Tail index estimation through the accomodation of bias in the weighted log-excesses*

M. Ivette Gomes
Universidade de Lisboa, Faculdade de Ciências (DEIO) and CEAUL

L. de Haan
Erasmus University of Rotterdam and CEAUL

and

Lígia Henriques
Instituto Politécnico de Tomar and CEAUL

July 26, 2004

Abstract. In this paper we are interested in the derivation of the asymptotic distributional properties of a weighted log-excesses' estimator of a positive tail index γ. One of the main objectives of such an estimator is the accomodation of the dominant component of asymptotic bias, together with the maintenance of the asymptotic variance of the maximum likelihood estimator of γ, under a strict Pareto model. We shall here consider the external estimation not only of a "shape" second order parameter ρ, but also of a "scale" second order parameter β, being then able to decrease the asymptotic variance of the final estimators under investigation, comparatively to the one of the "asymptotically unbiased" estimators already available in the literature. The "asymptotically unbiased" estimators herewith considered will also be studied for finite samples, through Monte Carlo techniques, as well as applied to real data in the

*Research partially supported by FCT / POCTI / FEDER.

field of finance.

AMS 2000 subject classification. Primary 62G32; Secondary 65C05.

Keywords and phrases. *Statistics of Extremes*, *Semi-parametric estimation*, *Log-excesses*, *Heavy tails*, *Maximum Likelihood.*

1 Introduction and motivation for the new class of estimators

Heavy-tailed models appear often in practice in fields like telecommunication traffic, insurance and finance. A model F is said to be heavy-tailed whenever the *tail function*, $1-F$, is a regularly varying function with a negative index of regular variation $\alpha = -1/\gamma$, i.e., for every $x > 0$,

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-1/\gamma}.$$

Then we are in the domain of attraction for maxima of an *Extreme Value* distribution function (d.f.),

$$EV_\gamma(x) = \exp(-(1+\gamma x)^{-1/\gamma}),\ 1+\gamma x \geq 0,\ \gamma > 0,$$

and we write $F \in \mathcal{D}_{\mathcal{M}}(EV_\gamma)$. The parameter γ is the *tail index*, one of the primary parameters of extreme or even rare events.

In a context of heavy tails, and with the notation $U(t) = F^{\leftarrow}(1-1/t),\ t \geq 1$, $F^{\leftarrow}(y) = \inf\{x : F(x) \geq y\}$ the generalized inverse function of the underlying model F, the first order parameter (or tail index) γ may also appear, for every $x > 0$, as the limiting value

$$\gamma = \lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t)}{\ln x},$$

i.e., with the usual notation RV_α for the class of regularly varying functions with index of regularly variation α,

$$F \in \mathcal{D}_{\mathcal{M}}(EV_\gamma)\ (\gamma > 0) \quad \text{iff} \quad 1-F \in RV_{-1/\gamma} \quad \text{iff} \quad U \in RV_\gamma. \tag{1.1}$$

The second order parameter ρ (≤ 0) is the non-positive value which appears in the limiting relation

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \frac{x^\rho - 1}{\rho}, \tag{1.2}$$

which we assume to hold for every $x > 0$, and where $|A(t)|$ is then of regular variation with index ρ (Geluk and de Haan, 1987).

For intermediate k, i.e., a sequence of integers $k = k_n$ between 1 and n such that

$$k = k_n \to \infty, \quad k_n = o(n), \quad \text{as } n \to \infty, \tag{1.3}$$

let us consider the log-excesses,

$$V_{ik} := \ln X_{n-i+1:n} - \ln X_{n-k:n}, \quad 1 \leq i \leq k < n, \tag{1.4}$$

where $X_{i:n}$ denotes, as usual, the i-th ascending order statistic (o.s.), $1 \leq i \leq n$, associated to a random sample $(X_1, X_2, \cdots, X_n)$.

From the definition of the function U and from the fact that, denoting R a uniform random variable (r.v.), $F^{\leftarrow}(R)$ is a r.v. with d.f. F, we get the representation $X_{i:n} \stackrel{d}{=} U(Y_{i:n})$ where Y is a unit Pareto r.v., i.e., $F_Y(y) = 1 - y^{-1},\ y \geq 1$. Indeed, $1 - 1/Y \stackrel{d}{=} R$. Since for $j > i$, $Y_{j:n}/Y_{i:n} \stackrel{d}{=} Y_{j-i:n-i}$, $\ln Y_{i:n} \stackrel{d}{=} E_{i:n}$, where $\{E_i\}$ denotes a sequence of independent, standard exponential r.v.'s and $Y_{n-k:n} \sim n/k$, as $n \to \infty$, we may indeed write, whenever we are under the first order framework in (1.1),

$$V_{ik} \stackrel{d}{=} \ln U(Y_{n-i+1:n}) - \ln U(Y_{n-k:n}) \ \sim\ \gamma\ \ln Y_{k-i+1:k} \stackrel{d}{=} \gamma\ E_{k-i+1:k},$$

i.e., the V_{ik}'s, $1 \leq i \leq k$, are, approximately, the k o.s.'s from an exponential random sample with mean value γ. This argument justifies the well-known Hill estimator (Hill, 1975):

$$H(k) := \frac{1}{k}\sum_{i=1}^{k}\{\ln X_{n-i+1:n} - \ln X_{n-k:n}\} \equiv \frac{1}{k}\sum_{i=1}^{k} V_{ik}. \tag{1.5}$$

More specifically, under the second order framework in (1.2), we may say that for intermediate k, i.e., whenever (1.3) holds true,

$$V_{ik} \stackrel{d}{=} \gamma\ \ln Y_{k-i+1:k} + \frac{Y_{k-i+1:k}^{\rho} - 1}{\rho}\ A(n/k)\ (1 + o_p(1)), \tag{1.6}$$

where the $o_p(1)$ term is uniform in i, $1 \le i \le k$.

Let us write now

$$\begin{aligned} V_{ik} &= \gamma \ln Y_{k-i+1:k}\left(1 + \frac{A(n/k)}{\gamma}\ \frac{Y_{k-i+1:k}^{\rho} - 1}{\rho\ \ln Y_{k-i+1:k}}\ (1 + o_p(1))\right) \\ &= \gamma\ e^{\frac{A(n/k)}{\gamma}\frac{Y_{k-i+1:k}^{\rho}-1}{\rho \ln Y_{k-i+1:k}}}\ E_{k-i+1:k} + o_p\left(A(n/k)\right). \end{aligned}$$

It thus follows that

$$V_{ik} - \gamma\ e^{\frac{A(n/k)}{\gamma}\frac{Y_{k-i+1:k}^{\rho}-1}{\rho \ln Y_{k-i+1:k}}}\ E_{k-i+1:k} = o_p\left(V_{ik} - \gamma\ E_{k-i+1:k}\right). \tag{1.7}$$

Note also that, for $1 \le i \le k$,

$$\frac{Y_{k-i+1:k}^{\rho} - 1}{\rho \ln Y_{k-i+1:k}} \approx -\frac{(i/k)^{-\rho} - 1}{\rho \ln(i/k)} =: \psi_{ik} \equiv \psi(i/k) \equiv \psi_{ik}(\rho)\ [\psi_{kk} \equiv 1], \tag{1.8}$$

with ψ a limited function (see Lemma 5.1).

The validity of (1.7), together with the approximation in (1.8), lead us to expect to be able to get a less biased estimator of the tail index γ if we assume that the random log-excess V_{ik}, in (1.4), comes from an exponential model with mean value not equal to γ, as it is done to support the estimator in (1.5), but dependent on i (and k), and more specifically given by

$$\gamma_{ik} = \gamma\ e^{A(n/k)\ \psi_{ik}/\gamma}, \quad 1 \le i \le k.$$

We shall herewith restrict ourselves to the case $\rho < 0$. We shall thus assume that we are in Hall's class of models (Hall, 1982; Hall and Welsh, 1985), with a tail function

$$1 - F(x) = \left(\frac{x}{C}\right)^{-1/\gamma} \left(1 + \frac{\beta}{\rho}\left(\frac{x}{C}\right)^{\rho/\gamma} + o\left(x^{\rho/\gamma}\right)\right), \quad \text{as } x \to \infty,$$

with $C > 0$, $\beta \neq 0$, $\rho < 0$. We may then choose $A(t) = \gamma\ \beta\ t^{\rho}$, dependent on the tail index γ, the "scale" second order parameter β and the "shape" second order parameter ρ.

Since we may write the approximation,

$$V_{ik} \ \approx\ \gamma e^{\beta\ (n/k)^{\rho}\ \psi_{ik}}\ E_{k-i+1:k}, \qquad \psi_{ik} = -\frac{(i/k)^{-\rho} - 1}{\rho\ \ln(i/k)}, \quad 1 \le i \le k,$$

the likelihood associated to the k log-excesses, $V_{ik}, 1 \le i \le k$, is proportional to

$$\mathcal{L}^*(\gamma, \beta, \rho) = \exp\left(-k \ln\gamma - \beta\left(\frac{n}{k}\right)^{\rho} \sum_{i=1}^{k} \psi_{ik} - \frac{1}{\gamma}\sum_{i=1}^{k} V_{ik}\ e^{-\beta\ (n/k)^{\rho}\ \psi_{ik}}\right). \quad (1.9)$$

Should we know β and ρ, the maximization of $\mathcal{L}^*$ would lead us to

$$\widehat{\gamma} = WLE_{\beta,\ \rho}(k) = \frac{1}{k}\sum_{i=1}^{k}\ e^{-\beta\ (n/k)^{\rho}\ \psi_{ik}}\ V_{ik} =: A_k^{(0)}.$$

More generally, and for $j \ge 0$, we shall denote

$$\begin{aligned} A_k^{(j)} \ &:=\ \frac{1}{k}\sum_{i=1}^{k}\ \psi_{ik}^{j} e^{-\beta\ (n/k)^{\rho}\ \psi_{ik}}\ V_{ik} && (1.10)\\ &\overset{d}{=}\ \frac{1}{k}\sum_{i=1}^{k}\ \psi_{ik}^{j}\ V_{ik} - \beta\left(\frac{n}{k}\right)^{\rho}\left(\frac{1}{k}\ \sum_{i=1}^{k}\ \psi_{ik}^{j+1}\ V_{ik}\right)(1 + o_p(1)) \\ &=:\ B_k^{(j)} - \beta\left(\frac{n}{k}\right)^{\rho} B_k^{(j+1)}(1 + o_p(1)), && (1.11)\end{aligned}$$

where the $o_p(1)$-terms are also uniform in i, $1 \le i \le k$.

It seems thus sensible to replace Hill's estimator in (1.5) by a weighted combination of the log-excesses, i.e., by

$$WLE_{\widehat{\beta},\ \widehat{\rho}}(k) := \frac{1}{k}\sum_{i=1}^{k} e^{-\widehat{\beta}\ (n/k)^{\widehat{\rho}}\ \widehat{\psi}_{ik}}\ \ln\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right), \qquad (1.12)$$

where with WLE we denote a Weighted Log-Excesses' estimator, being $\widehat{\beta}$ and $\widehat{\rho}$ any consistent estimators of the second order parameters β and ρ, respectively. We use the obvious notation $\widehat{\psi}_{ik} = -\left((i/k)^{-\widehat{\rho}} - 1\right) / \left(\widehat{\rho}\ \ln(i/k)\right),\ 1 \le i \le k$.

