulated samples as in Table 1 with $p = (1/10)(1/n) = 1/5000$, and the ratio R of RMSE($\hat{x}^A_{p,k_0}$) and the minimal empirical RMSE of $\hat{x}^A_{p,k}$ in the simulations.

4.3. Bias-corrected estimation of high quantiles

In order to construct a quantile estimator that reduces the bias of $\hat{x}^A_p$, especially in ill-behaved cases, we first need an estimator for the scaling factor $c_{n,k+1}$. As in Section 3, this can be derived from a refined exponential regression model for spacings of order statistics. On the basis of (B0) we obtain for $1 \le j < k < n$ the approximation

$$\begin{aligned} &X_{n-j+1,n} - X_{n-j,n} \\ &\stackrel{d}{\approx} c_{n,k+1}\left[V_{j,k}^{-\gamma}\exp\left(b_{n,k+1}\frac{V_{j,k}^{-\rho}-1}{\rho}\right) - V_{j+1,k}^{-\gamma}\exp\left(b_{n,k+1}\frac{V_{j+1,k}^{-\rho}-1}{\rho}\right)\right]. \end{aligned}$$

With the Mean Value Theorem applied to the left-hand side, we arrive at the following analogue to model (A2) for spacings of order statistics with $1 \le j \le k$:

$$\begin{aligned} &j(X_{n-j+1,n} - X_{n-j,n}) \\ &\stackrel{d}{\approx} c_{n,k+1}\left(\gamma + b_{n,k+1}\left(\frac{j}{k+1}\right)^{-\rho}\right)\left(\frac{j}{k+1}\right)^{-\gamma}\exp\left(b_{n,k+1}\frac{1-\left(\frac{j}{k+1}\right)^{-\rho}}{-\rho}\right) f_{k-j+1}, \end{aligned} \tag{B2}$$

where the f_j represent i.i.d. standard exponential random variables. If $b_{n,k+1} = 0$, (B2) reduces to (A2) with $a_{n,k+1} = \gamma c_{n,k+1}$.

Maximising the loglikelihood corresponding to model (B2) and estimating γ, $b_{n,k+1}$ and ρ by $\hat{\gamma}^B_{k+1}$, $\hat{b}^B_{n,k+1}$ and $\hat{\rho}^B_{k+1}$, respectively, we propose for $c_{n,k+1}$ the estimator

$$\hat{c}^B_{n,k+1} := \frac{1}{k}\sum_{j=1}^{k} \frac{j(X_{n-j+1,n} - X_{n-j,n})\left(\frac{j}{k+1}\right)^{\hat{\gamma}^B_{k+1}}}{\left(\hat{\gamma}^B_{k+1} + \hat{b}^B_{n,k+1}\left(\frac{j}{k+1}\right)^{-\hat{\rho}^B_{k+1}}\right)\exp\left(\hat{b}^B_{n,k+1}\frac{1-\left(\frac{j}{k+1}\right)^{-\hat{\rho}^B_{k+1}}}{-\hat{\rho}^B_{k+1}}\right)}.$$

Finally for high quantiles, (11) and (R_ℓ) lead to the approximation

$$\begin{aligned} x_p - X_{n-k,n} &\stackrel{d}{=} U((U_{k+1,n}/p)\, U^{-1}_{k+1,n}) - U(U^{-1}_{k+1,n}) \\ &\stackrel{d}{\approx} c_{n,k+1}\left[(U_{k+1,n}/p)^{\gamma}\exp\left(b_{n,k+1}\frac{(U_{k+1,n}/p)^{\rho}-1}{\rho}\right) - 1\right]. \end{aligned}$$

We estimate γ, $b_{n,k+1}$ and ρ by $\hat{\gamma}^B_{k+1}$, $\hat{b}^B_{n,k+1}$ and $\hat{\rho}^B_{k+1}$, $c_{n,k+1}$ by $\hat{c}^B_{n,k+1}$, and $U_{k+1,n}$ by its expected value $(k+1)/(n+1)$ to obtain for $k < n$ the estimator

$$\hat{x}^B_{p,k+1} := X_{n-k,n} + \hat{c}^B_{n,k+1}\left[\left(\frac{k+1}{p(n+1)}\right)^{\hat{\gamma}^B_{k+1}} \exp\left(\hat{b}^B_{n,k+1}\frac{\left(\frac{k+1}{p(n+1)}\right)^{\hat{\rho}^B_{k+1}}-1}{\hat{\rho}^B_{n,k+1}}\right)-1\right].$$

