

Assessing confidence intervals for the tail index by Edgeworth expansions for the Hill estimator

ERICH HAEUSLER¹ and JOHAN SEGERS²

¹*Mathematical Institute, University of Giessen, Arndtstrasse 2, D-35392 Giessen, Germany.
E-mail: erich.haeusler@math.uni-giessen.de*

²*Department of Econometrics and OR, Tilburg University, PO Box 90153, NL-5000 LE Tilburg, The Netherlands. E-mail: jsegers@uvt.nl*

We establish Edgeworth expansions for the distribution function of the standardized Hill estimator for the reciprocal of the index of regular variation of the tail of a distribution function. The expansions are used to derive expansions for coverage probabilities of confidence intervals for the tail index based on the Hill estimator.

Keywords: asymptotic normality; confidence intervals; Edgeworth expansions; extreme value index; Hill estimator; regular variation; tail index

1. Introduction

Three decades have passed since Hill's (1975) seminal paper on the estimation of the index of regular variation of the tail of a distribution function, thereby introducing what is now unanimously called the Hill estimator for the tail index, the latter being defined as the reciprocal of the index of regular variation. Since then, the tail-estimation literature has witnessed a true explosion featuring numerous alternative estimators, each claimed by its inventors to be better than its competitors in at least a number of more or less well-specified situations. Despite all this scientific vigour, Hill's estimator remains as popular as ever. Why? One reason may be that its expression is so elegant and its implementation so simple: extract the top $k + 1$ observations $X_{n-k:n} \leq X_{n-k+1:n} \leq \dots \leq X_{n:n}$ from a given sample X_1, \dots, X_n and compute

$$\hat{H}_n(k) = \frac{1}{k} \sum_{i=1}^k \log X_{n-i+1:n} - \log X_{n-k:n}.$$

Another reason may be that its interpretations are so convincing: (pseudo-)maximum likelihood estimator in an exponential model for log-excesses and least-squares estimator of the slope of the ultimately linear part in a Pareto quantile–quantile plot. A third, more sophisticated, reason may be that its asymptotic variance is minimal in carefully formulated settings of allowed models and estimators; see Reiss (1989: Section 9.4), Drees (1998),

Segers (2001a) and Beirlant *et al.* (2006). The Hill estimator is probably the most intensively studied statistic in the extreme-value literature, the first papers on its asymptotic properties dating back to Mason (1982) and Hall (1982). More recent contributions such as Resnick and Stărică (1995, 1998) treat the case of dependent data.

Our aim is to add to the understanding of the Hill estimator through the derivation of detailed asymptotic expansions of its distribution function. These Edgeworth expansions then serve to derive asymptotic expansions for the coverage probabilities of a number of two-sided confidence intervals for the tail index which involve the Hill estimator in a natural way. The confidence intervals under consideration are the Wald, score, likelihood ratio and Bartlett-corrected likelihood ratio confidence regions that arise from the Pareto pseudo-log-likelihood given the relative excesses $X_{n-i+1:n}/X_{n-k:n}$ for $i = 1, \dots, k$. The expansions take the form of a main term based upon the asymptotic normality of the Hill estimator plus a number of correction and remainder terms.

This line of research was initiated in Cheng and Pan (1998), featuring a one-term expansion in the case of zero asymptotic bias. In the same case, expansions of arbitrary length in terms of certain gamma distributions were established in Cheng and de Haan (2001) and Guillou and Hall (2001). In the more difficult case of non-zero asymptotic bias, the only relevant work we are aware of is Ferreira (2002: Chapter 4), containing a one-term expansion in the case where the number of order statistics is the one for which the asymptotic mean squared error of the Hill estimator is minimal. All these expansions, however, lack the accuracy or – as far as their technical assumptions are concerned – the flexibility to generate easily comprehensible coverage probability expansions for the aforementioned two-sided confidence intervals based on the Hill estimator.

The Hill-based confidence intervals for the tail index are described in Section 2. In Section 3, expansions are derived for intermediate sequences $k = k_n$ that grow to infinity sufficiently slowly with n so that the bias of the Hill estimator does not enter the main correction term in the coverage probability expansion. We will call this the case of negligible bias. For such k_n , the Bartlett likelihood ratio intervals achieve the highest accuracy. For intermediate sequences k_n growing to infinity at faster rates, even when converging to zero, the bias enters the coverage probability expansions as well, making the performance of the various intervals dependent on the sign of this bias; this is the case of non-negligible bias and the topic of Section 4. The proofs of the two main theorems are spelled out in Sections 5 and 6. The Appendix contains a number of auxiliary results.

2. Confidence intervals

Recall that a positive, measurable function a defined on a neighbourhood of infinity is called regularly varying (at infinity) with real index τ , $a \in \mathcal{R}_\tau$, if $a(ux)/a(u) \rightarrow x^\tau$ as $u \rightarrow \infty$, for all $0 < x < \infty$. A distribution function F on the real line with support unbounded above has a positive tail index γ if its tail function $\bar{F} = 1 - F$ is regularly varying with index $-1/\gamma$. A probabilistic interpretation is that the conditional distribution of the relative excess over a high threshold u is approximately Pareto distributed with parameter $1/\gamma$, that is, $\Pr[X/u \leq x | X > u] \rightarrow 1 - x^{-1/\gamma}$ as $u \rightarrow \infty$ for $1 \leq x < \infty$.

Let $X_{1:n} \leq \dots \leq X_{n:n}$ be the ordered values of a random sample from a distribution function F with positive tail index γ . For simplicity, assume $F(0) = 0$. Choosing the threshold u as the $(k + 1)$ th largest order statistic, $X_{n-k:n}$, and modelling the relative excesses $X_{n-k+i:n}/X_{n-k:n}$, $i = 1, \dots, k$, by the Pareto distributions yields the following pseudo-log-likelihood for γ :

$$\ell_n(\gamma, k) = -k(\gamma^{-1} \hat{H}_n(k) + \log \gamma). \tag{2.1}$$

This pseudo-log-likelihood is maximal for γ equal to the Hill estimator, $\hat{H}_n(k)$. The score function is $\dot{\ell}_n(\gamma, k) = \gamma^{-1} k \{\gamma^{-1} \hat{H}_n(k) - 1\}$, while the deviance statistic is

$$D_n(\gamma, k) = 2\{\ell_n(\hat{H}_n(k), k) - \ell_n(\gamma, k)\} = 2k \left(\frac{\hat{H}_n(k)}{\gamma} - 1 - \log \frac{\hat{H}_n(k)}{\gamma} \right).$$

The Fisher information in the Pareto model is $I(\gamma) = \gamma^{-2}$.

Standard theory on parametric inference now yields a number of confidence intervals for γ . Denote the p th tail quantile of the standard normal distribution by z_p , so $\Phi(z_p) = 1 - p$, with Φ the standard normal distribution function. Let α be the nominal type I error of the confidence interval, that is, the probability of covering the true value is equal to $1 - \alpha$ in the limit. The *Wald* confidence interval

$$I_n^{(1)}(\alpha, k) = \left[\left(1 - k^{-1/2} z_{\alpha/2}\right) \hat{H}_n(k), \left(1 + k^{-1/2} z_{\alpha/2}\right) \hat{H}_n(k) \right]$$

is based on the limiting normal distribution of the maximum likelihood estimator, $\hat{H}_n(k)$. The *score* confidence interval

$$I_n^{(2)}(\alpha, k) = \left[\left(1 + k^{-1/2} z_{\alpha/2}\right)^{-1} \hat{H}_n(k), \left(1 - k^{-1/2} z_{\alpha/2}\right)^{-1} \hat{H}_n(k) \right]$$

is based on the limiting normal distribution of the score statistic. The *likelihood ratio* (LR) confidence interval

$$I_n^{(3)}(\alpha, k) = \left\{ 0 < \gamma < \infty : D_n(\gamma, k) \leq z_{\alpha/2}^2 \right\}$$

is based on the limiting chi-squared distribution of the deviance statistic. Finally, the *Bartlett-corrected LR* confidence interval

$$I_n^{(4)}(\alpha, k) = \left\{ 0 < \gamma < \infty : D_n(\gamma, k) / (1 + (6k)^{-1}) \leq z_{\alpha/2}^2 \right\}$$

is the same as the ordinary LR confidence interval but with the deviance statistic divided by its asymptotic mean. Note that the Wald interval is symmetric around the Hill estimator, while the others are not.