Remark 1.1. *The class of estimators in* (1.12)*, although a linear combination of log-excesses, is not in the class of kernel's estimators of Csörgő et al.* (1985)*, because the weights of the log-excesses are not functions of i and k through the quotient $\{i/k\}$. It does not also belong to the more general class of Drees* (1998)*, now because the weights are dependent on the second order parameters' estimators $\widehat{\beta}$ and $\widehat{\rho}$, which may use a larger number of order statistics than the number k used in the estimation of the tail index γ. In Drees's class of functionals, the minimal asymptotic variance of an "asymptotically unbiased" estimator is given by $(\gamma(1-\rho)/\rho)^2 > \gamma^2$, whereas, as we shall see later on, we are here able to obtain estimators with an asymptotic variance equal to γ^2, the asymptotic variance of the Hill estimator in* (1.5)*.*

Remark 1.2. *If we work with the excesses over a high random threshold, $W_{ik} := X_{n-i+1:n} - X_{n-k:n}$, $1 \le i \le k$, and in a similar way try to accomodate bias in these excesses, assuming that W_{ik} comes from a Generalized Pareto model, with d.f. $GP_{\gamma_{ik},\alpha}(w) = 1 - (1+\alpha w)^{-1/\gamma_{ik}}$, $w \ge 0$, $\gamma_{ik} = \gamma\ e^{\beta(n/k)^{\rho}\psi_{ik}}$, ψ_{ik} given in* (1.8)*, $1 \le i \le k$, we arrive at the maximum likelihood tail index estimator,*

$$\widehat{\gamma}(k) := \frac{1}{k}\sum_{i=1}^{k}\ e^{-\widehat{\beta}(n/k)^{\widehat{\rho}}\widehat{\psi}_{ik}}\ \ln(1+\widehat{\alpha}\ W_{ik}),$$

obviously dependent on the maximum likelihood estimators $\widehat{\alpha}$, $\widehat{\beta}$ and $\widehat{\rho}$ of α, β and ρ, respectively. If we do not estimate α through maximum likelihood, but we further think that for heavy tails, a possible estimator of the scale parameter α in the GP model is $\{1/X_{n-k:n}\}$, we come again to the estimator in (1.12)*.*

Let us assume everything is known, apart from γ. We may state the following:

Theorem 1.1. *Let us consider the r.v.* $WLE(k) \equiv WLE_{\beta,\ \rho}(k)$*, given by the functional expression in* (1.12) *but with* $\left(\widehat{\beta},\ \widehat{\rho}\right)$ *replaced by* $(\beta,\ \rho)$*. Under the second order framework in* (1.2)*, and for levels* k *such that* (1.3) *holds, the asymptotic distributional representation,*

$$WLE(k) \equiv A_k^{(0)} \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\ N_k + R_k, \qquad R_k = o_p(A(n/k)), \tag{1.13}$$

holds true, with $N_k \stackrel{a}{\sim}$ *Normal(0,1).*
Consequently, $\sqrt{k}\ (WLE(k) - \gamma)$ *is asymptotically normal with variance equal to* γ^2*, and a null mean value not only when* $\sqrt{k}\ A(n/k) \to 0$*, but also when* $\sqrt{k}\ A(n/k) \longrightarrow \lambda \neq 0$*, finite, as* $n \to \infty$*. We may even guarantee the asymptotic normality of* $WLE(k)$ *if we further consider levels* k *such that* $\sqrt{k}\ A(n/k) \to \infty$*, provided that* $\sqrt{k}\ R_k \underset{n\to\infty}{\overset{p}{\longrightarrow}} \lambda_{_R}$*, finite.*

Remark 1.3. *Theorem 1.1 provides thus a technical motivation for the estimator in* (1.12) *when we assume that all the model parameters, but the tail index* γ*, are known. If we estimate* β *and* ρ *consistently, using perhaps a number of top order statistics larger than the one needed for the estimation of* γ *at sub-optimal levels, i.e., levels* k *such that* $\sqrt{k}\ R_k \to 0$*, being* R_k *the remainder in* (1.13)*, we hope to be able to get also an asymptotic variance equal to* γ^2*, which is smaller than the minimal asymptotic variance we have been able to reach so far with "asymptotically unbiased" estimators of* γ*.*

The main problems to be dealt with are then related to the estimation of β and ρ in order to get $WLE_{\widehat{\beta},\ \widehat{\rho}}(k)$ in (1.12). Computationally, we shall pay special attention to the external estimation of the second order parameters β and ρ. Such a decision is related to the discussion in Gomes and Martins (2002) on the advantages of an external estimation of the second order parameter ρ — or even their misspecification, as in Gomes *et al.* (2000) and Gomes and Martins (2004) — versus an internal estimation at the same level k, as done in Beirlant *et al.* (1999) and Feuerverger and Hall (1999). In section 2 we shall briefly review well-known estimators of the second order parameters β and ρ, providing additional information on the class of β-estimators. We also deal with an extra estimator of β based on the log-excesses, which is asymptotically equivalent to the first one. In section 3, we derive the asymptotic behaviour of the $WLE-$estimator in (1.12), estimating ρ externally and β both internally and externally. In section 4, we shall exhibit the performance of the $WLE-$estimator, comparatively to the classical Hill estimator and to one of the "asymptotically unbiased" estimators proposed in Gomes and Martins (2002), through the use of simulation techniques. We shall also consider a case-study related to the exchange rate of the Euro against the UK Pound, in order to illustrate the behaviour of this new estimator. Finally, in section 5, we shall provide an appendix, with the proofs of the main results in the paper.

2 The second order parameters' estimators

2.1 The estimation of ρ

We shall first address the estimation of ρ. We have nowadays easy access to classes of ρ-estimators which work well both theoretically and in practice, like the ones introduced in Gomes *et al.* (2002) and Fraga Alves *et al.* (2003). The estimators of ρ to be considered in this study are particular members of the class of estimators proposed by Fraga Alves *et al.* (2003), where an heuristic non-optimal choice of the threshold seems to provide interesting results for a large set of models. Under adequate general conditions, such a class provides semi-parametric asymptotically normal estimators of $\rho < 0$, which show highly stable sample paths as functions of k, the number of top o.s.'s used, for a wide range of large k-values. Such a class of estimators is parameterised in a tuning parameter τ, but we shall consider only, in the simulations, the statistics associated to $\tau = 0$ and to $\tau = 1$, usually preferable whenever $|\rho| \le 1$ and $|\rho| > 1$, respectively. Let us consider the statistics

$$T_n^{(\tau)}(k) := \begin{cases} \dfrac{\ln\left(M_n^{(1)}(k)\right) - \frac{1}{2}\ln\left(M_n^{(2)}(k)/2\right)}{\frac{1}{2}\ln\left(M_n^{(2)}(k)/2\right) - \frac{1}{3}\ln\left(M_n^{(3)}(k)/6\right)} & \text{if} \quad \tau = 0 \\[2ex] \dfrac{\left(M_n^{(1)}(k)\right)^{\tau} - \left(M_n^{(2)}(k)/2\right)^{\tau/2}}{\left(M_n^{(2)}(k)/2\right)^{\tau/2} - \left(M_n^{(3)}(k)/6\right)^{\tau/3}} & \text{if} \quad \tau > 0, \end{cases} \tag{2.1}$$

where

$$M_n^{(j)}(k) = \frac{1}{k}\sum_{i=1}^{k}\left[\ln\frac{X_{n-i+1:n}}{X_{n-k:n}}\right]^j, \; j \ge 1 \qquad [M_n^{(1)} \equiv H \text{ in } (1.5)].$$

The statistics in (2.1) converge towards $3(1-\rho)/(3-\rho)$ for every $\tau \ge 0$, whenever the second order condition (1.2) holds, k is such that (1.3) holds and, as $n \to \infty$, $\sqrt{k}\; A(n/k) \to \infty$. We may thus get a class of consistent estimators for ρ,

$$\widehat{\rho}(k) \equiv \widehat{\rho}_n^{(\tau)}(k) := -\left|\frac{3\left(T_n^{(\tau)}(k) - 1\right)}{T_n^{(\tau)}(k) - 3}\right|. \tag{2.2}$$

The theoretical and simulated results in Fraga Alves *et al.* (2003) led us to consider the ρ-estimators associated to a high level k_1, already used with success in Gomes and Martins (2002) to estimate the tail index γ. Such a level k_1, given by

$$k_1 = \min\left(n-1, \left[\frac{2n}{\ln\ln n}\right]\right), \tag{2.3}$$

where $[x]$ denotes, as usual, the integer part of x, has not been chosen in any optimal way, but works well in practice. We shall thus work with the ρ-estimators,

$$\widehat{\rho}_i := -\left|\frac{3(T_n^{(i)}(k_1) - 1)}{T_n^{(i)}(k_1) - 3}\right|, \quad i = 0,\; 1, \tag{2.4}$$

with $T_n^{(i)}(k)$ and k_1 given in (2.1) and (2.3), respectively. To denote generally any of the estimators $\widehat{\rho}_i,\ i=0,1$, or more generally any of the estimators in (2.2) computed at the level k_1 in (2.3), we shall often use the notation $\widehat{\rho}$.

Remark 2.1. *Note that $\widehat{\rho}(k)$ in* (2.2) *is consistent for the estimation of ρ whenever k is intermediate and $\sqrt{k}\ A(n/k) \to \infty$. Moreover, it is possible to prove* (*Fraga Alves et al., 2003*) *that $\widehat{\rho}(k) - \rho = O_p\left(1/\left(\sqrt{k}\ A(n/k)\right)\right)$. Consequently, with $\widehat{\rho}_i$ given in* (2.4), *$\widehat{\rho}_i - \rho = O_p\left(1/\left(\sqrt{k_1}\ A(n/k_1)\right)\right) = O_p\left((\ln_2 n)^{(1-2\rho)/2}/\sqrt{n}\right)$, $i=0,1$, with the obvious notation $\ln_2 n = \ln\ln n$.*

2.2 The estimation of β based on the scaled log-spacings

In the computational study in this paper, we have considered the estimator of β obtained in Gomes and Martins (2002), and based on the scaled log-spacings

$$U_i = i\left\{\ln X_{n-i+1:n} - \ln X_{n-i:n}\right\},\ 1 \le i \le k. \tag{2.5}$$

Such an estimator is given by

$$\widehat{\beta}_U(k;\widehat{\rho}) := \left(\frac{k}{n}\right)^{\widehat{\rho}} \frac{\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}}\right)\left(\frac{1}{k}\sum_{i=1}^{k}U_i\right) - \left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}}\ U_i\right)}{\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}}\right)\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}}\ U_i\right) - \left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-2\widehat{\rho}}\ U_i\right)}. \tag{2.6}$$

In Gomes and Martins (2002) has been derived, under the second order framework, the asymptotic behaviour of $\widehat{\beta}_U(k;\rho)$. We shall here summarize the results therewith presented, together with further results on $\widehat{\beta}_U(k;\widehat{\rho})$, $\widehat{\rho}$ any of the estimators in (2.2) computed at the level k_1 in (2.3), as well as on $\widehat{\beta}_U(k;\widehat{\rho}(k))$.

Theorem 2.1. *If the second order condition* (1.2) *holds, with $A(t) = \gamma\ \beta\ t^{\rho}$, $\rho < 0$, if $\widehat{\rho}$ is a consistent estimator of ρ, if $k = k_n$ is a sequence of intermediate* *positive integers, i.e.* (1.3) *holds, and if* $\sqrt{k}\ A(n/k) \underset{n\to\infty}{\longrightarrow} \infty$, *then* $\widehat{\beta}_U(k;\widehat{\rho})$ *in* (2.6) *converges in probability towards* β, *as* $n \to \infty$. *Moreover,*

$$\widehat{\beta}_U(k;\rho) \stackrel{d}{=} \beta + \frac{\sigma_{\widehat{\beta}_U}}{\sqrt{k}\ A(n/k)}\ B_k^U\ + R_k^U, \quad R_k^U = o_p(1), \tag{2.7}$$

where $B_k^U \stackrel{a}{\sim}$ *Normal(0,1) and*

$$\sigma_{\widehat{\beta}_U} = \frac{\gamma\ |\beta|(1-\rho)\sqrt{1-2\rho}}{|\rho|}. \tag{2.8}$$

The distributional representation (2.7) *remains true if we replace* $\widehat{\beta}_U(k,\rho)$ *by* $\widehat{\beta}_U(k;\widehat{\rho})$, *with* $\widehat{\rho}$ *any of the estimators in* (2.2) *computed at the level* k_1 *in* (2.3). *If* $\sqrt{k}\ A(n/k)\ R_k^U \to \lambda_U$, *finite, we may further guarantee the asymptotic normality of* $\widehat{\beta}_U(k;\widehat{\rho})$.

