Extremes
DOI 10.1007/s10687-009-0091-9

Uniform in bandwidth consistency of kernel estimators of the tail index

Julia Dony · David M. Mason

Received: 9 February 2009 / Revised: 21 September 2009 / Accepted: 24 September 2009

Abstract We consider a class of kernel estimators $\hat{\tau}_{n,h}$ of the tail index of a Pareto-type distribution, which generalizes and includes the classical Hill estimator $\hat{a}_{n,k}$. It is well-known that $\hat{a}_{n,k}$ is a consistent estimator of the tail index if and only if $k \to \infty$ and $k/n \to 0$. Under suitable assumptions on the kernel, $\hat{\tau}_{n,h}$ is consistent whenever the bandwidth is taken to be a sequence of non-random numbers satisfying $h_n \to 0$ and $nh_n \to \infty$. We extend this result and prove the consistency uniformly over a certain range of bandwidths. This permits the treatment of estimators of the tail index based upon data-dependent bandwidths, which are often used in practice. In the process, we establish a uniform in bandwidth result for kernel-type regression estimators with a fixed design, which will likely be of separate interest.

Keywords Pareto-type distributions · Hill estimator · Tail index · Uniform in bandwidth consistency · Kernel-type estimators · Fixed design · Plug-in estimator

AMS 2000 Subject Classification 62G32

This paper is dedicated to the memory of Sándor Csörgő.

Julia Dony's research is financed by a Postdoctoral Fellowship of the Fund for Scientific Research, Flanders, Belgium (FWO – Vlaanderen). David Mason's research was partially supported by an NSF Grant.

J. Dony (✉)
Department of Mathematics, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
e-mail: jdony@vub.ac.be

D. M. Mason
Food and Resource Economics, University of Delaware,
206 Townsend Hall, Newark, DE 19717, USA
e-mail: davidm@udel.edu

Published online: 15 October 2009

1 Introduction

Let Y be a random variable with cumulative distribution function F, and assume that it has a regularly varying upper tail of exponent $-1/\tau < 0$, meaning that there exists a $\tau > 0$ such that,

$$\lim_{y\to\infty} \frac{1-F(\lambda y)}{1-F(y)} = \lambda^{-1/\tau}, \quad \lambda > 0. \tag{$F.\tau$}$$

Such a cumulative distribution function is called *heavy tailed*. For a fixed location y, let

$$F_y(\lambda) = 1 - (1 - F(\lambda y))/(1 - F(y)), \quad \lambda \geq 1,$$

denote the so-called "excess distribution function over the threshold y". Then if $Y_1, \ldots, Y_n$ are $n \geq 1$ independent copies of Y, the tail condition ($F.\tau$) says that if y is large enough, $F_y(\lambda)$ can be approximated by $P_\tau(\lambda) = 1 - \lambda^{-1/\tau}$. This cumulative distribution function corresponds to a Pareto-type distribution with parameter $1/\tau$. We call $1/\tau$ the *tail index*.

A well-known estimator of τ is the so-called Hill estimator, introduced by Hill (1975). It is defined for $1 \leq k \leq n$ as

$$\hat{a}_{n,k} := k^{-1} \sum_{i=1}^{k} \log(Y_{n-i+1:n} \vee 1) - \log(Y_{n-k:n} \vee 1),$$

where $Y_{i:n}$, $1 \leq i \leq n$, denote the order statistics of $Y_1, \ldots, Y_n$. Here, we set $Y_{0:n} \equiv 0$. The consistency of $\hat{a}_{n,k}$ as an estimator of τ was established by Mason (1982), who showed that Eq. ($F.\tau$) was necessary and sufficient for $\hat{a}_{n,k} \to_{\mathbb{P}} \tau$ for every sequence $k \equiv k_n \to \infty$ and $k_n/n \to 0$. Later, Deheuvels et al. (1988) characterized those sequences k_n for which almost sure consistency holds.

Csörgő et al. (1985) introduced the following generalization of the Hill estimator. This is the kernel-type estimator for τ, defined as

$$\hat{\tau}_{n,h} := \frac{\sum_{j=1}^{n} \frac{j}{nh} K\left(\frac{j}{nh}\right) \left\{\log\left(Y_{n-j+1:n} \vee 1\right) - \log\left(Y_{n-j:n} \vee 1\right)\right\}}{\frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{j}{nh}\right)} =: \frac{\phi_{n,h}}{\kappa_{n,h}}, \tag{1}$$

where K is a kernel function and $0 < h < 1$ is a bandwidth. Note that for the choices

$$K(u) = \mathbf{I}\{0 \leq u \leq 1\} \quad \text{and} \quad h = k/n,$$

the kernel estimator $\hat{\tau}_{n,h}$ corresponds to the classic Hill estimator $\hat{a}_{n,k}$. In practice, one has to choose a bandwidth sequence h_n in such a way that the bias and the variance term of the estimator are reasonably balanced. Indeed, just as in the case of kernel-type estimators in general, $\hat{\tau}_{n,h}$ suffers from an unavoidable bias-variance trade-off.

The advantage of considering a kernel-type version of $\hat{a}_{n,k}$ is that it permits the minimization of a quantity that asymptotically acts like expected mean squared error by choosing an appropriate bandwidth and kernel function. In Csörgő et al. (1985), such optimal choices for h and K are determined for cumulative distributions F having a tail behavior of the form

$$1 - F(x) = C_1 x^{-1/\tau} \left[1 + C_2 x^{-\beta}(1 + o(1))\right], \quad \text{as } x \to \infty, \tag{2}$$

where $C_1 > 0$, $C_2 \neq 0$ and τ and β are positive constants. (We mention that this tail condition is also known to correspond to the so-called "Hall and Welsh model"; see Hall and Welsh 1985). More details about these choices will be given in the next section. The results in Csörgő et al. (1985) show, in particular, that for such cumulative distributions, the Hill estimator is far from being optimal.

Typically, one turns to optimal bandwidth sequences that depend on some quantity involving the unknown cumulative distribution. The replacement of this quantity by a suitable estimator results in a bandwidth sequence depending on the data. This means that one can no longer investigate the behavior of such estimators via the available results for estimators based on deterministic bandwidth sequences. Instead, one requires results that permit the study of the asymptotic properties of such optimal estimators based upon data-dependent bandwidth sequences.

To this end, we shall be interested in the uniform in bandwidth consistency of $\hat{\tau}_{n,h}$. This means that we shall be studying the consistency of $\hat{\tau}_{n,h}$ that holds on a wide range of bandwidths h, simultaneously. The main motivation to consider such an extended form of consistency is that it is immediately applicable to the treatment of estimators based on data-dependent bandwidth sequences, and hence to plug-in estimators of the tail index. Such uniform in bandwidth results have recently been obtained for a large variety of kernel-type estimators (see Dony 2008 for a detailed discussion of some of such results). Nolan and Marron (1989) first introduced the notion of uniform in bandwidth consistency for kernel density estimators. It was extended and expanded to other kernel-type nonparametric function estimators in Einmahl and Mason (2005), whose methods lead to the derivation of the precise order of convergence as well.

The major goal of this paper is to establish the uniform in bandwidth consistency of the kernel-based tail estimator $\hat{\tau}_{n,h}$ of τ. For our main result we shall assume the following conditions on K.

(K.i) K has support contained in $[0, M]$ for some $M > 0$ and is bounded by κ,
(K.ii) K is non-increasing and non-negative on $[0, M]$ and $\int_0^M K(u)du = 1$,
(K.iii) K satisfies a Lipschitz condition of order $0 < \vartheta \leq 1$ on $[0, M]$.

