PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 19, Fasc. 2 (1999), pp. 249–265

WEIGHTED LEAST-SQUARES ESTIMATORS OF TAIL INDICES

BY

LÁSZLÓ VIHAROS* (SZEGED)

Abstract. We propose a class of weighted least-squares estimators for the tail index of a regularly varying upper tail of a distribution. Universal asymptotic normality of the estimators is established over the whole model. Asymptotic mean square errors of these and earlier estimators are compared within a submodel of regular variation, more general than Hall's model. We also discuss the choice of the optimal weights and the choice of the number of extreme order statistics to be used.

1. Introduction and main result. Let $X, X_1, X_2, \ldots$ be independent random variables with a common distribution function $F(x) = P\{X \leqslant x\}$, $x \in \boldsymbol{R}$, and for each integer $n \geqslant 1$, let $X_{1,n} \leqslant \ldots \leqslant X_{n,n}$ denote the order statistics pertaining to the sample $X_1, \ldots, X_n$. We assume that

$$1 - F(x) = x^{-1/\alpha} l(x), \quad 0 < x < \infty, \tag{1.1}$$

where l is some function slowly varying at infinity and $\alpha > 0$ is a fixed unknown parameter to be estimated. The class of distribution functions satisfying (1.1) will be denoted by $\mathscr{R}_\alpha$. Several estimators exist for the tail index α among which Hill's estimator is the most classical (see Hill [15]):

$$\hat{\alpha}_n^{(H)}(k_n) := k_n^{-1} \sum_{i=1}^{k_n} \log^* X_{n+1-i,n} - \log^* X_{n-k_n,n},$$

where the k_n are some integers satisfying

$$1 \leqslant k_n < n, \quad k_n \to \infty \quad \text{and} \quad k_n/n \to 0 \quad \text{as } n \to \infty, \tag{1.2}$$

and $\log^* x = \log \max(x, 1)$, $x \in \boldsymbol{R}$. The class of kernel estimators of Csörgő et al. [5] generalizes the Hill estimator:

$$\hat{\alpha}_n^{(K)}(k_n) := \left(\sum_{i=1}^{n} \frac{i}{k_n} K\left(\frac{i}{k_n}\right) \{\log^* X_{n+1-i,n} - \log^* X_{n-i,n}\}\right) \Big(\int_0^{n/k_n} K(v)\, dv\Big)^{-1},$$

* Research partially supported by the Hungarian NSF, Grant F016226.

where $K(\cdot)$ is a non-increasing non-negative function such that $\int_0^\infty K(v)\,dv = 1$; the Hill estimator corresponds to $K(v) = 1_{[0<v\leqslant 1]}$. For simplicity of notation, we assume without loss of generality from now on that $F(0) = 0$ for all $F \in \mathscr{R}_\alpha$, i.e. X is positive; otherwise one only has to replace log by log* in what follows with some trivial extra reasonings in the proofs. The difficult problem of the asymptotic normality of $\hat{\alpha}_n^{(H)}$ has been investigated by Hall [13], Hall and Welsh [14], Haeusler and Teugels [12], Csörgő and Mason [7], Beirlant and Teugels [1] and their references, and, more generally for $\hat{\alpha}_n^{(K)}$, by Csörgő et al. [5]. However, as was shown in Csörgő and Viharos [8], $\hat{\alpha}_n^{(H)}$, is *not* universally asymptotically normal over the whole class $\mathscr{R}_\alpha$.

Recently, Schultze and Steinebach [17] proposed three new estimators of α, which are based on least-squares considerations under the restricted model in which $l(x) = e^c$ in (1.1), for all x beyond a threshold and for some constant c. (In a mathematically equivalent fashion, they in fact do this in a corresponding exponential model.) Taking the logarithm of (1.1) in the restricted case, substituting $x = X_{n+1-i,n}$ into $-\log(1-F(x)) = -c+\alpha^{-1}\log x$ and approximating the left-hand sides by

$$-\log(1-\hat{F}_n(X_{n+1-i,n}-)) = \log(n/i),$$

where $\hat{F}_n(\cdot)$ is the sample distribution function, for some $2 \leqslant k_n < n$ we have $\log X_{n+1-i,n} \approx d+\alpha\log(n/i)$, $i = 1, \ldots, k_n$, where $d = \alpha c$. Least-squares fit based on these approximative equations gives the first estimator of Schultze and Steinebach [17]:

$$\hat{\alpha}_n^{(S)}(k_n) := \frac{\sum_{i=1}^{k_n}(\log(n/i))(\log X_{n+1-i,n}) - k_n^{-1}\left(\sum_{i=1}^{k_n}\log X_{n+1-i,n}\right)\left(\sum_{i=1}^{k_n}\log(n/i)\right)}{\sum_{i=1}^{k_n}\log^2(n/i) - k_n^{-1}\left(\sum_{i=1}^{k_n}\log(n/i)\right)^2}.$$

Independently of Schultze and Steinebach [17], Kratz and Resnick [16] also proposed $\hat{\alpha}_n^{(S)}(k_n)$ at about the same time, using a mathematically equivalent heuristic argument. Csörgő and Viharos [9] proved that for suitable $\mu_n^{(S)}(k_n) \to \alpha$ the sequence $k_n^{1/2}\{\hat{\alpha}_n^{(S)}(k_n) - \mu_n^{(S)}(k_n)\}$ is universally asymptotically normal over $\mathscr{R}_\alpha$ for all $\{k_n\}$ satisfying $k_n/\log^4 n \to \infty$ as $n \to \infty$. In contrast, Csörgő and Viharos [8] constructed distribution functions $F \in \mathscr{R}_\alpha$ such that $\hat{\alpha}_n^{(H)}(\lfloor n^{2/3}\rfloor)$, where $\lfloor\cdot\rfloor$ denotes the integer part, does not have a non-degenerate asymptotic distribution for *any* centering and norming sequences. Since the Hill estimator was also constructed in the restricted model $l(x) = e^c$, it may be concluded that the asymptotic normality of $\hat{\alpha}_n^{(S)}(k_n)$ is more robust against deviations of l from a constant; see Csörgő and Viharos [9] for more on this aspect.

Beirlant et al. [2] obtained a subclass of kernel estimates of Csörgő et al. [5] using a weighted least-squares method fitting an appropriate regression line through a fixed data point. In the language above, this class of estimators, containing the Hill estimator, has a lack of robustness properties similar to that of the original Hill estimator. Beirlant et al. [2] also discussed the difficult problem of the choice of the parameter k_n in their estimators.

These results raise the question of weighted least-squares versions without forcing the regression line through any fixed point as a generalization of Schultze and Steinebach's [17] method. We show that, using appropriate weights, we obtain a weighted estimator $\hat{\alpha}_n^{(W)}$ which inherits the universal asymptotic normality of $\hat{\alpha}_n^{(S)}$, and hence the asymptotic normality of $\hat{\alpha}_n^{(W)}$ is also more robust than that of the Hill estimator and the generalized Hill estimator $\hat{\alpha}_n^{(K)}$. In the restricted Hall submodel of regular variation, discussed in Section 3 below, Csörgő and Viharos [9] compared $\hat{\alpha}_n^{(S)}$ to the optimal kernel estimator and to other proposals, through the investigation of the asymptotic mean square error, and the kernel estimator proved to be the best. In Section 2 we show that, in a more general submodel, the asymptotic mean square error of the optimal $\hat{\alpha}_n^{(W)}$ is the same as that of the optimal kernel estimator.