In Figure 5 the quantile estimator $\hat{x}^B_p$ is compared to the simplified version $\hat{x}^A_p$, the POT ML estimator $\hat{x}^{MLP}_p$ and the moment estimator $\hat{x}^M_p$, with the simulated data from Figure 4 and with $p = (1/10)(1/n) = 1/5000$. As with the refined EVI estimator $\hat{\gamma}^B$, $\hat{x}^B_p$ usually succeeds well in reducing the bias of $\hat{x}^A_p$. On the other hand, it has a higher variance and is often less efficient in MSE sense. Here again, we propose $\hat{x}^B_p$ as a data-analytical tool that can be used in combination with $\hat{x}^A_p$ to warn of ill-behaved cases, to assess the bias of $\hat{x}^A_p$, and to judge the position of a proposed threshold $X_{n-k,n}$ for $\hat{x}^A_p$.

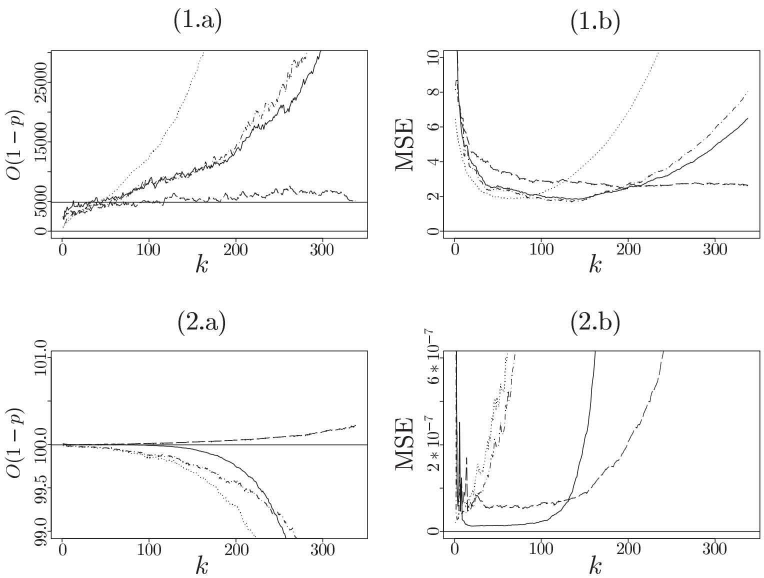

Figure 5. (a) Medians and (b) empirical MSE's of $\hat{x}^A_{p,k}$ (solid line), $\hat{x}^B_{p,k}$ (dashes), $\hat{x}^{MLP}_{p,k}$ (dots-dashes) and $\hat{x}^M_{p,k}$ (dots), with $p = 1/5000$, $k = 3, \ldots,$ 340, for 100 simulated samples of size $n = 500$ with (1) a Burr(1,0.5,2) distribution ($\gamma = 1$), and (2) a reversed Burr(1,0.5,3) distribution ($\gamma = -2/3$). Horizontal lines indicate the true value of x_p.

A. Appendix

A.1. Asymptotic properties of $\hat{\gamma}^A$

Theorem 1. (Asymptotic normality of $\hat{\gamma}^A$). *Suppose that*

- *in case* $\gamma > 0$, $U(r) = r^{\gamma}\ell(r)$ *with* ℓ *satisfying* (R_ℓ) *with* $\rho < 0$,
- *in case* $\gamma = 0$, *as* $r \to \infty$, $(U(tr) - U(r) - b_0(r)\log t)/\tilde{b}_0(r) \to \pm(\log t)^2/2$ *for some positive functions* b_0 *and* $\tilde{b}_0$ (*where* $\tilde{b}_0 \sim a_U$),
- *in case* $\gamma < 0$, $U(r) = x_+ - r^{\gamma}\ell(r)$ *with* ℓ *satisfying* (R_ℓ) *with* $\rho < 0$.

Suppose also that k, $n \to \infty$ *with* $k/n \to 0$, *and in case*

$$\begin{cases} \gamma \neq 0, & \sqrt{k}\, b(n/k) \to 0, \\ \gamma = 0, & \sqrt{k}\, \frac{\tilde{b}_0(n/k)}{b_0(n/k)} \to 0. \end{cases} \tag{12}$$