Our aim is to analyse the performance of the above confidence intervals. Note that, as the Hill estimator is a sufficient statistic for the pseudo-log-likelihood (2.1), the four confidence intervals considered above depend on the Hill estimator only. Denoting the standardized Hill estimator by $H_n(\gamma, k) = k^{1/2} \{\gamma^{-1} \hat{H}_n(k) - 1\}$, we can write the four types of confidence intervals as

$$I_n^{(i)}(\alpha, k) = \{0 < \gamma < \infty : q_{ki}(-z_{\alpha/2}) \leq H_n(\gamma, k) \leq q_{ki}(z_{\alpha/2})\} \tag{2.2}$$

with, for all real z , and the functions $a_{ij}(z)$ as in Table 1,

$$q_{ki}(z) = z + \sum_{j=1}^3 a_{ij}(z)k^{-j/2} + O(k^{-2}), \quad \text{as } k \rightarrow \infty. \tag{2.3}$$

The coverage probabilities $\Pr[\gamma \in I_n^{(i)}(\alpha, k)]$ of the four confidence intervals can thus be expressed in terms of the distribution function of the standardized Hill estimator. Edgeworth expansions for this distribution function combined with the expansion in (2.3) then lead to asymptotic expansions for these coverage probabilities. This is the programme for the next two sections.

Remark 2.1. From (2.2), it is clear how to define the one-sided analogues of the Wald, score, LR and Bartlett-corrected LR confidence intervals. To analyse the performance of such intervals, the one-term Edgeworth expansion in Cheng and Peng (2001: Proposition 2) is sufficiently accurate. In view of our exposition it is straightforward to extend that paper’s analysis of the one-sided score confidence interval to the other one-sided intervals.

3. Negligible bias

In this section we derive expansions for the coverage probabilities of the confidence intervals in the previous section for the case where the bias of the Hill estimator is so small that it does not appear in the main correction term in the expansion. Throughout, we make the following standing assumption.

Assumption 1. The distribution function F is supported on the positive half-line and has positive tail index γ . The random variables X_1, \dots, X_n are independent and have common distribution function F .

Table 1. The functions $a_{ij}(z)$ appearing in (2.3)

Confidence interval		$j = 1$	$j = 2$	$j = 3$
Wald	$i = 1$	z^2	z^3	z^4
Score	$i = 2$	0	0	0
LR	$i = 3$	$\frac{1}{3}z^2$	$\frac{1}{36}z^3$	$-\frac{1}{270}z^4$
Bartlett LR	$i = 4$	$\frac{1}{3}z^2$	$\frac{1}{36}z^3 + \frac{1}{12}z$	$-\frac{1}{270}z^4 + \frac{1}{18}z^2$

The tail quantile function, V , of a distribution function F is defined as $V(y) = \inf\{x : F(x) \geq 1 - 1/y\}$ for $1 < y < \infty$. The assumption that F has a positive tail index γ is equivalent to $V \in \mathcal{R}_\gamma$. In order to study the asymptotics of the Hill estimator, we need to quantify the speed of convergence in the limit relation embedded in the definition of regular variation of \bar{F} or V (Bingham *et al.* 1987: Chapter 3). This is the aim of the following assumption. For real τ and positive y put

$$h_\tau(y) = \int_1^y u^{\tau-1} du = \begin{cases} (y^\tau - 1)/\tau, & \text{if } \tau \neq 0, \\ \log y, & \text{if } \tau = 0. \end{cases}$$

Assumption 2. *There exist real constants $\rho \leq 0$ and $c \neq 0$ as well as a function $a \in \mathcal{R}_\rho$ vanishing at infinity such that the tail quantile function V of F satisfies*

$$\lim_{t \rightarrow \infty} \frac{1}{a(t)} \left(\frac{V(ty)}{V(t)} - y^\gamma \right) = cy^\gamma h_\rho(y), \quad \text{for all } y > 0. \tag{3.1}$$

For the Hill estimator to be consistent, the number of relative excesses, k , used in its definition should tend to infinity along with the sample size. On the other hand, for the Hill estimator to be asymptotically normal, the threshold, to be chosen as the $(k + 1)$ th largest order statistic, should also tend to infinity and should do so fast enough to validate the Pareto approximation to the distribution of relative excesses. To balance these requirements is the aim of the following assumption.

Assumption 3. *The positive integer sequence k_n is such that $k_n \rightarrow \infty$, $k_n/n \rightarrow 0$, and $\lambda_n = k_n^{1/2} a(n/k_n) \rightarrow 0$ as $n \rightarrow \infty$.*

Under Assumptions 1–3, the standardized Hill estimator

$$H_n = H_n(\gamma, k_n) = k_n^{1/2} \{ \gamma^{-1} \hat{H}_n(k_n) - 1 \}$$

is asymptotically standard normal; see, for instance, de Haan and Peng (1998: Theorem 1). Under a side condition on the behaviour of F near zero, the expectation of H_n is asymptotically equivalent to

$$\mu_n = c\lambda_n / \{ \gamma(1 - \rho) \}$$

(see Segers 2001b), so it is not surprising that this μ_n will show up in the expansions to come. See Remark 3.4 for a discussion of the case where λ_n is allowed to converge to some arbitrary real number.

Approximations of the distribution of H_n typically feature standardized sums of independent standard exponential random variables, and indeed our first result features the classical Edgeworth expansion for such sums. Let $(E_i)_{i \geq 1}$ be a sequence of independent random variables, exponentially distributed with mean one. There exist polynomials P_j indexed by a positive integer j such that, for every positive integer m , we have uniformly in $x \in \mathbb{R}$,

$$\Pr \left[\frac{1}{k^{1/2}} \sum_{i=1}^k (E_i - 1) \leq x \right] = \Phi(x) + \sum_{j=1}^m P_j(x) \varphi(x) k^{-j/2} + O(k^{-(m+1)/2}), \quad (3.2)$$

as $k \rightarrow \infty$; see Petrov (1975: Theorem VI.4). The polynomials P_j are defined in terms of Hermite polynomials and the cumulants of the standard exponential distribution. In general, P_j is a polynomial of degree $3j - 1$ and P_j is even (odd) if j is odd (even). Some explicit expressions are

$$P_1(x) = -\frac{1}{3}(x^2 - 1) \quad \text{and} \quad P_2(x) = -\frac{1}{36}x(2x^4 - 11x^2 + 3). \quad (3.3)$$

Theorem 3.1. *Under Assumptions 1–3, we have, for every integer $m \geq 1$ and uniformly in $x \in \mathbb{R}$,*

$$\Pr[H_n \leq x] = \Phi(x) + \sum_{j=1}^m P_j(x) \varphi(x) k_n^{-j/2} + O(k_n^{-(m+1)/2}) - \mu_n \varphi(x) + o(|\mu_n|),$$

as $n \rightarrow \infty$, where the polynomials P_j are the ones appearing in (3.2).

Combine the expansion in Theorem 3.1 for $m = 3$ with equations (2.2) and (2.3) to derive expansions for the coverage probabilities of the considered confidence intervals. Note that we do not need an explicit expression for P_3 : since P_3 is even, the corresponding correction terms cancel out.

Corollary 3.2. *Under Assumptions 1–3, the coverage probabilities of the Wald, score, LR and Bartlett-corrected LR confidence intervals at nominal coverage probability $1 - \alpha$ admit the expansion*

$$\Pr[\gamma \in I_n^{(i)}(\alpha, k_n)] = 1 - \alpha + z Q_i(z) \varphi(z) k_n^{-1} + O(k_n^{-2}) + o(|\mu_n|) \quad (3.4)$$

as $n \rightarrow \infty$, where $z = z_{\alpha/2}$ and

$$Q_i(z) = \begin{cases} -\frac{1}{18}(8z^4 - 11z^2 + 3), & \text{if } i = 1 \text{ (Wald),} \\ -\frac{1}{18}(2z^4 - 11z^2 + 3), & \text{if } i = 2 \text{ (score),} \\ -\frac{1}{6}, & \text{if } i = 3 \text{ (LR),} \\ 0, & \text{if } i = 4 \text{ (Bartlett LR).} \end{cases} \quad (3.5)$$

Example 3.1. The asymptotic normality of the Hill estimator was studied in Hall (1982) for distribution functions F with the property that there exist constants $\gamma > 0$, $\rho < 0$, $A > 0$ and $B \neq 0$ such that

$$\bar{F}(x) = Ax^{-1/\gamma} \left(1 + Bx^{\rho/\gamma} + o(x^{\rho/\gamma}) \right), \quad \text{as } x \rightarrow \infty. \quad (3.6)$$

For these distribution functions, Assumption 2 is satisfied with the same ρ and with $a(t) = t^\rho$ and $c = \gamma \rho A^\rho B$. An example is the distribution function

$$F(x) = (1 - x^{-1/\gamma})^\delta, \quad \text{for all } x \geq 1, \quad (3.7)$$

with parameters $\gamma > 0$ and $\delta > 0$. If $\delta \neq 1$, then (3.6) holds true with $A = \delta$, $B = (1 - \delta)/2$ and $\rho = -1$; if $\delta = 1$, then F is the Pareto distribution.

