INSTITUT DE STATISTIQUE

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION PAPER

0824

SECOND-ORDER REFINED PEAKS-OVER-THRESHOLD MODELLING FOR HEAVY-TAILED DISTRIBUTIONS

BEIRLANT, J., JOOSSENS, E. and J. SEGERS

Second-Order Refined Peaks-Over-Threshold Modelling for Heavy-Tailed Distributions

Jan Beirlant [a,1]

[a] *Department of Mathematics and Leuven Statistics Research Centre*
Katholieke Universiteit Leuven, Celestijnenlaan 200b, B-3001 Heverlee, Belgium

Elisabeth Joossens [b]

[b] *Joint Research Centre, European Commission*
Via Fermi 2749, 21027 Ispra (VA), Italy

Johan Segers [c,*,1]

[c] *Institut de statistique, Université catholique de Louvain*
Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium

Abstract

Modelling excesses over a high threshold using the Pareto or generalized Pareto distribution (PD/GPD) is the most popular approach in extreme value statistics. This method typically requires high thresholds in order for the (G)PD to fit well and in such a case applies only to a small upper fraction of the data. The extension of the (G)PD proposed in this paper is able to describe the excess distribution for lower thresholds in case of heavy tailed distributions. This yields a statistical model that can be fitted to a larger portion of the data. Moreover, estimates of tail parameters display stability for a larger range of thresholds. Our findings are supported by asymptotic results, simulations and a case study.

Key words: bias reduction, Hill estimator, extended Pareto distribution, extreme value index, heavy tails, regular variation, tail empirical process, tail probability, Weissman probability estimator

* Corresponding author.
Email addresses: jan.beirlant@wis.kuleuven.be (Jan Beirlant), elisabeth.joossens@jrc.it (Elisabeth Joossens), johan.segers@uclouvain.be (Johan Segers).

[1] Supported by IAP research network grant no. P6/03 of the Belgian government (Belgian Science Policy).

1 Introduction

It is well known that a distribution is in the max-domain of attraction of an extreme value distribution if and only if the distribution of excesses over high thresholds is asymptotically generalized Pareto (GP) (Balkema and de Haan, 1974; Pickands, 1975). This result gave rise to the peaks-over-threshold methodology introduced in Davison and Smith (1990); see also Coles (2001). The method consists of two components: modelling of clusters of high-threshold exceedances with a Poisson process and modelling of excesses associated to the cluster peaks with a GPD. In practice, a way to verify the validity of the model is to check whether the estimates of the GP shape parameter are stable when the model is fitted to excesses over a range of thresholds. The question then arises how to proceed if this threshold stability is not visible for a given data set. From a theoretical point of view, absence of the stability property can be explained by a slow rate of convergence in the Pickands–Balkema–de Haan theorem. In case of heavy-tailed distributions, the same issue arises when fitting a Pareto distribution (PD) to the relative excesses over high, positive thresholds.

A possible solution is to build a more flexible model capable of capturing the deviation between the true excess distribution and the asymptotic model. For heavy-tailed distributions, this deviation can be parametrized using a power series expansion of the tail function (Hall, 1982), or more generally via second-order regular variation (Geluk and de Haan, 1987; Bingham et al., 1987).

The aim of this paper is to propose such an extension, called the extended Pareto or extended generalized Pareto distribution (EPD/EGPD). A key distinction with other approaches is that although in previous papers the second-order approximation is used for adjusting the inference of the tail index, inference on the tail itself is still based on the GPD; in contrast, in our approach the EP(G)D is fitted directly to the high-threshold excesses. Indeed, as we will show later, even if the (G)PD parameters are estimated in an unbiased way, tail probability estimators may still exhibit asymptotic bias if based upon the (G)PD approximation.

The main advantages of the new model are a reduction of the bias of estimators of tail parameters and a good fit to excesses over a larger range of thresholds. In an actuarial context, the relevance of using more elaborate models has already been discussed for instance in Frigessi et al. (2002) and Cooray and Ananda (2005).

In case of heavy-tailed distributions, it is more convenient to work with relative excesses X/u rather than absolute excesses $X - u$. Under the domain of attraction condition the limit distribution of X/u given $X > u$ for $u \to \infty$

is the PD. The EPD and EGPD presented here are related through the same affine transformation that links these relative and absolute excesses. Building on the theory of generalized regular variation of second order in de Haan and Stadtmüller (1996), it is also possible to construct an extension of the GPD with comparable merits applicable to distributions in all max-domains of attraction. However, parameter estimation in this more general setting is numerically quite involved (Beirlant et al., 2002b): the model contains one additional parameter and the upper endpoint of the distribution depends in a complicated way on the parameters, which complicates both theory and computations.

Bias-reduction methods have already been proposed in, amongst others, Feuerverger and Hall (1999), Gomes et al. (2000), Beirlant et al. (1999), Beirlant et al. (2002a), Gomes and Martins (2002), and Gomes and Martins (2004). These methods focus on the distribution of log-spacings of high order statistics. Moreover, *ad hoc* construction methods for asymptotically unbiased estimators of the extreme value index were introduced in Peng (1998), Drees (1996) and Segers (2005). In contrast, next to providing bias-reduced tail index estimators, our model can be fitted directly to the excesses over a high threshold. The fitted model can then be used to estimate any tail-related risk measure, such as tail probabilities, tail quantiles (or value-at-risk), etc.

In the same spirit as in this paper, a mixture model with two Pareto components was proposed in Peng and Qi (2004). The advantage of our model is that it also incorporates the popular GPD. From our experience, this connection can assist in judging the quality of the GPD fit; see for instance the case study in Example 5.3.

The paper is structured as follows. The next section provides the definition of the E(G)PD, which is shown to yield a more accurate approximation to the distribution of absolute and relative excesses for a wide class of heavy-tailed distributions. Estimators of the EPD parameters are derived in Section 3 using the linearized score equations, and their asymptotic normality is formally stated. In Section 4, we compare the asymptotic distribution and the finite-sample behavior of the estimators of the extreme value index following from PD, GPD and EPD modelling. To illustrate how to apply the methodology to the estimation of general tail-related risk measures, we elaborate in Section 5 on tail probability estimation with theoretical results and a practical case. The appendices, finally, contain the statement and proof of an auxiliary result on a certain tail empirical process followed by the proofs of the main theorems.

2 The Extended (Generalized) Pareto Distribution

Definition 2.1 *The* Extended Pareto Distribution (EPD) *with parameter vector* (γ, δ, τ) *in the range* $\tau < 0 < \gamma$ *and* $\delta > \max(-1, 1/\tau)$ *is defined by its distribution function*

$$G_{\gamma,\delta,\tau}(y) = \begin{cases} 1 - \{y(1+\delta-\delta y^{\tau})\}^{-1/\gamma}, & \text{if } y > 1, \\ 0, & \text{if } y \leqslant 1. \end{cases}$$

The Extended Generalized Pareto Distribution (EGPD) *is defined by its distribution function*

$$H_{\gamma,\delta,\tau}(x) = G_{\gamma,\delta,\tau}(1+x), \qquad x \in \mathbb{R}.$$

The ordinary Pareto Distribution (PD) with shape parameter $\alpha > 0$ is a member of the EPD family: take $\gamma = 1/\alpha$ and $\delta = 0$ (arbitrary τ). The Generalized Pareto Distribution (GPD) with positive shape parameter $\gamma > 0$ and scale parameter $\sigma > 0$ is a member of the EGPD family: take $\tau = -1$ and $\delta = \gamma/\sigma - 1$. Finally, the distribution of the random variable Y is EPD(γ, δ, τ) if and only if the distribution of $Y - 1$ is EGPD(γ, δ, τ).

We will use the E(G)PD to model tails of heavy-tailed distributions that satisfy a certain second-order condition, to be described next. For a distribution function F, write $\overline{F} = 1 - F$. Recall that a positive, measurable function f defined in some right neighborhood of infinity is *regularly varying* with index $\beta \in \mathbb{R}$ if $\lim_{u\to\infty} f(ux)/f(u) = x^{\beta}$ for all $x \in (0, \infty)$; notation $f \in \mathscr{R}_{\beta}$. The following definition describes a subset of the class of distribution functions F for which $\overline{F} \in \mathscr{R}_{-1/\gamma}$, $\gamma > 0$. Note that the latter is precisely the class of distributions in the max-domain of attraction of the Fréchet distribution with shape parameter $1/\gamma$.

Definition 2.2 *Let* $\gamma > 0$ *and* $\tau < 0$ *be constants. A distribution function* F *is said to belong to the class* $\mathscr{F}(\gamma, \tau)$ *if* $x^{1/\gamma}\overline{F}(x) \to C \in (0, \infty)$ *as* $x \to \infty$ *and if the function* δ *defined via*

$$\overline{F}(x) = Cx^{-1/\gamma}\{1 + \gamma^{-1}\delta(x)\} \tag{2.1}$$

is eventually nonzero and of constant sign and such that $|\delta| \in \mathscr{R}_{\tau}$.

Note that $|\delta| \in \mathscr{R}_{\tau}$ with $\tau < 0$ implies $\delta(x) \to 0$ as $x \to \infty$. In many examples, the function δ in Definition 2.2 is actually of the form $\delta(x) \sim Dx^{\tau}$ as $x \to \infty$ for some nonzero constant D, a class of distributions which was first considered in Hall (1982). See Table 1 for examples; for later use, we also list $\rho = \gamma\tau$ (see Lemma 2.4 below).

distribution [parameters]	distribution function	γ	τ	$\rho = \gamma\tau$
Burr(γ, ρ, β) $[\gamma > 0,\ \rho < 0,\ \beta > 0]$	$1 - (1 + x^{-\rho/\gamma}/\beta)^{1/\rho}$	γ	ρ/γ	ρ
Fréchet(α) $[\alpha > 0]$	$\exp(-x^{-\alpha})$	$1/\alpha$	$-\alpha$	-1
GPD(γ, σ) $[\gamma > 0,\ \sigma > 0]$	$1 - (1 + \gamma x/\sigma)^{-1/\gamma}$	γ	-1	$-\gamma$
Student-t_ν $[\nu > 0]$	$C(\nu) \int_{-\infty}^{x} (1 + \frac{y^2}{\nu})^{-(\nu+1)/2} \, \mathrm{d}y$	$1/\nu$	-2	$-2/\nu$

Table 1
Extreme value index γ and second-order constants τ and $\rho = \gamma\tau$ for selected heavy-tailed distributions.

Let X be a random variable with distribution function F and let $u > 0$ be such that $F(u) < 1$. The conditional distributions of relative and absolute excesses of X over u are given by

$$\Pr(X/u > y \mid X > u) = \frac{\overline{F}(uy)}{\overline{F}(u)} \quad \text{and} \quad \Pr(X - u > x \mid X > u) = \frac{\overline{F}(u + x)}{\overline{F}(u)}$$

for $x \geqslant 0$ and $y \geqslant 1$. The next proposition shows that for $F \in \mathscr{F}(\gamma, \tau)$, the EPD and the EGPD improve the PD and GPD approximations to these excess distributions with an order of magnitude.