Here is our main result.

Theorem 1 *Let Y be a random variable satisfying the tail condition ($F.\tau$) and for each $n \geq 1$, let $Y_1, \ldots, Y_n$ be i.i.d. Y. Assume that K satisfies the conditions (K.i)–(K.iii). Then for any choice of sequences a_n and b_n such that $0 < a_n < b_n \leq 1/2$ and*

$$\frac{\log n}{\sqrt{n a_n}} \to 0 \quad \textit{and} \quad b_n \to 0, \tag{3}$$

we have, as $n \to \infty$,

$$\sup_{a_n \le h \le b_n} |\hat{\tau}_{n,h} - \tau| \xrightarrow{\mathbb{P}} 0. \tag{4}$$

If one is not interested in a uniform in bandwidths result but only in consistency for a fixed sequence of bandwidths h_n, then one only requires $h_n \to 0$ and $nh_n \to \infty$. Refer to Theorem 1 of Csörgő et al. (1985). The kernel function $K(u) = \mathbf{I}\{0 \le u \le 1\}$ satisfies all the above conditions. This means that the Hill estimator $\hat{a}_{n,k}$ satisfies all the required conditions to apply Theorem 1, as long as $k = k_n \to \infty$ is such that $k_n/n \to 0$ and $\log n/\sqrt{k_n} \to 0$.

In a recent paper, Grama and Spokoiny (2008) considered the related problem of estimating the excess distribution $F_y(\lambda)$. Their method is based on the idea that for any fixed value of $y \ge y_0$, F_y can be approximated by a Pareto-type distribution with parameter τ_y, which may depend on the location y (and which is typically different from τ). The estimation of τ_y is then based on a step-wise lack-of-fit test for the Pareto-type distribution. In contrast to many of the proposed procedures (see for instance the monograph by Beirlant et al. 2004 for a survey of such procedures), their method does not depend on the unknown parameter τ, and therefore does not require a prior estimator for τ. To determine the location of the tail y directly from the data, they consider a countable set of locations $\mathcal{L}$, and the family of estimators

$$\left\{ \hat{a}_{n,k(y)} = \frac{1}{k(y)} \sum_{i=1}^{k(y)} \log(Y_{n-i+1:n} \vee 1) - \log(Y_{n-k(y):n} \vee 1)\ ;\ y \in \mathcal{L}, k(y) > 0 \right\},$$

where $k(y) = \sum_{i=1}^{n} \mathbf{I}\{Y_i > y\}$. They propose an *adaptive* method for determining the best location threshold $y \in \mathcal{L}$ (or equivalently, the best number k of upper order statistics that should be taken into account in the Hill estimator $\hat{a}_{n,k}$). This method is based on an oracle inequality for the Kullback–Leibler loss.

The proof of Theorem 1 is based on empirical process methods, and is detailed in Section 4. In Section 3, we develop some tools needed to prove our main result. This includes a uniform in bandwidth consistency result for kernel-type regression estimators with a fixed design, which will likely be of separate interest. In Section 2, we discuss how our results can be applied to establish the consistency of plug-in estimators of the tail index.

2 Application to plug-in estimators

Define for $u \in\ ]0, 1[$ the inverse distribution $F^{-1}(u) = \inf\{x : F(x) \ge u\}$ of F. It turns out that Eq. ($F.\tau$) holds if and only if

$$F^{-1}\left(1 - y^{-1}\right) = y^{\tau} c(y) \exp\left(\int_1^y \frac{b(t)}{t} dt\right), \quad y > 1, \tag{$I.\tau$}$$

where $c(y) \to c_0 > 0$ and $b(y) \to 0$ as $y \to \infty$. In their 1985 paper, Csörgő et al. computed optimal choices for h and K in the case the cumulative distribution F has a tail behavior of the form Eq. 2. Their optimality criterion is based on minimizing the asymptotic expected mean squared error of $\hat{\tau}_{n,h}$, which they showed to be given by

$$M(n,h,K) = \frac{1}{nh}\int_0^\infty K^2(u)du + \left\{\int_0^{1/h} b(uh)K(u)du\right\}^2,$$

with $b(u)$ the function as in Eq. ($I.\tau$). Set $D_2 = \tau C_2/C_1^\rho$ with $\rho = \tau\beta$. Under a suitable set of conditions regarding the kernel and the behavior of the quantile function (and the function $b(u)$ in Eq. ($I.\tau$) in particular), they showed that for a fixed kernel K, the choice

$$\hat{h}_n = \left\{\frac{1}{n}\frac{\|K\|_2^2}{\left(2\rho^3 D_2^2\right)}\left(\int_0^\infty u^\rho K(u)du\right)^{-2}\right\}^{1/(2\rho+1)} \tag{5}$$

ensures that $M(n,\hat{h}_n,K) \sim \inf_{0<h<\Lambda_1} M(n,h,K)$ for some $\Lambda_1 > 0$ which depends on F only (see their Theorem 5 on p. 1055). This implies readily that

$$M\left(n,\hat{h}_n,K\right) \sim \left(1+\frac{1}{2\rho}\right)\left\{\frac{\left(2\rho^3 D_2^2\right)\|K\|_2^{4\rho}}{n^{2\rho}}\left(\int_0^\infty u^\rho K(u)du\right)^2\right\}^{1/(2\rho+1)},$$

and thus $M(n,\hat{h}_n,K) = O\left(n^{-2\rho/(2\rho+1)}\right)$, where the constant in the O-term only depends on K and ρ. The kernel function that minimizes $M(n,\hat{h}_n,K)$ turns out to be

$$\hat{K}(u) = \left(\frac{\rho+1}{\rho}\right)\left(\frac{2\rho+1}{2\rho+2}\right)^{\rho+1}\left[\left(\frac{2\rho+2}{2\rho+1}\right)^\rho - u^\rho\right]\mathbf{I}\left\{u \in \left]0,\frac{2\rho+2}{2\rho+1}\right[\right\}.$$

Hence, the combination of $\hat{K}(u)$ and $\hat{h}_n$ yields a minimal expected mean squared error satisfying

$$M\left(n,\hat{h}_n,\hat{K}\right) \sim \inf_{0<h<\Lambda_1}\left\{M(n,h,K) : \int_0^\infty K(u)du = \|K\|_2^2 = 1\right\},$$

and which is of order $O\left(n^{-2\rho/(2\rho+1)}\right)$.

Notice that the optimal kernel $\hat{K}(u)$ depends on the unknown value of $\rho = \tau\beta$. This makes it difficult to use in practice. Of course, one could replace ρ with a consistent estimator $\tilde{\rho}$. However the estimation of ρ is known to be difficult, and in practice, one usually sets $\rho = 1$ to avoid the estimation step.