Choosing some weights $w_{i,n} \geqslant 0$, $i = 1, \ldots, k_n$, and minimizing the corresponding weighted error sum of squares

$$\sum_{i=1}^{k_n} w_{i,n} \left(\log X_{n+1-i,n} - \alpha \log(n/i) - d\right)^2,$$

we obtain the weighted generalization of $\hat{\alpha}_n^{(S)}$:

$$\hat{\alpha}_n^{(W)}(k_n) := \frac{\sum_{i=1}^{k_n} w_{i,n} (\log(n/i))(\log X_{n+1-i,n}) - \left(\sum_{i=1}^{k_n} w_{i,n}\right)^{-1} \left(\sum_{i=1}^{k_n} w_{i,n} \log X_{n+1-i,n}\right)\left(\sum_{i=1}^{k_n} w_{i,n} \log(n/i)\right)}{\sum_{i=1}^{k_n} w_{i,n} \log^2(n/i) - \left(\sum_{i=1}^{k_n} w_{i,n}\right)^{-1} \left(\sum_{i=1}^{k_n} w_{i,n} \log(n/i)\right)^2}.$$

In what follows we shall study the class of weights of the form

$$w_{i,n} = w_{i,n}(J) := \frac{k_n \int_{(i-1)/k_n}^{i/k_n} J(x)\,dx}{-\log(i/k_n) + \log x_J},$$

where $J(\cdot)$ is a weight function specified by conditions (i)–(v) below and x_J is a shift parameter chosen so that all the weights $w_{i,n}$, $i = 1, \ldots, k_n$, are non-negative. Let $\lceil x \rceil$ denote the smallest integer not smaller than x and take the step function

$$J_S(x) := -\log \frac{\lceil x k_n \rceil}{k_n} - 1 = -\log \frac{i}{k_n} - 1, \quad \frac{i-1}{k_n} < x \leqslant \frac{i}{k_n}, \; i = 1, \ldots, k_n,$$

and set $x_{J_S} := e^{-1}$. Then $w_{i,n}(J_S) \equiv 1$, and therefore we get the original least-squares estimator $\hat{\alpha}_n^{(S)}$. For the sake of technical simplicity we will consider a class of smooth weight functions $\{J(x)\colon\ 0 < x \leqslant 1\}$, which preserves the properties of the function $J_0(x) := -\log x - 1 \approx J_S(x)$, described by the following

CONDITIONS. (i) $\int_0^1 J(x)\,dx = 0$.

(ii) $J(\cdot)$ is non-increasing on $(0, 1]$ such that $\lim_{x \downarrow 0} J(x) > 0$ and $J(1) < 0$.

(iii) The derivative $J'(\cdot)$ is monotone on $(0, 1]$.
(iv) $\sup_{0<x\leqslant 1} x|J'(x)|$ is finite.
(v) The integral $\int_0^1 x^{-(1/2)-\nu}|J(x)|\,dx$ is finite for some $\nu>0$.

Condition (ii) forces $\nu\in(0, 1/2)$ in (v). Assume that y_J satisfies $J(y_J)=0$. Then the x_J appearing in $w_{i,n}(J)$ has to be close to y_J. Observe that, for any number c_n,

$$\hat{\alpha}_n^{(W)}(k_n) = \frac{\sum_{i=1}^{k_n}\{w_{i,n}(c_n+\log(n/i))-w_{i,n}(\sum_{j=1}^{k_n} w_{j,n})^{-1}\sum_{j=1}^{k_n} w_{j,n}(c_n+\log(n/j))\}\log X_{n+1-i,n}}{\sum_{i=1}^{k_n} w_{i,n}(c_n+\log(n/i))^2-(\sum_{i=1}^{k_n} w_{i,n})^{-1}(\sum_{i=1}^{k_n} w_{i,n}(c_n+\log(n/i)))^2}.$$

Set

$$g_{n+1-i,n} := k_n \int_{(i-1)/k_n}^{i/k_n} J(x)\,dx$$

and choose

$$c_n := \log x_J + \log\frac{k_n}{n}.$$

Then we have

$$\sum_{i=1}^{k_n} w_{i,n}(c_n+\log(n/i)) = \sum_{i=1}^{k_n} g_{n+1-i,n} = k_n\int_0^1 J(x)\,dx = 0,$$

resulting in

$$\hat{\alpha}_n^{(W)}(k_n) = \frac{\sum_{i=1}^{k_n} g_{n+1-i,n}\log X_{n+1-i,n}}{\sum_{i=1}^{k_n} g_{n+1-i,n}\log(k_n/i)}.$$

So we see that the final estimate $\hat{\alpha}_n^{(W)}(k_n)$ does not depend on the shift parameter x_J within the class of shift parameters specified above.

Introducing $Q(s):=\inf\{x\colon F(x)\geqslant s\}$, $0<s\leqslant 1$, $Q(0):=Q(0+)$, the inverse or quantile function of F, and letting $Q(1-s-)$ denote the left-continuous version of the right-continuous function $Q(1-s)$, $0<s<1$, it is well known that $F\in\mathscr{R}_\alpha$ if and only if, for some function $L(\cdot)$ slowly varying at zero,

$$Q(1-s-)=s^{-\alpha}L(s), \quad 0<s<1. \tag{1.3}$$

Let $\xrightarrow{\mathscr{D}}$ denote convergence in distribution and let $\mathscr{N}(\mu, \sigma^2)$ be the normal distribution with mean $\mu\in\boldsymbol{R}$ and variance $\sigma^2>0$, and define $u\wedge v=\min(u, v)$ and $u\vee v=\max(u, v)$. Understanding limiting and order relations and asymptotic equalities $\sim$ as $n\to\infty$ throughout if not specified otherwise, we can state now the main limit theorem of the paper. All the proofs are in the fourth section.

THEOREM 1.1. *If k_n is any sequence of positive integers such that* (1.2) *holds, then, whenever $F\in\mathscr{R}_\alpha$ for some $\alpha\in(0, \infty)$,*

$$k_n^{1/2}\{\hat{\alpha}_n^{(W)}(k_n)-\mu_n^{(W)}(k_n)\}\xrightarrow{\mathscr{D}}\mathscr{N}(0, \alpha^2\sigma_J^2),$$

where, with $\tau(J) := -\int_0^1 J(v)\log v\,dv$,

$$\sigma_J^2 := \frac{1}{\tau^2(J)}\int_0^1\int_0^1 \frac{u\wedge v}{uv} J(u)J(v)\,du\,dv$$

and $\mu_n^{(W)}(k_n) := n\big(\tau(J)k_n\big)^{-1}\int_0^{k_n/n} J(nt/k_n)\log Q(1-t-)\,dt \to \alpha$.

In Theorem 1.1 we do not need the restriction $k_n/\log^4 n \to \infty$ as in Theorem 1.1 of Csörgő and Viharos [9]. The reason is that Theorem 1.1 above deals with estimators $\hat{\alpha}_n^{(W)}(k_n) = \hat{\alpha}_n^{(W)}(k_n, J)$ with smooth weight functions J satisfying conditions (i)–(v). On the other hand, Theorem 1.1 of [9] deals with $\hat{\alpha}_n^{(S)} = \hat{\alpha}_n^{(W)}(k_n, J_S)$ with the step function J_S not satisfying conditions (i)–(v). The proof of Theorem 1.1 of [9] is partly based on the step

$$k_n^{1/2}\big(\hat{\alpha}_n^{(W)}(k_n, J_S) - \hat{\alpha}_n^{(W)}(k_n, J_0)\big) = o_P(1),$$

where $J_0(x) = -\log x - 1$ satisfies conditions (i)–(v). This step requires the growth condition $k_n/\log^4 n \to \infty$ (cf. the notice preceding Lemma 5.4 of Csörgő and Viharos [9]). After this step, asymptotic normality of $\hat{\alpha}_n^{(W)}(k_n, J_0)$ is established, which is the main point of the proof of Theorem 1.1 of [9].