Proposition 2.3 *If $F \in \mathscr{F}(\gamma, \tau)$, then as $u \to \infty$,*

$$\sup_{y \geqslant 1} \left| \frac{\overline{F}(uy)}{\overline{F}(u)} - \overline{G}_{\gamma, \delta(u), \tau}(y) \right| = o\{|\delta(u)|\}, \tag{2.2}$$

$$\sup_{x \geqslant 0} \left| \frac{\overline{F}(u + x)}{\overline{F}(u)} - \overline{H}_{\gamma, \delta(u), \tau}(x/u) \right| = o\{|\delta(u)|\}. \tag{2.3}$$

Proof Equation (2.3) follows directly from (2.2) by writing $u + x = uy$ or $y = 1 + x/u$ and exploiting the link between the EPD and the EGPD. So let us show (2.2). On the one hand, we have

$$\frac{\overline{F}(uy)}{\overline{F}(u)} = y^{-1/\gamma} \frac{1 + \gamma^{-1}\delta(uy)}{1 + \gamma^{-1}\delta(u)} = y^{-1/\gamma} \left(1 - \gamma^{-1}\delta(u) \frac{1 - \frac{\delta(uy)}{\delta(u)}}{1 + \gamma^{-1}\delta(u)} \right).$$

On the other hand, since $0 \leqslant 1 - y^\tau \leqslant 1$ for $y \geqslant 1$ and since $\delta(u) \to 0$,

$$[y\{1 + \delta(u) - \delta(u)y^\tau\}]^{-1/\gamma} \\ = y^{-1/\gamma}\{1 - \gamma^{-1}\delta(u)(1 - y^\tau)\} + o\{|\delta(u)|\}, \qquad u \to \infty,$$

uniformly in $y \geqslant 1$. As a consequence,

$$\frac{\overline{F}(uy)}{\overline{F}(u)} - [y\{1 + \delta(u) - \delta(u)y^\tau\}]^{-1/\gamma} \\ = -\gamma^{-1}y^{-1/\gamma}\delta(u)\left(\frac{1 - \frac{\delta(uy)}{\delta(u)}}{1 + \gamma^{-1}\delta(u)} - (1 - y^\tau)\right) + o\{|\delta(u)|\}, \qquad u \to \infty,$$

uniformly in $y \geqslant 1$. The asymptotic relation (2.2) now follows from the uniform convergence theorem for regularly varying functions with negative index (Bingham et al., 1987, Theorem 1.5.2). □

If in (2.2) we would replace the EPD tail function $\overline{G}_{\gamma,\delta(u),\tau}(y)$ by the PD tail function $y^{-1/\gamma}$, the rate of convergence would be $O\{|\delta(u)|\}$ only. Similarly, if in (2.3) we would replace the EGPD tail function $\overline{H}_{\gamma,\delta(u),\tau}(x/u)$ by the GPD tail function $(1 + \gamma x/\sigma)^{-1/\gamma}$ for some $\sigma = \sigma(u)$, then, provided $\tau \neq -1$, the rate of convergence would again be $O\{|\delta(u)|\}$ only. If $\tau = -1$, the EGPD is just a reparametrization of the GPD, so that in that case, the GPD approximation is already of the order $o\{|\delta(u)|\}$.

It will be useful to rephrase our second-order assumption on F in terms of the tail quantile function U defined by

$$U(y) = Q(1 - 1/y) \quad \text{with} \quad Q(p) = \inf\{x \in \mathbb{R} : F(x) \geqslant p\}, \tag{2.4}$$

where $y \in (1, \infty)$ and $p \in (0, 1)$. Note that U is a (generalized) inverse of $1/\overline{F}$.

Lemma 2.4 *If $F \in \mathscr{F}(\gamma, \tau)$ with $\lim_{x\to\infty} x^{1/\gamma}\overline{F}(x) = C \in (0, \infty)$, then $\lim_{y\to\infty} y^{-\gamma}U(y) = C^\gamma$, and the function a defined implicitly by*

$$U(y) = C^\gamma y^\gamma\{1 + a(y)\} \tag{2.5}$$

satisfies $a(y) = \delta(U(y))\{1 + o(1)\} = \delta(C^\gamma y^\gamma)\{1 + o(1)\}$ as $y \to \infty$, with δ as in (2.1).

In particular, a is eventually nonzero and of constant sign and $|a| \in \mathscr{R}_\rho$ with $\rho = \gamma\tau < 0$. In addition, even if F is not continuous, then still $y\overline{F}(U(y)) = 1 + o\{|a(y)|\}$ as $y \to \infty$.

3 Parameter Estimation

Our aim is to make inference on the distribution function F on the region to the right of some high, positive threshold u. To this end, we assume $F \in \mathscr{F}(\gamma, \tau)$ and rewrite (2.2) as follows: as $u \to \infty$ and uniformly in $y \geqslant 1$,

$$\overline{F}(uy) = \overline{F}(u)\overline{G}_{\gamma,\delta(u),\tau}(y) + o\{\overline{F}(u)|\delta(u)|\}. \tag{3.1}$$

Omitting the remainder term leads to an approximation of $\overline{F}(x)$ for $x \geqslant u$ in terms of $\overline{F}(u)$ and the EPD parameters $(\gamma, \delta(u), \tau)$. Replacing these unknown quantities by estimates then yields our estimate for $\overline{F}(x)$.

The purpose of this section is to construct estimators of the E(G)PD parameters $(\gamma, \delta(u), \tau)$. As usual in extreme value statistics, the threshold exceedance probability $\overline{F}(u)$ will be estimated nonparametrically. Although the arguments leading to the estimators will be of a heuristic nature only, the asymptotic behaviour of the estimators will be stated and proved rigorously.

Let $X_1, \ldots, X_n$ be a random sample from F. In view of (2.2), the estimates of the EPD parameters will be based on the relative excesses X_i/u over u, for those $i \in \{1, \ldots, n\}$ such that $X_i > u$. In an extreme value asymptotic setting, the threshold u needs to tend to infinity to make the approximation valid; at the same time, in a statistical context, the number of excesses over u must be sufficiently large to make inference feasible. Denoting the order statistics by $X_{1:n} \leqslant \cdots \leqslant X_{n:n}$, we can ensure both criteria to be met by choosing a data-adaptive threshold $u = u_n = X_{n-k:n}$ where $k = k_n \in \{1, \ldots, n-1\}$ is an intermediate sequence of integers, that is, $k \to \infty$ and $k/n \to 0$ as $n \to \infty$. For convenience, assume $F(0) = 0$, so that all X_i are positive with probability one.

Recall the tail quantile function U in (2.4) and the auxiliary function a in Lemma 2.4. In addition to k being an intermediate sequence, we will assume that

$$\sqrt{k}a(n/k) \to \lambda \in \mathbb{R}, \qquad n \to \infty. \tag{3.2}$$

Writing $\delta_n = \delta(u_n) = \delta(X_{n-k:n})$, we will show later that (3.2) implies

$$\sqrt{k}\delta_n = \lambda + o_p(1), \qquad n \to \infty. \tag{3.3}$$

Since in the definition of the EPD the term x^τ is multiplied by δ, the previous display implies that the asymptotic distribution of tail estimators based on (3.1) will not depend on the asymptotic distribution of the estimator of τ, not even on its rate of convergence. Therefore, we will assume for the moment that τ (or ρ) is known. In the end, the unknown second-order parameters will be replaced by consistent estimators, a substitution which will be shown not to affect the asymptotic distributions of the other estimators. Note that

under the regime $\sqrt{k}|a(n/k)| \to \infty$ as $n \to \infty$, which will not be considered in this paper, the asymptotic distribution of the estimator of the second-order parameter does play a role.

The estimators of γ and δ_n will be found by maximizing an approximation to the EPD likelihood given the sample of k relative excesses $X_{n-k+i:n}/X_{n-k:n}$, $i \in \{1, \ldots, k\}$, over the random threshold $X_{n-k:n}$. The density function of the EPD is given by

$$g_{\gamma,\delta,\tau}(x) = \frac{1}{\gamma} x^{-1/\gamma-1} \{1 + \delta(1 - x^\tau)\}^{-1/\gamma-1} [1 + \delta\{1 - (1+\tau)x^\tau\}].$$

The score functions admit the following expansions in $\delta \to 0$:

$$\begin{aligned}
\frac{\partial}{\partial\gamma} \log g_{\gamma,\delta,\tau}(x) &= -\frac{1}{\gamma} + \frac{1}{\gamma^2} \log x + \frac{\delta}{\gamma^2}(1 - x^\tau) + O(\delta^2), \\
\frac{\partial}{\partial\delta} \log g_{\gamma,\delta,\tau}(x) &= \frac{1}{\gamma}\{(1-\gamma\tau)x^\tau - 1\} \\
&\qquad + \{1 - 2(1-\gamma\tau)x^\tau + (1 - 2\gamma\tau - \gamma\tau^2)x^{2\tau}\}\frac{\delta}{\gamma} + O(\delta^2).
\end{aligned}$$

Define

$$H_{k,n} = \frac{1}{k} \sum_{i=1}^{k} \log(X_{n-k+i:n}/X_{n-k:n}), \tag{3.4}$$

$$E_{k,n}(s) = \frac{1}{k} \sum_{i=1}^{k} (X_{n-k+i:n}/X_{n-k:n})^s, \qquad s \leqslant 0. \tag{3.5}$$

Note that $H_{k,n}$ is the Hill estimator (Hill, 1975). Assume for the moment that τ is known. Given the sample of excesses $X_{n-k+i:n}/X_{n-k:n}$, $i = 1, \ldots, k$, solving the linearized score equations yields the following equations for the pseudo-maximum likelihood estimators for γ and δ:

$$\hat{\gamma}_{k,n} = H_{k,n} + \hat{\delta}_{k,n}\{1 - E_{k,n}(\tau)\}, \tag{3.6}$$

$$\begin{aligned}
(\hat{\gamma}_{k,n}\tau - 1)E_{k,n}(\tau) + 1 = \{&1 - 2(1 - \hat{\gamma}_{n,k}\tau)E_{k,n}(\tau) \\
&+ (1 - 2\hat{\gamma}_{k,n}\tau - \hat{\gamma}_{k,n}\tau^2)E_{k,n}(2\tau)\}\hat{\delta}_{k,n}.
\end{aligned} \tag{3.7}$$