It is readily verified that the $\hat{h}_n$ in Eq. 5 fulfills all the conditions in Eq. 3. Furthermore, $\hat{K}$ clearly satisfies (K.i) and (K.ii). However, whether the Lipschitz condition in (K.iii) is satisfied or not, depends on the value of $\rho = \tau\beta$. When $\rho = 1$, the optimal kernel equals $\hat{K}(u) = \frac{9}{8}\left(\frac{4}{3} - u\right)$, which is Lipschitz of order $0 < \vartheta \leq 1$ for

u in the interval $[0, \frac{4}{3}]$, and thus satisfies all the conditions in (K.i)–(K.iii). Write $\hat{h}_n^*$ for the optimal bandwidth sequence $\hat{h}_n$ computed with the kernel $\hat{K}(u)$, and note that $\|\hat{K}\|_2^2 = 1$ and $\int u\hat{K}(u)du = \frac{4}{9}$. Then

$$\hat{h}_n^* = \left(\frac{1}{n}\frac{81}{32D_2^2}\right)^{1/3} = Cn^{-1/3}\tau^{-2/3}.$$

Thus the optimal bandwidth $\hat{h}_n^*$ still depends upon the unknown parameter τ. In practice, this parameter is replaced by an initial estimator, say $\tilde{\tau}$, which is plugged into $\hat{h}_n^*$, yielding a data-dependent bandwidth $\tilde{H}_n^*$ and resulting in a *plug-in estimator* $\hat{\tau}_{n,\tilde{H}_n^*}$ of τ. Theorem 1 ensures that the optimal tail index estimator $\hat{\tau}_{n,\hat{h}_n^*}$ is consistent for τ. Moreover, if the initial estimator $\tilde{\tau}$ is such that $\tilde{H}_n^*$ still satisfies the conditions in Eq. 3, Theorem 1 also implies the consistency of the corresponding plug-in estimator for τ.

It should be noted that the optimal choice of h, as just described, does not lead to a centered limiting normal distribution for $\sqrt{nh}(\hat{\tau}_{n,h}^{-1} - \tau^{-1})$. This is only attained when $\hat{h}_n = o(n^{-1/(2\rho+1)}) = o(n^{-1/(2\tau\beta+1)})$ (see Theorem 2 and Remarks 3 and 4 of Csörgő et al. 1985). Analogous conclusions are obtained by Haeusler and Teugels (1985) in their derivation of a central limit theorem for the Hill estimator $\hat{a}_{n,k}$. (See their examples.) For a related central limit theorem for the Hill estimator, where the centering is random, see Csörgő and Mason (1985).

Chapter 4 of the monograph by Beirlant et al. (2004) is devoted to tail estimation under Pareto-type models. It discusses the asymptotic properties of both the Hill estimator and the class of kernel estimators $\hat{\tau}_{n,h}$. In their Section 4.7 they treat the problem of choosing the optimal fraction needed to apply a tail index estimator. For more proposed solutions to this problem consult Section 7 of Groeneboom et al. (2003) and the references therein.

Finally, we should mention that Caeiro et al. (2005) investigated the asymptotic behavior of a "rescaled" Hill estimator (denoted by $\bar{\bar{H}}_{\beta,\rho}(k)$ with $k \equiv k_n \to \infty$ at a certain rate), where the scale is a function of the second order parameters β and ρ (related with but different from the parameters β and ρ we define here). They showed that by plugging appropriately chosen consistent estimators for β and ρ into this modified estimator, the bias of the resulting estimator can be significantly reduced. In particular, they establish that $\sqrt{k}(\bar{\bar{H}}_{\hat{\beta},\hat{\rho}}(k) - \tau)$ is asymptotically normal with variance τ^2. The preliminary estimators $\hat{\beta}$ and $\hat{\rho}$ that appear in their bias-corrected Hill estimator $\bar{\bar{H}}_{\hat{\beta},\hat{\rho}}(k)$, are those that were proposed and studied earlier by Fraga Alves et al. (2003). These are based on statistics of the form

$$M_n^{(\alpha)}(k) := \frac{1}{k}\sum_{i=1}^{k}(\log Y_{n-i+1:n} - \log Y_{n-i:n})^{\alpha},$$

for appropriate choices of $k \equiv k_n \to \infty$ and where $\alpha > 0$. (Notice that the choice $\alpha = 1$ corresponds to the classical Hill estimator, i.e. $M_n^{(1)}(k) = \hat{a}_{n,k}$.) Under some regularity conditions on the tail of the distribution function F, they establish that their estimators are consistent and asymptotically normal.

Moreover, their consistent estimators for ρ and β also yield bias-reduced estimators in several other situations. We mention briefly one application considered in Gomes and Pestana (2007), which concerns the estimation of extreme quantiles (also commonly referred to as being the value at risk, denoted VaR_p). They propose an alternative to the classical quantile estimator, and derive an improved estimator in the sense of reducing the asymptotic mean squared error. The main idea behind their technique is based on the fact that the classical quantile estimator is a function of the tail index of a distribution function satisfying the tail conditions in Eqs. ($F.\tau$) and 2. Hence, by replacing the Hill estimator $\hat{a}_{n,k}$ (which is typically used for estimating the tail index τ) by the aforementioned bias-corrected estimator $\bar{\bar{H}}_{\hat{\beta},\hat{\rho}}(k)$, they show that the resulting quantile estimator becomes asymptotically unbiased, though the variance remains unchanged.

Our Theorem 1 can be generalized to all of these plug-in estimators. We shall not go into the details of showing this here.

3 A UiB consistency result for fixed design regression

In this section we shall establish a uniform in bandwidth consistency result for kernel-type regression estimators with a fixed design, which is needed to prove Theorem 1. In order to state this result, we require a number of facts from empirical process theory. To do this we must introduce some notation, definitions and working assumptions.

Let $\mathcal{G}$ be a class of real valued measurable functions defined on a probability space $(\mathcal{X}, \mathcal{A}, P)$. We shall assume that $\mathcal{G}$ satisfies the following measurability condition:

(F) The class $\mathcal{G}$ is pointwise measurable.

This means that there exists a countable subclass $\mathcal{G}_\infty$ of $\mathcal{G}$ such that we can find for any function $g \in \mathcal{G}$ a sequence of functions $\{g_m, m \geq 1\}$ in $\mathcal{G}_\infty$ for which $\lim_{m\to\infty} g_m(x) = g(x)$ for all $x \in \mathcal{X}$. This requirement is imposed to avoid using outer probability measures in our statements (see Example 2.3.4 in van der Vaart and Wellner 1996). For any class of functions $\mathcal{G}$ and $\psi \in \ell_\infty(\mathcal{G})$ (the class of bounded functions on $\mathcal{G}$) we shall use the notation

$$\|\psi\|_{\mathcal{G}} = \sup_{g\in\mathcal{G}} |\psi(g)|. \tag{6}$$

By an *envelope function* for the class we mean a measurable function $G : \mathcal{X} \to [0, \infty]$, such that

$$G(y) \geq \sup_{g\in\mathcal{G}} |g(y)|, \quad y \in \mathcal{X}, \tag{7}$$

and we shall assume that $\mathbb{E}G^2(Z_1) < \infty$.

We shall be applying the following moment inequalities for

$$\mathbb{E}\left\|\alpha_n(g)\right\|_{\mathcal{G}} \quad \text{and} \quad \mathbb{E}\left\|\alpha_n(g \cdot \mathbf{I}_D)\right\|_{\mathcal{G}}, \quad D \in \mathcal{A},$$

where $\alpha_n(g)$ is the empirical process indexed by $\mathcal{G}$, based on $n \geq 1$ i.i.d. random variables $Z_1, \ldots, Z_n$ taking values in a measurable space $(\mathcal{X}, \mathcal{A})$, and each $g \in \mathcal{G}$ is a measurable function from $\mathcal{X} \to \mathbb{R}$. We have namely

$$\alpha_n(g) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{g(Z_i) - \mathbb{E}g(Z_i)\}, \quad g \in \mathcal{G}. \tag{8}$$

(The empirical process $\alpha_n(g \cdot \mathbf{I}_D)$ is defined analogously.)