Then $\sqrt{k}(\hat{\gamma}^A_{k+1} - \gamma) \xrightarrow{\mathcal{L}} N(0, \sigma^2_\gamma/a^2_\gamma)$, *where* $a_\gamma = \gamma^{-2}\int_0^1 (1 - u^\gamma + u^\gamma \log u^\gamma)^2\ (1 - u^\gamma)^{-2} du$ *and* σ^2_γ *equals the variance of* $K_\gamma(U)$ *with* U *uniformly* $(0,1)$ *distributed and with* $K_\gamma(u) = (\log u)/\gamma + (1+\gamma)\ (\mathrm{dilog}\, u^\gamma)/\gamma^2$, *where* $\mathrm{dilog}\, u = \int_1^u (\log t)/(1-t)\, dt$ $(u \geq 0)$ *denotes the dilogarithm function.*

Proof. We follow the classical approach using a Taylor expansion of $(\partial \mathcal{L}^{A1}_{(\gamma)}(\mathbf{Y}))/\partial\gamma$ around the correct population value γ computed at $\hat{\gamma}^A_{k+1}$:

$$0 = \frac{1}{k}\frac{\partial \mathcal{L}^{A1}_{(\gamma)}(\mathbf{Y})}{\partial\gamma} + (\hat{\gamma}^A_{k+1} - \gamma)\frac{1}{k}\left.\frac{\partial^2 \mathcal{L}^{A1}_{(\gamma)}(\mathbf{Y})}{\partial\gamma^2}\right|_{\gamma = \hat{\gamma}^*_{k+1}},$$

with $\hat{\gamma}^*_{k+1}$ denoting a value situated between $\hat{\gamma}^A_{k+1}$ and γ. The result is then obtained by proving the consistency of $\hat{\gamma}^A_{k+1}$ (see e.g., Lehmann (1983, p.413, Theorem 2.2)), followed by the asymptotic normality of $-k^{-1/2}\partial\mathcal{L}^{A1}_{(\gamma)}(\mathbf{Y})/\partial\gamma$, and the convergence in probability of $k^{-1}\partial^2\mathcal{L}^{A1}_{(\gamma)}(\mathbf{Y})/\partial\gamma^2$. We carry out this program in more detail for the case $\gamma > 0$, and end with some details of the proof for $\gamma = 0$ and $\gamma < 0$. First, remark that

$$-\frac{1}{\sqrt{k}}\frac{\partial \mathcal{L}^{A1}_{(\gamma)}(\mathbf{Y})}{\partial\gamma} = \frac{1}{\sqrt{k}}\sum_{j=1}^{k-1}\tilde{J}\Big(\frac{j}{k+1}\Big)\left(Y_j - \frac{\gamma}{1 - \left(\frac{j}{k+1}\right)^\gamma}\right)$$

with $\tilde{J}(u) = \gamma^{-2}(u^\gamma - 1 - \gamma u^\gamma \log u)$, $0 < u < 1$. Partial summation yields

$$\begin{aligned}
&-\frac{1}{\sqrt{k}}\frac{\partial \mathcal{L}^{A1}_{(\gamma)}(\mathbf{Y})}{\partial\gamma} \\
&= \frac{1}{\sqrt{k}}\sum_{j=1}^{k-1}\tilde{J}\Big(\frac{j}{k+1}\Big)\left(j\log\left[\frac{(X_{n-j+1,n} - X_{n-k,n})/X_{n-k,n}}{(X_{n-j,n} - X_{n-k,n})/X_{n-k,n}}\right] - \frac{\gamma}{1 - \left(\frac{j}{k+1}\right)^\gamma}\right) \\
&= \frac{1}{\sqrt{k}}\tilde{J}\left(\frac{1}{k+1}\right)\log\left(\frac{X_{n,n} - X_{n-k,n}}{X_{n-k,n}}\right) \\
&\quad -\frac{k-1}{\sqrt{k}}\tilde{J}\left(\frac{k-1}{k+1}\right)\log\left(\frac{X_{n-k+1,n} - X_{n-k,n}}{X_{n-k,n}}\right)
\end{aligned}$$

$$+\frac{1}{\sqrt{k}}\sum_{j=2}^{k-1}\left[(k+1)\int_{\frac{j-1}{k+1}}^{\frac{j}{k+1}} J(u)\,du\right]\log\left(\frac{X_{n-j+1,n}-X_{n-k,n}}{X_{n-k,n}}\right)$$
$$-\frac{1}{\sqrt{k}}\sum_{j=1}^{k-1}\tilde{J}\left(\frac{j}{k+1}\right)\frac{\gamma}{1-\left(\frac{j}{k+1}\right)^{\gamma}}$$
$$=: T_{1,n}-T_{2,n}+T_{3,n}-T_{4,n},$$

with $J(u)=\gamma^{-2}(u^{\gamma}-1-\gamma(1+\gamma)u^{\gamma}\log u)$, $0<u<1$. Now one shows that under the given conditions $T_{1,n}\stackrel{P}{\to}0$, $T_{2,n}\stackrel{P}{\to}0$ and $T_{3,n}-T_{4,n}\stackrel{\mathcal{L}}{\to}N(0,\sigma_{\gamma}^{2})$.