Substitute the expression for $\hat{\gamma}_{k,n}$ in (3.6) into the left-hand side of (3.7) and solve for $\hat{\delta}_{k,n}$ to get

$$\hat{\delta}_{k,n} = \frac{(H_{k,n}\tau - 1)E_{k,n}(\tau) + 1}{D_{k,n}} = \frac{H_{k,n}\tau - 1}{D_{k,n}} \left(E_{k,n}(\tau) - \frac{1}{1 - H_{k,n}\tau} \right),$$

the denominator being

$$\begin{aligned}
D_{k,n} = 1 - 2(1 - \hat{\gamma}_{n,k}\tau)E_{k,n}(\tau) + (1 - 2\hat{\gamma}_{k,n}\tau - \hat{\gamma}_{k,n}\tau^2)E_{k,n}(2\tau) \\
- \tau\{1 - E_{k,n}(\tau)\}E_{k,n}(\tau).
\end{aligned}$$

By (3.3), $\hat{\delta}_{k,n}$ can be expected to be of the order $O_p(k^{-1/2})$ as $n \to \infty$. This justifies the following simplifications. Since the distribution of relative excesses over a large threshold is approximately Pareto with shape parameter $1/\gamma$, for $s \leqslant 0$,

$$E_{k,n}(s) = \frac{1}{1-\gamma s} + o_p(1), \qquad n \to \infty;$$

see Theorem A.1. Hence, writing $\rho = \gamma\tau$, we have $E_{k,n}(\tau) = (1-\rho)^{-1} + o_p(1)$ and $E_{k,n}(2\tau) = (1-2\rho)^{-1} + o_p(1)$ as $n \to \infty$, so that

$$D_{k,n} = -\frac{\rho^4}{\gamma(1-2\rho)(1-\rho)^2} + o_p(1), \qquad n \to \infty.$$

This leads to the following simplified estimators:

$$\hat{\delta}_{k,n} = H_{k,n}(1-2\rho)(1-\rho)^3\rho^{-4}\left(E_{k,n}(\tau) - \frac{1}{1-H_{k,n}\tau}\right),$$
$$\hat{\gamma}_{k,n} = H_{k,n} - \hat{\delta}_{k,n}\frac{\rho}{1-\rho}.$$

Up to now we have assumed that ρ is known. Let $\hat{\rho}_n$ be a weakly consistent estimator sequence of $\rho = \gamma\tau$; see for instance Fraga Alves et al. (2003a), Fraga Alves et al. (2003b), and Peng and Qi (2004). Replace τ, which is unknown, by $\hat{\tau}_{k,n} = \hat{\rho}_n/H_{k,n}$, to finally get

$$\hat{\delta}_{k,n} = H_{k,n}(1-2\hat{\rho}_n)(1-\hat{\rho}_n)^3\hat{\rho}_n^{-4}\left(E_{k,n}(\hat{\rho}_n/H_{k,n}) - \frac{1}{1-\hat{\rho}_n}\right), \tag{3.8}$$

$$\hat{\gamma}_{k,n} = H_{k,n} - \hat{\delta}_{k,n}\frac{\hat{\rho}_n}{1-\hat{\rho}_n}. \tag{3.9}$$

Further, put

$$Z_{k,n} = \sqrt{k}\{n\overline{F}(X_{n-k:n})/k - 1\}. \tag{3.10}$$

The joint asymptotics of $Z_{k,n}$ with $(\hat{\gamma}_{k,n}, \hat{\delta}_{k,n})$ will become relevant in Section 5 when estimating tail probabilities on the basis of (3.1) with $u = X_{n-k:n}$. Let the arrow $\rightsquigarrow$ denote convergence in distribution.

Theorem 3.1 *Let $F \in \mathscr{F}(\gamma,\tau)$ and let $X_1,\ldots,X_n$ be independent random variables with common distribution function F. Let $k = k_n$ be an intermediate sequence satisfying (3.2). Recall $\delta_n = \delta(X_{n-k:n})$ and $Z_{k,n}$ in (3.10). If $\hat{\rho}_n = \rho + o_p(1)$ as $n \to \infty$, with $\rho = \gamma\tau$, then $\sqrt{k}\delta_n = \lambda + o_p(1)$ as $n \to \infty$ and*

$$\left(\sqrt{k}(\hat{\gamma}_{k,n} - \gamma), \sqrt{k}(\hat{\delta}_{k,n} - \delta_n), Z_{k,n}\right) \rightsquigarrow N_3(\mathbf{0}, \Sigma), \qquad n \to \infty, \tag{3.11}$$

a trivariate normal distribution with mean vector zero and covariance matrix

$$\Sigma = \begin{pmatrix} \gamma^2 \frac{(1-\rho)^2}{\rho^2} & -\gamma^2 \frac{(1-2\rho)(1-\rho)}{\rho^3} & 0 \\ -\gamma^2 \frac{(1-2\rho)(1-\rho)}{\rho^3} & \gamma^2 \frac{(1-2\rho)(1-\rho)^2}{\rho^4} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3.12}$$

An asymptotic confidence interval for γ of nominal level $1-\alpha$ is given by

$$\left[\hat{\gamma}_{k,n}\left(1+\frac{1-\hat{\rho}_n}{\hat{\rho}_n}\frac{z_{\alpha/2}}{\sqrt{k}}\right),\ \hat{\gamma}_{k,n}\left(1-\frac{1-\hat{\rho}_n}{\hat{\rho}_n}\frac{z_{\alpha/2}}{\sqrt{k}}\right)\right], \tag{3.13}$$

with $z_{\alpha/2}$ the $1-\alpha/2$ quantile of the standard normal distribution.

The proof of Theorem 3.1 is given in Appendix B. It is based on a functional central limit theorem for a certain tail empirical process, stated and proved in Appendix A. Note that the asymptotic distribution of $\hat{\rho}_{k,n}$ is unimportant; the only requirement is that the estimator is consistent for ρ.

The fact that the limit distribution in (3.11) is centered for any λ, is important for two reasons:

1 It makes possible the use of larger k and thus of lower thresholds compared to when the mean would be proportional to λ. In this way, the model can be fitted to a larger fraction of the data, leading to a reduction of the asymptotic variances and thus of the asymptotic mean squared errors of the parameter estimates.

2 Sample paths of the estimates as a function of k will exhibit larger regions of stability around the true value. As a consequence, the choice of k becomes easier.

These issues will be illustrated in the simulations in Section 4 and in the case study in Example 5.3.

4 Comparison of Extreme Value Index Estimators

Under the conditions of Theorem 3.1, we have

$$\sqrt{k}(\hat{\gamma}_{k,n}-\gamma) \rightsquigarrow N\left(0, \gamma^2\frac{(1-\rho)^2}{\rho^2}\right), \qquad n\to\infty. \tag{4.1}$$

According to Drees (1998), the asymptotic variance is minimal for scale-invariant, asymptotically unbiased estimators of γ of a certain form. The limit

distribution in (4.1) corresponds with the one of the estimators in Beirlant et al. (1999), Feuerverger and Hall (1999) and Gomes and Martins (2002).

The maximum likelihood estimator for γ arises from fitting the GPD to the excesses $X_{n-k+i:n} - X_{n-k:n}$, $i = 1, \ldots, k$. Its asymptotics have been studied in Smith (1987), Drees et al. (2004) and de Haan and Ferreira (2006, Theorem 3.4.2). From the latter theorem, it follows that under the conditions of our Theorem 3.1, we have

$$\sqrt{k}\left(\hat{\gamma}_{k,n}^{\mathrm{GPD}} - \gamma\right) \rightsquigarrow N\left(\lambda b(\gamma,\rho), (1+\gamma)^2\right), \qquad n \to \infty, \tag{4.2}$$

where

$$b(\gamma,\rho) = \frac{\rho(1+\gamma)(\gamma+\rho)}{\gamma(1-\rho)(1+\gamma-\rho)}.$$

Comparing (4.1) and (4.2), we see that if $\tau = -1$ and thus $\rho = -\gamma$, the asymptotic distributions of $\hat{\gamma}_{k,n}$ and $\hat{\gamma}_{k,n}^{\mathrm{GPD}}$ coincide. This is in correspondance with the fact that the EGPD with $\tau = -1$ is a reparametrization of the GPD and the fact that the EPD estimators were obtained by solving the linearized score equations.

Finally, under the conditions of Theorem 3.1, the asymptotic distribution of the Hill estimator is

$$\sqrt{k}(H_{k,n} - \gamma) \rightsquigarrow N\left(\lambda\frac{\rho}{1-\rho}, \gamma^2\right), \qquad n \to \infty; \tag{4.3}$$

see for instance Theorem A.1 below. Of the three estimators considered, the Hill estimator has the smallest asymptotic variance. Unless $\lambda = 0$, however, its asymptotic bias is never zero. The asymptotic distribution of the Hill estimator and its optimal variance property are of course well known; see for instance Reiss (1989, Section 9.4), Drees (1998) and Beirlant et al. (2006).

To illustrate the behavior of the three estimators, we generated samples from four different distributions. For each distribution, we generated $10,000$ samples of size $n = 1,000$ and computed the three extreme value index estimators for k up to 500. For the EPD estimator, we estimated the second-order parameter ρ using the estimator in Fraga Alves et al. (2003b). For each distribution and each estimator, we computed Monte Carlo estimates of the bias, variance and mean squared error by averaging out over the $10,000$ samples.

Comparing the asymptotic results to the graphs in Figures 1–2 we learn the following:

Fréchet distribution with $\alpha = 1$. We have $\gamma = 1/\alpha = 1$, $\tau = -\alpha = -1$, and $\rho = \gamma\tau = -1$. From (4.1) and (4.2), it follows that the asymptotic distributions of the EPD and the GPD estimators coincide, with zero asymptotic

bias and an asymptotic variance of $4/k$. The Hill estimator has an asymptotic variance of $1/k$ only, but its asymptotic bias is nonzero.

Student t *distribution* with $\nu = 4$. We have $\gamma = 1/\nu = 1/4$, $\tau = -2$, and $\rho = \gamma\tau = -1/2$. The asymptotic variances of the three estimators are σ^2/k with $\sigma^2 = \gamma^2 = 1/16$ for the Hill estimator, $\sigma^2 = \gamma^2(1-\rho)^2/\rho^2 = 9/16$ for the EPD estimator, and $\sigma^2 = (1+\gamma)^2 = 25/16$ for the GPD estimator. Of the three estimators, the EPD estimator is the only one which is asymptotically unbiased.