If we consider $\widehat{\beta}_U\ (k,\widehat{\rho}(k))$, *then*

$$\widehat{\beta}_U\ (k,\widehat{\rho}(k)) - \beta \sim -\beta\ \ln(n/k)\ \left(\widehat{\rho}(k) - \rho\right). \tag{2.9}$$

Remark 2.2. *As shown in Gomes and Martins* (2002), *under the second order framework in* (1.2), *and for levels* k *such that* (1.3) *holds,*

$$\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{\alpha-1}\ U_i\ \stackrel{d}{=}\ \frac{\gamma}{\alpha} + \left(\frac{\gamma\ Z_k^{(\alpha)}}{\sqrt{(2\alpha-1)\ k}} + \frac{A(n/k)}{\alpha-\rho}\right)(1+o_p(1)),$$

with

$$Z_k^{(\alpha)} = \sqrt{(2\alpha-1)\ k}\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{\alpha-1} E_i - \frac{1}{\alpha}\right), \quad \alpha \ge 1, \tag{2.10}$$

denoting $\{E_i\}$ *again independent, unit exponential r.v.'s. Since the denominator of* $\widehat{\beta}_U(k;\widehat{\rho})$, *in* (2.6), *converges towards* $\left\{-\gamma\ \rho^2/\left((1-\rho)^2(1-2\rho)\right)\right\}$, *we may write*

$$\begin{aligned}\widehat{\beta}_U(k;\rho)\ \stackrel{d}{=}\ -\frac{\beta(1-\rho)^2(1-2\rho)}{\rho^2\ A(n/k)}\Bigg(\frac{\gamma}{\sqrt{k}}\left(\frac{Z_k^{(1)}}{1-\rho} - \frac{Z_k^{(1-\rho)}}{\sqrt{1-2\rho}}\right) \\ -\frac{\rho^2\ A(n/k)}{(1-\rho)^2(1-2\rho)}\Bigg)(1+o_p(1)),\end{aligned}$$

and, if $\sqrt{k}\ A(n/k) \to \infty$*, we have*

$$\widehat{\beta}_U(k;\rho) \stackrel{d}{=} \beta + \frac{\gamma\ \beta}{\sqrt{k}\ A(n/k)} \left(\frac{Z_k^{(1)}}{1-\rho} - \frac{Z_k^{(1-\rho)}}{\sqrt{1-2\rho}} \right) + o_p(1).$$

Hence, since the asymptotic covariance between $Z_k^{(1)}$ *and* $Z_k^{(1-\rho)}$ *is* $\sqrt{1-2\rho}/(1-\rho)$*, we may choose* B_k^U *in* (2.7) *as*

$$B_k^U = \frac{(1-\rho)\sqrt{1-2\rho}}{|\rho|} \left(\frac{Z_k^{(1)}}{1-\rho} - \frac{Z_k^{(1-\rho)}}{\sqrt{1-2\rho}} \right),$$

with $Z_k^{(\alpha)}$ *given in* (2.10)*.*

Remark 2.3. *For* $i = 0,\ 1$*, let us denote* $\widehat{\beta}_{i1} := \widehat{\beta}_U(k_1,\ \widehat{\rho}_i)$*, with* $\widehat{\rho}_i$ *given in* (2.4)*. From* (2.9)*, we get* $\widehat{\beta}_{i1} - \beta = O_p\left(\ln(n/k_1)/\left(\sqrt{k_1}\ A(n/k_1)\right)\right) = O_p\left(\ln_3 n\ (\ln_2 n)^{(1-2\rho)/2}/\sqrt{n}\right)$*, again with the notation* $\ln_2 n = \ln\ln n$ *and* $\ln_3 n = \ln\ln\ln n$*.*

2.3 A first simulation experiment

We have here implemented a simulation experiment, with 1000 runs, for an underlying Burr parent, $F(x) = 1 - \left(1 + x^{-\rho/\gamma}\right)^{1/\rho},\ x \geq 0$, with $\rho = -0.5$ and $\gamma = 1$. For these Burr models, $\beta = \gamma$ for any ρ. We have estimated β through $\widehat{\beta}_U(k;\widehat{\rho}_0)$, computed at the level k used for the estimation of the tail index, as well as computed at the level $k_1 = \min(n-1, [2n/\ln\ln n])$ in (2.3), the one used for the estimator $\widehat{\rho}_0$ in (2.4), and again not chosen in any optimal way. We use the notation $\widehat{\beta}_{01} = \widehat{\beta}_U(k_1;\widehat{\rho}_0)$. The estimates of β and ρ have been incorporated in the WLE-estimator, leading to $WLE_{\widehat{\beta}_U(k;\ \widehat{\rho}_0),\ \widehat{\rho}_0}(k)$ and $WLE_{\widehat{\beta}_{01},\ \widehat{\rho}_0}(k)$. The simulations show that the tail index estimator $WLE_{\widehat{\beta}_{01},\ \widehat{\rho}_0}$ seems to work reasonably well, as illustrated in Figure 1.

The discrepancy between the behaviour of the estimator $WLE_{\widehat{\beta}_{01},\ \widehat{\rho}_0}(k)$ and the r.v. $WLE_{\beta,\ \rho}(k)$ suggests that some improvement in the estimation of second order parameters may be still welcome, but the behaviour of the mean

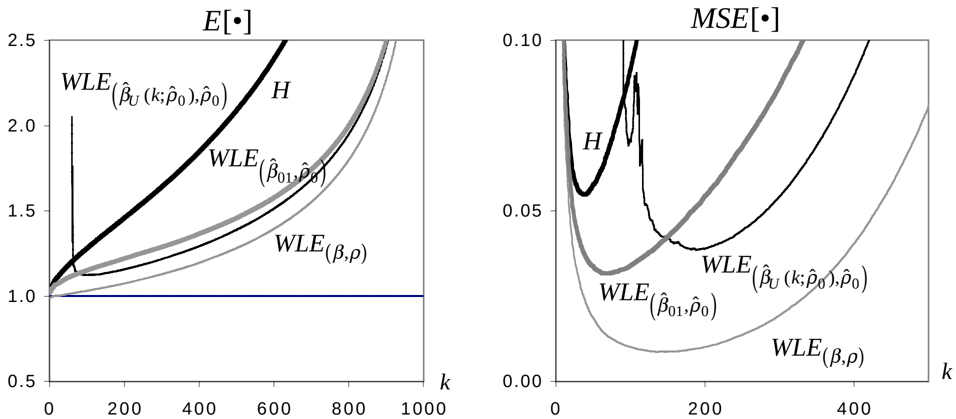


Figure 1: Mean values and Mean Squared errors of the estimators under study for samples of size $n = 1000$, from a *Burr* parent with $\gamma = 1$ and $\rho = -0.5$ ($\beta = 1$).

squared error of the $WLE-$estimator is rather interesting: the new estimator, $WLE_{\widehat{\beta}_{01},\ \widehat{\rho}_0}(k)$, is better than the Hill estimator not only when both are considered at their optimal levels, but also for every sub-optimal level k, and this contrarily to what happens with $WLE_{\widehat{\beta}_U(k;\ \widehat{\rho}_0),\ \widehat{\rho}_0}(k)$, as we may also see in this same figure.

2.4 Estimation of β based on the log-excesses

Since our tail index estimator is a linear combination of the log-excesses, we thought it would be sensible to consider here an approach for the estimation of β, based now on the log-excesses, $V_{ik},\ 1 \le i \le k$, in (1.4).

Let us introduce the notations:

$$s_k := \frac{1}{k}\sum_{i=1}^{k}\psi_{ik} \equiv \frac{1}{k}\sum_{i=1}^{k}\psi(i/k),\ 1 \le i \le k, \tag{2.11}$$

with ψ given in (1.8),

$$s_k^* := \frac{1}{k}\sum_{i=1}^{k}\psi_{ik}^* \equiv \frac{1}{k}\sum_{i=1}^{k}\psi^*(i/k),\ 1 \le i \le k, \quad \psi^*(u) = \frac{u^{-\rho}-1}{\rho}. \tag{2.12}$$

Remark 2.4. *The following limiting relations hold true:*

$$\lim_{k\to\infty} s_k \;=\; -\frac{1}{\rho}\int_0^1 \frac{x^{-\rho}-1}{\ln x}\; dx = -\frac{\ln(1-\rho)}{\rho}, \tag{2.13}$$

and

$$\lim_{k\to\infty} s_k^* \;=\; \frac{1}{\rho}\int_0^1 (x^{-\rho}-1)\; dx = \frac{1}{1-\rho}. \tag{2.14}$$

The derivative of the log-likelihood (1.9) in order to β leads to the maximum likelihood equation

$$\frac{1}{k}\sum_{i=1}^{k} \psi_{ik}\; V_{ik}\; e^{-\beta\;(n/k)^\rho\;\psi_{ik}} - \frac{\gamma}{k}\;\sum_{i=1}^{k}\psi_{ik} = 0.$$

But this equation does not lead to consistent estimators of β, because as we shall see later on, in Remark 5.1, the first member, equal to $A_k^{(1)} - \gamma\; s_k$, with $A_k^{(1)}$ and s_k given in (1.10) and (2.11), respectively, converges, as $k\to\infty$, towards $\gamma\,(1/(1-\rho)+\ln(1-\rho)/\rho)\neq 0$. In order to get convergence towards 0 we shall replace, in the second sum, ψ_{ik} by

$$\psi_{ik}^* = \psi^*(i/k) = -\psi_{ik}\;\ln(i/k) = \frac{(i/k)^{-\rho}-1}{\rho},\; 1\le i\le k.$$

The "quasi-maximum likelihood" β-estimator is thus solution of the implicit equation,

$$\left(\frac{1}{k}\sum_{i=1}^{k}\widehat{\psi}_{ik}\; e^{-\widehat{\beta}\;(n/k)^{\widehat{\rho}}\;\widehat{\psi}_{ik}}\;V_{ik}\right) - \left(\frac{1}{k}\sum_{i=1}^{k}\widehat{\psi}_{ik}^*\right)\;\left(\frac{1}{k}\sum_{i=1}^{k}\; e^{-\widehat{\beta}\;(n/k)^{\widehat{\rho}}\;\widehat{\psi}_{ik}}\;V_{ik}\right)$$
$$=: \widehat{A}_k^{(1)} - \widehat{s}_k^*\;\widehat{A}_k^{(0)} = 0. \tag{2.15}$$

If we use a first order approximation for $e^x = 1+x,$ as $x\to 0$, we come to the explicit β-estimator:

$$\widehat{\beta}_V(k;\widehat{\rho}) = \left(\frac{k}{n}\right)^{\widehat{\rho}} \frac{\left(\frac{1}{k}\sum_{i=1}^{k}\widehat{\psi}_{ik}^*\right)\left(\frac{1}{k}\sum_{i=1}^{k}V_{ik}\right) - \left(\frac{1}{k}\sum_{i=1}^{k}\widehat{\psi}_{ik}V_{ik}\right)}{\left(\frac{1}{k}\sum_{i=1}^{k}\widehat{\psi}_{ik}^*\right)\left(\frac{1}{k}\sum_{i=1}^{k}\widehat{\psi}_{ik}V_{ik}\right) - \left(\frac{1}{k}\sum_{i=1}^{k}\widehat{\psi}_{ik}^2V_{ik}\right)}, \tag{2.16}$$

quite similar to the estimator in (2.6), but with U_i replaced by V_{ik} and $(i/k)^{-\rho}$ replaced by ψ_{ik}.

