



**HAL**  
open science

## An estimator for the tail index of an integrated conditional Pareto–Weibull-type model

Yuri Goegebeur, Armelle Guillou, Michael Osmann

► **To cite this version:**

Yuri Goegebeur, Armelle Guillou, Michael Osmann. An estimator for the tail index of an integrated conditional Pareto–Weibull-type model. *Statistics and Probability Letters*, 2015, 103, pp.8-16. 10.1016/j.spl.2015.04.008 . hal-01312904

**HAL Id: hal-01312904**

**<https://hal.science/hal-01312904>**

Submitted on 11 May 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# An estimator for the tail index of an integrated conditional Pareto-Weibull-type model

Yuri Goegebeur <sup>\*</sup>

Armelle Guillou <sup>†</sup>

Michael Osmann <sup>‡</sup>

**Abstract.** We introduce a nonparametric regression estimator for a tail heaviness parameter in an integrated conditional Pareto-Weibull-type model. The estimator is based on local log excesses over a high random threshold. Asymptotic properties are derived under proper regularity conditions.

**Key words and phrases:** Extremes, local estimation, regression, tail index.

## 1 Introduction

In the recent years, a lot of attention in extreme value theory has been devoted to situations where the variable of interest  $Y$  is observed together with a random covariate  $X$ . Goegebeur *et al.* (2014) introduced an estimator for the conditional extreme value index  $\gamma(x)$  when  $\gamma(x) > 0$ , while de Wet *et al.* (2015) introduced an estimator for the conditional Weibull-tail coefficient. In both of these cases, a weighted average of the log-excesses over a threshold is used, where the threshold is considered to be non-random. The aim of the present paper is to construct an estimator that can be used for both conditional Weibull-tail distributions and Pareto-type

---

<sup>\*</sup>Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark (email: yuri.goegebeur@imada.sdu.dk).

<sup>†</sup>Institut Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg et CNRS, 7 rue René Descartes, 67084 Strasbourg cedex, France (email: armelle.guillou@math.unistra.fr).

<sup>‡</sup>Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark (email: mosma@imada.sdu.dk).

distributions. To this end, we use a two parameter family of distributions, which contain both the Pareto-type distributions and the Weibull-tail distributions. The estimator is based on a random threshold, as was also done in Stupfler (2013), who introduced an estimator for the conditional extreme value index  $\gamma(x)$  with  $\gamma(x) \in \mathbb{R}$ .

Let  $F(y; x) := \mathbb{P}(Y \leq y | X = x)$ , the conditional response distribution function, and  $\bar{F}(\cdot; x) := 1 - F(\cdot; x)$ . Assume

$$\bar{F}(y; x) = \exp\left(-D_{\tau(x)}^{\leftarrow}(\ln H(y; x))\right), \quad (1)$$

where

- $y > y^*(x)$  with  $y^*(x) > 0$ ,
- $D_{\tau(x)}(y) = \int_1^y u^{\tau(x)-1} du$ , with  $\tau(x) \in [0, 1]$ ,
- $H$  is an increasing function that satisfies  $H^{\leftarrow}(t; x) := \inf\{y : H(y; x) \geq t\} = t^{\theta(x)} \ell(t; x)$ , where  $\theta(x) > 0$ , and  $\ell$  is a slowly varying function at infinity, i.e.  $\frac{\ell(\lambda y; x)}{\ell(y; x)} \rightarrow 1$  as  $y \rightarrow \infty$  for all  $\lambda > 0$ .

As noted in Gardes *et al.* (2011), this model includes Weibull-tail distributions with Weibull-tail coefficient  $\theta(x)$  if  $\tau(x) = 0$ , and Pareto-type tails with extreme value index  $\theta(x)$  if  $\tau(x) = 1$ , while  $\tau(x) \in (0, 1)$  is an intermediate class of distributions. In the following, we let  $(X_i, Y_i)$ ,  $i = 1, \dots, n$ , be independent copies of the random vector  $(X, Y) \in \mathbb{R}^q \times \mathbb{R}_+$  with  $q \geq 1$ , where the conditional distribution of  $Y$  given  $X = x$  satisfies (1). Furthermore, let  $x \in \mathbb{R}^q$  be arbitrary and denote by  $B(x, h)$ , the ball with center  $x$  and radius  $h$ , i.e.  $B(x, h) := \{z \in \mathbb{R}^q : d(x, z) \leq h\}$ , with  $d(x, z)$  being the distance between  $x$  and  $z$ . The number of observations in the ball is given by  $N_{n,x,h} := \sum_{i=1}^n \mathbb{1}_{\{X_i \in B(x,h)\}}$ , where  $\mathbb{1}_{\{\cdot\}}$  is the indicator function, and denote by  $n_x$  the expected number of observations in  $B(x, h)$ , i.e.  $n_x := n\mathbb{P}(X \in B(x, h))$ .