Theorem 2 (Dony and Einmahl 2010) *Let $\mathcal{G}$ be a measurable function class satisfying the above assumptions. Assume further that for any sequence of i.i.d. $\mathcal{X}$-valued random variables $Z_1, Z_2, \ldots$ it holds that*

$$\mathbb{E}\left\|\sum_{i=1}^{n}\big(g(Z_i) - \mathbb{E}g(Z_1)\big)\right\|_{\mathcal{G}} \leq A\sqrt{n}\|G(Z_1)\|_2, \quad n \geq 1, \tag{9}$$

where $A > 0$ is a constant depending on $\mathcal{G}$ only. Then we have for any $D \in \mathcal{A}$,

$$\mathbb{E}\left\|\sqrt{n}\alpha_n(g \cdot \mathbf{I}_D)\right\|_{\mathcal{G}} \leq 2A\sqrt{n}\|G(Z_1)\mathbf{I}_D(Z_1)\|_2, \quad n \geq 1.$$

In order to apply this moment bound, we need the following reformulation of Theorem 6.2.2 of Dony (2008). The proof is identical. One only has to consider the case when $\|G(Z_1)\|_2$ is finite and set $\beta = \|G(Z_1)\|_2$ in the proof. Also see Theorem 3.2 of Dony and Einmahl (2010).

Before we can state this result, we must first introduce another definition. We call a class of functions $\mathcal{G}$ with envelope function $G : \mathcal{X} \to [0, \infty]$ a *VC-class*, if $\mathcal{N}(\epsilon, \mathcal{G}) \leq C\epsilon^{-\nu}$, $0 < \epsilon < 1$ for some constants $C, \nu > 0$. As usual we define

$$\mathcal{N}(\epsilon, \mathcal{G}) = \sup_{Q} \mathcal{N}(\epsilon\sqrt{Q(G^2)}, \mathcal{G}, d_Q),$$

where the supremum is taken over all probability measures Q on $(\mathcal{X}, \mathcal{A})$ with $Q(G^2) < \infty$. Here, d_Q is the $L_2(Q)$-metric and $\mathcal{N}(\epsilon, \mathcal{G}, d)$ is the minimal number of d-balls with radius ϵ which are needed to cover the function class $\mathcal{G}$.

Theorem 3 (Theorem 6.2.2 of Dony 2008) *Let $\mathcal{G}$ be a pointwise measurable VC-class with envelope function G such that $\mathcal{N}(\epsilon, \mathcal{G}) \leq C\epsilon^{-\nu}$, $0 < \epsilon < 1$, for constants $C > 1$ and $\nu > 1$. Then for any sequence of i.i.d. $\mathcal{X}$-valued random variables $Z_1, Z_2, \ldots$ the inequality in Eq. 9 holds.*

In particular, Theorem 3 implies that VC-classes satisfy Eq. 9. We shall be applying these results to the following class $\mathcal{G}$ of functions: let φ be a measurable function on $\mathbb{R}$ and K be a function of bounded variation on $\mathbb{R}$ with support contained in $[-M, M]$ for some $M > 0$. Suppose further that K is bounded by $\kappa > 0$. Fix $x_0 \in \mathbb{R}$. The class of measurable functions

$$\mathcal{G} = \{(x, y) \mapsto \varphi(y) K((x_0 - x)/h) : 0 < h \leq 1\},$$

has an envelope function $G(x, y) = \kappa\varphi(y)$ and satisfies for some $\nu > 1$ and $C > 1$,

$$\mathcal{N}(\epsilon, \mathcal{G}) \leq C\epsilon^{-\nu}, \quad 0 < \epsilon < 1.$$

(See for instance Lemma 4.3.9 in Dony 2008 for a comparable result.) Let now (X, Y) be a pair of random variables such that X and Y are *independent* and X has a marginal density f_X that satisfies for some $0 < \delta \leq M$ and $0 < \tilde{F} < \infty$,

$$f_X(x) \leq \tilde{F}, \quad \forall x \in [x_0 - \delta, x_0 + \delta]. \tag{10}$$

For any $0 < h \leq 1$, consider the class

$$\mathcal{G}_h = \{(x, y) \mapsto \varphi(y) K((x_0 - x)/\gamma) : 0 < \gamma \leq h\}.$$

Set further $D_h = \{(x, y) : x \in [x_0 - hM, x_0 + hM]\} =: E_h \times \mathbb{R}$. Clearly, if $g \in \mathcal{G}_h$ and $g(x, y) \neq 0$, then $(x, y) \in D_h$. Hence, $g(x, y) = g(x, y)\mathbf{I}_{D_h}(x, y)$ for all (x, y), and thus

$$\mathcal{G}_h \subseteq \left\{g \cdot \mathbf{I}_{D_h} : g \in \mathcal{G}\right\}.$$

Following the same argument as in Lemma 8.4.2 of Dony (2008) (see also Section 5.1 of Dony and Mason 2008), we can show that $\{g \cdot \mathbf{I}_{D_h} : g \in \mathcal{G}\}$ is a VC class of functions with values of $C > 0$ and $\nu > 1$ independent of n and h. Thus, Theorems 2 and 3 give us that

$$\mathbb{E}\left\|\sqrt{n}\alpha_n(g)\right\|_{\mathcal{G}_h} \leq \mathbb{E}\left\|\sqrt{n}\alpha_n(g \cdot \mathbf{I}_{D_h})\right\|_{\mathcal{G}} \leq 2A\sqrt{n}\|G(X, Y)\mathbf{I}_{D_h}(X, Y)\|_2.$$

Now, since X and Y are independent,

$$\|G(X, Y)\mathbf{I}_{D_h}(X, Y)\|_2 \leq \kappa\|\varphi(Y)\|_2 \left\|\mathbf{I}_{E_h}(X)\right\|_2,$$

and whenever $h \leq \delta/M$, it follows that

$$\|\mathbf{I}_{E_h}(X)\|_2^2 = \mathbb{P}\{X \in [x_0 - hM, x_0 + hM]\} \leq 2M\tilde{F}h.$$

Therefore, for some $A_1 \geq 2\sqrt{2}A\kappa\|\varphi(Y)\|_2\sqrt{M\tilde{F}} > 0$ and for all $0 < h \leq \delta/M$,

$$\mathbb{E}\left\|\sqrt{n}\alpha_n(g)\right\|_{\mathcal{G}_h} \leq A_1\sqrt{nh}, \quad \text{for all } n \geq 1. \tag{11}$$

We shall now derive our uniform in bandwidth result for a general class of regression estimators with a *fixed* design. Here is our basic setup. Let H be a measurable function defined on $[0,\infty[$ with support contained in $[0,M]$ for some $M>0$, and suppose that H satisfies a Lipschitz condition of order $0<\vartheta\le 1$ on $[0,M]$. Assume further that H is bounded by ρ. (Thus H satisfies (K.i) and (K.iii) with $K=H$ and $\kappa=\rho$.) We shall be studying estimators of the form

$$\hat{\xi}_{n,h} := \frac{1}{nh}\sum_{i=1}^{n}\varphi(Y_i)H\left(\frac{i}{(n+1)h}\right).$$

To do so, let (U,Y) be a pair of random variables such that U and Y are independent and U is a uniform$(0,1)$ random variable. For each $n\ge 1$, let $(U_1,Y_1),\ldots,(U_n,Y_n)$ be independent copies of (U,Y), and let $U_{1:n}\le\cdots\le U_{n:n}<U_{n+1:n}:=1$ denote the order statistics of $U_1,\ldots,U_n$. Next, let φ be a measurable function such that $\mathbb{E}|\varphi(Y)|<\infty$, and define

$$\Phi_n := \max_{1\le i\le n}|\varphi(Y_i)|. \tag{12}$$

Notice that in this setup, we can take $\delta=M$ and $x_0=0$ in Eq. 10. Here is our uniform in bandwidth result for the fixed design estimators $\hat{\xi}_{n,h}$.