With the obvious notation for $\widehat{B}_k^{(j)}$ and $\widehat{s}_k^*$, $j \geq 0$, with $B_k^{(j)}$ and s_k^* given in (1.11) and (2.12), respectively, we may write:

$$\widehat{\beta}_V(k; \widehat{\rho}) = \left(\frac{k}{n}\right)^{\widehat{\rho}} \frac{\widehat{s}_k^* \ \widehat{B}_k^{(0)} - \widehat{B}_k^{(1)}}{\widehat{s}_k^* \ \widehat{B}_k^{(1)} - \widehat{B}_k^{(2)}},$$

and we may further state the following result, similar to Theorem 2.1, but now related to the β-estimator herewith considered:

Theorem 2.2. *If the second order condition* (1.2) *holds, with* $A(t) = \gamma\ \beta\ t^{\rho}$, $\rho < 0$, *if* $\widehat{\rho}$ *is any consistent estimator of* ρ, *if* $k = k_n$ *is a sequence of intermediate integers, i.e.,* (1.3) *holds, and if we further have* $\lim_{n\to\infty} \sqrt{k}\ A(n/k) = \infty$, *then* $\widehat{\beta}_V(k; \widehat{\rho})$ *in* (2.16) *converges in probability towards* β, *as* $n \to \infty$. *We may write*

$$\widehat{\beta}_V(k; \rho) \stackrel{d}{=} \beta + \frac{\sigma_{\widehat{\beta}_V}}{\sqrt{k}\ A(n/k)}\ B_k^V + R_k^V, \quad \textit{with} \quad R_k^V = o_p(1), \tag{2.17}$$

where $B_k^V \stackrel{a}{\sim}$ *Normal*$(0,1)$, *being*

$$\sigma_{\widehat{\beta}_V} = \frac{\gamma\ |\beta| \sqrt{\sigma_1^2 - a_1^2}}{(a_1^2 - a_2)}, \tag{2.18}$$

with

$$a_1 = \frac{1}{1-\rho}, \quad a_2 = -\frac{\ln(1-2\rho) - 2\ln(1-\rho)}{\rho^2} \tag{2.19}$$

and

$$\sigma_1^2 = \frac{2}{\rho^2} \iint\limits_{0 \leq u < v \leq 1} \left(\frac{u^{-\rho} - 1}{\ln u}\ \frac{v^{-\rho} - 1}{\ln v}\right) \frac{1-v}{v}\ du\ dv. \tag{2.20}$$

The distributional representation (2.17) *remains true if we replace* $\widehat{\beta}_V(k, \rho)$ *by* $\widehat{\beta}_V(k; \widehat{\rho})$, *with* $\widehat{\rho}$ *any of the estimators in* (2.2) *computed at the level* k_1 *in* (2.3). *If* $\sqrt{k}\ A(n/k)\ R_k^V \to \lambda_V$, *finite, we may further guarantee the asymptotic normality of* $\widehat{\beta}_V(k; \widehat{\rho})$.

Again, if we consider $\widehat{\beta}_V\,(k,\widehat{\rho}(k))$*,*

$$\widehat{\beta}_V\,(k,\widehat{\rho}(k)) - \beta \sim -\beta\ \ln(n/k)\ \ (\widehat{\rho}(k)-\rho)\,, \quad \textit{as } n\to\infty, \tag{2.21}$$

i.e., the rate of convergence of $\widehat{\beta}_V\,(k,\widehat{\rho}(k))$ *towards* β *is of the order of* $\ln(n/k)/\Big(\sqrt{k}\ A(n/k)\Big)$.

Remark 2.5. *Note that the result in Remark 2.3 holds true if we replace* $\widehat{\beta}_U$ *by* $\widehat{\beta}_V$*. For the difference between* $\sigma_{\widehat{\beta}_U}$ *in* (2.8) *and* $\sigma_{\widehat{\beta}_V}$ *in* (2.18) *see Figure 2 in Remark 3.3.*

3 Asymptotic behaviour of the tail index estimator

3.1 The Weighted Log Excesses' estimator and the external estimation of β and ρ

If we estimate β externally, in an adequate way and at a larger level than the level k on which we are going to base the estimation of the tail index γ, we may be able to keep the asymptotic variance of the final tail index estimator equal to the asymptotic variance of Hill's estimator. Indeed, as a consequence of Theorem 1.1, the use of Cramèr's delta-method enables us to state the following:

Theorem 3.1. *Under the conditions of Theorem 1.1, the same distributional representation* (1.13) *holds true if we consider the tail index estimator* $WLE_{\widehat{\beta},\,\widehat{\rho}}(k)$ *for any consistent estimators* $\widehat{\beta}$ *and* $\widehat{\rho}$ *of* β *and* ρ*, respectively, such that both* $\{\widehat{\beta}-\beta\}$ *and* $(\widehat{\rho}-\rho)\ln(n/k)$ *are simulataneously* $o_p(1)$ *and* $o_p(1/(\sqrt{k}\ A(n/k)))$*, for the* k*-values on which we base the estimation of the tail index* γ*. These conditions for* $\widehat{\rho}$ *and* $\widehat{\beta}$ *hold true if we consider levels* k *such that* $\sqrt{k}\ A^2(n/k)\to\lambda$*, finite, and the estimators in* (2.2) *and* (2.6) [*or* (2.16)]*, respectively, both computed at the level* k_1 *in* (2.3)*.*

Remark 3.1. *We think that this is a remarkable result from a practical point of view, because we are able to reduce the dominant component of bias, without increasing the asymptotic variance. We may thus expect to obtain, for the new estimator, a mean squared error smaller than that of the Hill estimator for every level k, either sub-optimal or optimal. We have thus been able to overpass the old trade-off between variance and bias.*

Remark 3.2. *Note also that the levels k such that $\sqrt{k}\ A(n/k) \to \lambda$, finite, are sub-optimal for this estimator. To go further to the optimal level associated to this estimator, we should go into a third order framework, like the one considered in Gomes and de Haan* (1999), *Gomes et al.* (2002) *and Fraga Alves et al.* (2003), *considering levels k such that $\sqrt{k}\ A(n/k) \to \infty$, as $n \to \infty$.*

3.2 The joint estimation of γ and β at the same level

The following result follows straightforwardly from Theorems 1.1, 2.1 and 2.2:

Theorem 3.2. *If the second order condition* (1.2) *holds, with $A(t) = \gamma\ \beta\ t^\rho$, $\rho < 0$, if $k = k_n$ is a sequence of intermediate integers, i.e.,* (1.3) *holds, and if $\widehat{\rho}$ is any of the estimators in* (2.2) *computed at the level k_1 in* (2.3),

$$WLE_{\widehat{\beta}_U(k;\ \widehat{\rho}),\ \widehat{\rho}}(k) \stackrel{d}{=} \gamma + \frac{\gamma(1-\rho)}{|\rho|\ \sqrt{k}}\ B_k^{U} + o_p(A(n/k)) \tag{3.1}$$

and

$$WLE_{\widehat{\beta}_V(k;\ \widehat{\rho}),\ \widehat{\rho}}(k) \stackrel{d}{=} \gamma + \frac{\gamma\ \sqrt{a_1^2\sigma_1^2 + a_2^2 - 2a_1^2 a_2}}{(a_1^2 - a_2)\ \sqrt{k}}\ B_k^{V} + o_p(A(n/k)), \tag{3.2}$$

where B_k^U and B_k^V are asymptotically standard normal r.v.'s. Hence, $\sqrt{k}\ \left(WLE_{\widehat{\beta}_\bullet(k;\ \widehat{\rho}),\ \widehat{\rho}}(k) - \gamma\right)$ are both asymptotically normal with a null mean value whenever $\sqrt{k}\ A(n/k) \to \lambda$, finite, non-necessarily null. The asymptotic standard deviations of $WLE_{\widehat{\beta}_U(k;\ \widehat{\rho}),\ \widehat{\rho}}(k)$ and $WLE_{\widehat{\beta}_V(k;\ \widehat{\rho}),\ \widehat{\rho}}(k)$ are thus ruled by

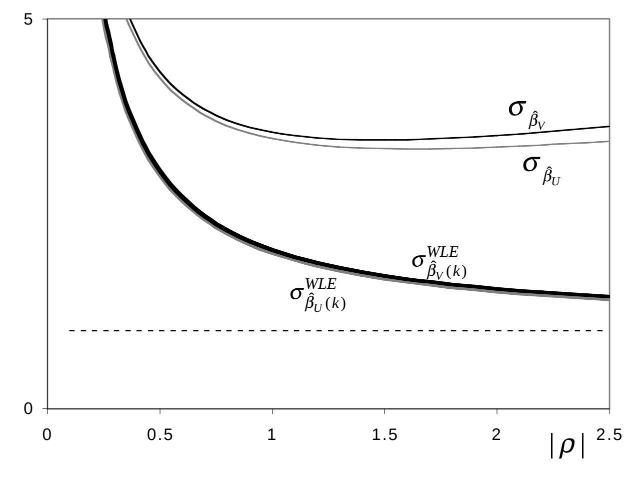
$$\sigma^{WLE}_{\widehat{\beta}_U(k;\ \widehat{\rho})} = \frac{\gamma\ (1-\rho)}{|\rho|} \tag{3.3}$$

and

$$\sigma^{WLE}_{\widehat{\beta}_V(k;\ \widehat{\rho})} = \frac{\gamma\ \sqrt{a_1^2\sigma_1^2 + a_2^2 - 2a_1^2 a_2}}{a_1^2 - a_2}, \tag{3.4}$$

respectively, with (a_1, a_2) *and* σ_1^2 *given in* (2.19) *and* (2.20), *respectively.*

Remark 3.3. *If we compare Theorem 3.1 and Theorem 3.2 we see that the estimation of* γ *and* β *at the same level* k *induces an increase in the asymptotic variance of our final* γ*-estimator of factors given by* $((1-\rho)/\rho)^2$ *or* $(a_1^2\sigma_1^2 + a_2^2 - 2a_1^2 a_2)/(a_1^2 - a_2)^2$, *both greater than 1, according as we base our estimation of* β *on the scaled log-spacings* U_i *or on the log-excesses* V_{ik}, *respectively. In Figure 2 we provide both a picture and some values of* $\sigma_{\widehat{\beta}_U}/(\beta\gamma)$, $\sigma_{\widehat{\beta}_V}/(\beta\gamma)$, $\sigma^{WLE}_{\widehat{\beta}_U(k;\ \widehat{\rho})}/\gamma$ *and* $\sigma^{WLE}_{\widehat{\beta}_V(k;\ \widehat{\rho})}/\gamma$, *as a function of* $|\rho|$, *with* $\sigma_{\widehat{\beta}_U}$, $\sigma_{\widehat{\beta}_V}$, $\sigma^{WLE}_{\widehat{\beta}_U(k;\ \widehat{\rho})}$ *and* $\sigma^{WLE}_{\widehat{\beta}_V(k;\ \widehat{\rho})}$ *given in* (2.8), (2.18), (3.3) *and* (3.4), *respectively.*



$\lvert\rho\rvert$	$\sigma^{WLE}_{\hat\beta_U(k)}$	$\sigma^{WLE}_{\hat\beta_V(k)}$	$\sigma_{\hat\beta_U}$	$\sigma_{\hat\beta_V}$
0.1	11.00	12.04	12.05	13.20
0.2	6.00	6.24	7.10	7.40
0.3	4.33	4.44	5.48	5.63
0.4	3.50	3.57	4.70	4.80
0.5	3.00	3.05	4.24	4.32
1.0	2.00	2.04	3.46	3.55
1.5	1.67	1.70	3.33	3.44
2.0	1.50	1.54	3.35	3.50
2.5	1.40	1.44	3.43	3.62

Figure 2: "Rulers" of the asymptotic standard deviations of $\widehat{\beta}_U(k;\ \rho)$, $\widehat{\beta}_V(k;\ \rho)$, $WLE_{\widehat{\beta}_U(k;\ \rho),\ \rho}$ and $WLE_{\widehat{\beta}_V(k;\ \rho),\ \rho}$, for $\gamma = \beta = 1$.