2. Asymptotic mean square errors. The theorem above suggests defining the asymptotic mean square error of $\hat{\alpha}_n^{(W)}$ as

$$M_n^{(W)} = M_n^{(W)}(k_n, J) := b_n^2(k_n) + \alpha^2\sigma_J^2 k_n^{-1} \quad \text{for some } b_n(k_n) \sim \{\mu_n^{(W)}(k_n) - \alpha\};$$

the sum of the asymptotic squared bias and variance (cf. Csörgő and Viharos [9], Section 4). We shall discuss the behavior of $M_n^{(W)}$ in a submodel of (1.3) studied by Beirlant et al. [2], restricting L to satisfy

$$\frac{\big(L(st)/L(s)\big)-1}{g(s)} = \frac{t^{-\varrho}-1}{\varrho} + \delta_t(s), \quad \text{where } \lim_{s\downarrow 0}\delta_t(s) = 0 \text{ for all } t > 0, \tag{2.1}$$

for some rate function $g(\cdot)$ of constant sign and some constant $\varrho \leqslant 0$, where $(t^{-\varrho}-1)/\varrho$ is understood as $-\log t$ if $\varrho = 0$. Condition (2.1) implies that $g(\cdot)$ is regularly varying at zero with index $-\varrho$ ($g(\cdot)\in RV_{-\varrho}^{0}$) and there exists a positive decreasing function $h(\cdot)$ regularly varying at infinity with index $2\varrho - 1$ ($h(\cdot)\in RV_{2\varrho-1}^{\infty}$) such that

$$g^2(1/t) \sim \int_t^\infty h(s)\,ds \quad \text{as } t\to\infty, \tag{2.2}$$

where for $\varrho = 0$ we assume $\lim_{s\downarrow 0} g(s) = 0$ and $g^2(1/t)$ is asymptotic to a non-increasing function as $t\to\infty$ (see Dekkers and de Haan [11], Lemma 2.8). To evaluate the asymptotic mean square error, we need further conditions on the underlying distribution which control the speed of convergence in (2.1):

$$\lim_{s\downarrow 0}\sup_{0<t\leqslant b} t^{\lambda}|\delta_t(s)| = 0 \quad \text{and} \quad \lim_{s\downarrow 0}\sup_{0<t\leqslant b} t^{\mu}\left|\log\frac{L(st)}{L(s)}\right| = 0 \tag{2.3}$$

for some constants λ, $\mu \geqslant 0$ and $b > 0$. Define

$$J_\vartheta(x) = (\vartheta+1)\vartheta^{-1} - (\vartheta+1)^2 \vartheta^{-1} x^\vartheta \quad \text{for any } \vartheta > 0,$$

and set

$$J_0(x) := \lim_{\vartheta \downarrow 0} J_\vartheta(x) = -\log x - 1.$$

Then $\{J_\vartheta \colon \vartheta \geqslant 0\}$ is a subclass of weight functions satisfying conditions (i)–(v) of Section 1. Note that $\tau(J_\vartheta) = 1$ and $\sigma^2_{J_\vartheta} = (2\vartheta+2)/(2\vartheta+1)$ for all $\vartheta \geqslant 0$. Set $q_\varrho(t) := (t^{-\varrho}-1)/\varrho$ for $\varrho > 0$ and $q_0(t) := -\log t$. The next theorem gives the optimal weight function J_ϑ in the model given in (2.1)–(2.3).

THEOREM 2.1. *Assume that* (2.1), (2.2) *and* (2.3) *hold with some* $\varrho \leqslant 0$, *some rate function* $g(\cdot)$, *and some constants* λ, $\mu \geqslant 0$, $\lambda+\mu \in [0, 2^{-1}+\nu)$, *and* $b = 1$, *where* ν *is specified in condition* (v). *For* $\varrho = 0$ *suppose that* $\lim_{s \downarrow 0} g(s) = 0$ *and* $g^2(1/t)$ *is asymptotic to a non-increasing function at* $t \to \infty$. *Then*

$$M_n^{(W)} = g^2\left(\frac{k_n}{n}\right) B_{J,\varrho} + \frac{\alpha^2 \sigma_J^2}{k_n} := g^2\left(\frac{k_n}{n}\right) \left(\frac{1}{\tau(J)} \int_0^1 q_\varrho(v) J(v)\, dv\right)^2 + \frac{\alpha^2 \sigma_J^2}{k_n},$$

and the optimal k_n *is determined asymptotically by*

$$k_n = k_n^* := \left\lfloor \frac{n}{h^{\leftarrow}\left(\alpha^2 \sigma_J^2/(n B_{J,\varrho})\right)} \right\rfloor, \tag{2.4}$$

where $\lfloor \cdot \rfloor$ *denotes integer part,* $h^{\leftarrow}(\cdot)$ *is the generalized inverse of* $h(\cdot)$. *Moreover, if* $\varrho < 0$, *then*

$$M_n^{(W)}(k_n^*) \sim (\alpha\sigma_J)^{4\varrho/(2\varrho-1)} B_{J,\varrho}^{1/(1-2\varrho)} \left(1 - \frac{1}{2\varrho}\right) \frac{h^{\leftarrow}(1/n)}{n} =: \tilde{M}_n^{(W)}(k_n^*) \tag{2.5}$$

and

$$\min_{\vartheta \geqslant 0} \tilde{M}_n^{(W)}(k_n^*, J_\vartheta) = \tilde{M}_n^{(W)}(k_n^*, J_{-\varrho}).$$

If $\varrho = 0$ *and* $\left(k_n g^2(k_n/n)\right)^{-1} \to 0$, *then* $M_n^{(W)} \sim g^2(k_n/n)$.

We conjecture that the $J_{-\varrho}$ in Theorem 2.1 is optimal over the whole class of weight functions specified by conditions (i)–(v) above. To compare $\hat{\alpha}_n^{(W)}$ to the class of kernel estimators, we generalize Theorem 5 of Csörgő et al. [5] for the model given in (2.1)–(2.3) (cf. the first example in the next section). We use almost the same conditions on the kernel function K assuming left-continuity instead of right-continuity in (H2) to involve the $K(v) = 1_{[0<v\leqslant 1]}$ Hill case in the restricted class (cf. Csörgő and Viharos [10], Section 2.2):

(H1) $K(u) \geqslant 0$ for $0 < u < \infty$.

(H2) $K(\cdot)$ is non-increasing and left-continuous on $(0, \infty)$.

(H3) $\int_0^\infty K(v)\, dv = 1$.

(H4) $\int_0^\infty v^{-1/2} K(v)\, dv < \infty$.

(H5) There exists a $0 < \Lambda < \infty$ such that $K(u) = 0$ for $u > \Lambda$.

(H6) There exists a $0 < \Lambda < \infty$ such that $dK(u)/du = k(u)$ is defined for $u > \Lambda$ and $\lim_{u \uparrow \infty} u^{3/2} k(u) = 0$.

Beside (2.1)–(2.3) we assume the following conditions on the underlying distribution:

(D1) The quantile function in (1.3) satisfies $Q(0) = 1$.

(D2) (i) In the Karamata representation given in (4.2) below, one has either (H5) is satisfied and $a(s) = a_0$ for $0 < s < \varepsilon$ for some $\varepsilon > 0$, or $a(s) = a_0$ for $0 < s \leqslant 1$.

(ii) One has either (H6) is satisfied, or the function $b(\cdot)$ in (4.2) may be chosen sueh that $b(\cdot)$ is bounded on (0, 1).