To this end, note that $(X_{n-j+1,n}-X_{n-k,n})/X_{n-k,n}\stackrel{d}{=}V_{j,k}^{-\gamma}\Delta_{j,k,n}-1$ with $\Delta_{j,k,n}=\ell(U_{k+1,n}^{-1}V_{j,k}^{-1})/\ell(U_{k+1,n}^{-1})$. Using the Mean Value Theorem, we find that

$$\left|\log\Big(\frac{X_{n-j+1,n}-X_{n-k,n}}{X_{n-k,n}}\Big)-\log(V_{j,k}^{-\gamma}-1)\right|\le\frac{|\Delta_{j,k,n}-1|}{|\Delta_{j,k,n}^{*}-V_{j,k}^{\gamma}|},$$

where $\Delta_{j,k,n}^{*}$ is situated between $\Delta_{j,k,n}$ and 1. Condition (R_{ℓ}) implies that for any $\epsilon>0$, we have for n, k large enough that

$$\frac{1-(1+\epsilon)V_{j,k}^{|\rho|-\epsilon}}{|\rho|}\le\frac{\Delta_{j,k,n}-1}{b(U_{k+1,n}^{-1})}\le\frac{1-(1-\epsilon)V_{j,k}^{|\rho|+\epsilon}}{|\rho|}\tag{13}$$

with arbitrary large probability (where we assume without loss of generality that $b(r)$ is ultimately positive for large values of r). Moreover,

$$\frac{1}{k}\sum_{j=2}^{k-1}\left[(k+1)\int_{\frac{j-1}{k+1}}^{\frac{j}{k+1}}|J(u)|du\right]\frac{1-(1-\epsilon)V_{j,k}^{|\rho|+\epsilon}}{\Delta_{j,k,n}^{*}-V_{j,k}^{\gamma}}\stackrel{P}{\to}$$
$$\int_{0}^{1}\frac{|J(v)|}{1-v^{\gamma}}[1-(1-\epsilon)v^{|\rho|+\epsilon}]dv<\infty,$$

and similarly for the lower bound in (13). Since b is regularly varying and $(n/k)U_{k+1,n}\stackrel{P}{\to}1$, we have under (12) that $\sqrt{k}\,b(U_{k+1,n}^{-1})\stackrel{P}{\to}0$, so that under the given conditions

$$T_{3,n}-\frac{1}{\sqrt{k}}\sum_{j=2}^{k-1}\left[(k+1)\int_{\frac{j-1}{k+1}}^{\frac{j}{k+1}}J(u)du\right]\log(V_{j,k}^{-\gamma}-1)\stackrel{P}{\to}0.\tag{14}$$

The asymptotic normality of

$$T_{3,n}^{*}:=\frac{1}{\sqrt{k}}\sum_{j=2}^{k-1}\left[(k+1)\int_{\frac{j-1}{k+1}}^{\frac{j}{k+1}}J(u)du\right]\log(V_{j,k}^{-\gamma}-1)$$

can be derived from Theorem 4.1 in Shorack (2000) on the asymptotic normality of L-statistics $k^{-1}\sum_{j=2}^{k-1} c_{k,j} H(V_{j,k})$, with, in our case, $H(t) = -\log(t^{-\gamma}-1)$ and

$$c_{k,j} = -\left[(k+1)\int_{\frac{j-1}{k+1}}^{\frac{j}{k+1}} J(u)du\right] = -\left[j\tilde{J}\Big(\frac{j}{k+1}\Big) - (j-1)\tilde{J}\Big(\frac{j-1}{k+1}\Big)\right].$$