We generated 100 000 pseudo-random samples of size 500 from this distribution and compared the coverage probability expansions in Corollary 3.2 (ignoring the remainder terms, of course) to the simulated coverage probabilities. Figure 1 shows the results for nominal type I error $\alpha = 0.1$ and parameter vectors $(\gamma, \delta) = (0.5, 1)$ (left) and $(\gamma, \rho) = (0.5, 2)$ (right). The simulated rejection probabilities are indicated by circles along the curves.

If $\delta = 1$, then the version of (3.4) without the $o(|\mu_n|)$ term holds true for every sequence k_n tending to infinity. Indeed, in the left-hand panel of Figure 1, the predicted rejection probabilities $\alpha - zQ_i(z)\varphi(z)k^{-1}$ are close to the simulated ones for all k . For $\delta \neq 1$, the $o(|\mu_n|)$ term ruins the expansion for larger k . In all cases, the LR confidence intervals are only slightly less accurate than their Bartlett-corrected versions. The score and in particular the Wald confidence intervals are much less reliable.

Remark 3.1. Using the expansion in Theorem 3.1 for $m = 2p + 1$, the coverage probability expansion (3.4) can be extended to include correction term of order $O(k_n^{-j})$ for $j = 1, \dots, p$ and a remainder term of order $O(k_n^{-p-1})$. However, these higher-order terms are likely to be blurred by the $o(|\mu_n|)$ remainder term, so that such an expansion is statistically not very relevant. A better idea is to try to make the $o(|\mu_n|)$ term explicit, as we will do in Corollary 4.2.

Remark 3.2. Under the assumptions of Corollary 3.2, the Wald and score confidence intervals can also be corrected to make the $O(k_n^{-1})$ term in (3.4) vanish, for example by changing the nominal level α to $\alpha_{k,i} = \alpha + zQ_i(z)\varphi(z)k^{-1}$ with $z = z_{\alpha/2}$, for $i = 1, 2$. However, the finite-sample properties of these confidence intervals are not as good as those of the Bartlett-

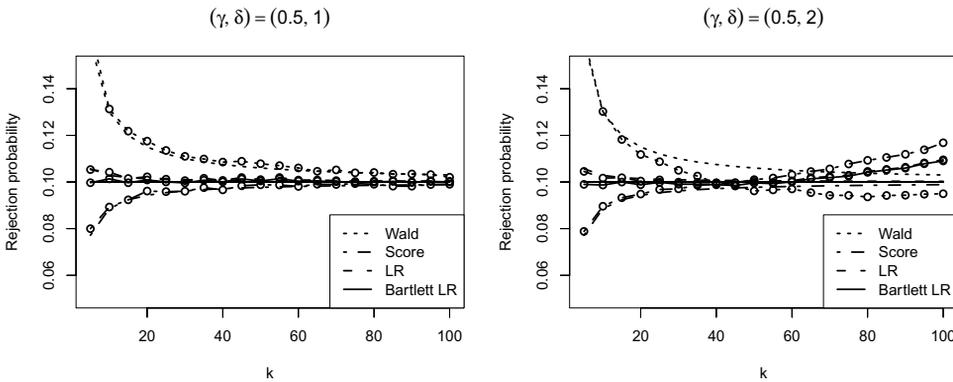


Figure 1. Simulated and predicted [by (3.4)] rejection probabilities of Wald, score, LR and Bartlett-corrected LR confidence intervals for the tail index at nominal rejection probability 0.1. Based on 100 000 samples of size 500 of the distribution function in (3.7) with $\gamma = 0.5$ and $\delta = 1$ (left) and $\delta = 2$ (right).

corrected LR confidence intervals, since the corrections to be made for the Wald and score intervals are much larger than for the LR interval.

Remark 3.3. The special case $m = 1$ of Theorem 3.1 has been proven in Cheng and Pan (1998: Theorem 1) under the assumption that $k_n a(n/k_n)$ converges to some non-negative constant, leading to a one-term Edgeworth expansion with an $O(k_n^{-1/2})$ correction term and an $o(k_n^{-1/2})$ remainder term; see also Cheng and Peng (2001: Proposition 2). The expansions in Cheng and de Haan (2001: Theorem 1) and Guillou and Hall (2001: Theorem 1) involve versions of gamma distributions depending on k_n instead of the limiting normal distribution. These approximations are stated under extra growth conditions on k_n and in Guillou and Hall (2001) for a submodel of Assumption 2.

Remark 3.4. If the limit of λ_n in Assumption 3 is allowed to be any real λ , then the standardized Hill estimator is asymptotically normal with mean $\mu = c\lambda/\{\gamma(1 - \rho)\}$ and variance one; see, for instance, de Haan and Peng (1998: Theorem 1). If λ and thus μ are different from zero, then confidence intervals for γ based on the postulated asymptotic standard normality of H_n are inconsistent in the sense that nominal and asymptotic coverage probabilities do not match. If $\rho < 0$, this situation arises if k_n is chosen to minimize the asymptotic mean squared error of the Hill estimator; see de Haan and Peng (1998: Theorem 2). For such k_n , bias-corrected confidence intervals are constructed in Ferreira and de Vries (2004). If $\rho = 0$, then the asymptotic mean squared error of the Hill estimator is minimized for sequences k_n such that λ_n tends to infinity at a certain rate, and for such k_n the asymptotic distribution of the Hill estimator is actually the same as that of a large class of estimators; see Drees (1998).

4. Non-negligible bias

In Corollary 3.2, if $\mu_n = O(k_n^{-1})$, then the $O(k_n^{-1})$ term on the right-hand side of (3.4) is indeed the main correction term in the expansion. However, if k_n is so large that μ_n is of larger order than k_n^{-1} , then the expansion, although correct, is not very informative as it does not say anything about the $o(|\mu_n|)$ remainder term. In order to derive a more detailed expansion, we need to refine Assumption 2. Note that (3.1) is equivalent to

$$\lim_{t \rightarrow \infty} \frac{\log V(ty) - \log V(t) - \gamma \log y}{a(t)} = ch_\rho(y), \quad \text{for all } y > 0. \tag{4.1}$$

The appropriate refinement corresponding to (4.1) is suggested by the theory of second-order regular variation; see de Haan and Stadtmüller (1996).

Assumption 4. *There exist real constants $\rho \leq 0$, $\tau \leq 0$, and $c \neq 0$ as well as functions $a \in \mathcal{R}_\rho$ and $b \in \mathcal{R}_\tau$ vanishing at infinity such that the limit*

$$B(y) := \lim_{t \rightarrow \infty} \frac{1}{b(t)} \left(\frac{\log V(ty) - \log V(t) - \gamma \log y}{a(t)} - ch_\rho(y) \right) \tag{4.2}$$

exists for all $y > 0$.

From the proof of Theorem 1 in de Haan and Stadtmüller (1996) applied to $f(t) = \log\{t^{-\gamma}V(t)\}$, it is immediate that the limit function B in (4.2) must be of the form

$$B(y) = \begin{cases} c_1(\log y)^2 + c_2 \log y, & \text{if } \rho = \tau = 0, \\ c_1 y^\rho \log y + c_2 h_\rho(y), & \text{if } \rho < 0 = \tau, \\ c_1 h_{\rho+\tau}(y) + c_2 h_\rho(y), & \text{if } \tau < 0, \end{cases}$$

for some $c_1, c_2 \in \mathbb{R}$; see also equation (2.9) in de Haan and Stadtmüller (1996). Moreover, necessarily

$$\lim_{t \rightarrow \infty} \frac{1}{b(t)} \left(\frac{a(ty)}{a(t)} - y^\rho \right) = dy^\rho h_\tau(y), \quad \text{for all } y > 0, \tag{4.3}$$

for some real constant d determined by ρ, τ, c, c_1 and c_2 . Set $B_0 = \int_0^1 B(1/u)du$ and recall the polynomials P_1 and P_2 in (3.3).