THEOREM 2.2. *Assume that the conditions* (H1)–(H5) *and* (D1)–(D2), *and all the conditions of Theorem* 2.1 *above are satisfied for the underlying distribution, restricting this time* $\lambda + \mu$ *from* $[0, 2^{-1} + \nu)$ *to the interval* $[0, 2^{-1})$ *and changing* $b = 1$ *to* $b = \Lambda$. *Then the asymptotic mean square error of* $\hat{\alpha}_n^{(K)}$ *is given by*

$$M_n^{(K)} = g^2\left(\frac{k_n}{n}\right) D_{K,\varrho} + \frac{\alpha^2 \sigma_K^2}{k_n} := g^2\left(\frac{k_n}{n}\right)\left(\int_0^{\Lambda} v^{-\varrho} K(v)\, dv\right)^2 + \frac{\alpha^2}{k_n} \int_0^{\Lambda} K^2(v)\, dv.$$

The optimal k_n *is determined asymptotically by*

$$k_n = k_n^{\diamond} := \left\lfloor \frac{n}{h^{\leftarrow}\left((\alpha^2 \sigma_K^2)/nD_{K,\varrho}\right)} \right\rfloor .$$

Moreover, if $\varrho < 0$, *then*

$$M_n^{(K)}(k_n^{\diamond}) \sim (\alpha\sigma_K)^{4\varrho/(2\varrho-1)}\, D_{K,\varrho}^{1/(1-2\varrho)} \left(1 - \frac{1}{2\varrho}\right) \frac{h^{\leftarrow}(1/n)}{n},$$

and the optimal $K(\cdot)$ *is given by*

$$K_\varrho^*(v) = \frac{\varrho - 1}{\varrho}\left(\frac{2\varrho - 1}{2\varrho - 2}\right)^{1-\varrho} \left\{\left(\frac{2\varrho - 2}{2\varrho - 1}\right)^{-\varrho} - v^{-\varrho}\right\}$$

if $0 < v < (2 - 2\varrho)/(1 - 2\varrho)$, $K_\varrho^*(v) = 0$ *otherwise. If* $\varrho = 0$ *and* $\left(k_n g^2(k_n/n)\right)^{-1} \to 0$, *then* $M_n^{(K)} \sim g^2(k_n/n)$.

The optimal kernel K_ϱ^* has the same analytical form as the one obtained in the Hall model given in (3.1) below (cf. Theorem 5 of Csörgő et al. [5]). The optimal $J_9 = J_{-\varrho}$ also depends on the unknown nuisance parameter ϱ so that further study is needed to find methods to pre-estimate ϱ (cf. the discussion of the Hall example in the next section). Using Theorems 2.1 and 2.2, it is possible to compare $\hat{\alpha}_n^{(W)}(k_n^*, J)$ to $\hat{\alpha}_n^{(S)}(k_n^*)$, the Hill estimator and the optimal kernel estimator. For each of these, the corresponding smallest possible asymptotic mean square error M_n under (2.1)–(2.3) is of the same order with the same leading constant $\alpha^{4\varrho/(2\varrho-1)}$, so the comparison can be made by means of the corresponding functions

$$m(\varrho) := \lim_{n\to\infty} M_n n / \{\alpha^{4\varrho/(2\varrho-1)}\, h^{\leftarrow}(1/n)\}, \qquad \varrho < 0.$$

Figure 1, drawn by the mathematical program package Maple V, depicts the $m(\varrho)$ of $\hat{\alpha}_n^{(S)}(k_n^*)$, of the optimal Hill estimator $\hat{\alpha}_n^{(H)}(k_n^\diamond)$ and of $\hat{\alpha}^{(W)}(k_n^*, J_\vartheta)$ for $\vartheta = 1/2, 1, 2, 3, -\varrho$ tagged by the corresponding values of ϑ of the five weight functions J_ϑ considered, as they leave the picture. As was pointed out by Csörgő and Viharos [10], the $m(\varrho)$ of $\hat{\alpha}^{(W)}(k_n^*, J_{-\varrho})$ is identical to that of the optimal kernel estimator so that the latter is not included separately in Figure 1. It can be calculated that $\lim_{\varrho\to-\infty} m(\varrho) = 2$ for $\hat{\alpha}_n^{(S)}(k_n^*)$, $\lim_{\varrho\to-\infty} m(\varrho) = (2\vartheta+2)/(2\vartheta+1)$ for $\hat{\alpha}_n^{(W)}(k_n^*, J_\vartheta)$, $\vartheta \geq 0$, and $\lim_{\varrho\to-\infty} m(\varrho) = 1$ for $\hat{\alpha}_n^{(W)}(k_n^*, J_{-\varrho})$, $\hat{\alpha}_n^{(H)}(k_n^\diamond)$ and for the optimal kernel estimator. The optimal estimator $\hat{\alpha}^{(W)}(k_n^*, J_{-\varrho})$ is uniformly better than the other estimators studied in Figure 1. These coincide with the finding in Csörgő and Viharos [9].

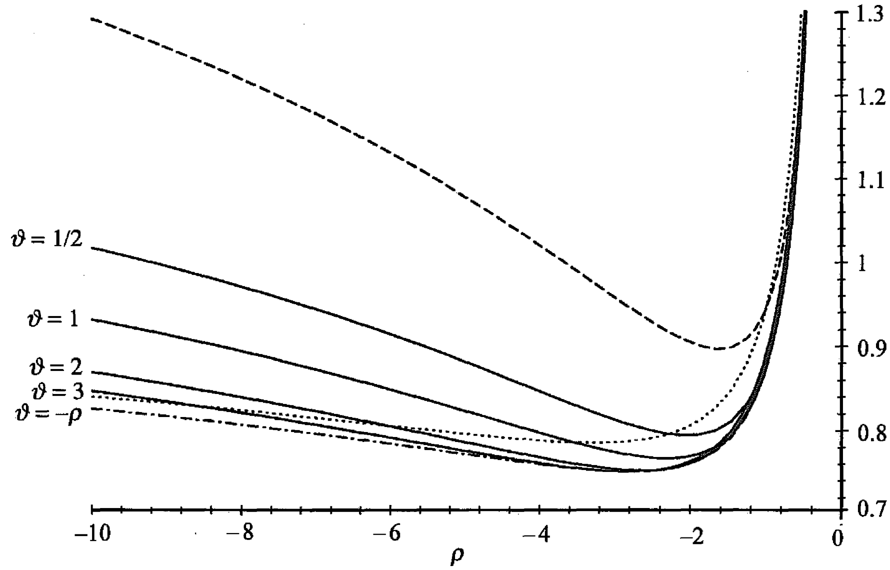

Fig. 1. The curves of $m(\varrho)$ for $\hat{\alpha}_n^{(S)}(k_n^*)$ (dashed), $\hat{\alpha}_n^{(H)}(k_n^\diamond)$ (dotted), $\hat{\alpha}^{(W)}(k_n^*, J_\vartheta)$, $\vartheta = 1/2, 1, 2, 3$ (solid) and $\vartheta = -\varrho$ (dash-dotted)

3. Examples. In this section we study two submodels satisfying the conditions of Theorems 2.1 and 2.2. Hall [13] and Csörgő et al. [5], Theorem 5, investigated the asymptotic normality of the Hill estimator in the model given by

$$L(s) = D_1[1+D_2 s^\beta\{1+o(1)\}] \quad \text{as } s \to 0, \tag{3.1}$$

where $D_1 > 0$, $D_2 \neq 0$ and $\beta > 0$ are constants. In some practical situations the parameters D_1, D_2, β can be previously estimated from a given sample (cf. Csörgő et al. [5] and Hall [13]). This is a submodel of (2.1) corresponding to $g(s) = -\beta D_2 s^\beta$ and $\varrho = -\beta$, and, by a routine calculation, satisfying (2.3) with any $b > 0$ and λ, $\mu \geq 0$. This restricted model contains those distributions for which the corresponding function $L(s)$ (or equivalently $l(x)$) is nearly constant for small s (for large x). Plots of the graph of $x^{1/\hat{\alpha}_n}\{1-\hat{F}_n(x)\}$ for large x, with preliminary estimates $\hat{\alpha}_n$, maybe helpful in exploring this. In this case, based on Figure 1, it is reasonable to use $\hat{\alpha}_n^{(W)}$ with some $\vartheta \geqslant 1$.