Notice that there is only a very slight difference between the asymptotic variances of $\widehat{\beta}_U(k;\ \rho)$, *based on the scaled log-spacings, and* $\widehat{\beta}_V(k;\ \rho)$, *based on the log-excesses, but such a difference is not at all relevant in practice, and the two estimators provide practically the same results when incorporated in the estima-*

tion of the tail index γ. Due to the slightly smaller asymptotic variances associated to the use of the scaled log-spacings U_i, we shall use, in the simulations, such an estimator of β.

4 Simulated behaviour of the estimators and an application to real data

4.1 The simulation experiment

In Figures 3 and 4 we present the mean value and the mean squared error patterns of the WLE-estimator for two typical heavy-tailed distributions, the Fréchet d.f., $F(x) = \exp\left(-x^{-1/\gamma}\right),\ x \geq 0$ and the Burr d.f., respectively, both with γ and $|\rho|$ equal to 1.

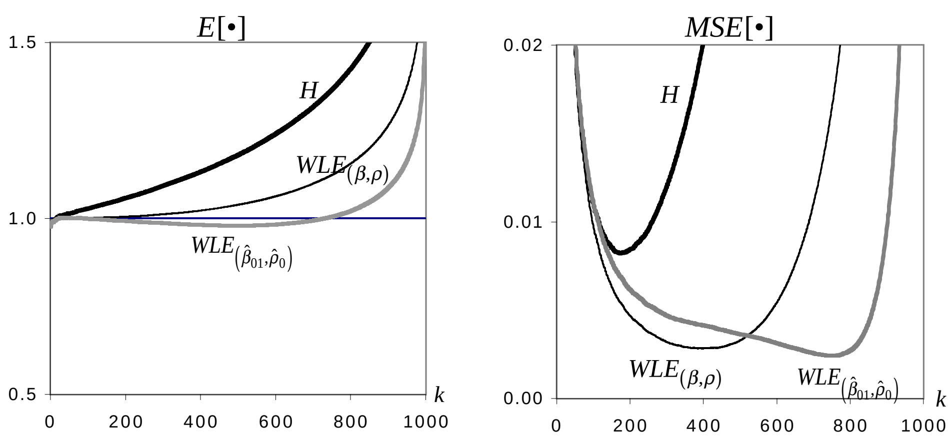


Figure 3: Mean values and Mean Squared errors of the estimator under study for a sample size $n = 1000$, from a standard *Fréchet* parent with γ =1 ($\beta = 0.5,\ \rho = -1$).

The interesting pattern we have got before appears here as well. As said before, we think that the most important feature of this estimator lies on the fact that its mean squared error is smaller than the mean squared error of the Hill estimator for all values of k. For values of $|\rho| > 1$, illustrated here in Figure

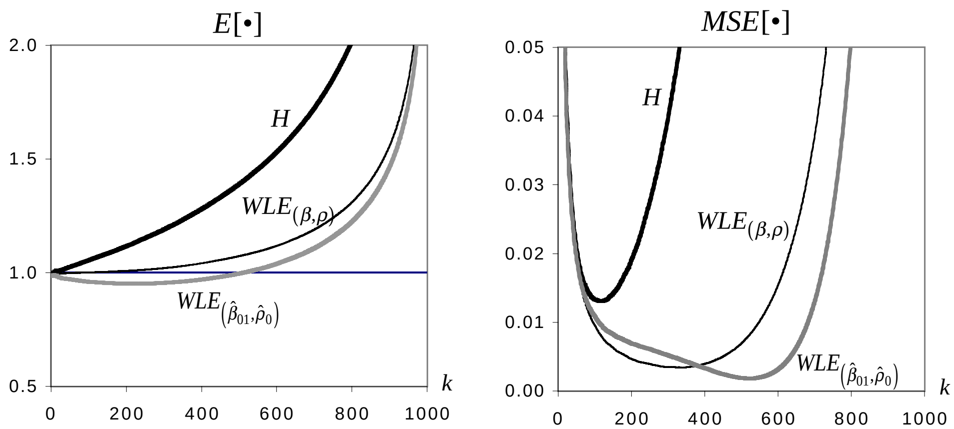


Figure 4: Mean values and Mean Squared errors of the estimator under study for a sample size $n = 1000$, from a *Burr* parent $\gamma = 1$ and $\rho = -1$ ($\beta = 1$).

5, with a Burr parent with $(\gamma, \rho) = (1, -2)$ such a nice feature disappears when we use $\widehat{\rho}_0$, but it is kept if we use $\widehat{\rho}_1$ instead.

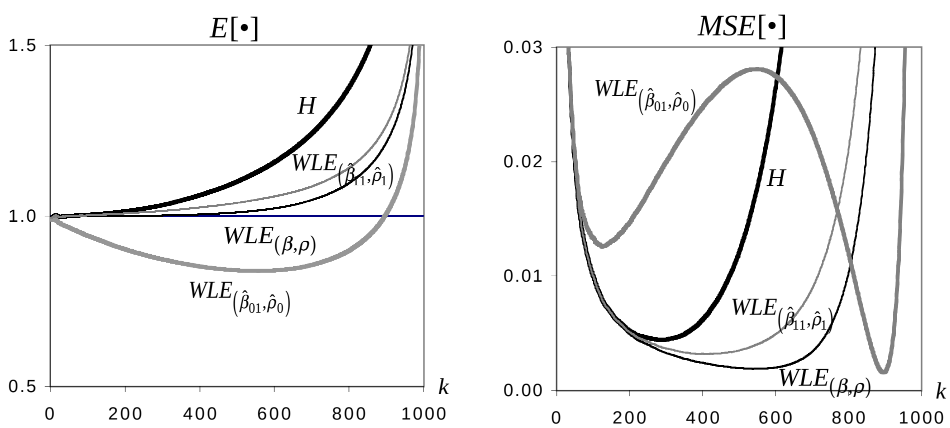


Figure 5: Mean values and Mean Squared errors of the estimator under study for a sample size $n = 1000$, from a *Burr* parent with $\gamma = 1$ and $\rho = -2$ ($\beta = 1$).

The simulation performed for other models enables us to say that it is always safe to use the new estimator $WLE_{\widehat{\beta}_{01},\ \widehat{\rho}_0}$, whenever we are in Hall's class of models and the Hill estimator clearly exhibits a reasonably high bias, either positive or negative, — and this means that we are for sure in a region of ρ-values such that $|\rho| \leq 1$. If $|\rho| > 1$ we shall then use $WLE_{\widehat{\beta}_{11},\ \widehat{\rho}_1}$. Anyway, to achieve the MSE pattern of the r.v. $WLE_{\beta,\ \rho}$, further work on the estimation of the second order parameters, or more generally of the bias' function still needs to be developed.

4.2 An illustration

We shall herewith consider an illustration of the performance of the above mentioned class of estimators, through the analysis of the Euro-UK Pound daily exchange rates from January 4, 1999 till December 15, 2003. In Figure 5, working with the $n_0 = 593$ positive log-returns, we present the sample path of the $\widehat{\rho}_n^{(\tau)}$ estimates in (2.2) (*left*), as function of k, for $\tau = 0$ and $\tau = 1$, together with the sample paths of the classical Hill estimator, H, and of $WLE_{\widehat{\beta}_{01},\ \widehat{\rho}_0}$ (*right*).

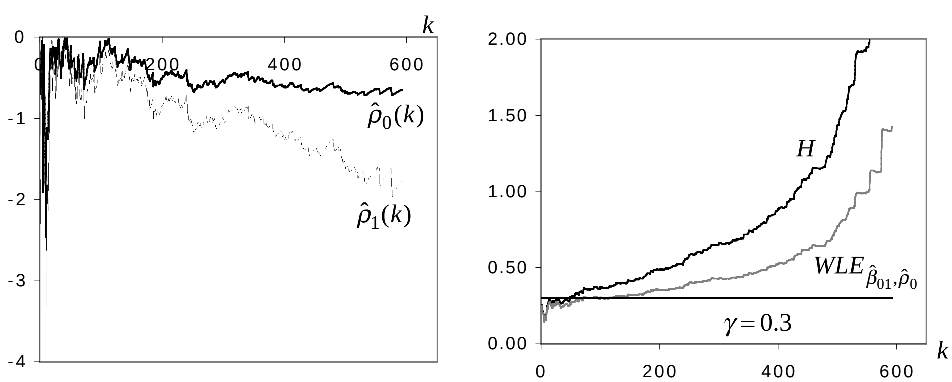


Figure 6: Estimates of the second order parameter ρ (*left*) and of the tail index γ (*right*) for the Daily Log-Returns on the Euro-UK Pound.

The sample paths of the ρ-estimates associated to $\tau = 0$ and $\tau = 1$ lead us to choose, on the basis of any stability criterion for large k, the estimate associated to $\tau = 0$. From previous experience with this type of estimates, we conclude that the underlying ρ-value is larger or equal to -1, and the consideration of $\tau = 0$ is then advisable. The estimate of ρ is in this case $\widehat{\rho}_0 = -0.66$. We further get $\widehat{\beta}_0 = 1.03$.

Regarding the tail index estimation, note that the Hill estimator exhibits a relevant positive bias, as may be seen from Figure 5 (*right*). The other estimator, $WLE_{\widehat{\beta}_{01},\,\widehat{\rho}_0}(k)$, which is "asymptotically unbiased", reveals without doubt a smaller bias, and, in an easier way, enables us to take a decision upon the estimate of γ to be used, with the help of any stability criterion. Indeed, for any level k, any estimate considered on the basis of $WLE_{\widehat{\beta}_{01},\,\widehat{\rho}_0}(k)$ performs for sure better than the estimate based on $H(k)$. In Figure 6, we represent the estimate $\widehat{\gamma} = 0.3$, the median of the $WLE_{\widehat{\beta}_{01},\,\widehat{\rho}_0}(k)$ estimates, for thresholds k between $\left[n^{-2\widehat{\rho}_0/(1-2\widehat{\rho}_0)}\right]/4 = 9$ and $\left[4 \times n^{-2\widehat{\rho}_0/(1-2\widehat{\rho}_0)}\right] = 150$. It is worth noticing that in Gomes *et al.* (2003), the use of a *Best Linear Unbiased Estimator* and an heuristic stablity criterion led us also to an estimate $\widehat{\gamma} = 0.3$. We have there used the following rule: given a set of tail index estimates $\widehat{\gamma}(k)$, $1 \le k < n$, based on the observed sample of size n, consider those estimates with a small number r of decimal figures, and denote them $\widehat{\gamma}_r(k)$.

1. For any possible value a in the domain of $\widehat{\gamma}_r(k)$, consider the largest run associated with a, i.e., $R(a)$, the maximum number of consecutive k values such that $\widehat{\gamma}_r(k) = a$;
2. Consider as a data-driven estimate of the tail index, $\widehat{\gamma} = \arg\max_a R(a)$.