Theorem 4 *Let H be a measurable function defined on $[0,\infty[$ with support contained in $[0,M]$ for some $M>0$, and suppose that H satisfies a Lipschitz condition of order $0<\vartheta\le 1$ on $[0,M]$. Assume further that H is bounded by $\rho>1$, and let φ be a measurable function such that $\mathbb{E}|\varphi(Y)|<\infty$. If in addition H is non-increasing and $\mathbb{E}\varphi^2(Y)<\infty$, it follows for any choice of a sequence $0<a_n\le 1/2$ which satisfies with Φ_n as in Eq. 12 the conditions*

$$a_n\to 0,\quad na_n\to\infty\quad\textit{and}\quad \mathbb{E}\Phi_n/\sqrt{na_n}\to 0, \tag{13}$$

that

$$\mathbb{E}\left[\sup_{a_n\le h\le\frac{1}{2}}\left|\hat{\beta}_{n,h}-\mathbb{E}\varphi(Y)\int_0^{1/h}H(u)du\right|\right] = O\left(\frac{1}{(na_n)^{\vartheta/2}}\right)+O\left(\frac{\mathbb{E}\Phi_n}{\sqrt{na_n}}\right).$$

Theorem 4 will be a consequence of two results, each of which will treat one of the processes defined in Eq. 14 below. To obtain this decomposition, note that, as a process in $h>0$,

$$\sum_{i=1}^{n}\varphi(Y_i)H(U_{i:n}/h)\stackrel{d}{=}\sum_{i=1}^{n}\varphi(Y_i)H(U_i/h),$$

where here and elsewhere $\stackrel{d}{=}$ denotes *equal in distribution*, and that by independence,

$$\frac{1}{nh}\sum_{i=1}^{n}\mathbb{E}[\varphi(Y_i)H(U_{i:n}/h)] = \mathbb{E}\varphi(Y)\int_0^{1/h}H(u)du,$$

so that for any $n \geq 1$ and $0 < h \leq 1$,

$$\left|\hat{\xi}_{n,h} - \mathbb{E}\varphi(Y)\int_0^{1/h} H(u)du\right| \leq \left|\frac{1}{nh}\sum_{i=1}^{n}\varphi(Y_i)H\left(\frac{U_{i:n}}{h}\right) - \mathbb{E}\varphi(Y)\int_0^{1/h} H(u)du\right|$$
$$+\left|\frac{1}{nh}\sum_{i=1}^{n}\varphi(Y_i)\left\{H\left(\frac{U_{i:n}}{h}\right) - H\left(\frac{i}{(n+1)h}\right)\right\}\right|$$
$$=: \left|\frac{1}{nh}\mathcal{S}_n(h)\right| + \left|\frac{1}{nh}\mathcal{T}_n(h)\right|. \tag{14}$$

The following proposition says that the first process in Eq. 14 converges to zero in mean, uniformly in $a_n \leq h \leq b_0 := \frac{1}{2}(1 \wedge \delta/M) = \frac{1}{2}$. (Recall that we assume that $\delta = M$ in Eq. 10.)

Proposition 1 *Under the above assumptions, we have for any sequence* $0 < a_n < 1/2$ *that,*

$$\mathbb{E}\left[\sup_{a_n \leq h \leq 1/2}\left|\frac{1}{nh}\mathcal{S}_n(h)\right|\right] = O\left(\frac{1}{\sqrt{na_n}}\right), \quad n \geq 1. \tag{15}$$

Consequently, for any sequence a_n *such that* $a_n \to 0$ *and* $na_n \to \infty$,

$$\mathbb{E}\left[\sup_{a_n \leq h \leq 1/2}\left|\frac{1}{nh}\mathcal{S}_n(h)\right|\right] \longrightarrow 0, \quad \textit{as } n \to \infty. \tag{16}$$

Proof Define for $0 < \gamma \leq 1$ and $0 < h \leq 1$ the related process

$$\mathcal{S}_n(h,\gamma) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\varphi(Y_i)H\left(\frac{U_i}{\gamma h}\right) - \mathbb{E}\left[\varphi(Y)H\left(\frac{U}{\gamma h}\right)\right]\right\},$$

and set $\mathcal{S}_n(h, 0) = 0$. Note that $\mathcal{S}_n(h) = \sqrt{n}\mathcal{S}_n(h, 1)$. Furthermore, the assumptions on H permit us to apply Eq. 11 to get that for $0 < h \leq \delta/M = 1$,

$$\mathbb{E}\left[\sup_{0\leq\gamma\leq 1}\left|\sqrt{n}\mathcal{S}_n(h,\gamma)\right|\right] \leq A_1\sqrt{nh}, \quad n \geq 1. \tag{17}$$

Let $k_n = \min\left\{k \geq 1 : 2^k a_n \geq \frac{1}{2}\right\}$. Clearly, $\left[a_n, \frac{1}{2}\right] \subseteq \left[2^0 a_n, 2^{k_n} a_n\right]$ and

$$\sup_{a_n \leq h \leq 1/2}\left|\frac{1}{nh}\mathcal{S}_n(h)\right| = \sup_{a_n \leq h \leq 1/2}\frac{|\sqrt{n}\mathcal{S}_n(h,1)|}{nh}$$
$$\leq \max_{1\leq k\leq k_n}\sup_{2^{k-1}a_n \leq h \leq 2^k a_n}\frac{|\sqrt{n}\mathcal{S}_n(h,1)|}{nh}$$
$$\leq \max_{1\leq k\leq k_n}\sup_{\frac{1}{2}\leq\gamma\leq 1}\frac{|\sqrt{n}\mathcal{S}_n(2^k a_n,\gamma)|}{n\gamma 2^k a_n},$$

where we used the fact that $\mathcal{S}_n\big(2^k\gamma a_n, 1\big) = \mathcal{S}_n\big(2^k a_n, \gamma\big)$. From this last inequality and Eq. 17, we get for some constants $B_1, B_2 > 0$,

$$\mathbb{E}\left[\sup_{a_n \le h \le 1/2}\left|\frac{1}{nh}\mathcal{S}_n(h)\right|\right] \le \sum_{k=1}^{k_n}\mathbb{E}\left[\sup_{\frac{1}{2}\le\gamma\le 1}\frac{|\sqrt{n}\mathcal{S}_n(2^k a_n,\gamma)|}{n\gamma 2^k a_n}\right] \le \sum_{k=1}^{k_n}\frac{B_1}{\sqrt{n2^k a_n}}$$
$$\le \frac{B_1}{\sqrt{na_n}}\sum_{k=1}^{\infty}2^{-k/2} =: \frac{B_2}{\sqrt{na_n}}.$$

□

Proposition 2 *Under the aforementioned assumptions of Theorem* 4*, we have for any choice of a sequence a_n which satisfies Eq.* 13*, that*

$$\mathbb{E}\left[\sup_{a_n\le h\le 1/2}\left|\frac{1}{nh}\mathcal{T}_n(h)\right|\right] = O\left(\frac{1}{(na_n)^{\vartheta/2}}\right) + O\left(\frac{\mathbb{E}\Phi_n}{\sqrt{na_n}}\right),$$

where $0 < \vartheta \le 1$ is the order of the Lipschitz condition on $H(\cdot)$.