Pareto mixture distribution defined by $\overline{F}(x) = (1+c)^{-1}x^{-\alpha}(1+cx^{-\alpha})$, $x \geqslant 1$, with shape parameter $\alpha = 2$ and mixing parameter $c = 2$. We have $\gamma = 1/\alpha = 1/2$, $\tau = -\alpha = -2$, and $\rho = \gamma\tau = -1$. The weight of the second-order component is equal to $c = 2$ times the weight of the first-order component, inducing a severe bias to the Hill and GPD estimators; the EPD estimator is much less affected by this. The asymptotic variances of the three estimators are σ^2/k with $\sigma^2 = \gamma^2 = 1/4$ for the Hill estimator, $\sigma^2 = \gamma^2(1-\rho)^2/\rho^2 = 1$ for the EPD estimator, and $\sigma^2 = (1+\gamma)^2 = 9/4$ for the GPD estimator.

Loggamma distribution with shape parameter $\alpha = 4$ and scale parameter $\beta = 2$. Although this distribution has positive extreme-value index $\gamma = 1/\beta$, it is not in any of the classes $\mathscr{F}(\gamma, \tau)$, since $\overline{F}(x) \sim \text{constant} \times x^{-1/\beta}(\log x)^{\alpha-1}$. Nevertheless, the EPD estimator performs reasonably well when compared to the Hill and GPD estimators.

5 Tail Probability Estimation

Let us return to the tail estimation problem raised in the beginning of Section 3. Given the order statistics $X_{1:n} \leqslant \cdots \leqslant X_{n:n}$ of an independent sample from an unknown distribution function $F \in \mathscr{F}(\gamma, \tau)$, we want to estimate the tail probability $p_n = \overline{F}(x_n)$, where $x_n \to \infty$ and thus $p_n \to 0$ as $n \to \infty$. As before, let $k = k_n \in \{1, \ldots, n-1\}$ be an intermediate integer sequence, that is, $k \to \infty$ and $k/n \to 0$. Assume that $p_n = \overline{F}(x_n)$ satisfies

$$np_n/k \to q \in [0, 1), \qquad n \to \infty. \tag{5.1}$$

Let $\hat{\gamma}_n$, $\hat{\delta}_n$, and $\hat{\tau}_n$ denote general estimator sequences and put $\delta_n = \delta(X_{n-k:n})$ as well as

$$\Gamma_{k,n} = \sqrt{k}(\hat{\gamma}_n - \gamma) \qquad \text{and} \qquad \Delta_{k,n} = \sqrt{k}(\hat{\delta}_n - \delta_n). \tag{5.2}$$

Recall $Z_{k,n}$ in (3.10) and assume that

$$\hat{\tau}_n = \tau + o_p(1) \quad \text{and} \quad (\Gamma_{k,n}, \Delta_{k,n}, Z_{k,n}) \rightsquigarrow (\Gamma, \Delta, Z), \qquad n \to \infty, \tag{5.3}$$

a trivariate random vector. A possible choice for the estimators of γ and δ_n are the ones studied in Theorem 3.1. However, we will formulate our results so

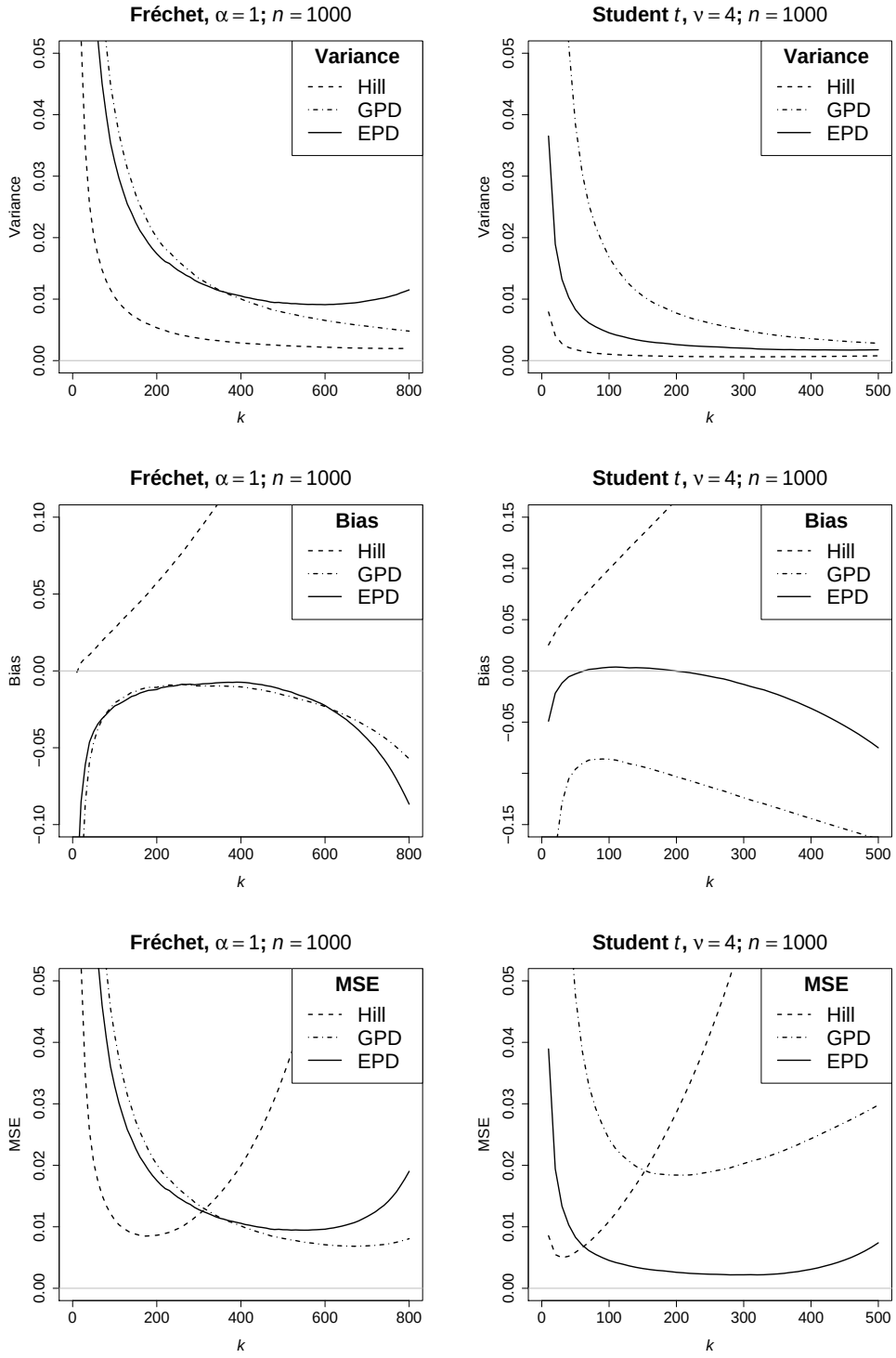

Fig. 1. *Variance (top), bias (middle), and mean squared error (bottom) of the Hill (dashed), GPD (dotdashed), and EPD (solid) estimator in case of the unit Fréchet distribution (left) and the Student t distribution with* $\nu = 4$ *degrees of freedom (right). The sample size was* $n = 1,000$ *and the plots were obtained by averaging out over* $10,000$ *samples.*

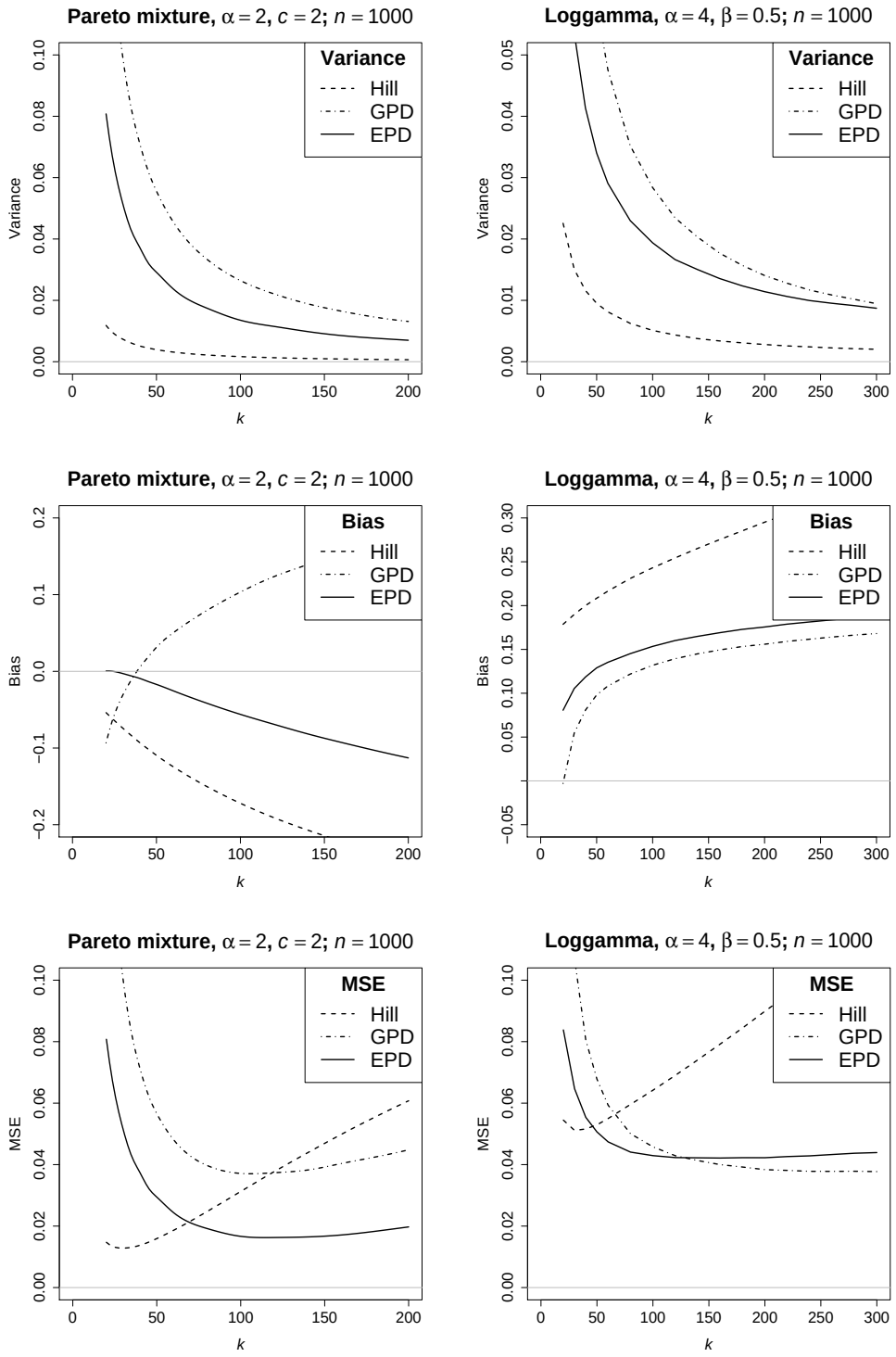

Fig. 2. *Variance (top), bias (middle), and mean squared error (bottom) of the Hill (dashed), GPD (dotdashed), and EPD (solid) estimator in case of a Pareto mixture distribution ($\alpha = 2$, $c = 2$; left) and a Loggamma distribution ($\alpha = 4$, $\beta = 2$; right). The sample size was $n = 1,000$ and the plots were obtained by averaging out over $10,000$ samples.*

as to allow for general estimator sequences satisfying (5.3). For the estimator of τ, one can for instance take $\hat{\tau}_n = \hat{\rho}_n/\hat{\gamma}_n$, where $\hat{\rho}_n$ is an estimator of $\rho = \gamma\tau$, see for instance Fraga Alves et al. (2003b). As in Theorem 3.1, the asymptotic distribution of $\hat{\tau}_n$ plays no role.