Here, if we consider the tail index estimates $WLE_{\widehat{\beta}_{01},\,\widehat{\rho}_0}(k)$ with one decimal figure, the largest run is also associated to the value 0.3. Such a largest run has a size equal to 131 ($49 \le k \le 189$). If we began counting the run from the first time the value a appears, even if after that we get some values smaller than a, we get a run of size 177 ($13 \le k \le 189$).

5 Appendix

5.1 Proof of Theorem 1.1

We shall base the proof of Theorem 1.1 on the following lemmas:

Lemma 5.1. *For every $\rho < 0$, $\psi(u) = -\left(u^{-\rho} - 1\right)/(\rho \ln u)$, in* (1.8)*, is a limited increasing function in $u \in [0,1]$, assuming values in $[0,1]$. Moreover, $\psi^*(u) = -\ln u \ \psi(u)$, in* (2.12)*, is decreasing in $u \in [0,1]$, assuming values in $[0, -1/\rho]$.*

Proof. From the definition of ψ in (1.8), $\psi(u) = -(u^{-\rho} - 1)/(\rho \ln u)$, $d\psi(u)/du = (1 - u^{-\rho} + u^{-\rho} \ln u^{-\rho})/(-\rho \ u \ln^2 u)$. From the inequality, $(u^{-\rho} - 1)/u^{-\rho} < \ln u^{-\rho}$, $0 < u < 1$, we get $d\psi(u)/du > 0$. We have trivially $\psi(0) = 0$ and $\psi(1) = \lim_{u \to 1} \rho \ u^{-\rho-1}/(\rho \ u^{-1}) = \lim_{u \to 1} u^{-\rho} = 1$. On the other side, since $\psi^*(u) = (u^{-\rho} - 1)/\rho$, $d\psi^*(u)/du = -u^{-\rho-1} < 0$, if $0 < u < 1$, and $\psi^*(0) = -1/\rho$, $\psi^*(1) = 0$.

□

Lemma 5.2. *For integer values j, let us consider*

$$P_k^{(j)} := \frac{1}{k} \sum_{i=1}^{k} \psi_{ik}^{j} \ E_{k-i+1:k}, \ j \geq 0, \tag{5.1}$$

and

$$Q_k^{(j)} := \frac{1}{k} \sum_{i=1}^{k} \psi_{ik}^{j-1} \ \frac{Y_{k-i+1:k}^{\rho} - 1}{\rho}, \ j \geq 1, \tag{5.2}$$

with ψ given in (1.8)*, being $\{E_i\}$ and $\{Y_i\}$ sequences of i.i.d. standard exponential and Pareto r.v.'s, respectively. Denoting $\mathbb{E}$ the mean value operator, both $\mathbb{E}\left(P_k^{(j)}\right)$ and $\mathbb{E}\left(Q_k^{(j)}\right)$ converge towards*

$$a_j := \frac{(-1)^{j-1}}{\rho^j} \int_0^1 \frac{(v^{-\rho} - 1)^j}{\ln^{j-1} v} \ dv = -\int_0^1 \psi^j(v) \ lnv \ dv < \infty, \quad j \geq 1, \tag{5.3}$$

with $a_0 = \mathbb{E}\left(P_k^{(0)}\right) = 1$. For the particular cases $j = 1, 2$, a_1 and a_2 are explicitly given in (2.19)*. Moreover, for $j \geq 0$,*

$$\begin{aligned} \sigma_j^2 &= \lim_{n \to \infty} k \ \mathbb{V}ar\left(P_k^{(j)}\right) \\ &= \frac{2}{\rho^{2j}} \iint_{0 \leq u < v \leq 1} \left(\frac{u^{-\rho} - 1}{\ln u} \ \frac{v^{-\rho} - 1}{\ln v}\right)^j \frac{1 - v}{v} du \ dv < \infty \quad [\sigma_0 = 1]. \end{aligned} \tag{5.4}$$

Consequently, for $j \geq 0$, $P_k^{(j)}$ in (5.1) converges in probability towards a_j, as $k \to \infty$, with $a_j,\ j \geq 1$, given in (5.3), $a_0 = 1$.

Also

$$k\ \mathbb{C}ov\left(P_k^{(0)}, P_k^{(1)}\right) \sim a_1 = 1/(1-\rho). \tag{5.5}$$

Proof. For $j = 0$, the law of large numbers enables us to guarantee immediately that $P_k^{(0)} = \frac{1}{k}\sum_{i=1}^{k} E_i$ converges in probability towards $\mathbb{E}\left(P_k^{(0)}\right) = 1 =: a_0$. For $j \geq 1$, the use of Rényi's representation of exponential o.s.'s as linear combinations of independent standard exponential r.v.'s (Rényi, 1953), $E_{i:n} = \sum_{j=1}^{i} E_j/(n-j+1)$, $1 \leq i \leq n$, enables us to write

$$\begin{aligned}
\mathbb{E}\left(P_k^{(j)}\right) &= \frac{1}{k}\sum_{i=1}^{k}\psi_{ik}^{j}\ \mathbb{E}\left(E_{k-i+1:k}\right) = \frac{1}{k}\sum_{i=1}^{k}\psi_{ik}^{j}\ \mathbb{E}\left(\sum_{r=i}^{k}\frac{E_r}{r}\right) \\
&= \left(\frac{1}{k}\sum_{i=1}^{k}\psi^j(i/k)\left(\frac{1}{k}\sum_{r=i}^{k}\frac{1}{r/k}\right)\right) \\
&\underset{k\to\infty}{\longrightarrow} \int_0^1 \psi^j(v)\ dv\int_v^1 \frac{1}{u}\ du = -\int_0^1 \psi^j(v)\ln(v)dv = a_j,
\end{aligned}$$

as given in (5.3). Note that, from Lemma 5.1, $|\psi| \leq 1$, and consequently, $0 \leq a_j \leq a_1 = 1/(1-\rho)$, for any $j \geq 1$, being thus finite. Since

$$\mathbb{C}ov\left(E_{k-i+1:k}, E_{k-r+1:k}\right) = \mathbb{V}ar\left(E_{k-\max(i,r)+1:k}\right) = \sum_{s=\max(i,r)}^{k}\frac{1}{s^2},$$

$$\begin{aligned}
k\ \mathbb{V}ar\left(P_k^{(j)}\right) &= \frac{2}{k}\sum_{1\leq i\leq r\leq k}\psi_{ik}^{j}\psi_{rk}^{j}\mathbb{V}ar\left(E_{k-r+1:k}\right) \\
&= \frac{2}{k}\sum_{i=1}^{k}\psi^j(i/k)\left(\frac{1}{k}\sum_{r=i}^{k}\psi^j(r/k)\left(\frac{1}{k}\sum_{s=r}^{k}\frac{1}{(s/k)^2}\right)\right) \\
&\underset{k\to\infty}{\longrightarrow} 2\iint\limits_{0\leq u\leq v\leq 1}\psi^j(u)\ \psi^j(v)\ \left(\frac{1-v}{v}\right)du\ dv \\
&=: \sigma_j^2(\rho), \quad \text{as given in (5.4).}
\end{aligned}$$

Again from Lemma 5.1, for every ρ, $\sigma_j^2(\rho)$ is non-increasing in j and bounded by $\sigma_0^2 = k\ \mathbb{V}ar\left(P_k^{(0)}\right) = 1$, being consequently finite. Since $\mathbb{E}\left(P_k^{(j)}\right) \to a_j$ and $\mathbb{V}ar\left(P_k^{(j)}\right) \to 0$, as $k \to \infty$, $P_k^{(j)}$ converges in probability towards a_j, for every $j \geq 0$. We also have

$$\begin{aligned}
k\ \mathbb{C}ov\left(P_k^{(0)}, P_k^{(1)}\right) &= \frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{k}\psi_{jk}\mathbb{C}ov\left(E_{k-i+1:k}, E_{k-j+1:k}\right) \\
&= \frac{1}{k}\left\{\sum_{i=1}^{k}\sum_{j=i}^{k}\psi_{jk}\sum_{s=j}^{k}\frac{1}{s^2} + \sum_{i=1}^{k}\sum_{j=1}^{i-1}\psi_{jk}\sum_{s=i}^{k}\frac{1}{s^2}\right\} \\
&= \frac{1}{k}\left\{\sum_{j=1}^{k}\psi_{jk}\sum_{i=1}^{j}\left(\sum_{s=j}^{k}\frac{1}{s^2}\right) + \sum_{j=1}^{k-1}\psi_{jk}\sum_{s=j+1}^{k}\frac{s-j}{s^2}\right\} \\
&= \frac{1}{k}\left\{\sum_{j=1}^{k}\psi_{jk}\sum_{s=j}^{k}\frac{j}{s^2} + \sum_{j=1}^{k-1}\psi_{jk}\sum_{s=j+1}^{k}\frac{s-j}{s^2}\right\} \\
&\underset{k\to\infty}{\longrightarrow} \int_0^1 \psi(u)du\int_u^1\frac{1}{v}dv = a_1 = 1/(1-\rho),
\end{aligned}$$

and (5.5) follows. □

Proof. (Theorem 1.1). As seen before in (1.6),

$$V_{ik} \stackrel{d}{=} \gamma\ E_{k-i+1:k} + \frac{Y_{k-i+1:k}^{\rho} - 1}{\rho}\ A(n/k)\ (1 + o_p(1)),$$

with the $o_p(1)$ uniform in i, $1 \leq i \leq k$. Also,

$$\exp\left(-\beta\left(\frac{n}{k}\right)^{\rho}\psi_{ik}\right) = 1 - \beta\left(\frac{n}{k}\right)^{\rho}\psi_{ik}\left(1 + o(1)\right),$$

with the $o(1)$-term again uniform in i, $1 \leq i \leq k$. We may then write

$$\begin{aligned}
WLE(k) &= \frac{1}{k}\sum_{i=1}^{k} e^{-\beta(n/k)^{\rho}\psi_{ik}}\ V_{ik} \\
&= \frac{1}{k}\sum_{i=1}^{k} V_{ik} - \beta\left(\frac{n}{k}\right)^{\rho}\left(\frac{1}{k}\sum_{i=1}^{k}\psi_{ik}\ V_{ik}\right)\left(1 + o_p(1)\right), \\
&\stackrel{d}{=} \gamma\ P_k^{(0)} + A(n/k)\left(Q_k^{(1)} - P_k^{(1)}\right)(1 + o_p(1)),
\end{aligned}$$

with $P_k^{(j)}$ and $Q_k^{(j)}$ given in (5.1) and (5.2), respectively. The law of large numbers enable us to guarantee that $Q_k^{(1)}$ converges in probability towards $\mathbb{E}\left((Y^\rho-1)/\rho\right) = 1/(1-\rho) \equiv a_1$, and from Lemma 5.2, $P_k^{(1)}$ converges also in probability towards a_1. Since both $P_k^{(1)}$ and $Q_k^{(1)}$ converge in probability towards $a_1 = 1/(1-\rho)$, (1.13) as well as the remaining of the theorem follow, with $N_k = \sqrt{k}\left(P_k^{(0)}-1\right)$, i.e., the usual dominant component of bias, which is for the classical estimators of the tail index of the order of $A(n/k)$, is now of smaller order. □

5.2 Proof of theorems in Section 2

Proof. (Theorem 2.1). Consistency of $\widehat{\beta}_U(k,\widehat{\rho})$, with $\widehat{\rho}$ any consistent estimator of ρ, together with the result in (2.7) have been proved in Gomes and Martins (2002). Next, note that

$$\frac{d^j}{d\rho^j}\widehat{\beta}_U(k;\rho) = -\widehat{\beta}_U(k;\rho)\ln^j(n/k)(1+o_p(1)), \quad j \geq 1,$$

and, provided that $(\widehat{\rho}-\rho)\ln(n/k) = o_p(1)$,

$$\widehat{\beta}_U(k;\widehat{\rho}) = \widehat{\beta}_U(k;\rho) - \widehat{\beta}_U(k;\rho)\ln(n/k)(\widehat{\rho}-\rho)(1+o_p(1)).$$

Consequently, (2.7) holds true, with ρ replaced by $\widehat{\rho}$. If $\widehat{\rho}$ is any of the estimators in (2.2) computed at the level k_1 in (2.3), $\widehat{\rho}-\rho = O_p\left((\ln\ln n)^{(1-2\rho)/2}/\sqrt{n}\right)$, as noticed in Remark 2.1, and consequently $(\widehat{\rho}-\rho)\ln(n/k) = o_p(1)$.