Conditional on  $N_{n,x,h} = p$ ,  $p \geq 1$ , we introduce  $Z_j$ ,  $j = 1, \dots, p$ , as the response variables for which the covariate  $X_j$  is in the ball  $B(x, h)$ , and denote by  $Z_{1,p} \leq \dots \leq Z_{p,p}$  the associated

order statistics. In this setting we define our estimator of  $\theta(x)$  as

$$\widehat{\theta}(k_x; x) := \frac{1}{\mu_{\tau(x)} \left( \ln \frac{p}{k_x} \right)} \frac{1}{k_x} \sum_{i=1}^{k_x} [\ln Z_{p-i+1,p} - \ln Z_{p-k_x,p}]$$

with

$$\mu_{\tau(x)}(t) := \int_0^\infty (D_{\tau(x)}(u+t) - D_{\tau(x)}(t)) \exp(-u) du,$$

and assuming that  $k_x \in \{1, \dots, p-1\}$ . This estimator is an adaptation of the estimator proposed by Gardes *et al.* (2011) to the regression context. It consists mainly in averaging the log-spacings between the upper order statistics of the response variables for which the covariates are in the ball centered at  $x$ .

In the following, we will let  $U_h(t; x)$  and  $U(t; x)$  be the tail quantile functions corresponding to the conditional distribution function  $F_h(y; x) := \mathbb{P}(Y \leq y | X \in B(x, h))$  and  $F(y; x)$ , respectively, i.e.  $U_h(\cdot; x) := (1/\overline{F}_h(\cdot; x))^\leftarrow$  and  $U(\cdot; x) := (1/\overline{F}(\cdot; x))^\leftarrow$ , where the superscript  $\leftarrow$  denotes the generalised inverse as introduced above. In order to control the difference between  $U_h(t; x)$  and  $U(t; x)$ , we define  $\omega(u, v, x, h) := \sup_{z \in [u, v]} |\log U_h(z; x) - \log U(z; x)|$ , with  $u \leq v$ . The asymptotic properties of  $\widehat{\theta}(k_x; x)$  will be examined under the following second order condition.

**Assumption**  $A(\rho(x))$  *There exist  $\rho(x) < 0$  and  $b(y; x) \rightarrow 0$  for  $y \rightarrow \infty$  such that*

$$\ln \frac{\ell(\lambda y; x)}{\ell(y; x)} = b(y; x) D_{\rho(x)}(\lambda) (1 + o(1)),$$

where  $o(1)$  is uniform on  $\lambda \in [1, \infty)$ .

Note that this assumption immediately implies that the function  $|b(y; x)|$  is regularly varying with index  $\rho(x)$ .

## 2 Asymptotic properties

In this section we examine the asymptotic properties of our estimator. We start by establishing the consistency of  $\widehat{\theta}(k_x; x)$ .

**Theorem 1** Assume that  $\overline{F}(\cdot; x)$  satisfies (1) and that  $A(\rho(x))$  holds. If  $n_x \rightarrow \infty$ ,  $k_x \rightarrow \infty$  and  $\frac{k_x}{n_x} \rightarrow 0$  in such a way that for some  $\delta > 0$ ,

$$\frac{1}{\mu_{\tau(x)} \left( \ln \frac{n_x}{k_x} \right)} \omega \left( \frac{n_x}{(1+\delta)k_x}, n_x^{1+\delta}, x, h \right) \rightarrow 0,$$

then

$$\widehat{\theta}(k_x; x) \xrightarrow{\mathbb{P}} \theta(x).$$

**Proof:** Let  $I_x := \mathbb{N} \cap [(1 - n_x^{-1/4})n_x, (1 + n_x^{-1/4})n_x]$ . According to Lemma 1 in Stupfler (2013), one has that  $\mathbb{P}(N_{n,x,h} \in I_x) \rightarrow 1$  as  $n_x \rightarrow \infty$ . For any  $t > 0$ , define the event

$$S(t; x) := \left\{ \left| \widehat{\theta}(k_x; x) - \theta(x) \right| > t \right\}.$$

Note that after applying the law of total probability one obtains the inequality

$$\mathbb{P}(S(t; x)) \leq \sup_{p \in I_x} \mathbb{P}(S(t; x) | N_{n,x,h} = p) + \mathbb{P}(N_{n,x,h} \notin I_x).$$

We have thus to show that  $\sup_{p \in I_x} \mathbb{P}(S(t; x) | N_{n,x,h} = p) \rightarrow 0$ .