Theorem 4.1. *Under Assumptions 1, 3 and 4, we have, as $n \rightarrow \infty$ and uniformly in $x \in \mathbb{R}$,*

$$\begin{aligned} \Pr[H_n \leq x] &= \Phi(x) + P_1(x)\varphi(x)k_n^{-1/2} + P_2(x)\varphi(x)k_n^{-1} + o(k_n^{-1}) \\ &\quad - \varphi(x)\mu_n - \frac{1}{2}x\varphi(x)\mu_n^2 + o(\mu_n^2) \\ &\quad - x\left(\frac{1}{3}x^2 + \frac{\rho}{1-\rho}\right)\varphi(x)k_n^{-1/2}\mu_n \\ &\quad - c^{-1}(1-\rho)B_0\varphi(x)\mu_n b\left(\frac{n}{k_n}\right) + o\left\{|\mu_n|b\left(\frac{n}{k_n}\right)\right\}. \end{aligned}$$

Combine Theorem 4.1 with equations (2.2) and (2.3) to obtain the following coverage probability expansions for the confidence intervals considered in Section 2.

Corollary 4.2. *Under Assumptions 1, 3 and 4, the coverage probabilities of the Wald, score, LR and Bartlett-corrected LR confidence intervals at nominal coverage probability $1 - \alpha$ admit the expansions*

$$\begin{aligned} \Pr[\gamma \in I_n^{(i)}(\alpha, k_n)] &= 1 - \alpha + z\varphi(z) \left\{ Q_i(z)k_n^{-1} + \left(a_i z^2 - \frac{2\rho}{1-\rho} \right) k_n^{-1/2} \mu_n - \mu_n^2 \right\} \\ &\quad + o(k_n^{-1}) + o(\mu_n^2) + o\left\{|\mu_n|b\left(\frac{n}{k_n}\right)\right\} \end{aligned} \tag{4.4}$$

as $n \rightarrow \infty$, with $z = z_{\alpha/2}$, $Q_i(z)$ as in (3.5) and

$$a_i = \begin{cases} 4/3, & \text{if } i = 1 \text{ (Wald)}, \\ -2/3, & \text{if } i = 2 \text{ (score)}, \\ 0, & \text{if } i = 3, 4 \text{ (LR and Bartlett LR)}. \end{cases}$$

Example 4.1. A distribution function F , the tail function of which admits the expansion, as $x \rightarrow \infty$,

$$\bar{F}(x) = Ax^{-1/\gamma} \left(1 + Bx^{\rho/\gamma} + Cx^{(\rho+\beta)/\gamma} + o(x^{(\rho+\beta)/\gamma}) \right), \tag{4.5}$$

with some C and $\beta < 0$ and the other constants as in Example 3.1, satisfies Assumption 4 with $\tau = \max(\rho, \beta)$, $b(t) = t^\tau$ and $B(y) = dh_{\rho+\tau}(y)$, where

$$d = \begin{cases} \gamma\rho(2\rho - 1)A^{2\rho}B^2, & \text{if } \beta < \rho < 0, \\ \gamma\rho A^{2\rho} \{ (2\rho - 1)B^2 + 2C \}, & \text{if } \beta = \rho < 0, \\ \gamma(\rho + \beta)A^{\rho+\beta}C, & \text{if } \rho < \beta < 0. \end{cases}$$

Expansion (4.5) is valid for, among others, the Fréchet, Burr, F and Student t distributions as well as the distribution in (3.7).

We compared the coverage probability expansions in Corollary 4.2 with Monte Carlo approximations based on 100 000 samples of size 500 of the distribution in (3.7). Figure 2 shows the results for nominal rejection probability $\alpha = 0.1$ and parameters $\gamma = 0.5$ and $\delta = 0.5$ (top left) and $\delta = 2$ (top right). As $c = \gamma(\delta - 1)/(2\delta)$, the sign of c and hence of μ_n is positive or negative according to whether δ is larger or smaller than one. This sign determines the way in which the two components of the correction term, the classical Edgeworth expansion for standardized gamma distributions and the bias term, interact (see also Remark 4.2 below). In particular, the two components may reinforce or neutralize each other, a phenomenon which is clearly visible for the Wald confidence intervals.

We also ran simulations for the Burr distribution, $F(x) = 1 - (x^{-\rho/\gamma} + 1)^{1/\rho}$, for $x \geq 0$, which satisfies (4.5) with $A = 1$, $B = 1/\rho$, $C = (1 - \rho)/(2\rho^2)$, and $\tau = \rho$. The sign of $c = \gamma$ and thus of the bias term μ_n is always positive. Indeed, the two bottom plots in Figure 2 have the same qualitative features as the top right-hand plot in the same figure (see the previous paragraph). The value of ρ determines the speed at which the bias term tends to zero, with ρ closer to zero implying a larger bias. This is clearly visible from the difference in the range of k -values with reasonable simulated rejection probabilities in the plots on the left ($\rho = -1$) and the right ($\rho = -0.5$).

Example 4.2. Let X be a random variable so that $\log X$ has a gamma distribution with shape parameter δ and scale parameter γ , that is, the probability density function of X is given by $f(x) = \gamma^{-\delta} \Gamma(\delta)^{-1} (\log x)^{\delta-1} x^{-1/\gamma-1}$, for $x > 1$. If $\delta \neq 1$, then Assumption 4 is satisfied with $\rho = \tau = 0$, $c = \gamma(\delta - 1)$, and $a(t) \sim (\log t)^{-1}$ as $t \rightarrow \infty$. As the rate function a disappears only at a logarithmic rate, astronomical sample sizes are needed for the asymptotics in the coverage probability expansions to become visible.

Remark 4.1. Unlike the expansion in Corollary 3.2, the expansion in Corollary 4.2 cannot be used directly to improve the accuracy of the confidence intervals as in Remark 3.2, because

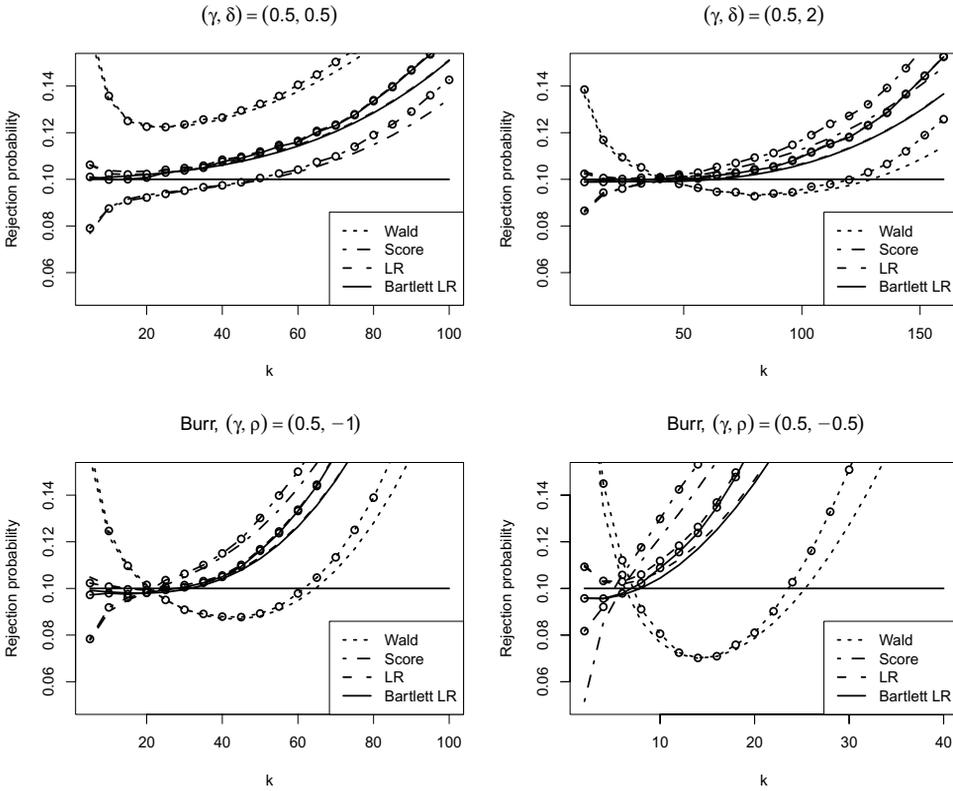


Figure 2. Simulated and predicted [by (4.4)] rejection probabilities of Wald, score, LR and Bartlett-corrected LR confidence intervals for the tail index at nominal rejection probability 0.1, based on 100 000 samples of size 500. Top: distribution function in (3.7) with $\gamma = 0.5$ and $\delta = 0.5$ (left) and $\delta = 2$ (right). Bottom: Burr distribution with $\gamma = 0.5$ and $\rho = -1$ (left) and $\rho = -0.5$ (right).

this time the correction term involves the unknown quantities ρ and μ_n . Of course, one could estimate these second-order quantities and use them to estimate the correction term in (4.4). However, given such estimates, a better idea is to compute the Hill estimator at the value for k_n that minimizes the asymptotic mean squared error and then to subtract the estimated bias; see, for instance, Gomes and Martins (2002) and Ferreira and de Vries (2004).