The result related to $\widehat{\beta}_U\,(k,\ \widehat{\rho}(k))$ comes from the fact that, since $d\widehat{\beta}_U(k,\ \rho)/d\rho = -\ln(n/k)\ \widehat{\beta}_U(k,\ \rho)$,

$$\widehat{\beta}_U\,(k,\ \widehat{\rho}(k)) = \beta - \beta\ \left(\widehat{\rho}(k)-\rho\right)\ \ln(n/k)(1+o_p(1)),$$

and (2.9) follows. □

Before the proof of Theorem 2.2, we first state a lemma, proved in Chernoff *et al.* (1967):

Lemma 5.3. (*Chernoff et al.*, 1967). *Let* $Z_k = \sum_{r=1}^{k} \alpha_{rk} \left(E_r - 1\right)/k$. *Let* $v_k = \sqrt{\frac{1}{k}\sum_{r=1}^{k} \alpha_{rk}^2} = \sqrt{k\ \mathbb{V}ar(Z_k)}$. *Then* $\sqrt{k}\ Z_k/v_k \underset{n\to\infty}{\overset{d}{\longrightarrow}}$ *Normal* $(0,1)$ *if and only if* $\max_{1\le r\le k} |\alpha_{rk}| = o\left(\sqrt{k}\ v_k\right)$, *as* $k \to \infty$.

We shall next prove the following lemmas:

Lemma 5.4. *For any* $j \ge 0$, *and with* $P_k^{(j)}$, a_j *and* σ_j^2 *given in* (5.1), (5.3) *and* (5.4), *respectively,*

$$\overline{P}_k^{(j)} := \sqrt{k}\left(P_k^{(j)} - \mathbb{E}\left(P_k^{(j)}\right)\right)/\sigma_j \overset{a}{\sim} \text{Normal}(0,1). \tag{5.6}$$

Proof. We may write,

$$\begin{aligned} P_k^{(j)} &= \frac{1}{k}\sum_{i=1}^{k} \psi_{ik}^j E_{k-i+1:k} = \frac{1}{k}\sum_{i=1}^{k}\psi_{ik}^j \sum_{r=i}^{k}\frac{E_r}{r} = \frac{1}{k}\sum_{r=1}^{k}\left(\frac{1}{r}\sum_{i=1}^{r}\psi_{ik}^j\right)E_r \\ &=: \frac{1}{k}\sum_{r=1}^{k}\alpha_{rk}^{(j)}E_r = \frac{1}{k}\sum_{r=1}^{k}\alpha_{rk}^{(j)}\left(E_r - 1\right) + \frac{1}{k}\sum_{r=1}^{k}\alpha_{rk}^{(j)} \\ &=: Z_k^{(j)} + \frac{1}{k}\sum_{r=1}^{k}\alpha_{rk}^{(j)} = Z_k^{(j)} + \mathbb{E}\left(P_k^{(j)}\right). \end{aligned}$$

Since $\alpha_{rk}^{(j)} = \sum_{i=1}^{r}\psi_{ik}^j/r$, with $\psi_{ik} = \psi(i/k)$, $\psi(u)$ increasing in $u \in [0,1]$, and varying between 0 and 1, we have

$$\max_{1\le r\le k}\left|\alpha_{rk}^{(j)}\right| = \alpha_{kk}^{(j)} = \frac{1}{k}\sum_{i=1}^{k}\psi_{ik}^j \underset{k\to\infty}{\longrightarrow} \int_0^1 \left(\frac{x^{-\rho}-1}{-\rho\ln x}\right)^j dx < \infty.$$

Also, $v_k^{(j)} = \sqrt{k\ \mathbb{V}ar\left(Z_k^{(j)}\right)} = \sqrt{k\ \mathbb{V}ar\left(P_k^{(j)}\right)} \longrightarrow \sigma_j(\rho)$, finite and given in (5.4). From Lemma 5.3, $\sqrt{k}\ Z_k^{(j)}/v_k^{(j)}$ is asymptotically standard normal, and consequently $\sqrt{k}\ Z_k^{(j)}/\sigma_j(\rho)$ and $\sqrt{k}\left(P_k^{(j)} - \mathbb{E}\left(P_k^{(j)}\right)\right)/\sigma_j(\rho)$ are also asymptotically standard normal, as $k \to \infty$. $\square$

Lemma 5.5. *For every $j > 1$, and with $\psi(u)$ given in* (1.8)*, the function $\varphi_j(u) := -\ln u\ \psi^j(u)$, $0 \le u \le 1$, is bounded, $0 \le \varphi_j(u) \le -1/\rho$, $0 \le u \le 1$, with a unique maximum at $k_0 \in (0,1)$, being $\varphi_j(0) = \varphi_j(1) = 0$.*

Proof. From Lemma 5.1, $|\psi| \le 1$, and consequently φ_j is non-increasing in j, and bounded by $\varphi_1(u) \equiv \psi^*(u) \le -1/\rho$, for every $u \in [0,1]$. Let us think on the function $g(u) = -\rho\ \varphi_j(u^{-1/\rho}) = -\ln u\,((u-1)/\ln u)^j$, with the same type of behaviour of $\varphi_j(u)$ for $0 < u < 1$. We easily get $g'(u) = -\left((u-1)/\ln u\right)^{j-1}\left(j\ u\ln u - (j-1)(u-1)\right)/(u\ln u) = 0$ if and only if $j\ u\ln u - (j-1)(u-1) = 0$, $0 < u < 1$. Equivalently, with $t = 1/u$, we get $\ln t = \left(\frac{j-1}{j}\right)(t-1)$, $t > 1$. For $j > 1$, this equation has a unique solution, and the result follows. □

Lemma 5.6. *For any integer $j \ge 0$, and with $P_k^{(j)}$, a_j and σ_j^2 given in* (5.1)*,* (5.3) *and* (5.4)*, respectively,*

$$\overline{P}_k^{(j)} := \sqrt{k}\left(P_k^{(j)} - a_j\right)/\sigma_j \ \overset{a}{\sim}\ \mathit{Normal}(0,1). \tag{5.7}$$

Proof. From Lemma 5.4, $\sqrt{k}\left(P_k^{(j)} - \mathbb{E}\left(P_k^{(j)}\right)\right)/\sigma_j \ \overset{a}{\sim}\ \text{Normal}(0,1)$. It is thus enough showing that

$$\mathbb{E}\left(P_k^{(j)}\right) - a_j = \frac{1}{k}\sum_{i=1}^{k}\psi_{ik}^j\ \sum_{r=i}^{k}\frac{1}{r} - \int_0^1 \varphi_j(v)dv = o\left(\frac{1}{\sqrt{k}}\right),$$

where $\varphi_j(v) := -\ln v\ \psi^j(v)$ was studied in Lemma 5.1 and in Lemma 5.5, for $j = 1$ and $j > 1$, respectively. Note that

$$\sum_{r=i+1}^{k}\frac{1}{r} \le -\ln(i/k) \le \sum_{r=i}^{k-1}\frac{1}{r}, \qquad \text{i.e.,} \qquad \frac{1}{k} \le \sum_{r=i}^{k}\frac{1}{r} + \ln(i/k) \le \frac{1}{i},$$

and consequently, since $\mathbb{E}\left(P_k^{(j)}\right) - \frac{1}{k}\sum_{i=1}^{k}\varphi_j\left(i/k\right) = \frac{1}{k}\sum_{i=1}^{k}\psi_{ik}^j\left[\sum_{r=i}^{k}\frac{1}{r} + \ln\frac{i}{k}\right]$,

$$\frac{1}{k^2}\sum_{i=1}^{k}\psi_{ik}^j \ \le\ \mathbb{E}\left(P_k^{(j)}\right) - \frac{1}{k}\sum_{i=1}^{k}\varphi_j\left(i/k\right) \ \le\ \frac{1}{k}\sum_{i=1}^{k}\frac{\psi_{ik}^j}{i} \ \le\ \frac{1}{k}\sum_{i=1}^{k}\frac{1}{i}.$$

Hence

$$\left|\mathbb{E}\left(P_k^{(j)}\right) - \frac{1}{k}\sum_{i=1}^{k}\varphi_j\left(i/k\right)\right| \;\le \frac{1}{k}\sum_{i=1}^{k}\frac{1}{i} \;\le \frac{1+\ln k}{k}.$$

On the other side, for $j = 1$, $\varphi_1(u) = \psi^*(u)$ is a decreasing function, and consequently,

$$\frac{1}{k}\sum_{i=1}^{k}\varphi_1(i/k) \;\le\; \int_0^1 \varphi_1(v)\ dv \;\le\; \frac{1}{k}\sum_{i=0}^{k-1}\varphi_1(i/k) = \frac{1}{k}\sum_{i=1}^{k}\varphi_1(i/k) - \frac{1}{\rho\ k},$$

i.e.,

$$0 \;\le\; \int_0^1 \varphi_1(v)\ dv - \frac{1}{k}\sum_{i=1}^{k}\varphi_1(i/k) \;\le\; -\frac{1}{\rho\ k}.$$

For any integer $j > 1$, and as seen in Lema 5.5, $0 \le \varphi_j(u) \le -1/\rho$, $\varphi_j(0) = \varphi_j(1) = 0$, first increasing on $[0, s_0]$ and next decreasing on $[s_0, 1]$. For each k, define $k_0 \in [1, k]$ such that $k_0/k \le s_0 \le (k_0+1)/k$. Then

$$\frac{1}{k}\left(\sum_{i=0}^{k_0}\ \varphi_j(i/k) + \sum_{i=k_0+2}^{k}\ \varphi_j(i/k)\right) \le \int_0^1 \varphi_j(v)\ dv \le \frac{1}{k}\left(\sum_{i=1}^{k-1}\ \varphi_j(i/k) + \varphi_j(s_0)\right).$$

Hence,

$$\frac{1}{k}\left(\sum_{i=1}^{k}\ \varphi_j(i/k) - \varphi_j(s_0)\right) \le \int_0^1 \varphi_j(v)\ dv \le \frac{1}{k}\left(\sum_{i=1}^{k}\ \varphi_j(i/k) + \varphi_j(s_0)\right).$$

and consequently,

$$\left|\frac{1}{k}\sum_{i=1}^{k}\varphi_j(i/k) - \int_0^1 \varphi_j(v)\ dv\right| \;\le\; \frac{\varphi_j(s_0)}{k} \le -\frac{1}{\rho\ k}.$$

Then, for any $j \ge 1$,

$$\left|\mathbb{E}\left(P_k^{(j)}\right) - a_j\right| \;\le\; \frac{1+\ln k}{k} - \frac{1}{\rho\ k}.$$

Consequently, for any $\epsilon > 0$, $k^{1-\epsilon}\left(\mathbb{E}\left(P_k^{(j)}\right) - a_j\right) \to 0$, and the lemma follows.