In Shorack (2000) it is shown that for smooth weight functions J (at 0 and 1) and increasing H, the asymptotic distribution of $T^*_{3,n} - \sqrt{k}\mu_k$ with $\mu_k = \sum_{j=2}^{k-1} c_{k,j}\int_{(j-1)/k}^{j/k} H(t)dt$, is identical to the one of $\sqrt{k}(1/k\sum_{j=1}^{k-1} K_\gamma^{(1)}(U_j) - \mathrm{E}\,[K_\gamma^{(1)}(U)])$, with $U, U_1, \ldots, U_k$ i.i.d. uniform(0,1) random variables and

$$\begin{aligned}
K_\gamma^{(1)}(u) &= -\int_{(1/2)^{1/\gamma}}^{u} J(t)dH(t) \\
&= \int_{(1/2)^{1/\gamma}}^{u} \frac{t^\gamma - 1 - \gamma(1+\gamma)t^\gamma \log t}{\gamma(t^\gamma - 1)}\frac{dt}{t} \\
&= \tfrac{1}{\gamma}\left(\log u + \frac{\log 2}{\gamma}\right) + \left(\frac{1+\gamma}{\gamma^2}\right)\left(\operatorname{dilog} u^\gamma - \operatorname{dilog}\tfrac{1}{2}\right),
\end{aligned}$$

which converges to a normal distribution with mean 0 and variance σ_γ^2. It is easily verified that $T_{4,n}$ is asymptotically equivalent to the required centring sequence $\sqrt{k}\,\mu_k$. Hence we find that

$$T^*_{3,n} - T_{4,n} \xrightarrow{\mathcal{L}} N(0,\sigma_\gamma^2). \tag{15}$$

Combining (14) and (15) then yields the same asymptotic limit law for $T_{3,n} - T_{4,n}$.

Next, concerning $T_{1,n}$, one shows as above that

$$\frac{1}{\sqrt{k}}\tilde{J}\Big(\frac{1}{k+1}\Big)\left\{\log\left(\frac{X_{n,n} - X_{n-k,n}}{X_{n-k,n}}\right) - \log(V_{1,k}^{-\gamma} - 1)\right\} \xrightarrow{P} 0$$

and one verifies that $k^{-1/2}\tilde{J}(1/(k+1))\log(V_{1,k}^{-\gamma} - 1) \xrightarrow{P} 0$. Similarly one shows that $T_{2,n} \xrightarrow{P} 0$.

Concerning the convergence in probability of $k^{-1}\partial^2\mathcal{L}^{A1}_{(\gamma)}(\mathbf{Y})/\partial\gamma^2$ note that

$$\begin{aligned}
\frac{1}{k}\frac{\partial^2\mathcal{L}^{A1}_{(\gamma)}(\mathbf{Y})}{\partial\gamma^2} &= \frac{1}{k}\sum_{j=1}^{k}\tilde{J}\Big(\frac{j}{k+1}\Big)\frac{\left[1 - \left(\frac{j}{k+1}\right)^\gamma + \gamma\left(\frac{j}{k+1}\right)^\gamma \log\left(\frac{j}{k+1}\right)\right]}{\left[1 - \left(\frac{j}{k+1}\right)^\gamma\right]^2} \\
&\quad -\frac{1}{k}\sum_{j=1}^{k}\Big(Y_j - \frac{\gamma}{1 - \left(\frac{j}{k+1}\right)^\gamma}\Big)\cdot\left.\frac{\partial\tilde{J}(u)}{\partial u}\right|_{u=\frac{j}{k+1}}.
\end{aligned}$$

The second term on the right-hand side tends to 0 in probability, while the first term tends in probability to $-a_\gamma$.

In case $\gamma < 0$, the proof follows the same lines. Here the analysis of $T_{j,n}$ ($j = 1, 2, 3, 4$) involves the negligibility of $k^{-1/2}\tilde{J}(1/(k+1))\log(1 - V_{1,k}^{|\gamma|})$ and $(k-1)^{\sqrt{k}}\tilde{J}((k-1)/(k+1))\log(1-V_{k,k}^{|\gamma|})$, together with the asymptotic normality of

$$\frac{1}{\sqrt{k}}\sum_{j=2}^{k-1}\left[(k+1)\int_{\frac{j-1}{k+1}}^{\frac{j}{k+1}} J(u)du\right]\log\left(1 - V_{j,k}^{|\gamma|}\right).$$

Finally, when $\gamma = 0$, one shows the asymptotic equivalence of

$$-\frac{1}{\sqrt{k}}\frac{\partial\mathcal{L}_{(\gamma)}^{A1}(\mathbf{Y})}{\partial\gamma} \quad \text{and} \quad \frac{1}{\sqrt{k}}\sum_{j=2}^{k-1}\left[(k+1)\int_{\frac{j-1}{k+1}}^{\frac{j}{k+1}} J(u)du\right]\log\log V_{j,k}^{-1}.$$

To this end one considers $\log((X_{n-j+1,n} - X_{n-k,n})/b_0(U_{k+1,n}^{-1})) = \log(\log V_{j,k}^{-1} + \Delta_{j,k,n}^{(0)})$ with

$$|\Delta_{j,k,n}^{(0)}| \le \frac{\tilde{b}_0(U_{k+1,n}^{-1})}{b_0(U_{k+1,n}^{-1})}\max\left\{(1+\epsilon)^2 V_{j,k}^{-\epsilon}\frac{(\log V_{j,k}^2)^2}{2} + 2\epsilon\log V_{j,k}^{-1} + \epsilon;\right.$$
$$\left. -(1-\epsilon^2)\frac{(\log V_{j,k}^2)^2}{2} + 2\epsilon\log V_{j,k}^{-1} + \epsilon\right\}$$

(see Dekkers, Einmahl and de Haan (1989, Lemma 3.5)).