Remark 4.2. Since the term in curly brackets on the right-hand side of (4.4) is equal to k_n^{-1} times a quadratic polynomial in $k_n a(n/k_n)$, there may be zero, one, or two values for k_n for which it vanishes. A possible threshold selection method then might be to try to locate such k_n , provided it exists. But as in Remark 4.1, this would require estimates of the second-order parameters. For one-sided score confidence intervals, this programme was carried out in Cheng and Peng (2001).

Remark 4.3. In Ferreira (2002: Appendix 4.B), a one-term expansion for the distribution function of H_n is derived in the case where the intermediate sequence k_n is the one for which the asymptotic mean squared error of the Hill estimator is minimal (see Remark 3.4) and for $\tau < \rho < 0$, forcing $k_n^{-1/2} = o\{b(n/k_n)\}$. The expansion of $\Pr[H_n \leq x]$ takes the form $\Phi(x - \mu_n)$ plus a correction term of the order $b(n/k_n)$.

Remark 4.4. Theorem 4.1 can be extended to expansions of arbitrary order and for intermediate sequences k_n such that λ_n remains bounded but does not necessarily converge to zero; see Cuntz *et al.* (2003: Theorem 2). However, the statement and proof of this result are rather intricate. We believe that the statistically relevant expansions, at least for asymptotic bias zero, are already covered by Theorems 3.1 and 4.1.

5. Proof of Theorem 3.1

Let Y_1, \dots, Y_n be independent, standard Pareto distributed random variables, that is, $\Pr[Y_i \leq t] = 1 - 1/t$ for $t \geq 1$. The corresponding order statistics are $Y_{1:n} \leq \dots \leq Y_{n:n}$. By the probability integral transform, the vectors $\{X_{i:n}\}_{i=1}^n$ and $\{V(Y_{i:n})\}_{i=1}^n$ have the same joint distribution. Writing $R(y, t) = \log V(ty) - \log V(t) - \gamma \log(y)$, for $y \geq 1, t \geq 1$, we arrive after some algebra at the distributional representation

$$k^{1/2} \frac{\hat{H}_n(k) - \gamma}{\gamma} \stackrel{d}{=} \frac{1}{k^{1/2}} \sum_{i=1}^k \left\{ \log \left(\frac{Y_{n-k+i:n}}{Y_{n-k:n}} \right) - 1 \right\} + \frac{1}{k^{1/2}} \sum_{i=1}^k \gamma^{-1} R \left(\frac{Y_{n-k+i:n}}{Y_{n-k:n}}, Y_{n-k:n} \right). \tag{5.1}$$

By the Markov property of order statistics, the joint distribution of $\{Y_{n-k+i:n}\}_{i=1}^k$ conditionally on $Y_{n-k:n} = t$ is the same as the joint distribution of $\{tY_{i:k}\}_{i=1}^k$. Hence, from (5.1),

$$\Pr[H_n \leq x] = E[f_{k_n}(Y_{n-k_n:n})], \tag{5.2}$$

where $f_k(t) = \Pr[Z_k + R_k(t) \leq x]$, $Z_k = k^{-1/2} \sum_{i=1}^k \{\log(Y_i) - 1\}$ and $R_k(t) = k^{-1/2} \times \sum_{i=1}^k \gamma^{-1} R(Y_i, t)$.

Since $(k_n)_n$ is an intermediate sequence, the order statistic $Y_{n-k_n:n}$ is with large probability contained in the interval

$$I_n = \left[nk_n^{-1} \left(1 - k_n^{-1/2} \log k_n \right), nk_n^{-1} \left(1 + k_n^{-1/2} \log k_n \right) \right].$$

More precisely, exponential bounds on tail probabilities of binomial random variables (inequalities (4) and (6) in Shorack and Wellner 1986: 440) imply that for every p ,

$$\Pr[Y_{n-k_n:n} \notin I_n] = O(k_n^{-p}), \quad \text{as } n \rightarrow \infty. \tag{5.3}$$

Since $|f_k(t)| \leq 1$ for every positive integer k and all $t \geq 1$, the bound in (5.3) combined with (5.2) implies, for every p , uniformly in $x \in \mathbb{R}$,

$$\Pr[H_n \leq x] = E[f_{k_n}(Y_{n-k_n:n}) | Y_{n-k_n:n} \in I_n] + O(k_n^{-p}), \quad \text{as } n \rightarrow \infty. \quad (5.4)$$

The idea of the proof now is to show that the random variables $R_{k_n}(t)$, $t \in I_n$, are sufficiently close to μ_n . More precisely, suppose that we can find a sequence $(c_n)_n$ of positive numbers such that, for every p ,

$$c_n = o\left\{k_n^{1/2} a\left(\frac{n}{k_n}\right)\right\} \quad \text{and} \quad \sup_{t \in I_n} \Pr[|R_{k_n}(t) - \mu_n| > c_n] = O(k_n^{-p}), \quad (5.5)$$

as $n \rightarrow \infty$. Then also

$$\begin{aligned} & \sup_{t \in I_n} |\Pr[Z_{k_n} \leq x - R_{k_n}(t)] - \Pr[Z_{k_n} \leq x - \mu_n]| \\ & \leq \Pr[|Z_{k_n} - (x - \mu_n)| \leq c_n] + O(k_n^{-p}), \quad \text{as } n \rightarrow \infty. \end{aligned} \quad (5.6)$$

By the Edgeworth expansion (3.2) for the standard exponential random variables $\log(Y_i)$, $i = 1, \dots, k_n$, we have, as $n \rightarrow \infty$,

$$\begin{aligned} \Pr[|Z_{k_n} - (x - \mu_n)| \leq c_n] &= \Phi(x - \mu_n + c_n) - \Phi(x - \mu_n - c_n) \\ &+ \sum_{j=1}^m \{P_j \varphi(x - \mu_n + c_n) - P_j \varphi(x - \mu_n - c_n)\} k_n^{-j/2} + O(k_n^{-(m+1)/2}). \end{aligned}$$

Since the derivatives of Φ and $P_j \varphi$ are uniformly bounded over \mathbb{R} , we obtain

$$\Pr[|Z_{k_n} - (x - \mu_n)| \leq c_n] = o\left\{k_n^{1/2} a\left(\frac{n}{k_n}\right)\right\} + O(k_n^{-(m+1)/2}), \quad (5.7)$$

uniformly in $x \in \mathbb{R}$, as $n \rightarrow \infty$. Combine (5.7) with (5.4) and the bound in (5.6) to obtain, uniformly in $x \in \mathbb{R}$, as $n \rightarrow \infty$,

$$\Pr[H_n \leq x] = \Pr[Z_{k_n} \leq x - \mu_n] + O(k_n^{-(m+1)/2}) + o\left\{k_n^{1/2} a\left(\frac{n}{k_n}\right)\right\}.$$

Applying (3.2) again gives, uniformly in $x \in \mathbb{R}$ and as $n \rightarrow \infty$,

$$\begin{aligned} \Pr[Z_{k_n} \leq x - \mu_n] &= \Phi(x - \mu_n) + \sum_{j=1}^m P_j \varphi(x - \mu_n) k_n^{-j/2} + O(k_n^{-(m+1)/2}) \\ &= \Phi(x) - \mu_n \varphi(x) + \sum_{j=1}^m P_j(x) \varphi(x) k_n^{-j/2} + O(k_n^{-(m+1)/2}) + o(|\mu_n|), \end{aligned}$$

once more by the uniform boundedness of the derivatives of the functions $P_j \varphi$. Combining the last two displays then yields the desired conclusion.