Proof In our proof, for notational convenience, we shall assume that $M = 1$. Recall that H has support in [0, 1] and is bounded by ρ. Set further

$$G_n(h) := \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\{U_i \le h\}, \quad h > 0.$$

Then for each $0 < h < 1$, it holds that $H(i/(n+1)h) = 0$ whenever $i > (n+1)h$, and that $H(U_{i:n}/h) = 0$ for all $i > nG_n(h)$. Hence, for each $0 < h < 1$ we can write

$$\left|\frac{1}{nh}\mathcal{T}_n(h)\right| \le \frac{1}{nh}\sum_{i=1}^{nG_n(h)\wedge\lfloor(n+1)h\rfloor}|\varphi(Y_i)|\left|H\left(\frac{i}{(n+1)h}\right) - H\left(\frac{U_{i:n}}{h}\right)\right|$$
$$+\frac{2\rho\Phi_n|nG_n(h) - \lfloor(n+1)h\rfloor|}{nh}$$
$$=: \Delta_{n,h}^{(1)} + \Delta_{n,h}^{(2)},$$

where $\lfloor x\rfloor$ denotes the integer part of x. First, we claim that

$$\mathbb{E}\left[\sup_{a_n\le h\le\frac{1}{2}}\Delta_{n,h}^{(2)}\right] = O\left(\frac{\mathbb{E}\Phi_n}{\sqrt{na_n}}\right) = o(1). \tag{18}$$

Clearly, by the independence of $Y_1, \ldots, Y_n$ and $U_1, \ldots, U_n$,

$$\mathbb{E}\left[\sup_{a_n\le h\le\frac{1}{2}}\Delta_{n,h}^{(2)}\right] \le 2\rho\mathbb{E}\Phi_n\,\mathbb{E}\left[\sup_{a_n\le h\le\frac{1}{2}}\frac{|nG_n(h) - \lfloor(n+1)h\rfloor|}{nh}\right]. \tag{19}$$

Firstly, note that $\mathbb{E}\Phi_n = o(\sqrt{na_n})$ by the assumption in Eq. 13. Note further that

$$(nh)^{-1}|nG_n(h) - \lfloor (n+1)h \rfloor| \le h^{-1}|G_n(h) - h| + 1/n,$$

and using Doob's inequality, one can show that

$$\mathbb{E}\left[\left(\sup_{a_n \le h \le \frac{1}{2}} \frac{|G_n(h) - h|}{h}\right)^2\right] \le 4\mathbb{E}\left[\frac{(G_n(a_n) - a_n)^2}{a_n^2}\right] = \frac{4(1-a_n)}{na_n}.$$

Hence,

$$\mathbb{E}\left[\sup_{a_n \le h \le \frac{1}{2}} \frac{|G_n(h) - h|}{h} + \frac{1}{n}\right] \le \frac{2}{\sqrt{na_n}},$$

which by Eq. 19 and the assumption Eq. 13 gives Eq. 18. To handle $\Delta^{(1)}_{n,h}$, let $k_n = \min\left\{k \ge 1 : 2^k a_n \ge 1/2\right\}$ and choose any $h \in \left[2^{k-1}a_n, 2^k a_n\right]$, $1 \le k \le k_n$. Since H is Lipschitz of order $0 < \vartheta \le 1$ on $[0, 1]$, we have for some $C_1 > 0$ that for every $1 \le k \le k_n$ and all $h \in \left[2^{k-1}a_n, 2^k a_n\right]$,

$$\begin{aligned}
\Delta^{(1)}_{n,h} &\le \frac{C_1}{nh} \sum_{i=1}^{nG_n(h)\wedge\lfloor (n+1)h\rfloor} |\varphi(Y_i)| \left|\frac{i}{(n+1)h} - \frac{U_{i:n}}{h}\right|^{\vartheta} \\
&\le \frac{C_1}{nh^{\vartheta+1}} \sum_{i=1}^{\lfloor (n+1)h\rfloor} |\varphi(Y_i)| \left|\frac{i}{n+1} - U_{i:n}\right|^{\vartheta} \\
&\le \frac{C_1}{n(2^{k-1}a_n)^{\vartheta+1}} \sum_{i=1}^{\lfloor 2^k a_n(n+1)\rfloor} |\varphi(Y_i)| \left|\frac{i}{n+1} - U_{i:n}\right|^{\vartheta}.
\end{aligned}$$

Now, since $\mathbb{E}U_{i:n} = i/(n+1)$,

$$\mathbb{E}\left|\frac{i}{(n+1)} - U_{i:n}\right|^{\vartheta} \le (\mathrm{Var}(U_{i:n}))^{\vartheta/2},$$

and it is readily verified that

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{\lfloor 2^k a_n(n+1)\rfloor} (\mathrm{Var}(U_{i:n}))^{\vartheta/2} &\le \frac{1}{n^{\vartheta/2}} \sum_{i=1}^{\lfloor 2^k a_n(n+1)\rfloor} \left(\frac{i}{n+1}\left(1 - \frac{i}{n+1}\right)\right)^{\vartheta/2} \\
&\le \frac{1}{n^{\vartheta/2}} \int_0^{2^k a_n} u^{\vartheta/2} du \le \frac{(2^k a_n)^{1+\vartheta/2}}{n^{\vartheta/2}\left(1 + \frac{\vartheta}{2}\right)}.
\end{aligned}$$

Therefore

$$\mathbb{E}\left[\sup_{2^{k-1}a_n \le h \le 2^k a_n} \Delta_{n,h}^{(1)}\right] \le 2^{\vartheta+1}\mathbb{E}\,|\varphi(Y)|\,\frac{C_1}{(n2^k a_n)^{\vartheta/2}(1+\vartheta/2)}.$$

Thus, recalling that $[a_n, \frac{1}{2}] \subseteq \bigcup_{k=1}^{k_n}[2^{k-1}a_n, 2^k a_n]$, we have uniformly in $a_n \le h \le 1/2$ that

$$\mathbb{E}\left[\sup_{a_n \le h \le \frac{1}{2}} \Delta_{n,h}^{(1)}\right] \le \sum_{k=1}^{k_n}\mathbb{E}\left[\sup_{2^{k-1}a_n \le h \le 2^k a_n} \Delta_{n,h}^{(1)}\right]$$

$$\le 2^{\vartheta+1}\mathbb{E}|\varphi(Y)|\sum_{k=1}^{k_n}\frac{C_1}{(n2^k a_n)^{\vartheta/2}(1+\vartheta/2)},$$

which for some $C_2 \ge 2^{\vartheta+1}\mathbb{E}|\varphi(Y)|C_1/(1+\vartheta/2) > 0$ is

$$< \frac{C_2}{(na_n)^{\vartheta/2}}\sum_{k=1}^{\infty}\frac{1}{(2^{\vartheta/2})^k} = O\left(\frac{1}{(na_n)^{\vartheta/2}}\right).$$

□

Proof of Theorem 4 Recall from Eq. 14 that

$$\mathbb{E}\left[\sup_{a_n \le h \le \frac{1}{2}}\left|\hat{\xi}_{n,h} - \mathbb{E}\varphi(Y)\int_0^{1/h} H(u)du\right|\right] \le \mathbb{E}\left[\sup_{a_n \le h \le \frac{1}{2}}\left|\frac{1}{nh}\mathcal{S}_n(h)\right|\right] + \mathbb{E}\left[\sup_{a_n \le h \le \frac{1}{2}}\left|\frac{1}{nh}\mathcal{T}_n(h)\right|\right].$$

Hence, by combining Propositions 1 and 2 we get Theorem 4. □

Remark An easy argument shows that the statements of Theorem 4 and Proposition 2 remain true with $H\left(\frac{i}{(n+1)h}\right)$ replaced by $H\left(\frac{i}{nh}\right)$.