Asymptotic bias of $\hat{\gamma}_{k+1}^{A}$. Combining $T_{3,n}$ in the proof of the asymptotic normality of $\hat{\gamma}^A$ with the asymptotic representation

$$V_{j,k}^{-\gamma}\left[1 + b(n/k)\frac{V_{j,k}^{-\rho} - 1}{\rho} + o_P\left(b(n/k)\right)\right] - 1 \quad \text{of} \quad \frac{X_{n-k+1,n} - X_{n-k,n}}{X_{n-k,n}} \quad (1 \le j \le k)$$

the main term of the asymptotic bias is found to be $-I_{\gamma,\rho}b(n/k)/a_\gamma$ with $I_{\gamma,\rho} = -\rho^{-1}\int_0^1(J(u)/(1-u^\gamma))(1-u^{-\rho})du$. Hence, the asymptotic mean squared error to be minimised with respect to k in order to find the optimal threshold for $\hat{\gamma}^A$ is proportional to $k^{-1}\sigma_\gamma^2 + b^2(n/k)I_{\gamma,\rho}^2$.

A.2. Asymptotic properties of $\hat{a}_{n,k}^{A}$ and $\hat{x}_{p,k}^{A}$

Using similar arguments as for the asymptotic normality of $\hat{\gamma}^A$ one can prove the following.

Theorem 2. (Asymptotic normality of $\hat{a}_{n,k}^A$). *Under the same conditions as in Theorem* 1, $\sqrt{k}((\hat{a}_{n,k}^A/a_{n,k}) - 1) \xrightarrow{\mathcal{L}} -G + H + N$, *with*

- $N \sim N(0, \gamma^2)$, *independent of* (G, H),
- G *representing the limit law from Theorem* 1,

- *$H \sim N(0, (1+\gamma)^2)$, and*
- *$\mathrm{Cov}\,(G, H) = a_\gamma^{-1}\mathrm{Cov}\,(K_\gamma^{(1)}(U), K_\gamma^{(2)}(U))$, with $K_\gamma^{(1)}$ as in the proof of Theorem 1, $K_\gamma^{(2)}(u) = (\gamma+1)\log u$, and U uniformly $(0,1)$ distributed.*

Turning to the asymptotics of the quantile estimator $\hat{x}_{p,k}^A$, one can use the method of proof of de Haan and Rootzén (1993) to verify the following result. In fact, $\hat{x}_{p,k}^A$ is of the same form as $\hat{x}_{p,k+1}^M$ (from Section 2.3 above) the asymptotics of which were considered in de Haan and Rootzén (1993). The conditions of the results are identical and it suffices to replace the asymptotic distribution of $(\hat{\gamma}_{k+1}^M, \hat{a}_{k+1}^M)$ by the joint asymptotic distribution of $(\hat{\gamma}_{k+1}^A, \hat{a}_{n,k+1}^A)$, which follows from Theorems 1 and 2 above.

Theorem 3. (Asymptotic normality of $\hat{x}_{p,k}^A$). *Under the conditions of Theorem 1 and for $np_n \to 0$, $k = k_n \to \infty$, $k/n \to 0$ and $\log(np)/\sqrt{k} \to 0$ $(n \to \infty)$, we have with $a_n = k/(np)$ and $f(x) = \exp((1-\gamma)x)U'(\exp x)$ that, if $\sqrt{k}(f'/f)(\log(n/k))(\int_1^{a_n} s^{\gamma-1}\int_1^s u^{\alpha-1}duds)/(\int_1^{a_n} s^{\gamma-1}(\log s)ds) \to 0$ for some $\alpha > 0$, and $(\log a_n)\sup_{v\geq \log n/k}|(f''(v)/f'(v)) - \alpha| \to 0$, then*

$$\frac{\sqrt{k}\left(\hat{x}_{p,k}^A - x_p\right)}{\hat{a}_{n,k}^A\int_1^{a_n} s^{\hat{\gamma}-1}(\log s)\,ds} \xrightarrow{\mathcal{L}} \begin{cases} G, & \text{if} \quad \gamma \geq 0, \\ (1-|\gamma|)G + |\gamma|\,H, & \text{if} \quad \gamma < 0, \end{cases}$$

with G, H and N as in Theorem 2.