Next we discuss another model containing a typical family of L functions that are not bounded near zero, given by $L(s) = L_\gamma(s) := \exp\{\log^\gamma(1/s)\}$, $0 < \gamma < 1$, taken from Bingham et al. [3], p. 16. This model, by a routine calculation, corresponds to $g(s) = \gamma\log^{\gamma-1}(1/s)$ and $\varrho = 0$ in (2.1). This submodel also satisfies (2.3) with any positive constants b, λ and μ, as a tedious calculation shows. From Theorems 2.1 and 2.2 we know that this time

$$M_n \sim \gamma^2 \log^{2\gamma-2}(n/k_n) \quad \text{if } \log^{2-2\gamma}(n/k_n)/k_n \to 0$$

for all the estimators under investigation. Define

$$A_t(L) := (\tau(J))^{-1}\int_0^1 J(v)\log L(vt)\,dv.$$

By calculation from (1.3) we obtain

$$\mu_n^{(W)}(k_n) - \alpha = \frac{1}{\tau(J)}\int_0^1 J(v)\log L\left(v\frac{k_n}{n}\right)dv = A_{k_n/n}(L). \tag{3.2}$$

To see differences within this model between the original least-squares estimator and the weighted versions, we shall study for finite n the behavior of the asymptotic mean square error

$$\hat{M}_n^{(W)}(k_n, J, L, \alpha) := A^2_{k_n/n}(L) + \alpha^2\sigma_J^2 k_n^{-1}.$$

Introduce the function $M_n^*(t; J, L, \alpha) := A_t^2(L) + \alpha^2\sigma_J^2 t^{-1}n^{-1}$, $0 < t < 1$, and notice that

$$\hat{M}_n^{(W)}(k_n, J, L, \alpha) = M_n^*(k_n/n; J, L, \alpha)$$

for any L in (1.3) and any k_n as in (1.2).

Figure 2 contains the graphs of $M_{500}^*(t; J_0, L_\gamma, 1)$, $t \in (0.01, 0.36]$, for $\gamma = 1/4, 1/2, 2/3, 3/4$ in dotted curves, tagged by the corresponding γ values of the four slowly varying functions considered. These curves correspond to $\hat{\alpha}_n^{(S)}$. Figure 2 also contains the graphs of $M_{500}^*(t; J_\vartheta, L_\gamma, 1)$, $\gamma = 1/4, 1/2, 2/3, 3/4$, for $\vartheta = 0.05, 0.2, 0.5, 1$ in solid curves. For each γ investigated, the corresponding curves for the different ϑ values are close to each other. As numerical evaluation shows, for $\gamma = 1/2$ and $\gamma = 2/3$, the absolute minima of the curves become smaller and smaller as the value of ϑ increases, albeit the decreases are hardly noticeable. For $\gamma = 1/4$ the minima are approximately the same, while for $\gamma = 3/4$ the estimator $\hat{\alpha}_n^{(S)}$ is slightly better than $\hat{\alpha}_n^{(W)}$. So there is a slight advantage of the use of the weights for γ values less than 3/4. For $\vartheta > 1$ the minima shift more and more upwards as the value of ϑ increases. From these two examples we conclude that the choice $\vartheta = 1$ appears to be reasonable for bounded and unbounded functions L as well.

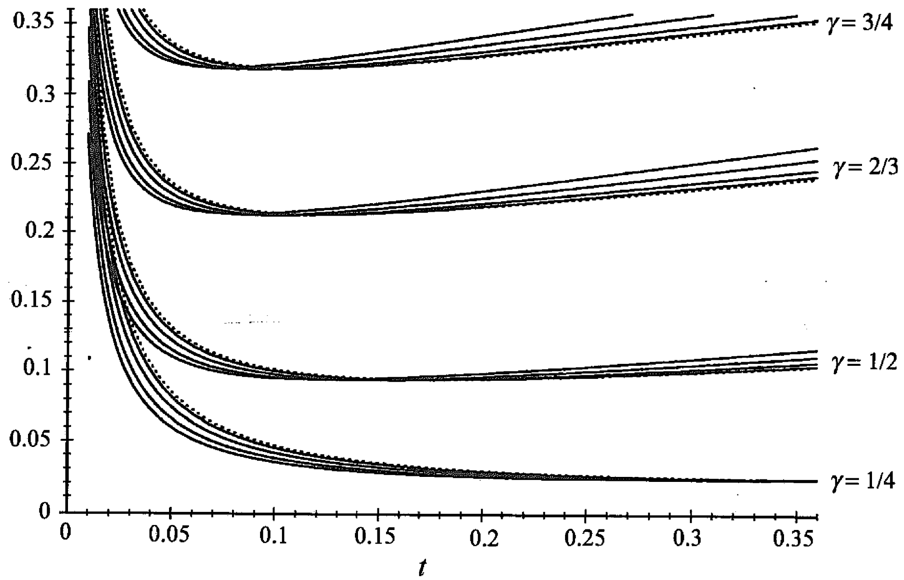

Fig. 2. The curves of $M^*_{500}(t;\, J_\vartheta,\, L_\gamma,\, 1)$ for $\vartheta = 0.05,\ 0.2,\ 0.5,\ 1$ (solid) and $\vartheta = 0$ (dotted), $\gamma = 1/4,\ 1/2,\ 2/3,\ 3/4$

For sample sizes $n = 400, 300, 200, 100$ the curves shift more and more upwards, but the whole global picture remains the same. Based on these and additional figures belonging to further slowly varying functions as in Csörgő and Viharos [9], the following pragmatic rule of thumb appears reasonable for estimating α near 1 using the weight function $J_\vartheta = J_1$: For $100 \leqslant n < 300$ use $\hat{\alpha}_n^{(W)}(k_n)$ with $7n/100 \leqslant k_n \leqslant 26n/100$, and for $300 \leqslant n \leqslant 3000$ use $\hat{\alpha}_n^{(W)}(k_n)$ with $6n/100 \leqslant k_n \leqslant 25n/100$. Several rules may be formulated for the adaptive choice of k_n within the indicated ranges. For example, choose that k_n for which $\hat{\alpha}_n^{(W)}(k_n)$ and $\hat{\alpha}_n^{(S)}(k_n)$ are the closest to each other; this is a part of the suggested rule and it uniquely determines the choice of a data-driven k_n for $n \geqslant 100$. From a further study as in Csörgő and Viharos [9], Section 4, it follows that this rule of thumb can be extended to the range $0 < \alpha \leqslant 2$, which range appears to cover all cases that are important in practice.

Various adaptive versions of both the present weighted least-squares and the kernel estimators are proposed in Csörgő and Viharos [10] and are investigated in an extensive simulation study. These are based on the respective classes $\{J_\vartheta\colon\ \vartheta \geqslant 0\}$ and $\{K_\vartheta^*\colon\ \vartheta \leqslant 0\}$ of weight and kernel functions, suggested by Theorems 2.1 and 2.2, and the procedures provide data-driven choices of both k_n and the shape parameter ϑ in both cases.

4. Proofs. Throughout, Csörgő and Viharos are abbreviated to CsV, and $\stackrel{\mathscr{D}}{=}$ denotes distributional equality. We use the convention $\int_a^b = \int_{[a,b)}$ when we integrate with respect to a left-continuous integrator if not specified otherwise.

To prove Theorem 1.1, we use the probability space constructed by Csörgő et al. [4]. It carries a sequence $\{U_n\}_{n=1}^{\infty}$ of independent random variables uniformly distributed on (0, 1), with $U_{1,n} \leqslant \ldots \leqslant U_{n,n}$ as the order statistics of $U_1, \ldots, U_n$, and a sequence $\{B_n(t)\colon 0 \leqslant t \leqslant 1\}_{n=1}^{\infty}$ of Brownian bridges such that, letting

$$G_n(s) := n^{-1} \sum_{k=1}^{n} I\{U_k \leqslant s\}, \quad 0 < s < 1,$$

and

$$U_n(s) := U_{k,n}, \quad (k-1)/n < s \leqslant k/n,\ k = 1, \ldots, n,$$

and putting

$$\beta_n(s) := n^{1/2}\{G_n(s) - s\} \quad \text{and} \quad \gamma_n(s) := n^{1/2}\{s - U_n(s)\}$$

for the respective uniform empirical and quantile processes, we have

$$\tag{4.1} \begin{aligned} \sup_{\varsigma/n \leqslant s \leqslant 1-(\varsigma/n)} \frac{|\beta_n(s) - B_n(s)|}{\big(s(1-s)\big)^{1/2-\theta}} &= \mathcal{O}_P\left(\frac{1}{n^{\theta}}\right), \\ \sup_{\varsigma/n \leqslant s \leqslant 1-(\varsigma/n)} \frac{|\gamma_n(s) - B_n(s)|}{\big(s(1-s)\big)^{1/2-\varkappa}} &= \mathcal{O}_P\left(\frac{1}{n^{\varkappa}}\right) \end{aligned}$$

for any fixed $\theta \in [0, 1/4)$, $\varkappa \in [0, 1/2)$ and $\varsigma > 0$.