4 Proof of Theorem 1

We are now ready to prove our main result, namely the uniform in bandwidth consistency of the tail index estimator $\hat{\tau}_{n,h}$ in Eq. 1 to τ. Without loss of generality we shall assume that $Y \ge 1$. To begin, recall that

$$\frac{1}{nh}\sum_{i=1}^{n}\mathbb{E}\left[\varphi(Y_i)H\left(\frac{U_{i:n}}{h}\right)\right] = \mathbb{E}\varphi(Y)\int_0^{1/h} H(u)du,$$

and observe that Proposition 2 along with Eq. 13 implies that

$$\sup_{a_n \le h \le \frac{1}{2}} \frac{1}{nh} |\mathbb{E}\mathcal{T}_n(h)| =$$

$$\sup_{a_n \le h \le 1/2} \left| \mathbb{E}\varphi(Y) \left\{ \frac{1}{nh} \sum_{i=1}^{n} H\left(\frac{i}{(n+1)h} \right) - \int_0^{1/h} H(u)du \right\} \right| \longrightarrow 0.$$

Applying this with $H = K$ and $\varphi \equiv 1$, we readily obtain that

$$\sup_{a_n \le h \le \frac{1}{2}} |\kappa_{n,h} - 1| \to 0, \quad \text{as } n \to \infty, \tag{20}$$

where $\kappa_{n,h} = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{j}{nh}\right)$ is the denominator of $\hat{\tau}_{n,h}$, as in Eq. 1. Furthermore, Eq. $(I.\tau)$ implies via Karamata's representation that $Y_i \stackrel{d}{=} (1 - U_i)^{-\tau} L((1 - U_i)^{-1})$, where $U_1, \ldots, U_n$, $n \ge 1$, are independent uniform(0, 1) variables, and L is a slowly varying function at infinity. It thus holds that

$$\begin{aligned} \log Y_{n-j+1:n} - \log Y_{n-j:n} &\stackrel{d}{=} -\tau \left\{ \log\left(1 - U_{n-j+1:n}\right) - \log\left(1 - U_{n-j:n}\right) \right\} \\ &\quad + \int_{(1-U_{n-j:n})^{-1}}^{(1-U_{n-j+1:n})^{-1}} \frac{b(u)}{u} du \\ &\quad + \log \frac{c\left(\left(1 - U_{n-j+1:n}\right)^{-1}\right)}{c\left(\left(1 - U_{n-j:n}\right)^{-1}\right)} \\ &=: -\tau A_{n,j}^{(1)} + A_{n,j}^{(2)} + A_{n,j}^{(3)}, \end{aligned}$$

where $c(y) \to c_0 > 0$ and $b(y) \to 0$ as $y \to \infty$. Consequently, $\phi_{n,h}$ (the numerator in Eq. 1) can be decomposed into three terms as follows:

$$\phi_{n,h} \stackrel{d}{=} \sum_{j=1}^{n} \frac{j}{nh} K\left(\frac{j}{nh}\right) \left\{ -\tau A_{n,j}^{(1)} + A_{n,j}^{(2)} + A_{n,j}^{(3)} \right\} =: \tau \phi_{n,h}^{(1)} + \phi_{n,h}^{(2)} + \phi_{n,h}^{(3)}. \tag{21}$$

Using Rényi's representation for the order statistics of n independent exponential random variables $\omega_1, \ldots, \omega_n$ with mean 1, we can show that

$$\left\{ \phi_{n,h}^{(1)} : a_n \le h \le b_n \right\} \stackrel{d}{=} \left\{ \frac{1}{nh} \sum_{j=1}^{n} \omega_j K\left(\frac{j}{nh}\right) : a_n \le h \le b_n \right\},$$

where $b_n \to 0$. Now, when $\varphi(Y)$ is an exponential random variable with mean γ, it is well-known that $\mathbb{E}\Phi_n \sim \gamma \log n$. Consequently, in Eq. 13 we require that $na_n/(\log n)^2 \to \infty$, corresponding to the condition in Eq. 3. Thus we can apply Theorem 4 to get that

$$\sup_{a_n \le h \le b_n} \left|\phi_{n,h}^{(1)} - 1\right| \xrightarrow{\mathbb{P}} 0, \quad \text{as } n \to \infty. \tag{22}$$

To show that $|\phi_{n,h}^{(2)}|$ converges uniformly in $a_n \le h \le b_n$ to zero, in probability, notice that since K is assumed to have support contained in $[0, M]$, only the first $\lfloor Mnh \rfloor$ terms of $\phi_{n,h}^{(2)}$ will contribute. This means that

$$\begin{aligned}
\sup_{a_n \le h \le b_n} |\phi_{n,h}^{(2)}| &= \sup_{a_n \le h \le b_n} \left| \sum_{j=1}^{\lfloor Mnh \rfloor} \frac{j}{nh} K\left(\frac{j}{nh}\right) A_{n,j}^{(2)} \right| \\
&= \sup_{a_n \le h \le b_n} \left| \sum_{j=1}^{\lfloor Mnh \rfloor} \frac{j}{nh} K\left(\frac{j}{nh}\right) \int_{(1-U_{n-j:n})^{-1}}^{(1-U_{n-j+1:n})^{-1}} \frac{b(u)}{u} du \right| \\
&\le B(n) \sup_{a_n \le h \le b_n} \left|\phi_{n,h}^{(1)} - 1\right| + B(n),
\end{aligned}$$

where

$$B(n) = \sup\left\{ |b(u)| : u \ge (1 - U_{n-\lfloor Mnb_n \rfloor:n})^{-1} \right\}.$$

Since $b(y) \to 0$ as $y \to \infty$ and $U_{n-\lfloor Mnb_n \rfloor:n} \xrightarrow{\mathbb{P}} 1$, we see that $B(n) \xrightarrow{\mathbb{P}} 0$. This combined with Eq. 22 implies readily that

$$\sup_{a_n \le h \le b_n} |\phi_{n,h}^{(2)}| \xrightarrow{\mathbb{P}} 0, \quad \text{as } n \to \infty. \tag{23}$$

To finish the proof of Theorem 1, we shall show that $\sup_{a_n \le h \le b_n} |\phi_{n,h}^{(3)}| \xrightarrow{\mathbb{P}} 0$. To this end, recall that

$$\phi_{n,h}^{(3)} = \sum_{j=1}^{n} \frac{j}{nh} K\left(\frac{j}{nh}\right) A_{n,j}^{(3)},$$

where $A^{(3)}_{n,j} = \log c((1-U_{n-j+1:n})^{-1}) - \log c((1-U_{n-j:n})^{-1})$ with c a measurable function such that $c(u) \to c_0 > 0$ as $u \to \infty$. Set $\zeta_{k,n} = \zeta_k := \log c((1-U_{k:n})^{-1})$. Then

$$\begin{aligned}
\phi^{(3)}_{n,h} &= \sum_{j=1}^{\lfloor Mnh \rfloor} \frac{j}{nh} K\left(\frac{j}{nh}\right) \{\zeta_{n-j+1} - \zeta_{n-j}\} \\
&= \sum_{j=1}^{\lfloor Mnh \rfloor} \frac{j}{nh} K\left(\frac{j}{nh}\right) \zeta_{n-j+1} - \sum_{j=2}^{\lfloor Mnh \rfloor+1} \frac{j}{nh} K\left(\frac{j-1}{nh}\right) \zeta_{n-j+1} \\
&\quad + \sum_{j=2}^{\lfloor Mnh \rfloor+1} \frac{j}{nh} K\left(\frac{j-1}{nh}\right) \zeta_{n-j+1} - \sum_{j=1}^{\lfloor Mnh \rfloor} \frac{j}{nh} K\left(\frac{j}{nh}\right) \zeta_{n-j} \\
&= \frac{1}{nh} K\left(\frac{1}{nh}\right) \zeta_n + \sum_{j=1}^{\lfloor Mnh \rfloor} \left\{\frac{j+1}{nh} - \frac{j}{nh}\right\} K\left(\frac{j}{nh}\right) \zeta_{n-j} \\
&\quad + \sum_{j=2}^{\lfloor Mnh \rfloor} \frac{j}{nh} \left\{K\left(\frac{j}{nh}\right) - K\left(\frac{j-1}{nh}\right)\right\} \zeta_{n-j+1} \\
&\quad - \frac{\lfloor Mnh \rfloor + 1}{nh} K\left(\frac{\lfloor Mnh \rfloor}{nh}\right) \zeta_{n-\lfloor Mnh \rfloor}.
\end{aligned}$$