Remark. In case $\gamma = -1$ one can show that $(k/\hat{a}_{n,k}^A)(\hat{x}_{p,k}^A - x_p)$ converges weakly to a non-degenerate limit.

Asymptotic mean squared error (AMSE) of $\hat{x}_{p,k}^A$. Following the approach from the proof of Proposition 4.12 in Ferreira, de Haan and Peng (2003) one can derive the minimiser for the asymptotic $\mathrm{E}\,(\hat{x}_{p_n,k}^A - x_{p_n})^2$ in case $b(x) = Cx^\rho(1+o(1))(x\to\infty)$ for some $C > 0$, $np_n \to c$ (finite, ≥ 0) and $\log p_n = o(n^\epsilon)$ for $\epsilon > 0$ $(n \to \infty)$. In case $\gamma > 0$ this AMSE is proportional to the AMSE of $\hat{\gamma}_k^A$, whereas for $\gamma < 0$ one finds that

$$\frac{\text{asymptotic E}\left(\hat{x}_{p_n,k}^A - x_{p_n}\right)^2}{a_{n,k}^2}$$
$$= \frac{1}{k}\left\{\frac{\sigma_\gamma^2}{a_\gamma^2}\frac{(1+\gamma)^2}{\gamma^4} + \frac{(1+\gamma)^2}{\gamma^2} + 2\mathrm{Cov}(G,H)\frac{1+\gamma}{\gamma^3}\right\} + b^2(n/k)\frac{(1+\gamma)^2}{\gamma^4}\left[\frac{I_{\gamma,\rho}}{a_\gamma} + \frac{1}{1-\rho}\right]^2.$$

Since $a_{n,k} \sim D(n/k)^\gamma$ for some $D > 0$, it follows that a minimum can be found in k when $1 + 2\gamma > 0$.

The algorithm to adaptively choose the threshold when estimating x_p with $\hat{x}^A_{p,k}$ is then based on finding a minimum w.r.t. k for an estimate of the function

$$\begin{cases} \frac{1}{k}\frac{\sigma_\gamma^2}{a_\gamma^2} + b^2(n/k)\,\frac{I^2_{\gamma,\rho}}{a^2_\gamma}, & \gamma > 0, \\ \begin{array}{l} k^{-(1+2\gamma)}\left\{\frac{\sigma^2_\gamma}{a^2_\gamma}(1+\gamma)^2 + \gamma^2(1+\gamma)^2 + 2\,\mathrm{Cov}\,(G,H)\,(1+\gamma)\gamma\right\} \\ \quad + b^2(n/k)(1+\gamma)^2\left[\frac{I_{\gamma,\rho}}{a_\gamma} + \frac{1}{1-\rho}\right]^2. \end{array} & \gamma < 0. \end{cases}$$

A.3. Distributions used in simulation study

We performed a simulation study to compare the finite sample behaviour of the novel EVI and quantile estimators $\hat{\gamma}^A$ and $\hat{\gamma}^B$, respectively $\hat{x}^A_p$ and $\hat{x}^B_p$, with the classical estimators that were described in Section 2. For a number of distributions from each of the three EVI classes, 100 random samples of size 500 were generated and the estimators for these samples were calculated. The distributions from which we simulated were as follows for $\gamma > 0$:

- Burr(β,τ,λ) distributions given by $1-F(x) = (1+(x^\tau/\beta))^{-\lambda}$ so that $\gamma = 1/\tau\lambda$ and $\rho = -1/\lambda$. We have chosen $(\beta,\tau,\lambda) = (1,1,1)$, $(1,0.5,2)$, and $(1,0.25,4)$.
- the Fréchet(γ) distribution given by $1-F(x) = 1-\exp(-x^{-1/\gamma})$ so that $\rho = -1$; we have chosen $\gamma = 1$.
- the absolute value of t distributions with ν degrees of freedom (denoted by $|t_\nu|$) for which $\gamma = 1/\nu$ and $\rho = -2/\nu$; we have chosen $\nu = 2$.
- the loggamma distribution with density $\Gamma^{-1}(\beta)x^{-2}(\log x)^{\beta-1}$, $x > 1$, with $\gamma = 1$ and $\rho = 0$. We have chosen $\beta = 2$.