Set $\bar{\alpha}_n := \big(\tau(J)\,k_n\big)^{-1} \sum_{i=1}^{k_n} g_{n+1-i,n} \log X_{n+1-i,n}$, where $\tau(J)$ is defined in Theorem 1.1. The next proposition states that $\bar{\alpha}_n$ is an equivalent version of $\hat{\alpha}_n^{(W)}$ in Theorem 1.1.

PROPOSITION 4.1. *Assume that the conditions of Theorem 1.1 are satisfied. Then*

$$k_n^{1/2}\{\bar{\alpha}_n - \mu_n^{(W)}(k_n)\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \alpha^2 \sigma_J^2)$$

and

$$\frac{1}{\tau(J)\,k_n} \sum_{i=1}^{k_n} g_{n+1-i,n} \log \frac{k_n}{i} = 1 + o(k_n^{-1/2}).$$

Proof. The proof of the first statement follows the general outline of the proof of Theorem 1.1 of CsV [9]. The main difficult point in the present case is to handle a whole class of weight functions $\{J\}$ instead of the single one

$$J_S(x) = -\log \frac{\lceil x k_n \rceil}{k_n} - 1.$$

Set $g(t) := \log Q(1-t-)$ and introduce the modified centering sequence

$$\bar{\mu}_n := \big(\tau(J)\big)^{-1}\Big\{n k_n^{-1} \int_{1/n}^{k_n/n} g(t)\, J(nt/k_n)\, dt + k_n^{-1} g_{n,n}\, g(1/n)\Big\}.$$

Further, let $G_n^*(s) := \{G_n(s) \wedge (1-n^{-1})\} \vee n^{-1}$ and, for $1 \leqslant m \leqslant r \leqslant n$,

$$\Theta_n(m, r) := -\int_{m/n}^{r/n} \int_{nt/k_n}^{nG_n^*(t)/k_n} J(s)\, ds\, dg(t), \qquad V_n(m) := \int_{U_{m,n}}^{m/n} \int_{m/k_n}^{nG_n^*(t)/k_n} J(s)\, ds\, dg(t).$$

Fix any integers m and l such that $1 \leqslant m \leqslant l \leqslant k_n$. Then we obtain

$$\tau(J)(\bar{\alpha}_n - \bar{\mu}_n) \stackrel{\mathcal{D}}{=} \Theta_n(1, m) - V_n(1) + R_n + \Theta_n(m, l) + V_n(k_n) + \Theta_n(l, k_n)$$
$$=: \Delta_n(m, l) + V_n(k_n) + \Theta_n(l, k_n)$$

as an analogue of (5.6) in CsV [9], where

$$R_n := g_{n,n}\{g(U_{1,n}) - g(1/n)\}/k_n = \mathcal{O}_P(g_{n,n}/k_n)$$

by an application of Lemma 5.4 (iv) in CsV [9]. Using condition (v), we obtain

$$g_{n,n} = k_n \int_0^{1/k_n} J(x)\,dx = k_n \int_0^{1/k_n} x^{(1/2)+\nu} x^{-(1/2)-\nu} J(x)\,dx$$
$$\leqslant k_n \frac{1}{k_n^{(1/2)+\nu}} \int_0^{1/k_n} x^{-(1/2)-\nu} J(x)\,dx = o(k_n^{(1/2)-\nu}),$$

whence $R_n = \mathcal{O}_P(k_n^{-(1/2)-\nu})$. Next we show that $\Theta_n(m, l) = \mathcal{O}_P(J(1/k_n)/k_n)$ for any fixed $1 \leqslant m < l$. It is well known that

$$(U_{i,n}\colon\ i = 1, \ldots, n) \stackrel{\mathcal{D}}{=} (S_i/S_{n+1}\colon\ i = 1, \ldots, n),$$

where $S_i = Y_1 + \ldots + Y_i$ for some $Y_1, Y_2, \ldots$ independent exponentially distributed random variables with mean one. Then

$$\{nG_n(v/n)\colon\ 0 \leqslant v < n\} \stackrel{\mathcal{D}}{=} \{N(vS_{n+1}/n)\colon\ 0 \leqslant v < n\},$$

where $N(\cdot)$ is a Poisson process with jump points $S_1, S_2, \ldots$ Set $f_s(v) := g(sv) - g(s)$ and notice that (1.3) implies $f_s(v) \to \log v^{-\alpha}$ as $s \downarrow 0$ for every $v > 0$ by applying Lemma 5.4 (iv) in CsV [9]. Let us define $y_J := \inf\{y\colon\ J(y) = 0\} > 0$. On the event

$$E_n := \{nG_n^*(l/n)/k_n < y_J\} \cap \{l/n < U_{n,n}\},$$

for some $\iota_n(v)$ between $nG_n^*(v/n)/k_n$ and v/k_n, for which necessarily $J(\iota_n(v)) > 0$, we have

$$|\Theta_n(m, l)| = \Big|\int_m^l \int_{v/k_n}^{nk_n^{-1}G_n^*(v/n)} J(s)\,ds\,dg(v/n)\Big| \leqslant \int_m^l \left|\frac{n}{k_n} G_n^*\left(\frac{v}{n}\right) - \frac{v}{k_n}\right| J(\iota_n(v))\,d(-g(v/n))$$
$$\leqslant J(1/k_n) \int_m^l \left|\frac{n}{k_n} G_n^*\left(\frac{v}{n}\right) - \frac{v}{k_n}\right| d(-g(v/n)),$$

where the right-hand side equals in distribution

$$\frac{J(1/k_n)}{k_n} \int_m^l |(N(vS_{n+1}/n) \vee 1) - v|\,d(-f_{1/n}(v)) = \mathcal{O}_P\left(\frac{J(1/k_n)}{k_n}\right).$$

Since $P\{E_n\} \to 1$, we indeed have $\Theta_n(m, l) = \mathcal{O}_P(J(1/k_n)/k_n)$, and similarly one can prove that $V_n(1) = \mathcal{O}_P(J(1/k_n)/k_n)$. Using an argument similar to that in the proof of Theorem 1.1 of CsV [9], we also have $V_n(k_n) = o_P(k_n^{-1/2})$ and $k_n^{1/2}(\bar{\mu}_n - \mu_n^{(W)}) = o(1)$. These steps heavily rely on conditions (i)–(v) described for the weight function J. Next observe that $t^{1/2} J(t) = o(1)$ as $t \downarrow 0$ by condition (v).

Then by a diagonal selection procedure, similar to the one used in the proof of Theorem 1 in Csörgő et al. [6], we can construct sequences $1 \leqslant m_n \leqslant l_n \leqslant k_n$ such that $m_n \to \infty$, $l_n/m_n \to \infty$, $k_n/l_n \to \infty$ and $k_n^{1/2} \Delta_n(m_n, l_n) = o_P(1)$. It follows that

$$k_n^{1/2} \tau(J)(\bar{\alpha}_n - \bar{\mu}_n) = k_n^{1/2} \Theta_n(l_n, k_n) + o_P(1),$$

and an argument as in the proof of Lemma 2 in Viharos [18] and the approximations (4.1) give

$$k_n^{1/2} \Theta_n(l_n, k_n) = N_n + o_P(1),$$

where $N_n := -(n/k_n)^{1/2} \int_{l_n/n}^{k_n/n} B_n(t) J(nt/k_n) dg(t)$ is a normal random variable with mean zero and variance

$$E(N_n^2) \to \int_0^1 \int_0^1 (u \wedge v) J(u) J(v) d\log u^{-\alpha} d\log v^{-\alpha}$$

(cf. the proof of Theorem 1.1 of CsV [9]). Thus

$$k_n^{1/2} \{\bar{\alpha}_n(k_n) - \mu_n^{(W)}(k_n)\} = (\tau(J))^{-1} N_n + o_P(1)$$

by using $k_n^{1/2}(\bar{\mu}_n - \mu_n^{(W)}) = o(1)$, which proves the first statement of the proposition.