Hence it remains to show that we can find a positive sequence $(c_n)_n$ satisfying (5.5). We claim that the sequence

$$c_n = 2 \max\left\{k_n^{1/4} a\left(\frac{n}{k_n}\right), \sup_{t \in I_n} |E[R_{k_n}(t)] - \mu_n|\right\}$$

meets the requirements. First of all, since $E[h_\rho(Y_1)] = (1 - \rho)^{-1}$,

$$\begin{aligned} \mathbb{E}[R_{k_n}(t)] - \mu_n &= \gamma^{-1} k_n^{1/2} \mathbb{E}[R(Y_1, t)] - \mu_n \\ &= \gamma^{-1} k_n^{1/2} a\left(\frac{n}{k_n}\right) \left(\frac{a(t)}{a(n/k_n)} \mathbb{E}\left[\frac{R(Y_1, t)}{a(t)}\right] - c\mathbb{E}[h_\rho(Y_1)] \right). \end{aligned}$$

The uniform convergence theorem for regularly varying functions (Bingham *et al.* 1987: Theorem 1.5.2) implies

$$\sup_{t \in I_n} \left| \frac{a(t)}{a(n/k_n)} - 1 \right| \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

Moreover, for every $\varepsilon > 0$ we can find $C_\varepsilon > 0$ and $t_\varepsilon \geq 1$ such that

$$\left| \frac{R(y, t)}{a(t)} \right| \leq C_\varepsilon y^\varepsilon, \quad \text{for all } y \geq 1, t \geq t_\varepsilon, \tag{5.8}$$

(see Bingham *et al.* 1987: Theorem 3.1.3). Since $R(y, t)/a(t) \rightarrow ch_\rho(y)$ as $t \rightarrow \infty$, by the dominated convergence theorem $\mathbb{E}[R(Y_1, t)/a(t)] \rightarrow c\mathbb{E}[h_\rho(Y_1)]$ as $t \rightarrow \infty$. Hence, $c_n = o\{k_n^{1/2} a(n/k_n)\}$, which is the first part of (5.5).

To prove the second part of (5.5), observe that by definition of c_n ,

$$\Pr[|R_{k_n}(t) - \mu_n| > c_n] \leq \Pr\left[|R_{k_n}(t) - \mathbb{E}[R_{k_n}(t)]| > k_n^{1/4} a\left(\frac{n}{k_n}\right)\right].$$

Fix an arbitrary $p \geq 2$. Choose $0 < \varepsilon < 1/p$ and let t_ε and C_ε be as in (5.8). For n large enough, we have $I_n \subset [t_\varepsilon, \infty)$ and $\sup_{t \in I_n} a(t)/a(n/k_n) \leq 2$. Applying Lemma A.1 in the Appendix yields for such large n a constant c_p depending only on p such that, for all $t \in I_n$,

$$\begin{aligned} &\Pr\left[|R_{k_n}(t) - \mathbb{E}[R_{k_n}(t)]| > k_n^{1/4} a\left(\frac{n}{k_n}\right)\right] \\ &\leq c_p \left(\frac{2}{a(t)\gamma}\right)^p \mathbb{E}[|R(Y_1, t)|^p] k_n^{-p/4} \leq c_p \left(\frac{2}{\gamma}\right)^p \frac{C_\varepsilon^p}{1 - \varepsilon p} k_n^{-p/4}. \end{aligned}$$

Since p can be chosen arbitrarily large, the last display now implies the second part of (5.5), as required. This completes the proof of Theorem 3.1.

6. Proof of Theorem 4.1

The proof of Theorem 4.1 starts in the same way as the proof of Theorem 3.1 in Section 5 up to and including equation (5.4). Define

$$B(y, t) = \frac{1}{b(t)} \left(\frac{R(y, t)}{a(t)} - ch_\rho(y) \right), \quad \text{for all } t \geq 1, y \geq 1,$$

with $R(y, t)$ as in Section 5 and the other ingredients as in Assumption 4. Then we can decompose the term $R_k(t)$ as $R_k(t) = S_k(t) + T_k(t)$, where

$$S_k(t) = c\gamma^{-1}k^{-1/2}a(t)\sum_{i=1}^k h_\rho(Y_i), \tag{6.1}$$

$$T_k(t) = \gamma^{-1}k^{-1/2}a(t)b(t)\sum_{i=1}^k B(Y_i, t). \tag{6.2}$$

First, we treat the term $S_k(t)$ in (6.1). Note that

$$S_k(t) = c\gamma^{-1}k^{-1/2}a\left(\frac{n}{k}\right)\left(\frac{t}{n/k}\right)^\rho\sum_{i=1}^k h_\rho(Y_i) + S_{n,k}(t),$$

where

$$S_{n,k}(t) = c\gamma^{-1}k^{-1/2}\left\{a(t) - a\left(\frac{n}{k}\right)\left(\frac{t}{n/k}\right)^\rho\right\}\sum_{i=1}^k h_\rho(Y_i).$$

Recall that $\lambda_n = k_n^{1/2}a(n/k_n)$. Since the convergence in (4.3) is necessarily locally uniform in $0 < y < \infty$ (Bingham *et al.* 1987: Theorem 3.1.16) and since $0 \leq h_\rho(y) \leq \log(y)$ for all $y \geq 1$, we find that there exists a positive sequence Δ_n^S such that $\Delta_n^S = o\{\lambda_n b(n/k_n)\}$ and

$$\Pr\left[\sup_{t \in I_n} |S_{n,k_n}(t)| \geq \Delta_n^S\right] = o(k_n^{-1}), \quad \text{as } n \rightarrow \infty. \tag{6.3}$$

Secondly, we treat $T_k(t)$ in (6.2). Note that $B_0 = E[B(Y_1)]$. We have

$$T_k(t) = \gamma^{-1}k^{1/2}a\left(\frac{n}{k}\right)b\left(\frac{n}{k}\right)B_0 + \sum_{\ell=1}^3 T_{n,k}^{(\ell)}(t),$$

with

$$T_{n,k}^{(\ell)}(t) = \begin{cases} \gamma^{-1}k^{1/2}\left\{a(t)b(t) - a\left(\frac{n}{k}\right)b\left(\frac{n}{k}\right)\right\}B_0, & \text{if } \ell = 1, \\ \gamma^{-1}k^{1/2}a(t)b(t)\{E[B(Y_1, t)] - B_0\}, & \text{if } \ell = 2, \\ \gamma^{-1}k^{1/2}a(t)b(t)\left\{k^{-1}\sum_{i=1}^k B(Y_i, t) - E[B(Y_1, t)]\right\}, & \text{if } \ell = 3. \end{cases}$$

By the uniform convergence theorem for regularly varying functions and the Potter bound for $B(y, t)$ in Lemma A.2 below at $\varepsilon = 1/2$, we have

$$\sup_{t \in I_n} |T_{n,k_n}^{(\ell)}(t)| = o\left\{\lambda_n b\left(\frac{n}{k_n}\right)\right\}, \quad \text{as } n \rightarrow \infty,$$

for $\ell = 1, 2$. Further, by Lemma A.1 with $\delta = 1/4$ and $p = 5$, there exists a positive constant c_5 such that

$$\Pr\left[\left|\frac{1}{k}\sum_{i=1}^k B(Y_i, t) - E[B(Y_1, t)]\right| \geq k^{-1/4}\right] \leq c_5 E[|B(Y_1, t)|^5]k^{-5/4}$$

for all $t \geq 1$ and all integers $k > 0$. Apply Lemma A.2 with $\varepsilon = 1/6$ to deduce that

$$\Pr \left[\sup_{t \in I_n} |T_{n,k_n}^{(3)}(t)| > 2\gamma^{-1}k_n^{-1/4}\lambda_n b\left(\frac{n}{k_n}\right) \right] = o(k_n^{-1}), \quad \text{as } n \rightarrow \infty.$$

All in all, we find that there exists a positive sequence Δ_n^T such that $\Delta_n^T = o\{\lambda_n b(n/k_n)\}$ and, as $n \rightarrow \infty$,

$$\Pr \left[\sup_{t \in I_n} |T_{k_n}(t) - \gamma^{-1}\lambda_n b\left(\frac{n}{k_n}\right)B_0| \geq \Delta_n^T \right] = o(k_n^{-1}). \tag{6.4}$$