For $\gamma = 0$ (and $\rho = 0$):

- the standard lognormal distribution.
- the gamma(β) distribution with density $\Gamma^{-1}(\beta)x^{\beta-1}e^{-x}$, $x > 0$; we have chosen $\beta = 2$.
- the Weibull(λ,τ) distribution given by $1-F(x) = \exp(-\lambda x^\tau)$. We have chosen $(\lambda,\tau) = (1,2)$.

For $\gamma < 0$:

- the uniform (0,1) distribution with $x_+ = 1$, $\gamma = -1$ and $\rho = -1$.
- reversed Burr(β,τ,λ) distributions given by $1-F(x) = (1+((x_+-x)^{-\tau}/\beta))^{-\lambda}$, so that $\gamma = -1/\tau\lambda$ and $\rho = -1/\lambda$. We have chosen $x_+ = 2$ and $(\beta,\tau,\lambda) = (1,1,1)$, $(1,0.5,2)$, $(1,0.5,3)$, $(1,0.25,4)$ in Table 1 and $(1,2,2)$ in Table 2.

References

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., Matthys, G. and Guillou, A. (2000). Estimation of the extreme value index and regression on generalized quantile plots. Preprint, Center for Statistics, University of Leuven.

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

Castillo, E. and Hadi, A. S. (1997). Fitting the generalized Pareto distribution to data. *J. Amer. Statist. Assoc.* **92**, 1604-1620.

Coles, S. and Powell, E. (1996). Bayesian methods in extreme value modelling: a review and new developments. *Internat. Statist. Rev.* **64**, 119-136.

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

Dekkers, A. L. M. and de Haan, L. (1989). On the estimation of the extreme-value index and large quantile estimation. *Ann. Statist.* **17**, 1795-1832.

Drees, H. (1996). Refined estimation of the extreme value index. Ph.D. thesis, Department of Mathematics, University of Siegen.

Drees, H. and Kaufmann, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Process. Appl.* **75**, 149-172.

Ferreira, A., de Haan, L. and Peng, L. (2003). On optimising the estimation of high quantiles of a probability distribution. *Statistics*. To appear.

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

Geluk, J. and de Haan, L. (1987). *Regular Variation, Extensions and Tauberian Theorems*. Math. Centre Tract 40, Amsterdam.

Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* **44**, 423-453.

Gomes, M. I. and Oliveira, O. (2001). The use of bootstrap methodology in Statistics of Extremes - choice of the optimal sample fraction. *Extremes* **4**, 331-358.

Guillou, A. and Hall, P. (2001). A diagnostic for selecting the threshold in extreme-value analysis. *J. Roy. Statist. Soc. Ser. B* **63**, 293-305.

de Haan, L. (1970). *On Regular Variation and its Application to the Weak Convergence of Sample Extremes*. Math. Centre Tract 32, Amsterdam.

de Haan, L. and Rootzén H. (1993). On the estimation of high quantiles. *J. Statist. Plann. Inference* **35**, 1-13.

de Haan, L. and Stadtmüller, U. (1996). Generalized regular variation of second order. *J. Austral. Math. Soc. Ser. A* **61**, 381-395.

Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *J. Multivariate Anal.* **32**, 177-203.

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

Hosking, J. R. M. and Wallis, J. R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* **29**, 339-349.

Lehmann, E. L. (1983). *Theory of Point Estimation*. Wiley, New York.

Matthys, G. and Beirlant, J. (2000a). Adaptive threshold selection in tail index estimation. In *Extremes and Integrated Risk Management* (Edited by P. Embrechts), 37-49. Risk Books, London.

Matthys, G. and Beirlant, J. (2000b). Extreme quantile estimation for heavy-tailed distributions. Preprint, Center for Statistics, University of Leuven.

Pickands III, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3**, 119-131.

Segers, J. (2001). Extremes of a random sample: limit theorems and statistical applications. Ph.D. thesis, Department of Mathematics, University of Leuven.

Shorack, G. R. (2000). *Probability for Statisticians*. Springer-Verlag, New York.

Smith, R. L. (1987). Estimating tails of probability distributions. *Ann. Statist.* **15**, 1174-1207.

Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Assoc.* **73**, 812-815.

Department of Mathematics, Katholieke Universiteit Leuven, 3001 Leuven, Belgium.

E-mail: gunther.matthys@ucs.kuleuven.ac.be

Department of Mathematics, Katholieke Universiteit Leuven, 3001 Leuven, Belgium.

E-mail: beirlant@wis.kuleuven.ac.be

(Received January 2002; accepted April 2003)