To prove the second half of the statement write

$$\frac{\sum_{i=1}^{k_n} g_{n+1-i,n} \log(k_n/i)}{\tau(J) k_n} - 1$$

$$= -\frac{1}{\tau(J)} \sum_{i=2}^{k_n} \int_{(i-1)/k_n}^{i/k_n} J(x) \log \frac{i}{x k_n} dx - \frac{1}{\tau(J)} \int_0^{1/k_n} J(x) \log \frac{1}{x k_n} dx =: -\frac{\xi_n}{\tau(J)} - \frac{\zeta_n}{\tau(J)}$$

and note that $\sup_{k_n^{-1} \leqslant x \leqslant 1} |J(x)| = \mathcal{O}(J(1/k_n))$ and $\log x \leqslant x - 1$ for all $x > 0$. Thus

$$|\xi_n| \leqslant \mathcal{O}(J(1/k_n)) \sum_{i=2}^{k_n} \int_{(i-1)/k_n}^{i/k_n} \left(\frac{i}{x k_n} - 1\right) dx = \mathcal{O}(J(1/k_n)) \sum_{i=2}^{k_n} \left(\frac{i}{k_n} \log \frac{i}{i-1} - \frac{1}{k_n}\right)$$

$$\leqslant \mathcal{O}(J(1/k_n)) \sum_{i=2}^{k_n} \left(\frac{i}{k_n}\left(\frac{i}{i-1} - 1\right) - \frac{1}{k_n}\right) = \mathcal{O}\left(\frac{\log k_n}{k_n} J(1/k_n)\right).$$

By conditions (ii) and (v) we have

$$o(1) = \int_0^{1/k_n} x^{-(1/2)-\nu} J(x) dx \geqslant J(1/k_n) \int_0^{1/k_n} x^{-(1/2)-\nu} dx > J(1/k_n) \frac{1}{k_n^{1/2}} \log k_n$$

for all n large enough. Hence $\xi_n = o(k_n^{-1/2})$. Finally, decompose ζ_n as

$$\zeta_n = \int_0^{1/k_n} J(x) \log \frac{1}{k_n} dx - \int_0^{1/k_n} J(x) \log x dx =: \zeta_n^{(1)} - \zeta_n^{(2)}.$$

Then, for all n large enough,

$$\frac{|\zeta_n^{(1)}|}{\log k_n} = \int_0^{1/k_n} J(x)\,dx = \int_0^{1/k_n} x^{(1/2)+\nu} x^{-(1/2)-\nu} J(x)\,dx$$

$$\leqslant \left(\frac{1}{k_n}\right)^{(1/2)+\nu} \int_0^{1/k_n} x^{-(1/2)-\nu} J(x)\,dx = o\left(k_n^{-(1/2)-\nu}\right).$$

A similar argument yields that $\zeta_n^{(2)} = o(k_n^{-(1/2)-\nu} \log k_n)$, which completes the proof. ■

Proof of Theorem 1.1. Based on Proposition 4.1, it is enough to prove that $\mu_n^{(W)}(k_n) \to \alpha$. For $L(s)$ in (1.3) we use the Karamata representation for slowly varying functions:

$$L(s) = a(s) \exp\left\{\int_s^1 \frac{b(u)}{u}\,du\right\}, \quad 0 < s < 1, \tag{4.2}$$

where $\lim_{s \downarrow 0} a(s) = a_0$ for some $a_0 \in (0, \infty)$ and $\lim_{s \downarrow 0} b(s) = 0$. The first equation in (3.2) in combination with condition (i) yields

$$\tau(J)\left(\mu_n^{(W)}(k_n) - \alpha\right) = \int_0^1 J(s) \log \frac{a(sk_n/n)}{a_0}\,ds + \int_0^1 J(s) \left\{\int_{sk_n/n}^{k_n/n} \frac{b(u)}{u}\,du\right\} ds =: I_1 + I_2.$$

From the equality

$$I_1 + I_2 = \mathcal{O}\Big(\sup_{0 < t \leqslant k_n/n} \left|\log\left(a(t)/a_0\right)\right| + \sup_{0 < t \leqslant k_n/n} |b(t)|\Big)$$

the theorem follows. ■

Proof of Theorem 2.1. In order to prove the first statement, we establish

$$\tau(J)\left(\mu_n^{(W)}(k_n) - \alpha\right) \sim g(k_n/n) \int_0^1 q_\varrho(v) J(v)\,dv. \tag{4.3}$$

Set

$$\varepsilon(x) := \left(\log(1+x) - x\right)/x \quad \text{and} \quad f_n(v) := \{L(k_n v/n)/L(k_n/n)\} - 1.$$

Then, using (1.3), condition (i) and (2.1), we obtain

$$\tau(J)\left(\mu_n^{(W)}(k_n) - \alpha\right) = \int_0^1 J(v) \log\left(1 + f_n(v)\right) dv = \int_0^1 J(v) f_n(v) \left[1 + \varepsilon\left(f_n(v)\right)\right] dv$$

$$= g(k_n/n) \int_0^1 J(v) \left(q_\varrho(v) + \delta_v(k_n/n)\right) \left[1 + \varepsilon\left(f_n(v)\right)\right] dv.$$

Set $\eta_n(v) := \delta_v(k_n/n) + q_\varrho(v)\,\varepsilon\left(f_n(v)\right) + \delta_v(k_n/n)\,\varepsilon\left(f_n(v)\right)$. Then we get

$$\left|\frac{\tau(J)\left(\mu_n^{(W)}(k_n) - \alpha\right)}{g(k_n/n) \int_0^1 q_\varrho(v)\, J(v)\,dv} - 1\right| = \left|\frac{\int_0^1 \eta_n(v)\, J(v)\,dv}{\int_0^1 q_\varrho(v)\, J(v)\,dv}\right| \leqslant C \sup_{0 < t \leqslant 1} t^{(1/2)+\nu} |\eta_n(t)|$$

using condition (v) in the last step, where $C > 0$ is a constant. Noticing that $|\varepsilon(x)| \leqslant |\log(1+x)|$, $x > -1$, we see that condition (2.3) implies $\sup_{0<t\leqslant 1} t^{(1/2)+\nu}|\eta_n(t)| \to 0$, which proves (4.3). To obtain the optimal k_n replace n/k_n by t and write

$$M_n^{(W)} = M_n^{(W)}(t) = g^2(1/t)\, B_{J,\varrho} + (t\alpha^2 \sigma_J^2/n).$$

By Lemma 2.8 of Dekkers and de Haan [11], $M_n^{(W)}(t)$ is minimized by

$$t = t^* = n/k_n^* \sim h^{\leftarrow}\big(\alpha^2 \sigma_J^2/(nB_{J,\varrho})\big),$$

from which (2.4) follows. For $\varrho < 0$, by $h(\cdot) \in RV_{2\varrho-1}^{\infty}$, we have

$$g^2\left(\frac{1}{t^*}\right) \sim -\frac{h(t^*)\,t^*}{2\varrho} \sim -\frac{1}{2\varrho}\frac{\alpha^2\sigma_J^2}{nB_{J,\varrho}}\, h^{\leftarrow}\left(\frac{\alpha^2\sigma_J^2}{nB_{J,\varrho}}\right) \sim -\frac{\alpha^2\sigma_J^2}{2\varrho\, B_{J,\varrho}\, k_n^*}.$$