Combining (6.3) and (6.4) for $S_k(t)$ and $T_k(t)$, respectively, we find that

$$\Pr \left[\sup_{t \in I_n} |R_{k_n}(t) - \tilde{R}_{n,k_n}(t)| \geq \Delta_n \right] = o(k_n^{-1}), \quad \text{as } n \rightarrow \infty,$$

with

$$\tilde{R}_{n,k}(t) = c\gamma^{-1}k^{-1/2}a\left(\frac{n}{k}\right)\left(\frac{t}{n/k}\right)^\rho \sum_{i=1}^k h_\rho(Y_i) + \gamma^{-1}k^{1/2}a\left(\frac{n}{k}\right)b\left(\frac{n}{k}\right)B_0,$$

while $\Delta_n = \Delta_n^S + \Delta_n^T = o\{\lambda_n b(n/k_n)\}$ as $n \rightarrow \infty$. With Z_k as introduced in Section 5, we obtain, uniformly in $t \in I_n$ and $x \in \mathbb{R}$, and as $n \rightarrow \infty$,

$$\begin{aligned} \Pr [Z_{k_n} + \tilde{R}_{n,k_n}(t) \leq x - \Delta_n] &= o(k_n^{-1}) \\ &\leq \Pr [Z_{k_n} + R_{k_n}(t) \leq x] \leq \Pr [Z_{k_n} + \tilde{R}_{n,k_n}(t) \leq x + \Delta_n] + o(k_n^{-1}). \end{aligned} \tag{6.5}$$

We can write the random variable $Z_k + \tilde{R}_{n,k}(t)$ as

$$\frac{1}{k^{1/2}} \sum_{i=1}^k \xi_i\left(t, \frac{n}{k}\right) + \frac{c}{\gamma(1-\rho)} k^{1/2} a\left(\frac{n}{k}\right) \left(\frac{t}{n/k}\right)^\rho + \gamma^{-1} k^{1/2} a\left(\frac{n}{k}\right) b\left(\frac{n}{k}\right) B_0,$$

where

$$\xi_i(t, u) = \log(Y_i) - 1 + c\gamma^{-1}a(u)\left(\frac{t}{u}\right)^\rho \{h_\rho(Y_i) - (1-\rho)^{-1}\}, \quad \text{for } i = 1, \dots, k.$$

Note that $E[\xi_1(t, u)] = 0$.

The distribution function of the standardized sum of the random variables $\xi_i(t, u)$ can be expanded by a special case of Petrov (1975: Theorem VI.3.1), for the reader's convenience stated explicitly as Theorem A.3 below. In order to apply Theorem A.3, we need to compute some characteristics of the distribution of $\xi_1(t, u)$. The variance of $\xi_1(t, u)$ equals

$$\sigma^2(t, u) = 1 + 2c\gamma^{-1}(1-\rho)^{-2}a(u)\left(\frac{t}{u}\right)^\rho + O\{a(u)^2\}, \quad \text{as } u \rightarrow \infty, \tag{6.6}$$

uniformly in $t \in [u/2, 2u]$. Further, since the distribution of $\log(Y_1)$ is standard exponential, the cumulants $\kappa_m(t, u)$ of $\xi_1(t, u)$ satisfy

$$\kappa_m(t, u) = (m-1)! + O\{a(u)\}, \quad \text{as } u \rightarrow \infty,$$

uniformly in $t \in [u/2, 2u]$ for positive integer m . Also, $E[|\xi_1(t, u)|^p] \rightarrow E[|\log(Y_1) - 1|^p]$ for positive p as $u \rightarrow \infty$ and uniformly in $t \in [u/2, 2u]$. Finally, if u and t are such that

$\eta(u, t) = c\gamma^{-1}a(u)(t/u)^\rho$ is larger than -1 , then the probability density of $\xi_1(t, u)$ is uniformly bounded by $\max[1, \{1 + \eta(u, t)\}^{-1}]$. By an inequality due to Statulevičius (1965) and cited in Petrov (1975: 21–22, supplement I.5.22), this bound on the density of $\xi_1(u, t)$ implies that for $z \neq 0$ the characteristic function of $\xi_1(u, t)$ is bounded by

$$|E[e^{iz\xi_1(u,t)}]| \leq \exp\left\{-\frac{\min[1, \{1 + \eta(u, t)\}^2]}{96\{2\sigma(u, t) + \pi/|z|\}^2}\right\}.$$

The calculations in the previous paragraph served to demonstrate that we may apply Petrov's Theorem A.3 to derive that, as $n \rightarrow \infty$,

$$\Pr\left[\frac{1}{\sigma(t, n/k_n)k_n^{1/2}}\sum_{i=1}^{k_n}\xi_i\left(t, \frac{n}{k_n}\right) \leq x\right] = \Phi(x) + \sum_{j=1}^2(P_j\varphi)(x)k_n^{-j/2} + o(k_n^{-1})$$

uniformly in $t \in I_n$ and $x \in \mathbb{R}$, with P_1 and P_2 as in (3.3). Here we used the asymptotic relation $k_n^{-1/2}a(n/k_n) = k_n^{-1}\lambda_n = o(k_n^{-1})$ as $n \rightarrow \infty$ as well as the fact that $x \mapsto x^m\varphi(x)$ is uniformly bounded in x for positive m . Writing

$$v_n(x, t) = \sigma^{-1}\left(t, \frac{n}{k_n}\right)\left(x - \mu_n\left(\frac{t}{n/k_n}\right)^\rho - \gamma^{-1}\lambda_nb\left(\frac{n}{k_n}\right)B_0\right), \tag{6.7}$$

we obtain, as $n \rightarrow \infty$ and uniformly in $t \in I_n$ and $x \in \mathbb{R}$,

$$\Pr[Z_{k_n} + \tilde{R}_{n,k_n}(t) \leq x] = \Phi\{v_n(x, t)\} + \sum_{j=1}^2(P_j\varphi)\{v_n(x, t)\}k_n^{-j/2} + o(k_n^{-1}). \tag{6.8}$$

Combine (6.6) and (6.7) to find, as $n \rightarrow \infty$, uniformly in $t \in I_n$ and $x \in \mathbb{R}$,

$$\begin{aligned} v_n(x, t) = & x - \frac{1}{1-\rho}\left(\frac{t}{n/k_n}\right)^\rho xk_n^{-1/2}\mu_n - \mu_n\left(\frac{t}{n/k_n}\right)^\rho \\ & - \gamma^{-1}\lambda_nb\left(\frac{n}{k_n}\right)B_0 + o(\mu_n^2) + o\left\{\lambda_nb\left(\frac{n}{k_n}\right)\right\}. \end{aligned} \tag{6.9}$$

Combine (6.8) and (6.9) to see that $\Pr[Z_{k_n} + \tilde{R}_{n,k_n}(t) \leq x]$ is equal to

$$\begin{aligned} & \Phi(x) - \frac{1}{1-\rho}\left(\frac{t}{n/k_n}\right)^\rho x\varphi(x)k_n^{-1/2}\mu_n - \left(\frac{t}{n/k_n}\right)^\rho \varphi(x)\mu_n \\ & - \frac{1}{2}\left(\frac{t}{n/k_n}\right)^{2\rho} x\varphi(x)\mu_n^2 - \gamma^{-1}\varphi(x)\lambda_nb\left(\frac{n}{k_n}\right)B_0 + (P_1\varphi)(x)k_n^{-1/2} \\ & - \left(\frac{t}{n/k_n}\right)^\rho (P_1\varphi)'(x)k_n^{-1/2}\mu_n + (P_2\varphi)(x)k_n^{-1} \end{aligned} \tag{6.10}$$

plus a remainder of the form $o(k_n^{-1}) + o(\mu_n^2) + o\{\lambda_nb(n/k_n)\}$ as $n \rightarrow \infty$ and uniformly in $t \in I_n$ and $x \in \mathbb{R}$. Because of the inequalities for $\Pr[Z_{k_n} + R_{n,k_n}(t) \leq x]$ in (6.5) and the fact that the term Δ_n in (6.5) is of the order $o\{\lambda_nb(n/k_n)\}$ as $n \rightarrow \infty$, we must also have that

$\Pr[Z_{k_n} + R_{n,k_n}(t) \leq x]$ can be written as in (6.10) plus a remainder term that is again of the form $o(k_n^{-1}) + o(\mu_n^2) + o\{\lambda_n b(n/k_n)\}$ as $n \rightarrow \infty$ and uniformly in $t \in I_n$ and $x \in \mathbb{R}$.