Omitting the remainder term in (3.1) and replacing the unknown quantities $\overline{F}(u)$ and $(\gamma, \delta(u), \tau)$ at the random threshold $u = X_{n-k:n}$ by k/n and $(\hat{\gamma}_n, \hat{\delta}_n, \hat{\tau}_n)$, respectively, yields the estimator

$$\hat{p}_{k,n} = \hat{\overline{F}}_n(x_n) = \frac{k}{n}\overline{G}_{\hat{\gamma}_n,\hat{\delta}_n,\hat{\tau}_n}(x_n/X_{n-k:n}).$$

In the same way, one can construct estimators for other tail quantities: return levels, expected shortfall, etc. For brevity, we focus here on tail probabilities.

In order to describe the asymptotics of $\hat{p}_{k,n}$, we need to make a distinction between the case $0 < q < 1$ in (5.1) and $q = 0$. The proofs of the following two theorems are to be found in Appendix C. Results for tail probability estimators based on the PD and GPD approximations can be found in de Haan and Ferreira (2006, Section 4.4).

Theorem 5.1 *Let $F \in \mathscr{F}(\gamma, \tau)$, let k_n be an intermediate sequence satisfying* (3.2) *and let p_n be such that* (5.1) *holds for some $0 < q < 1$. If* (5.3)*, then*

$$\sqrt{k}\left(\frac{\hat{p}_{k,n}}{p_n} - 1\right) \rightsquigarrow -\gamma^{-1}\Gamma\log q - \gamma^{-1}\Delta(1 - q^{-\rho}) - Z, \qquad n \to \infty. \tag{5.4}$$

Theorem 5.2 *In Theorem 5.1, if* (5.1) *is replaced by*

$$np_n/k \to 0 \qquad \textit{and} \qquad \log(np_n)/\sqrt{k} \to 0, \qquad n \to \infty,$$

then

$$\frac{\sqrt{k}}{\log\{k/(np_n)\}}\left(\frac{\hat{p}_{k,n}}{p_n} - 1\right) \rightsquigarrow \gamma^{-1}\Gamma, \qquad n \to \infty.$$

For the EPD estimators $\hat{\gamma}_n = \hat{\gamma}_{k,n}$ and $\hat{\delta}_n = \hat{\delta}_{k,n}$, Theorems 3.1 and 5.1 lead to

$$\sqrt{k}\left(\frac{\hat{p}_n}{p_n} - 1\right) \rightsquigarrow N\left(0, \sigma^2(q,\rho)\right), \qquad n \to \infty, \tag{5.5}$$

with asymptotic variance given by

$$\sigma^2(q,\rho) = (\log q)^2\frac{(1-\rho)^2}{\rho^2} + \left(\frac{1-q^{-\rho}}{\rho}\right)^2\frac{(1-2\rho)(1-\rho)^2}{\rho^2} - 2\log(q)\frac{1-q^{-\rho}}{\rho}\frac{(1-2\rho)(1-\rho)}{\rho^2} + 1.$$

The importance of the fact that the limit distribution in (5.5) has mean zero was already discussed after Theorem 3.1. An asymptotic confidence interval of nominal level $1-\alpha$ is given by

$$\left[\hat{p}_n\left(1-\sigma(\hat{q}_n,\hat{\rho}_n)\frac{z_{\alpha/2}}{\sqrt{k}}\right),\ \hat{p}_n\left(1+\sigma(\hat{q}_n,\hat{\rho}_n)\frac{z_{\alpha/2}}{\sqrt{k}}\right)\right] \tag{5.6}$$

where $\hat{q}_n = n\hat{p}_n/k_n$ and with $z_{\alpha/2}$ the $1-\alpha/2$ quantile of the standard normal distribution.

If we simply define $\hat{\delta}_n = 0$, then $\Delta_{k,n} = -\sqrt{k}\delta_n$ in (5.2) and thus $\Delta = -\lambda$ in (5.3). The tail probability estimator $\hat{p}_n$ then reduces to the Weissman estimator (Weissman, 1978)

$$\hat{p}_n^{\mathrm{W}} = \frac{k}{n}\left(\frac{x_n}{X_{n-k:n}}\right)^{-1/\hat{\gamma}_n}. \tag{5.7}$$

Theorem 5.1 then implies

$$\sqrt{k}\left(\frac{\hat{p}_n^{\mathrm{W}}}{p_n}-1\right) \rightsquigarrow -\gamma^{-1}\Gamma\log q + \gamma^{-1}\lambda(1-q^{-\gamma\tau}) - Z, \qquad n\to\infty. \tag{5.8}$$

For instance, if we estimate γ by the Hill estimator, then in view of Theorem A.1,

$$\sqrt{k}\left(\frac{k}{np_n}\left(\frac{x_n}{X_{n-k:n}}\right)^{-1/H_{k,n}}-1\right) \rightsquigarrow N\left(-\lambda\frac{\rho}{\gamma}\left(\frac{q^{-\rho}-1}{\rho}+\frac{\log q}{1-\rho}\right), 1+(\log q)^2\right), \qquad n\to\infty.$$

Even if the extreme value index estimator $\hat{\gamma}_n$ is such that the asymptotic distribution of Γ_n has mean zero, then still the asymptotic distribution (5.8) of the Weissman estimator will have a mean which is proportional to λ. In other words, unbiased tail estimation requires more than unbiased estimation of the extreme value index alone.

From Theorem 5.2 and its proof, we learn that for estimation of tail probabilities p_n of smaller order than k/n, the difference between the Pareto approximation and the EPD approximation does not matter asymptotically. Still, for $\hat{p}_n$ to be an asymptotically unbiased estimator of p_n, the estimator $\hat{\gamma}_n$ needs to be asymptotically unbiased for γ. For instance, if we use the EPD estimator $\hat{\gamma}_{k,n}$, then

$$\frac{\sqrt{k}}{\log\{k/(np_n)\}}\left(\frac{\hat{p}_n}{p_n}-1\right) \rightsquigarrow N\left(0,\frac{(1-\rho)^2}{\rho^2}\right), \qquad n\to\infty.$$

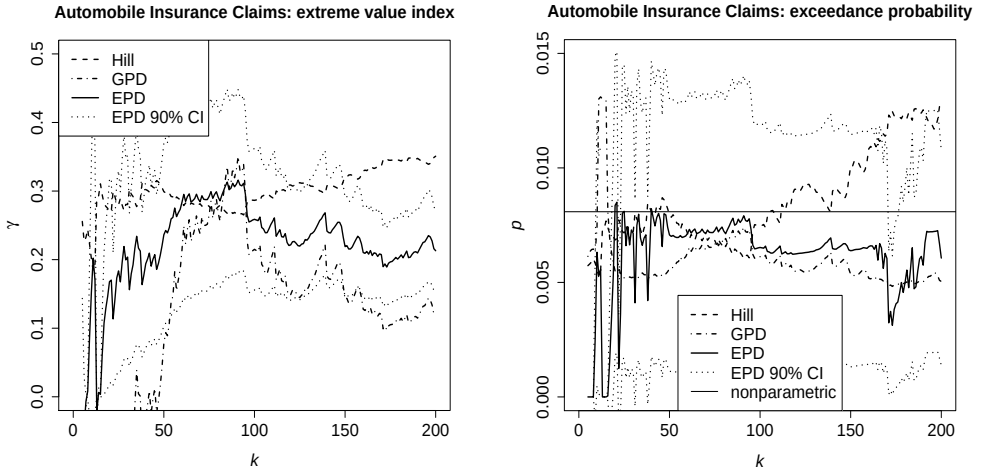

Fig. 3. *Trajectories of estimates of γ (left) and of the exceedance probability over €7 million (right) for the Secura Belgian Re data in Example 5.3.*

Example 5.3 The Secura Belgian Re data in Beirlant et al. (2004, Section 1.3.3) comprise 371 automobile claims not smaller than €1.2 million. The data span the period 1988–2001 and have been gathered from several European insurance companies. Figure 3 shows the estimates of γ (left) and of the probability of a claim to exceed €7 million (right). Nominal 90 % confidence intervals for the EPD estimates are added too, see (3.13) and (5.6). In the data-set, there were actually 3 exceedances over €1.2 million, yielding a nonparametric estimate of $3/371 = 0.81\%$. In comparison to the Weissman (Hill) and POT (GPD) estimates, the trajectories of the EPD estimates are relatively stable, with $\hat{\gamma}$ around 0.3 and p around 0.75%. By way of comparison, in Beirlant et al. (2004, Section 6.2.4) it is suggested to model the complete distribution by a mixture of two components, an exponential and a Pareto distribution, with the knot at about €2.6 million, which corresponds to the order statistic $X_{n-k:n}$ with $k = 95$. Although this knot is detected by the EPD estimator, it does not cause the tail parameter estimates to change dramatically.

A Tail Empirical Processes

Recall $H_{k,n}$, $E_{k,n}(s)$ and $Z_{k,n}$ from equations (3.4), (3.5) and (3.10), respectively, and define

$$\Gamma_{k,n} = \sqrt{k}(H_{k,n} - \gamma), \tag{A.1}$$

$$\mathbb{E}_{k,n}(s) = \sqrt{k}\left(E_{k,n}(s) - \frac{1}{1-s\gamma}\right), \qquad s \leqslant 0. \tag{A.2}$$

Our proof of Theorem 3.1 will be based on the fact that $(\Gamma_{k,n}, \mathbb{E}_{k,n}, Z_{k,n})$ converges weakly in the space $\mathbb{R} \times \mathscr{C}[s_0, 0] \times \mathbb{R}$; here $s_0 < 0$ and $\mathscr{C}[a, b]$ is the Banach space of continuous functions $f : [a, b] \to \mathbb{R}$ equipped with the topology of uniform convergence. Of course, the asymptotic distribution of the normalized Hill estimator $\Gamma_{k,n}$ has been established in numerous other papers; in the following theorem, it is the joint convergence which is our main concern.