Thus we obtain the decomposition of $\phi^{(3)}_{n,h}$ into three terms:

$$\begin{aligned}
\phi^{(3)}_{n,h} &= \frac{1}{nh} K\left(\frac{1}{nh}\right) \zeta_n + \frac{1}{nh} \sum_{j=1}^{\lfloor Mnh \rfloor} K\left(\frac{j}{nh}\right) \zeta_{n-j} \\
&\quad + \left(\sum_{j=2}^{\lfloor Mnh \rfloor} \frac{j}{nh} \left\{K\left(\frac{j}{nh}\right) - K\left(\frac{j-1}{nh}\right)\right\} \zeta_{n-j+1} \right. \\
&\qquad \left. - \frac{\lfloor Mnh \rfloor + 1}{nh} K\left(\frac{\lfloor Mnh \rfloor}{nh}\right) \zeta_{n-\lfloor Mnh \rfloor} \right) \\
&=: Z^{(1)}_{n,h} + Z^{(2)}_{n,h} + Z^{(3)}_{n,h}.
\end{aligned}$$

We now treat the three parts $Z^{(i)}_{n,h}$, $i = 1, 2, 3$ separately. First, note that uniformly in $a_n \le h \le b_n$, as $n \to \infty$,

$$\zeta_{n-\lfloor nh \rfloor+1} \xrightarrow{\mathbb{P}} \log c_0. \tag{24}$$

Therefore, by using (K.i) and (K.ii), we can argue that uniformly in $a_n \le h \le b_n$, as $n \to \infty$,

$$Z_{n,h}^{(3)} \xrightarrow{\mathbb{P}} \log c_0 \left\{ \int_0^M t dK(t) - MK(M) \right\}.$$

Furthermore, (K.i) and (K.ii) ensure that

$$\int_0^M t dK(t) = [tK(t)]_0^M - \int_0^M K(t)dt = MK(M) - 1.$$

Therefore, it follows that uniformly in $a_n \le h \le b_n$, as $n \to \infty$,

$$Z_{n,h}^{(3)} \xrightarrow{\mathbb{P}} -\log c_0. \tag{25}$$

Next, since K is assumed to be bounded, it is readily verified that

$$\sup_{a_n \le h \le b_n} |Z_{n,h}^{(1)}| = \sup_{a_n \le h \le b_n} \frac{1}{nh} K\left(\frac{1}{nh}\right) |\zeta_n| \xrightarrow{\mathbb{P}} 0. \tag{26}$$

Finally, recalling Eqs. 20 and 24, we get uniformly in $a_n \le h \le b_n$, as $n \to \infty$,

$$Z_{n,h}^{(2)} = \frac{1}{nh} \sum_{j=1}^{\lfloor Mnh \rfloor} K\left(\frac{j}{nh}\right) \zeta_{n-j} \xrightarrow{\mathbb{P}} \log c_0. \tag{27}$$

We conclude now from Eqs. 25, 26 and 27 that uniformly in $a_n \le h \le b_n$, as $n \to \infty$,

$$\phi_{n,h}^{(3)} \xrightarrow{\mathbb{P}} 0 + \log c_0 - \log c_0 = 0,$$

which with Eqs. 21, 22 and 23 completes the proof of Theorem 1. □

Acknowledgments The authors are grateful to the late Sándor Csörgő for proposing the problem that led to this research and to the two referees for a careful reading of the manuscript and helpful suggestions.

References

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J.: Statistics of extremes: theory and applications. In: Wiley Series in Probability and Statistics. Wiley, New York (2004)

Caeiro, F., Gomes, M.I., Pestana, D.: Direct reduction of bias of the classical Hill estimator. Revstat **3**(2), 113–136 (2005)

Csörgő, S., Mason, D.M.: Central limit theorems for sums of extreme values. Math. Proc. Camb. Philos. Soc. **98**, 547–558 (1985)

Csörgő, S., Deheuvels, P., Mason, D.M.: Kernel estimates of the tail index of a distribution. Ann. Stat. **13**, 1050–1077 (1985)

Deheuvels, P., Haeusler, E., Mason, D.M.: Almost sure convergence of the Hill estimator. Math. Proc. Camb. Philos. Soc. **104**, 371–381 (1988)

Dony, J.: Nonparametric regression estimation: an empirical process approach to the uniform in bandwidth consistency of kernel-type estimators and conditional U-statistics. Ph.D. thesis, Vrije Universiteit Brussel, Belgium (2008)

Dony, J., Einmahl, U.: Uniform in bandwidth consistency of kernel-type estimators at a fixed point. In: Houdré, C., Koltchinskii, V., Peligrad, M., Mason, D. (eds.) Proceedings of High Dimensional Probability V, Luminy, France, 2008. I.M.S. Publications (2010, to appear)

Dony, J., Mason, D.M.: Uniform in bandwidth consistency of conditional U-statistics. Bernoulli **14**, 1108–1133 (2008)

Einmahl, U., Mason, D.M.: Uniform in bandwidth consistency of kernel-type function estimators. Ann. Stat. **33**, 1380–1403 (2005)

Fraga Alves, M.I., Gomes, M.I., de Haan, L.: A new class of semi-parametric estimators of the second order parameter. Port. Math. **60**(2), 193–213 (2003)

Gomes, M.I., Pestana, D.: A sturdy reduced bias extreme quantile (VaR) estimator. J. Am. Stat. Assoc. **102**, 280–292 (2007)

Grama, I., Spokoiny, V.: Statistics of extremes by oracle inequalities. Ann. Stat. **36**, 1619–1648 (2008)

Groeneboom, P., Lopuhaä, H.P., de Wolf, P.P.: Kernel-type estimators for the extreme value index. Ann. Stat. **31**, 1956–1995 (2003)

Haeusler, E., Teugels, J.L.: On asymptotic normality of Hill's estimator for the exponent of regular variation. Ann. Stat. **13**, 743–756 (1985)

Hall, P., Welsh, A.W.: Adaptive estimates of parameters of regular variation. Ann. Stat. **13**, 331–341 (1985)

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

Mason, D.M.: Laws of large numbers for sums of extreme values. Ann. Probab. **10**, 754–764 (1982)

Nolan, D., Marron, J.S.: Uniform consistency of automatic and location-adaptive delta-sequence estimators. Probab. Theory Relat. Fields **80**, 619–632 (1989)

van der Vaart, A.W., Wellner, J.A.: Weak convergence and empirical processes. With applications to statistics. In: Springer Series in Statistics. Springer, New York (1996)