In view of the representation of $\Pr[H_n \leq x]$ in equation (5.4), all that is left to do is to integrate out the variable t in (6.10) with respect to the conditional distribution of $Y_{n-k_n:n}$ given $Y_{n-k_n:n} \in I_n$. First of all, note that for any intermediate sequence k_n and any real p , as $n \rightarrow \infty$,

$$E\left[Y_{n-k_n:n}^p\right] = \frac{\Gamma(n+1)\Gamma(k_n+1-p)}{\Gamma(n+1-p)\Gamma(k_n+1)} = \left(\frac{n}{k_n}\right)^p \left(1 + \frac{p(p-1)}{2} \frac{1}{k_n} + o\left(\frac{1}{k_n}\right)\right). \tag{6.11}$$

The first relation follows from elementary calculus, while the second one is a consequence of the asymptotic expansion of the gamma function. The large-deviation result for $Y_{n-k_n:n}$ in (5.3), together with Chebyshev’s inequality, then implies that the asymptotic expansion in (6.11) also holds for $E[Y_{n-k_n:n}^p | Y_{n-k_n:n} \in I_n]$. Hence, if in (6.10) we integrate out the variable t with respect to the conditional distribution of $Y_{n-k_n:n}$ given $Y_{n-k_n:n} \in I_n$, the result is the same expression but with every t replaced by n/k_n and up to a remainder term of the order $o(k_n^{-1})$ as $n \rightarrow \infty$ uniformly in $x \in \mathbb{R}$. Finally, collect the terms of the same order to arrive at the expansion for $\Pr[H_n \leq x]$ stated in Theorem 4.1.

Appendix: Auxiliary results

Lemma A.1. *Let $(\xi_i)_{i \geq 1}$ be a sequence of independent, identically distributed random variables. If $E[|\xi_1|^p] < \infty$ for some $p \geq 2$, then there exists a positive constant c_p , depending only on p , such that for every real number δ and every positive integer n ,*

$$\Pr\left[\left|\frac{1}{n} \sum_{i=1}^n \xi_i - E[\xi_1]\right| \geq n^{-1/2+\delta}\right] \leq c_p E[|\xi_1|^p] n^{-\delta p}.$$

Lemma A.1 is a consequence of Markov’s inequality and the Marcinkiewicz–Zygmund inequality; see Chung (1951: 348–349). From Theorem 3.1.4 in Bingham *et al.* (1987) one can derive the Potter bound used in Section 6; see Lemma 2 in Cuntz *et al.* (2003) for details.

Lemma A.2. *Under Assumption 4, for every $\varepsilon > 0$ there exist $K_\varepsilon > 0$ and $t_\varepsilon > 1$ such that for all $1 \leq y < \infty$ and $t \geq t_\varepsilon$,*

$$\left|\frac{1}{b(t)} \left(\frac{\log V(ty) - \log V(t) - \gamma \log y}{a(t)} - ch_\rho(y)\right)\right| \leq K_\varepsilon y^\varepsilon.$$

Recall that the monic Chebyshev–Hermite polynomials are given by

$$H_m(x) = (-1)^m \exp(x^2/2) \frac{d^m}{dx^m} \exp(-x^2/2), \quad \text{for } m = 1, 2, \dots$$

The following theorem is a special case of Petrov (1975: Theorem VI.3.1).

Theorem A.3. *Let $(\xi_i)_{i \geq 1}$ be a sequence of independent and identically distributed*

random variables with zero mean and finite fifth absolute moment. Write $\sigma^2 = E[\xi_1^2]$, $\kappa_3 = E[\xi_1^3]$, and $\kappa_4 = E[\xi_1^4] - 3\sigma^2$. There exists an absolute positive constant C such that, for all real x ,

$$\left| \Pr \left[\frac{1}{\sigma k^{1/2}} \sum_{i=1}^k \xi_i \leq x \right] - \Phi(x) - \sum_{j=1}^2 f_j(x) k^{-j/2} \right| \\ \leq C \left(2\sigma^{-5} E[|\xi_1|^5] k^{-3/2} + k^{10} \left(\sup_{|z| \geq \delta} |E[e^{iz\xi_1}]| + (2k)^{-1} \right)^k \right)$$

with $\delta = \sigma^2 / \{12E[|\xi_1|^3]\}$, and

$$f_1(x) = -\frac{\kappa_3}{6\sigma^3} H_2(x) \varphi(x), \\ f_2(x) = -\left(\frac{\kappa_4}{24\sigma^4} H_3(x) + \frac{\kappa_3^2}{72\sigma^6} H_5(x) \right) \varphi(x).$$

Explicit expressions for the Hermite polynomials appearing in Theorem A.3 are $H_2(x) = x^2 - 1$, $H_3(x) = x^3 - 3x$, and $H_5(x) = x^5 - 10x^3 + 15x$.

Acknowledgement

This paper was partially written while Erich Haeusler was visiting the Catholic University Leuven and the University of Tilburg. He is very grateful to Jef Teugels and John Einmahl for their hospitality. Johan Segers gratefully acknowledges financial support in the form of a post-doctoral grant by the Fund for Scientific Research, Flanders, and in the form of a VENI grant by the Netherlands Organisation for Scientific Research (NWO). Both authors thank Björn Vandewalle for helpful comments.

References

- Beirlant, J., Bouquiaux, C. and Werker, B.J.M. (2006) Semiparametric lower bounds for tail index estimation. *J. Statist. Plann. Inference*, **136**, 705–729.
- Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987) *Regular Variation*. Cambridge: Cambridge University Press.
- Cheng, S. and de Haan, L. (2001) Penultimate approximation for Hill's estimator. *Scand. J. Statist.*, **28**, 569–575.
- Cheng, S. and Pan, J. (1998) Asymptotic expansions of estimators for the tail index with applications. *Scand. J. Statist.*, **25**, 717–728.
- Cheng, S. and Peng, L. (2001) Confidence intervals for the tail index. *Bernoulli*, **7**, 751–760.
- Chung, K.L. (1951) The strong law of large numbers. J. Neyman (ed.), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pp. 341–352. Berkeley: University of California Press.

- Cuntz, A., Haeusler, E. and Segers, J. (2003) Edgeworth expansions for the distribution function of the Hill estimator. CentER Discussion Paper 2003-08, Tilburg University. <http://center.uvt.nl/pub/dp/>.
- de Haan, L. and Peng, L. (1998) Comparison of tail index estimators. *Statist. Neerlandica*, **52**, 60–70.
- de Haan, L. and Stadtmüller, U. (1996) Generalized regular variation of second order. *J. Austral. Math. Soc. Ser. A*, **61**, 381–395.
- Drees, H. (1998) A general class of estimators of the extreme value index. *J. Statist. Plann. Inference*, **66**, 95–112.
- Ferreira, A. (2002) *Statistics of Extremes: Estimation and Optimality*. PhD dissertation, Tilburg University.
- Ferreira, A. and de Vries, C.G. (2004) Optimal confidence intervals for the tail index and high quantiles. Discussion Paper 2004-090/2, Tinbergen Institute. <http://www.tinbergen.nl>.
- 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**, 5–31.
- Guillou, A. and Hall, P. (2001) A diagnostic for selecting the threshold in extreme value analysis. *J. R. Statist. Soc. Ser. B*, **63**, 293–305.
- Hall, P. (1982) On some simple estimates of an exponent of regular variation. *J. Roy. Statist. Soc. Ser. B*, **44**, 37–42.
- Hill, B.M. (1975) A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, **3**, 1163–1174.
- Mason, D.M. (1982) Laws of large numbers for sums of extreme values. *Ann. Probab.*, **10**, 754–764.
- Petrov, V.V. (1975) *Sums of Independent Random Variables*. Berlin: Springer-Verlag.
- Reiss, R.-D. (1989) *Approximate Distributions of Order Statistics*. New York: Springer-Verlag.
- Resnick, S.I. and Stărică, C. (1995) Consistency of Hill’s estimator for dependent data. *J. Appl. Probab.*, **32**, 239–267.
- Resnick, S.I. and Stărică, C. (1998) Tail index estimation for dependent data. *Ann. Appl. Probab.*, **8**, 1156–1183.
- Segers, J. (2001a) Residual estimators. *J. Statist. Plann. Inference*, **98**, 15–27.
- Segers, J. (2001b) Abelian and Tauberian theorems for the bias of the Hill estimator. *Scand. J. Statist.*, **29**, 461–483.
- Shorack, G.R. and Wellner, J.A. (1986) *Empirical Processes*. New York: Wiley.
- Statulevičius, V.A. (1965) Limit theorems for densities and asymptotic expansions of distributions of sums of independent random variables. *Theory Probab. Appl.*, **10**, 582–595.

Received December 2005 and revised June 2006