Theorem A.1 *Let $F \in \mathscr{F}(\gamma, \tau)$. If $k = k_n$ is an intermediate integer sequence satisfying (3.2), then for every $s_0 < 0$, in $\mathbb{R} \times \mathscr{C}[-s_0, 0] \times \mathbb{R}$,*

$$(\Gamma_{k,n}, \mathbb{E}_{k,n}, Z_{k,n}) \rightsquigarrow (\Gamma, \mathbb{E}, Z), \qquad n \to \infty,$$

a Gaussian process with the following distribution: Z is standard normal and is independent of $(\Gamma, \mathbb{E})$, and for $s, s_1, s_2 \in [s_0, 0]$,

$$\begin{aligned} \mathrm{E}[\mathbb{E}(s)] &= \lambda \frac{s\rho}{(1 - s\gamma - \rho)(1 - s\gamma)}, & \mathrm{E}[\Gamma] &= \lambda \frac{\rho}{1 - \rho}, \\ \mathrm{cov}\{\mathbb{E}(s_1), \mathbb{E}(s_2)\} &= \frac{s_1 s_2 \gamma^2}{(1 - s_1\gamma - s_2\gamma)(1 - s_1\gamma)(1 - s_2\gamma)}, & \mathrm{var}(\Gamma) &= \gamma^2, \\ \mathrm{cov}\{\Gamma, \mathbb{E}(s)\} &= \frac{s\gamma^2}{(1 - s\gamma)^2}. \end{aligned}$$

Proof Let $Y_1, \check{Y}_1, Y_2, \check{Y}_2, \ldots$ be independent Pareto(1) random variables. For positive integer k, denote the order statistics of $Y_1, \ldots, Y_k$ by $Y_{1:k} < \cdots < Y_{k:k}$; also, let $Y_{0:k} = 1$. Similarly, denote the order statistics of $\check{Y}_1, \ldots, \check{Y}_n$ by $\check{Y}_{1:n} < \cdots < \check{Y}_{n:n}$. Then the following three vectors are equal in distribution:

$$\begin{aligned} (X_{n-k+i:n} : i = 0, \ldots, k) &\stackrel{d}{=} (U(\check{Y}_{n-k+i:n}) : i = 0, \ldots, k) \\ &\stackrel{d}{=} (U(Y_{i:k}\check{Y}_{n-k:n}) : i = 0, \ldots, k). \end{aligned}$$

Since we are only interested in the asymptotic distribution of $(\Gamma_{k,n}, \mathbb{E}_{k,n}, Z_{k,n})$, we may without loss of generality assume that actually

$$(X_{n-k+i:n} : i = 0, \ldots, k) = (U(Y_{i:k}\check{Y}_{n-k:n}) : i = 0, \ldots, k).$$

The following property is well-known: if k is an intermediate sequence, then

$$\sqrt{k}\{(n/k)\check{Y}_{n-k:n}^{-1} - 1\} \rightsquigarrow N(0, 1), \qquad n \to \infty. \tag{A.3}$$

[A quick proof is to employ the distributional representation $\check{Y}_{n-k:n} \stackrel{d}{=} (E_1 + \cdots + E_{n+1})/(E_1 + \ldots + E_k)$, with $E_1, \ldots, E_n$ independent standard exponential random variables.] As a consequence, we have $\check{Y}_{n-k:n} = (n/k)\{1 + o_p(1)\}$ as $n \to \infty$, and therefore, by (3.2) and the Uniform Convergence Theorem for

$\mathscr{R}_\rho$ (Bingham et al., 1987, Theorem 1.5.2),

$$\sqrt{k}a(\check{Y}_{n-k:n}) = \sqrt{k}a(n/k)\frac{a(\check{Y}_{n-k:n})}{a(n/k)} = \lambda + o_p(1), \qquad n \to \infty. \tag{A.4}$$

Since $a(y) \sim \delta(U(y))$ as $n \to \infty$, this also shows that $\sqrt{k}\delta(X_{n-k:n}) = \lambda + o_p(1)$ as $n \to \infty$.

In the next three paragraphs, we will analyse the components $\Gamma_{k,n}$, $\mathbb{E}_{k,n}$ and $Z_{k,n}$ separately. In the fourth and final paragraph, these analyses will be combined.

1. The component $\Gamma_{k,n}$. Let the function a be as in (2.5) and define $\eta(y) = \log\{1+a(y)\}$. Since $\lim_{y\to\infty} a(y) = 0$, we have $\eta(y) = a(y)\{1+o(1)\}$ as $y \to \infty$, and hence

$$\sqrt{k}\eta(\check{Y}_{n-k:n}) = \lambda + o_p(1), \qquad n \to \infty. \tag{A.5}$$

In particular, η is eventually nonzero and of constant sign, and $|\eta| \in \mathscr{R}_\rho$. We have

$$\begin{aligned} H_{k,n} &= \frac{1}{k}\sum_{i=1}^{k} \log X_{n-k+i:n} - \log X_{n-k:n} \\ &= \frac{1}{k}\sum_{i=1}^{k} \log U(Y_i\check{Y}_{n-k:n}) - \log U(\check{Y}_{n-k:n}) \\ &= \frac{1}{k}\sum_{i=1}^{k} \{\gamma \log Y_i + \eta(Y_i\check{Y}_{n-k:n}) - \eta(\check{Y}_{n-k:n})\}. \end{aligned}$$

As a consequence,

$$\begin{aligned} \Gamma_{k,n} &= \sqrt{k}(H_{k,n} - \gamma) \\ &= \frac{\gamma}{\sqrt{k}}\sum_{i=1}^{k}(\log Y_i - 1) + \sqrt{k}\eta(\check{Y}_{n-k:n})\frac{1}{k}\sum_{i=1}^{k}\left(\frac{\eta(Y_i\check{Y}_{n-k:n})}{\eta(\check{Y}_{n-k:n})} - 1\right). \end{aligned}$$

By the Uniform Convergence Theorem for $\mathscr{R}_\rho$, for every $x_0 > 0$,

$$\lim_{y\to\infty} \sup_{x \geqslant x_0} \left| \frac{\eta(xy)}{\eta(y)} - x^\rho \right| = 0.$$

By the last two displays and in view of (A.5),

$$\max_{i=1,\ldots,k} \left| \frac{\eta(Y_i\check{Y}_{n-k:n})}{\eta(\check{Y}_{n-k:n})} - Y_i^\rho \right| = o_p(1), \qquad n \to \infty. \tag{A.6}$$

By (A.5) and since $k^{-1}\sum_{i=1}^{k} Y_i^\rho = (1-\rho)^{-1} + o_p(1)$ as $k \to \infty$, we find

$$\Gamma_{k,n} = \frac{\gamma}{\sqrt{k}}\sum_{i=1}^{k}(\log Y_i - 1) + \lambda\left(\frac{1}{1-\rho} - 1\right) + o_p(1), \qquad n \to \infty. \tag{A.7}$$

2. The component $\mathbb{E}_{k,n}$. Recall the notation $\eta(y) = \log\{1 + a(y)\}$, so that $U(y) = C^\gamma y^\gamma \exp\{\eta(y)\}$. We have

$$E_{k,n}(s) = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{X_{n-k+i:n}}{X_{n-k:n}}\right)^s = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{U(Y_i\check{Y}_{n-k:n})}{U(\check{Y}_{n-k:n})}\right)^s$$
$$= \frac{1}{k}\sum_{i=1}^{k} Y_i^{\gamma s}\exp[s\{\eta(Y_i\check{Y}_{n-k:n}) - \eta(\check{Y}_{n-k:n})\}].$$

Writing $\varepsilon_{i,n} = \eta(Y_i\check{Y}_{n-k:n})/\eta(\check{Y}_{n-k:n}) - Y_i^\rho$, we find

$$E_{k,n}(s) = \frac{1}{k}\sum_{i=1}^{k} Y_i^{\gamma s}\exp\{s\eta(\check{Y}_{n-k:n})(Y_i^\rho - 1 + \varepsilon_{i,n})\}.$$

Recall the elementary inequality $|e^z - 1 - z| \leqslant (z^2/2)\max(e^z, 1)$, $z \in \mathbb{R}$. Since $0 < Y_i^{\gamma s} \leqslant 1$, $0 < Y_i^\rho \leqslant 1$ and $\max_{i=1,\ldots,n}|\varepsilon_{i,n}| = o_p(1)$ [see (A.6)], we get by (A.5),

$$\sup_{s\in[s_0,0]}\left|E_{k,n}(s) - \frac{1}{k}\sum_{i=1}^{k} Y_i^{\gamma s}\{1 + s\eta(\check{Y}_{n-k:n})(Y_i^\rho - 1)\}\right|$$
$$= o_p\{|\eta(\check{Y}_{n-k:n})|\} = o_p(k^{-1/2}), \qquad n \to \infty.$$

For $\theta_0 < 0$, the class of functions $\{f_\theta : \theta \in [\theta_0, 0]\}$ from $[1, \infty)$ to $(0, 1]$ defined by $f_\theta(y) = y^\theta$, $y \geqslant 1$, satisfies the Glivenko-Cantelli property

$$\sup_{\theta\in[\theta_0,0]}\left|\frac{1}{k}\sum_{i=1}^{k} Y_i^\theta - \frac{1}{1-\theta}\right| = o_p(1), \qquad k \to \infty;$$

see for instance Example 19.8 in van der Vaart (1998) or just use the monotonicity and continuity of y^θ in θ. In view of (A.5), we obtain

$$\sup_{s\in[s_0,0]}\left|E_{k,n}(s) - \frac{1}{k}\sum_{i=1}^{k} Y_i^{\gamma s} - s\eta(\check{Y}_{n-k:n})\left(\frac{1}{1-\gamma s-\rho} - \frac{1}{1-\gamma s}\right)\right|$$
$$= o_p\{|\eta(\check{Y}_{n-k:n})|\} = o_p(k^{-1/2}), \qquad n \to \infty.$$

Using (A.5) again, we find

$$\mathbb{E}_{k,n}(s) = \sqrt{k}\left(E_{k,n}(s) - \frac{1}{1-\gamma s}\right)$$
$$= \frac{1}{\sqrt{k}}\sum_{i=1}^{k}\left(Y_i^{\gamma s} - \frac{1}{1-\gamma s}\right) + s\lambda\left(\frac{1}{1-\gamma s-\rho} - \frac{1}{1-\gamma s}\right) + \varepsilon_n(s),$$

with

$$\sup_{s\in[s_0,0]}|\varepsilon_n(s)| = o_p(1), \qquad n \to \infty. \tag{A.8}$$