Hence

$$M_n^{(W)}(k_n^*) = -\frac{\alpha^2\sigma_J^2}{2\varrho\, k_n^*}\big(1+o(1)\big) + \frac{\alpha^2\sigma_J^2}{k_n^*} \sim \frac{\alpha^2\sigma_J^2}{k_n^*}\left(1-\frac{1}{2\varrho}\right).$$

Using $h^{\leftarrow}(\cdot) \in RV_{(2\varrho-1)^{-1}}^{0}$, we easily get (2.5). The optimal J_ϑ can be obtained by calculating the first and second derivatives of $\tilde{M}_n^{(W)}(k_n^*, J_\vartheta)$ in (2.5) with respect to ϑ. If $\varrho = 0$, then $B_{J,\varrho} = 1$, which implies the last statement. ∎

Proof of Theorem 2.2. The asymptotic squared bias and variance of $\hat{\alpha}_n^{(K)}$ are given by $\big(\int_0^{\Delta} K(v)\, b(k_n v/n)\, dv\big)^2$ and $\alpha^2 k_n^{-1} \int_0^{\Delta} K^2(v)\, dv$, respectively, where $b(\cdot)$ is defined in (4.2) (cf. Theorems 1 and 5 in Csörgő et al. [5]; they do not formally include the factor α^2 in their variance term). In the rest of the paper we use the integral convention $\int_0^{\Delta} = \int_{(0,\Delta]}$. For some constant $a_0 \in (0, \infty)$ we may assume $a(s) \equiv a_0$ in (4.2) as in condition (D2) (i). Hence $b(s) = -sL'(s)/L(s)$ and

$$\int_0^{\Delta} K(v)\, b(k_n v/n)\, dv = -\int_0^{\Delta} K(v)\, \frac{k_n v n^{-1} L'(k_n v/n)}{L(k_n v/n)}\, dv = -\int_0^{\Delta} vK(v)\, d\log\frac{L(k_n v/n)}{L(k_n/n)}$$
$$= \int_0^{\Delta} \log\frac{L(k_n v/n)}{L(k_n/n)}\, d\big(vK(v)\big),$$

where partial integration with the convention $\int_0^{\Delta} = \int_{(0,\Delta]}$ and Lemma 1 of Csörgő et al. [5] were used in the last step. Similarly, by partial integration,

$$\int_0^{\Delta} v^{-\varrho} K(v)\, dv = -\int_0^{\Delta} vK(v)\, dq_\varrho(v) = \int_0^{\Delta} q_\varrho(v)\, d\big(vK(v)\big).$$

Hence, as in the proof of Theorem 2.1,

$$\left|\frac{\int_0^{\Delta} K(v)\, b(k_n v/n)\, dv}{g(k_n/n)\int_0^{\Delta} v^{-\varrho} K(v)\, dv} - 1\right| = \left|\frac{\int_0^{\Delta} \log\dfrac{L(k_n v/n)}{L(k_n/n)}\, d\big(vK(v)\big)}{g(k_n/n)\int_0^{\Delta} q_\varrho(v)\, d\big(vK(v)\big)} - 1\right| = \left|\frac{\int_0^{\Delta} \eta_n(v)\big(vK(v)\big)}{\int_0^{\Delta} q_\varrho(v)\, d\big(vK(v)\big)}\right|$$

$$\leqslant \frac{\left|\int_0^{\Delta} \eta_n(v)\, K(v)\, dv\right| + \left|\int_0^{\Delta} \eta_n(v)\, v dK(v)\right|}{\left|\int_0^{\Delta} q_\varrho(v)\, d\big(vK(v)\big)\right|} =: \frac{A_{n1} + A_{n2}}{\left|\int_0^{\Delta} q_\varrho(v)\, d\big(vK(v)\big)\right|},$$

where we used the equation $d(vK(v)) = K(v)\,dv + v\,dK(v)$. Using condition (H4), we obtain $A_{n1} \leqslant C \sup_{0<t\leqslant A} t^{1/2}|\eta_n(t)|$ for some constant C. Partial integration, condition (H4) and Lemma 1 of Csörgő et al. [5] imply that the integral $\int_0^A v^{1/2}\,dK(v)$ is finite, and hence $A_{n2} \leqslant C' \sup_{0<t\leqslant A} t^{1/2}|\eta_n(t)|$ for some constant C'. As in the proof of Theorem 2.1, by condition (2.3) we have $\sup_{0<t\leqslant A} t^{1/2}|\eta_n(t)| \to 0$. The rest of the proof is the same as the proof of Theorem 2.1 above and the proof of Theorem 5 in Csörgő et al. [5] and, therefore, is omitted. ∎

Acknowledgement. I thank Professor Sándor Csörgő for his suggestions. I also thank the referee for the valuable remarks which resulted in an improved form of the paper.

REFERENCES

[1] J. Beirlant and J. L. Teugels, *Asymptotic normality of Hill's estimator*, in: *Extreme Value Theory*, Oberwolfach, 1987. Lecture Notes in Statist. 51 (1989), pp. 148–155.

[2] J. Beirlant, P. Vynckier and J. L. Teugels, *Tail index estimation, Pareto quantile plots, and regression diagnostics*, J. Amer. Statist. Assoc. 91 (1996), pp. 1659–1667.

[3] N. H. Bingham, C. M. Goldie and J. L. Teugels, *Regular Variation*, Cambridge University Press, Cambridge 1987.

[4] M. Csörgő, S. Csörgő, L. Horváth and D. M. Mason, *Weighted empirical and quantile processes*, Ann. Probab. 14 (1986), pp. 31–85.

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

[6] S. Csörgő, E. Haeusler and D. M. Mason, *A probabilistic approach to the asymptotic distribution of sums of independent, identically distributed random variables*, Adv. in Appl. Math. 9 (1988), pp. 259–333.

[7] S. Csörgő and D. M. Mason, *Central limit theorems for sums of extreme values*, Math. Proc. Cambridge Philos. Soc. 98 (1985), pp. 547–558.

[8] S. Csörgő and L. Viharos, *On the asymptotic normality of Hill's estimator*, ibidem 118 (1995), pp. 375–382.

[9] S. Csörgő and L. Viharos, *Asymptotic normality of least-squares estimators of tail indices*, Bernoulli 3 (1997), pp. 351–370.

[10] S. Csörgő and L. Viharos, *Estimating the tail index*, in: *Asymptotic Methods in Probability and Statistics*, B. Szyszkowicz (Ed.), North-Holland, Amsterdam 1998, pp. 833–881.

[11] A. L. M. Dekkers and L. de Haan, *Optimal choice of sample fraction in extreme-value estimation*, J. Multivariate Anal. 47 (1993), pp. 173–195.

[12] E. Haeusler and J. L. Teugels, *On the asymptotic normality of Hill's estimator for the exponent of regular variation*, Ann. Statist. 13 (1985), pp. 743–756.

[13] P. Hall, *On some simple estimates of an exponent of regular variation*, J. Roy. Statist. Soc. Ser. B 44 (1982), pp. 37–42.

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

[15] B. M. Hill, *A simple general approach to inference about the tail of a distribution*, ibidem 3 (1975), pp. 1163–1174.

[16] M. Kratz and S. Resnick, *The qq-estimator of the index of regular variation*, Comm. Statist. Stochastic Models 12 (1996), pp. 699–724.
[17] J. Schultze and J. Steinebach, *On least squares estimates of an exponential tail coefficient*, Statist. Decisions 14 (1996), pp. 353–372.
[18] L. Viharos, *Asymptotic distributions of linear combinations of extreme values*, Acta Sci. Math. (Szeged) 58 (1993), pp. 211–231.

Bolyai Institute
University of Szeged
Aradi vértanúk tere 1
6720 Szeged, Hungary
E-mail: viharos@math.u-szeged.hu

Received on 29.3.1998;
revised version on 10.9.1999