□

Lemma 5.7. *For $j \ge 0$, and for $B_k^{(j)}$ in* (1.11),

$$B_k^{(j)} \;\stackrel{d}{=}\; \gamma\ a_j + \frac{\gamma\ \sigma_j}{\sqrt{k}}\ \overline{P}_k^{(j)} + a_{j+1}A(n/k)(1+o_p(1)),$$

with a_j, σ_j and $\overline{P}_k^{(j)}$ given in (5.3), (5.4) *and* (5.7), *respectively.*

Proof. We have for $j \geq 0$,

$$\begin{aligned} B_k^{(j)} &\stackrel{d}{=} \frac{\gamma}{k}\sum_{i=1}^{k} \psi_{ik}^{j}\ E_{k-i+1:k} + A(n/k)\ \frac{1}{k}\sum_{i=1}^{k} \frac{\psi_{ik}^{j}\ \left(Y_{k-i+1:k}^{\rho}-1\right)}{\rho}(1+o_p(1)) \\ &\stackrel{d}{=} \gamma\ a_j + \frac{\gamma\ \sigma_j}{\sqrt{k}}\ \overline{P}_k^{(j)} + a_{j+1}\ A(n/k)(1+o_p(1)). \end{aligned}$$

□

Proof. (Theorem 2.2). Notice that

$$\frac{\widehat{\beta}_V(k;\rho)}{\beta} = \frac{\gamma}{A(n/k)}\ \frac{s_k^*\ B_k^{(0)} - B_k^{(1)}}{s_k^*\ B_k^{(1)} - B_k^{(2)}},$$

with $B_k^{(j)}$ given in (1.11). Noticing that $a_1 = 1/(1-\rho)$ is the limiting value of s_k^*, as $n \to \infty$, and assuming ρ known, we have

$$\begin{aligned} \frac{A(n/k)\ \widehat{\beta}_V(k;\rho)}{\gamma\ \beta} &\stackrel{d}{=} \frac{a_1\ B_k^{(0)}(1+o(1)) - B_k^{(1)}}{a_1\ B_k^{(1)}(1+o(1)) - B_k^{(2)}} \\ &\stackrel{d}{=} \frac{\frac{\gamma}{\sqrt{k}}\left(a_1\overline{P}_k^{(0)} - \sigma_1\overline{P}_k^{(1)}\right) + (a_1^2-a_2)\ A(n/k) + o_p(A(n/k))}{\gamma(a_1^2-a_2)(1+o_p(1))}, \end{aligned}$$

and consequently,

$$\frac{\widehat{\beta}_V(k;\rho)}{\beta} \stackrel{d}{=} 1 + \frac{\gamma}{\sqrt{k}\ A(n/k)}\left(\frac{a_1\overline{P}_k^{(0)} - \sigma_1\overline{P}_k^{(1)}}{a_1^2-a_2}\right) + R_k^{V}, \quad \text{with} \quad R_k^{V} = o_p(1),$$

i.e., $\widehat{\beta}_V(k;\rho)$ converges in probability towards β, provided that $\sqrt{k}\ A(n/k) \to \infty$, as $n \to \infty$. The same result is true if we replace ρ by $\widehat{\rho}$, any consistent estimator of the second order parameter ρ.

Next, note that also here

$$\frac{d}{d\rho}\widehat{\beta}_V(k;\rho) = -\widehat{\beta}_V(k;\rho)\ln(n/k)(1+o_p(1)).$$

Consequently, (2.17), (2.21), as well as the remaining of the theorem follow, as in the proof of Theorem 2.1. □

Remark 5.1. *Note that, more generally than Theorem 1.1, and with the same notation as before, we may say that, for every $j \geq 0$,*

$$\begin{aligned} A_k^{(j)} &= B_k^{(j)} - \beta\left(\frac{n}{k}\right)^{\rho} B_k^{(j+1)}(1+o_p(1)) = B_k^{(j)} - \frac{A(n/k)}{\gamma}\ B_k^{(j+1)}(1+o_p(1)) \\ &\stackrel{d}{=} \gamma\ a_j + \frac{\gamma\ \sigma_j}{\sqrt{k}}\ \overline{P}_k^{(j)} + a_{j+1}\ A(n/k) - \frac{A(n/k)}{\gamma}\ (\gamma\ a_{j+1} + o_p(1)) \\ &\stackrel{d}{=} \gamma\ a_j + \frac{\gamma\ \sigma_j}{\sqrt{k}}\ \overline{P}_k^{(j)} + o_p(A(n/k)). \end{aligned}$$

5.3 Proof of theorems in Section 3

Proof. (Theorem 3.1). Denoting $W := WLE(k)$, it is enough noticing that

$$\frac{\partial W}{\partial \beta} = O_p\left(A(n/k)\right),\ \frac{\partial^2 W}{\partial \beta^2} = O_p\left(A^2(n/k)\right),\ \frac{\partial W}{\partial \rho} = O_p\left(A(n/k)\ln(n/k)\right),$$

$$\frac{\partial^2 W}{\partial \beta \partial \rho} = O_p\left(A(n/k)\ln(n/k)\right),\ \frac{\partial^2 W}{\partial \rho^2} = O_p\left(A(n/k)\ln^2(n/k)\right),$$

Consequently, the use of the δ-method, enables us to write

$$WLE_{\widehat{\beta},\ \widehat{\rho}}(k) - WLE_{\beta,\rho}(k) \sim \left(\widehat{\beta} - \beta\right) \times O_p(A(n/k)) + (\widehat{\rho} - \rho) \times O_p\left(A(n/k)\ln(n/k)\right). \quad (5.8)$$

If $\widehat{\beta}$ and $\widehat{\rho}$ are consistent for the estimation of β and ρ, respectively, and $(\widehat{\rho} - \rho)\ \ln(n/k) = o_p(1)$, the two terms in the second member of the previous relation are $o_p(A(n/k))$. In order to have no kind of modification in the asymptotic variance of $WLE_{\beta,\ \rho}(k)$, we need to impose the two extra conditions in the theorem, i.e., $\left\{\widehat{\beta} - \beta\right\}$ and $(\widehat{\rho} - \rho)\ln(n/k)$ need both to be $o_p\left(1/\left(\sqrt{k}\ A(n/k)\right)\right)$. If we consider the estimators $\widehat{\rho}$ and $\widehat{\beta}$ in (2.2) and (2.6) [or (2.16)], respectively, computed at the level k_1 in (2.3), $\widehat{\rho} - \rho = O_p\left((\ln_2 n)^{(1-2\rho)/2}/\sqrt{n}\right)$, and we merely need to guarantee that

$$\sqrt{k}\ A(n/k)\ \ln(n/k) \times (\widehat{\rho} - \rho) \underset{n\to\infty}{\overset{p}{\longrightarrow}} 0.$$

Such a result is obviously true if $\sqrt{k}\ A(n/k) \to \lambda$, finite, but we may even have the same result if $\sqrt{k}\ A(n/k) \to \infty$. Indeed, if $\sqrt{k}\ A(n/k) \to \infty$, k is of a larger order than $n^{-2\rho/(1-2\rho)}$. Then $n/k < O\left(n^{1/(1-2\rho)}\right)$, $\ln(n/k)/A(n/k) < O\left(n^{-\rho/(1-2\rho)} \ln n\right)$, and consequently,

$$0 \le \left| \frac{(\widehat{\rho}-\rho)\ln(n/k)}{A(n/k)} \right| < O\left(\frac{(\ln\ln n)^{(1-2\rho)/2}\ \ln n}{n^{1/(2(1-2\rho))}} \right) \underset{n\to\infty}{\longrightarrow} 0, \qquad (5.9)$$

i.e., $\sqrt{k}\ A(n/k)\,(\widehat{\rho}-\rho)\ln(n/k) \to 0$, as $n \to \infty$, provided that $\sqrt{k}\ A^2(n/k) \to \lambda$, finite, as assumed. □

Proof. (Theorem 3.2). Since (5.8) holds true with $\widehat{\beta}$ replaced by $\widehat{\beta}_U(k)$, but $\partial WLE_{\beta,\ \rho}(k)/\partial\beta = O_p\left(A(n/k)\right)$ and $\widehat{\beta}_U(k) - \beta = O_p\left(1/\left(\sqrt{k}\ A(n/k)\right)\right)$, we are going to get an extra term of the order of $1/\sqrt{k}$, which is going to modify the asymptotic variance of our final tail index estimator. When we base the estimation of β on the scaled log-spacings in (2.5), with the same notation as before, and noticing that $N_k \equiv Z_k^{(1)}$, $Z_k^{(\alpha)}$ given in (2.10), the term of the order of $1/\sqrt{k}$, in $WLE_{\widehat{\beta}_U(k),\ \widehat{\rho}}(k)$, is going to be

$$\frac{\gamma}{\sqrt{k}} \left(Z_k^{(1)} + \frac{(1-\rho)(1-2\rho)}{\rho^2} \left(\frac{Z_k^{(1)}}{1-\rho} - \frac{Z_k^{(1-\rho)}}{\sqrt{1-2\rho}} \right) \right),$$

which may be written as

$$\frac{\gamma}{\sqrt{k}} \left(\left(\frac{1-\rho}{\rho} \right)^2 Z_k^{(1)} - \frac{(1-\rho)\sqrt{1-2\rho}}{\rho^2}\ Z_k^{(1-\rho)} \right).$$

Since $Z_k^{(\alpha)}$ are asymptotically standard normal r.v.'s and the asymptotic variance between $Z_k^{(1)}$ and $Z_k^{(1-\rho)}$ is given by $\sqrt{1-2\rho}/(1-\rho)$, (3.1) follows. The result in (3.2) follows in a similar way. □

References

[1] Beirlant, J., Dierckx, G., Goegebeur, Y. and Matthys, G. (1999). Tail index estimation and an exponential regression model. *Extremes* **2**, 177-200.

[2] Csörgő, S., Deheuvels, P. and Mason, D. (1985). Kernel estimates of the tail index of a distribution. *Ann. Statist.* **13**, 1050-1077.

[3] Drees, H. (1998). A general class of estimators of the extreme value index. *J. Statist. Planning and Inference* **98**, 95-112.

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

[5] Fraga Alves, M. I., Gomes, M. I. and de Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica* **60**:1, 193-213.

[6] Geluk, J. and de Haan, L. (1987). *Regular Variation, Extensions and Tauberian Theorems*. CWI Tract 40, Center for Mathematics and Computer Science, Amsterdam, Netherlands.

[7] Gomes, M. I., Figueiredo, F. and Mendonça, S. (2003). Asymptotically best linear unbiased tail estimators under a second order regular variation condition. Accepted at *J. Statist. Planning and Inference*.

[8] Gomes, M. I. and de Haan, L. (1999). Approximation by penultimate extreme value distributions. *Extremes* **2**:1, 71-85.

[9] Gomes, M. I., de Haan, L. and Peng, L. (2002). Semi-parametric estimation of the second order parameter — asymptotic and finite sample behaviour. *Extremes* **5**:4, 387-414.

[10] Gomes, M. I. and Martins, M. J. (2002). "Asymptotically unbiased" estimators of the tail index based on external estimation of the second order parameter. *Extremes* **5**:1, 5-31.

[11] Gomes, M. I. and Martins, M. J. (2004). Bias reduction and explicit efficient estimation of the tail index. *J. Statist. Planning and Inference* **126**:2.

[12] Gomes, M. I., Martins, M. J. and Neves, M. (2000). *Alternatives to a semi-parametric estimator of parameters of rare events — the Jackknife methodology. Extremes* **3**:3, 207-229.

[13] Hall, P. and Welsh, A. H. (1985). Adaptive estimates of parameters of regular variation. *Ann. Statist.* **13**, 331-341.

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

[15] Rényi, A. (1953). On the theory of order statistics. *Acta Math. Acad. Sci. Hung.* **4**, 191-231.