3. The component $Z_{k,n}$. By (2.1) and (2.5), we find

$$y\overline{F}(U(y)) = 1 + o\{|a(y)|\}, \qquad y \to \infty.$$

As a consequence,

$$\begin{aligned}\overline{F}(X_{n-k:n}) &= \overline{F}(U(\check{Y}_{n-k:n})) \\ &= \check{Y}_{n-k:n}^{-1}[1 + o_p\{|a(\check{Y}_{n-k:n})|\}] \\ &= \check{Y}_{n-k:n}^{-1}\{1 + o_p(k^{-1/2})\}, \qquad n \to \infty,\end{aligned}$$

where we used (A.4) in the last step. We obtain

$$\begin{aligned}Z_{k,n} &= \sqrt{k}\{(n/k)\overline{F}(X_{n-k:n}) - 1\} \\ &= \sqrt{k}\{(n/k)\check{Y}_{n-k:n}^{-1} - 1\} + o_p(1), \qquad n \to \infty. \end{aligned} \tag{A.9}$$

4. Joint convergence. Define

$$\begin{aligned}\tilde{\Gamma}_k &= \frac{\gamma}{\sqrt{k}} \sum_{i=1}^{k} (\log Y_i - 1), \\ \tilde{\mathbb{E}}_k(s) &= \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \left(Y_i^{\gamma s} - \frac{1}{1-\gamma s} \right), \qquad s \in [s_0, 0].\end{aligned}$$

For $\theta_0 < 0$, the class of functions $\{f_\theta : \theta \in [\theta_0, 0]\}$ defined by $f_\theta(y) = y^\theta$, $y \geqslant 1$, is Donsker with respect to the Pareto(1) distribution; this follows from Example 19.7 in van der Vaart (1998) upon noting that $|y^{\theta_1} - y^{\theta_2}| \leqslant |\theta_1 - \theta_2| \log y$ for $\theta_1 \leqslant 0$, $\theta_2 \leqslant 0$ and $y \geqslant 1$. As a consequence, in $\mathbb{R} \times \mathscr{C}[s_0, 0]$,

$$(\tilde{\Gamma}_k, \tilde{\mathbb{E}}_k) \rightsquigarrow (\tilde{\Gamma}, \tilde{\mathbb{E}}), \qquad k \to \infty, \tag{A.10}$$

a centered Gaussian process with covariance function

$$\begin{aligned}\operatorname{var} \tilde{\Gamma} &= \operatorname{var}(\gamma \log Y_1) = \gamma^2, \\ \operatorname{cov}\{\tilde{\mathbb{E}}(s_1), \tilde{\mathbb{E}}(s_2)\} &= \operatorname{cov}(Y_1^{\gamma s_1}, Y_1^{\gamma s_2}) = \frac{1}{1 - s_1\gamma - s_2\gamma} - \frac{1}{(1 - s_1\gamma)(1 - s_2\gamma)}, \\ \operatorname{cov}\{\tilde{\Gamma}, \tilde{\mathbb{E}}(s)\} &= \operatorname{cov}(\gamma \log Y_1, Y_1^{\gamma s}) = \frac{s\gamma^2}{(1 - s\gamma)^2}.\end{aligned}$$

By (A.7) and (A.8), it follows that in $\mathbb{R} \times \mathscr{C}[s_0, 0]$,

$$(\Gamma_{k,n}, \mathbb{E}_{k,n}) \rightsquigarrow (\Gamma, \mathbb{E}), \qquad n \to \infty.$$

Finally, from (A.3) and (A.9) it follows that $(\Gamma_{k,n}, \mathbb{E}_{k,n}, Z_{k,n}) \rightsquigarrow (\Gamma, \mathbb{E}, Z)$ as $n \to \infty$, where Z is standard normally distributed and is independent of $(\Gamma, \mathbb{E})$. □

B Proof of Theorem 3.1

The fact that $\sqrt{k}\delta(X_{n-k:n}) = \lambda + o_p(1)$ as $n \to \infty$ has already been shown in the proof of Theorem A.1; in particular, see (A.4). Recall $\Gamma_{k,n}$ and $\mathbb{E}_{k,n}(s)$ in equations (A.1) and (A.2), respectively, and write $\hat{\tau}_{k,n} = \hat{\rho}_n / H_{k,n}$. We have

$$\begin{aligned}
&\sqrt{k}\left(E_{k,n}(\hat{\tau}_{k,n}) - \frac{1}{1-\hat{\rho}_n}\right) \\
&= \sqrt{k}\left(E_{k,n}(\hat{\tau}_{k,n}) - \frac{1}{1-\gamma\hat{\tau}_n}\right) + \sqrt{k}\left(\frac{1}{1-\gamma\hat{\rho}_n/H_{k,n}} - \frac{1}{1-\hat{\rho}_n}\right) \\
&= \mathbb{E}_{k,n}(\hat{\tau}_{k,n}) - \frac{1}{H_{k,n}}\frac{\hat{\rho}_n}{(1-\gamma\hat{\rho}_{k,n}/H_{k,n})(1-\hat{\rho}_n)}\Gamma_{k,n}.
\end{aligned}$$

By Theorem A.1, $H_{k,n} = \gamma + k^{-1/2}\Gamma_{k,n} = \gamma + o_p(1)$ and thus $\hat{\tau}_{k,n} = \tau + o_p(1)$ as $n \to \infty$. It follows that

$$\sqrt{k}\left(E_{k,n}(\hat{\tau}_{k,n}) - \frac{1}{1-\hat{\rho}_n}\right) = \mathbb{E}_{k,n}(\hat{\tau}_{k,n}) - \frac{\rho}{\gamma(1-\rho)^2}\Gamma_{k,n} + o_p(1), \qquad n \to \infty.$$

Substituting this into the definition of $\hat{\delta}_{k,n}$ yields

$$\sqrt{k}\hat{\delta}_{k,n} = \gamma(1-2\rho)(1-\rho)^3\rho^{-4}\left(\mathbb{E}_{k,n}(\hat{\tau}_{k,n}) - \frac{\rho}{\gamma(1-\rho)^2}\Gamma_{k,n}\right) + o_p(1), \qquad n \to \infty, \quad \text{(B.1)}$$

as well as

$$\begin{aligned}
\sqrt{k}(\hat{\gamma}_{k,n} - \gamma) &= \sqrt{k}\left(H_{k,n} - \hat{\delta}_{k,n}\frac{\hat{\rho}_n}{1-\hat{\rho}_n} - \gamma\right) \\
&= \Gamma_{k,n} - \sqrt{k}\hat{\delta}_{k,n}\frac{\rho}{1-\rho} + o_p(1) \\
&= \frac{(1-\rho)^2}{\rho^2}\left(\Gamma_{k,n} - \gamma\frac{1-2\rho}{\rho}\mathbb{E}_{k,n}(\hat{\tau}_{k,n})\right) + o_p(1), \qquad n \to \infty.
\end{aligned} \quad \text{(B.2)}$$

From $\hat{\tau}_{k,n} = \tau + o_p(1)$ and Theorem A.1, it follows that in $\mathbb{R} \times \mathscr{C}[s_0, 0] \times \mathbb{R} \times \mathbb{R}$,

$$(\Gamma_{k,n}, \mathbb{E}_{k,n}, Z_{k,n}, \hat{\tau}_{n,k}) \rightsquigarrow (\Gamma, \mathbb{E}, Z, \tau), \qquad n \to \infty.$$

For $s_0 < \tau$, we have $\Pr(s_0 \leqslant \hat{\tau}_{n,k} \leqslant 0) \to 1$ as $n \to \infty$, and thus, by the previous display and the continuous mapping theorem,

$$(\Gamma_{k,n}, \mathbb{E}_{k,n}(\hat{\tau}_{k,n}), Z_{k,n}) \rightsquigarrow (\Gamma, \mathbb{E}(\tau), Z), \qquad n \to \infty.$$

In view of (B.1) and (B.2), as $n \to \infty$,

$$\begin{aligned}\Big(\sqrt{k}(\hat{\gamma}_{k,n} - \gamma), \sqrt{k}\hat{\delta}_{k,n}, Z_{k,n}\Big) \\ \rightsquigarrow \left(\frac{(1-\rho)^2}{\rho^2} \left(\Gamma - \gamma \frac{1-2\rho}{\rho} \mathbb{E}(\tau) \right), \right. \\ \left. \frac{(1-2\rho)(1-\rho)}{\rho^3} \left(-\Gamma + \gamma \frac{(1-\rho)^2}{\rho} \mathbb{E}(\tau) \right), Z \right). \quad \text{(B.3)}\end{aligned}$$

The vector $(\Gamma, \mathbb{E}(\tau), Z)$ is trivariate normal, with Z standard normal and independent of $(\Gamma, \mathbb{E}(\tau))$, with Γ as in Theorem A.1, and with

$$\begin{aligned}\mathrm{E}[\mathbb{E}(\tau)] = \lambda \frac{\rho^2}{\gamma(1-2\rho)(1-\rho)}, \qquad \mathrm{var}\{\mathbb{E}(\tau)\} = \frac{\rho^2}{(1-2\rho)(1-\rho)^2}, \\ \mathrm{cov}\{\Gamma, \mathbb{E}(\tau)\} = \gamma \frac{\rho}{(1-\rho)^2}.\end{aligned}$$

As a consequence, the distribution of the limit vector in (B.3) is trivariate normal with mean vector $(0, \lambda, 0)'$ and covariance matrix Σ as in (3.12).

C Proofs for Section 5

Proof of Theorem 5.1 Put $y_n = x_n / X_{n-k:n}$, recall $\delta_n = \delta(X_{n-k:n})$, and define

$$\tilde{p}_n = \overline{F}(X_{n-k:n}) \overline{G}_{\gamma,\delta_n,\tau}(y_n).$$

Since $k \to \infty$ and $p_n \to 0$ as $n \to \infty$, it is sufficient to prove (5.4) with $\hat{p}_n / p_n - 1$ replaced by $\log \hat{p}_n - \log p_n$. Let us write

$$\sqrt{k}(\log \hat{p}_n - \log p_n) = \sqrt{k}(\log \hat{p}_n - \log \tilde{p}_n) + \sqrt{k}(\log \tilde{p}_n - \log p_n)$$

and treat the two terms on the right-hand side separately.

1. The term $\sqrt{k}(\log \tilde{p}_n - \log p_n)$. We have

$$\log \tilde{p}_n - \log p_n = \log \overline{G}_{\gamma,\delta_n,\tau}(y_n) - \log \frac{\overline{F}(y_n X_{n-k:n})}{\overline{F}(X_{n-k:n})}.$$

Since $(n/k_n)\overline{F}(y_n X_{n-k:n}) \to q$ and $(n/k_n)\overline{F}(X_{n-k:n}) = 1 + o_p(1)$ as $n \to \infty$,

$$\frac{\overline{F}(y_n X_{n-k:n})}{\overline{F}(X_{n-k:n})} = \frac{\overline{F}(x_n)}{\overline{F}(X_{n-k:n})} = q + o_p(1), \qquad n \to \infty.$$

Since moreover $\overline{F}$ is monotone and regularly varying of index $-1/\gamma$, this forces $y_n \to y$ as $n \to \infty$ with $y^{-1/\gamma} = q$, or $y = q^{-\gamma} \in (1, \infty)$. By Proposition 2.3,

we find

$$\frac{\log \tilde{p}_n - \log p_n}{\delta_n} = o_p(1), \qquad n \to \infty.$$

Finally, from $\sqrt{k}\delta_n = \lambda + o_p(1)$ as $n \to \infty$, we can conclude that

$$\sqrt{k}(\log \tilde{p}_n - \log p_n) = \sqrt{k}\delta_n \frac{\log \tilde{p}_n - \log p_n}{\delta_n} = o_p(1), \qquad n \to \infty. \tag{C.1}$$

2. The term $\sqrt{k}(\log \hat{p}_n - \log \tilde{p}_n)$. We have

$$\begin{aligned}\log \hat{p}_n - \log \tilde{p}_n = {} & \{\log(k/n) - \log \overline{F}(X_{n-k:n})\} \\ & + \{\log \overline{G}_{\hat{\gamma}_n, \hat{\delta}_n, \hat{\tau}_n}(y_n) - \log \overline{G}_{\gamma, \delta_n, \tau}(y_n)\}. \end{aligned} \tag{C.2}$$

The first term on the right-hand side is

$$\begin{aligned}\log(k/n) - \log \overline{F}(X_{n-k:n}) &= -\log\{n\overline{F}(X_{n-k:n})/k\} \\ &= -\log(1 + k^{-1/2}Z_n) \\ &= -k^{-1/2}Z_n + o_p(k^{-1/2}), \qquad n \to \infty.\end{aligned}$$

For the second term on the right-hand side in (C.2), we proceed as follows. Since $y_n = y + o_p(1)$ as $n \to \infty$ and $y > 1$, it is sufficient to work on the event $y_n > 1$. Then

$$\begin{aligned}&\log \overline{G}_{\hat{\gamma}_n, \hat{\delta}_n, \hat{\tau}_n}(y_n) - \log \overline{G}_{\gamma, \delta_n, \tau}(y_n) \\ &= \log[\{y_n(1 + \hat{\delta}_n - \hat{\delta}_n y_n^{\hat{\tau}_n})\}^{-1/\hat{\gamma}_n}] - \log[\{y_n(1 + \delta_n - \delta_n y_n^{\tau})\}^{-1/\gamma}] \\ &= \left((-\frac{1}{\hat{\gamma}_n}) - (-\frac{1}{\gamma})\right) \log y_n \\ &\quad + (-\frac{1}{\hat{\gamma}_n})\{\log(1 + \hat{\delta}_n - \hat{\delta}_n y_n^{\hat{\tau}_n}) - \log(1 + \delta_n - \delta_n y_n^{\tau})\} \\ &\quad + \left((-\frac{1}{\hat{\gamma}_n}) - (-\frac{1}{\gamma})\right) \log(1 + \delta_n - \delta_n y_n^{\tau}).\end{aligned} \tag{C.3}$$

We treat the three terms on the right-hand side of (C.3) in turn. First,

$$\begin{aligned}(-\frac{1}{\hat{\gamma}_n}) - (-\frac{1}{\gamma}) &= \frac{\hat{\gamma}_n - \gamma}{\hat{\gamma}_n \gamma} \\ &= k^{-1/2}\gamma^{-2}\Gamma_n + O_p(k^{-1}), \qquad n \to \infty.\end{aligned}$$

Second, $\delta_n = O_p(k^{-1/2})$ and therefore also $\hat{\delta}_n = O_p(k^{-1/2})$ as $n \to \infty$. Hence the second term on the right-hand side of (C.3) is

$$\begin{aligned}&-\hat{\gamma}_n^{-1}\{\log(1 + \hat{\delta}_n - \hat{\delta}_n y_n^{\hat{\tau}_n}) - \log(1 + \delta_n - \delta_n y_n^{\tau})\} \\ &= \{-\gamma^{-1} + O_p(k^{-1/2})\}\{\hat{\delta}_n - \hat{\delta}_n y_n^{\hat{\tau}_n} - \delta_n + \delta_n y_n^{\tau} + O_p(k^{-1})\} \\ &= -k^{-1/2}\gamma^{-1}\Delta_n(1 - y_n^{\tau}) + o_p(k^{-1/2}), \qquad n \to \infty.\end{aligned}$$

The third term on the right-hand side of (C.3) is $O_p(k^{-1/2})O_p(k^{-1/2}) = O_p(k^{-1})$. All in all, we find

$$\sqrt{k}(\log \hat{p}_n - \log \tilde{p}_n) = -Z_n + \gamma^{-2}\Gamma_n \log y - \gamma^{-1}\Delta_n(1 - y^\tau) + o_p(1) \qquad \text{(C.4)}$$

as $n \to \infty$. Combine (C.3) and (C.4) and recall $y = q^{-\gamma}$ and $\rho = \gamma\tau$ to find the result. □

Proof of Theorem 5.2 Recall the Weissman estimator $\hat{p}_n^{\mathrm{W}}$ in (5.7) and put $d_n = k/(np_n)$. From Theorem 4.4.7 in de Haan and Ferreira (2006), it follows that

$$\frac{\sqrt{k}}{\log d_n}\left(\frac{\hat{p}_n^{\mathrm{W}}}{p_n} - 1\right) \rightsquigarrow \gamma^{-1}\Gamma, \qquad n \to \infty.$$

Moreover, writing $y_n = x_n/X_{n-k:n}$,

$$\frac{\hat{p}_n}{\hat{p}_n^{\mathrm{W}}} = \{1 + \hat{\delta}_n - \hat{\delta}_n y_n^{\hat{\tau}_n}\}^{-1/\hat{\gamma}_n} = 1 + O_p(k^{-1/2}), \qquad n \to \infty.$$

As $\log d_n \to \infty$, we find that $\hat{p}_n$ and $\hat{p}_n^{\mathrm{W}_n}$ have the same asymptotic distribution. □

Acknowledgments

We are grateful to two referees for their speedy reports containing thoughtful and constructive remarks.

References

Balkema, A. A., de Haan, L., 1974. Residual life time at great age. The Annals of Probability 2, 792–804.

Beirlant, J., Bouquiaux, C., Werker, B. J., 2006. Semiparametric lower bounds for tail index estimation. Journal of Statistical Planning and Inference 136, 705–729.

Beirlant, J., Dierckx, G., Goegebeur, Y., Matthys, G., 1999. Tail index estimation and an exponential regression model. Extremes 2, 177–200.

Beirlant, J., Dierckx, G., Guillou, A., Stărică, C., 2002a. On exponential representations of log-spacings of extreme order statistics. Extremes 5, 157–180.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2004. Statistics of Extremes: Theory and Applications. Wiley, Chichester.

Beirlant, J., Joossens, E., Segers, J., 2002b. Modelling excesses over high thresholds by perturbed generalized Pareto distributions. Tech. Rep. 2002-030, EURANDOM, Eindhoven.

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

Coles, S. G., 2001. An Introduction to Statistical Modelling of Extreme Values. Springer series in statistics. Springer-Verlag, London.

Cooray, K., Ananda, M. A., 2005. Modeling actuarial data with a composite lognormal-Pareto model. Scandinavian Actuarial Journal 5, 321–334.

Davison, A. C., Smith, R. L., 1990. Models for exceedances over high thresholds. Journal of the Royal Statistical Society, Series B 52, 393–442.

de Haan, L., Ferreira, A., 2006. Extreme Value Theory: An Introduction. Springer, New York.

de Haan, L., Stadtmüller, U., 1996. Generalized regular variation of second order. Journal of the Australian Mathematical Society (Series A) 61, 381–395.

Drees, H., 1996. Refined pickands estimators with bias correction. Communications in Statistics – Theory and Methods 25, 837–851.

Drees, H., 1998. A general class of estimators of the extreme value index. Journal of Statistical Planning and Inference 66, 95–112.

Drees, H., Ferreira, A., de Haan, L., 2004. On maximum likelihood estimation of the extreme value index. The Annals of Applied Probability 14, 1179–1201.

Feuerverger, A., Hall, P., 1999. Estimating a tail exponent by modelling departure from a Pareto distribution. The Annals of Statistics 27, 760–781.

Fraga Alves, M. I., de Haan, L., Lin, T., 2003a. Estimation of the parameter controlling the speed of convergence in extreme value theory. Mathematical Methods in Statistics 12, 155–176.

Fraga Alves, M. I., Gomes, M. I., de Haan, L., 2003b. A new class of semi-parametric estimators of the second order parameter. Portugaliae Mathematica 60, 193–214.

Frigessi, A., Haug, O., Rue, H., 2002. A dynamic mixture model for unsupervised tail estimation without threshold selection. Extremes 5, 219–235.

Geluk, J., de Haan, L., 1987. Regular variation, extensions and tauberian theorems. Tech. Rep. 40, CWI tract, Amsterdam.

Gomes, M. I., Martins, M. J., 2002. Asymptotically unbiased estimators of the tail index based on external estimation of the second order parameter. Extremes 5, 5–31.

Gomes, M. I., Martins, M. J., 2004. Bias reduction and explicit semi-parametric estimation of the tail index. Journal of Statistical Planning and Inference 124, 361–378.

Gomes, M. I., Martins, M. J., Neves, M., 2000. Alternatives to a semi-parametric estimator of parameters of rare events – the Jackknife methodology. Extremes 3, 207–229.

Hall, P., 1982. On some simple estimates of an exponent of regular variation.

Journal of the Royal Statistical Society, Series B 44, 37–42.

Hill, B. M., 1975. A simple approach to inference about the tail of a distribution. The Annals of Statistics 3, 1163–1174.

Peng, L., 1998. Asymptotically unbiased estimators for the extreme-value index. Statistics & Probability Letters 38, 107–115.

Peng, L., Qi, Y., 2004. Estimating the first- and second-order parameters of a heavy-tailed distribution. Australian and New-Zealand Journal of Statistics 46 (2), 305–312.

Pickands, J., 1975. Statistical inference using extreme order statistics. The Annals of Statistics 3, 119–131.

Reiss, R.-D., 1989. Approximate Distributions of Order Statistics. Springer Series in Statistics. Springer, New York.

Segers, J., 2005. Generalized Pickands estimators for the extreme value index. Journal of Statistical Planning and Inference 28, 381–396.

Smith, R. L., 1987. Estimating tails of probability distributions. The Annals of Statistics 15, 1174–1207.

van der Vaart, A., 1998. Asymptotic Statistics. Cambridge University Press, Cambridge.

Weissman, I., 1978. Estimation of parameters and larger quantiles based on the k largest observations. Journal of the American Statistical Association 73, 812–815.