To this aim, let  $T_i, i = 1, \dots, p$ , be unit Pareto random variables, with  $T_{1,p} \leq \dots \leq T_{p,p}$  the associated order statistics. Given  $N_{n,x,h} = p \geq 1$ , the distribution of the random vector  $(Z_1, \dots, Z_p)$ , is the same as that of the random vector  $(U_h(T_1; x), \dots, U_h(T_p; x))$ ; see Lemma 2 in Stupfler (2013). Thus, denoting

$$\begin{aligned} \check{\theta}(k_x; x) &:= \frac{1}{\mu_{\tau(x)} \left( \ln \frac{p}{k_x} \right)} \frac{1}{k_x} \sum_{i=1}^{k_x} [\ln U_h(T_{p-i+1,p}; x) - \ln U_h(T_{p-k_x,p}; x)], \\ \tilde{\theta}(k_x; x) &:= \frac{1}{\mu_{\tau(x)} \left( \ln \frac{p}{k_x} \right)} \frac{1}{k_x} \sum_{i=1}^{k_x} [\ln U(T_{p-i+1,p}; x) - \ln U(T_{p-k_x,p}; x)], \end{aligned}$$

and

$$R_p(x) := \frac{1}{\mu_{\tau(x)} \left( \ln \frac{p}{k_x} \right)} \frac{1}{k_x} \sum_{i=1}^{k_x} [\ln U_h(T_{p-i+1,p}; x) - \ln U_h(T_{p-k_x,p}; x) - (\ln U(T_{p-i+1,p}; x) - \ln U(T_{p-k_x,p}; x))],$$

we have

$$\mathbb{P}(S(t; x) | N_{n,x,h} = p) = \mathbb{P} \left( \left| \check{\theta}(k_x; x) - \theta(x) \right| > t \right) \leq \mathbb{P} \left( \left| \tilde{\theta}(k_x; x) - \theta(x) \right| > \frac{t}{2} \right) + \mathbb{P} \left( |R_p(x)| > \frac{t}{2} \right). \quad (2)$$

The two probabilities on the right-hand side of (2) are now studied separately. Concerning the first one, note that, with  $T_i^*(p) := \frac{T_{p-i+1,p}}{T_{p-k_x,p}}$ ,  $i = 1, \dots, k_x$ ,

$$\begin{aligned}\tilde{\theta}(k_x; x) &= \theta(x) \frac{1}{\mu_{\tau(x)} \left( \ln \frac{p}{k_x} \right)} \frac{1}{k_x} \sum_{i=1}^{k_x} [D_{\tau(x)}(\ln T_{p-k_x,p} + \ln T_i^*(p)) - D_{\tau(x)}(\ln T_{p-k_x,p})] \\ &\quad + \frac{1}{\mu_{\tau(x)} \left( \ln \frac{p}{k_x} \right)} \frac{1}{k_x} \sum_{i=1}^{k_x} \ln \frac{\ell(\exp(D_{\tau(x)}(\ln T_{p-k_x,p} + \ln T_i^*(p))); x)}{\ell(\exp(D_{\tau(x)}(\ln T_{p-k_x,p}))); x)} \\ &=: \tilde{\theta}_1(k_x; x) + \tilde{\theta}_2(k_x; x).\end{aligned}$$

For the sequel, it is important to keep in mind that  $(T_{k_x-i+1}^*(p), i = 1, \dots, k_x) \stackrel{D}{=} (T_{1,k_x}, \dots, T_{k_x,k_x})$ , independently of  $T_{p-k_x,p}$ . Application of a Taylor series expansion to  $\tilde{\theta}_1(k_x; x)$  gives

$$\begin{aligned}\tilde{\theta}_1(k_x; x) &= \theta(x) \frac{(\ln T_{p-k_x,p})^{\tau(x)-1} \left( \ln \frac{p}{k_x} \right)^{\tau(x)-1}}{\left( \ln \frac{p}{k_x} \right)^{\tau(x)-1} \mu_{\tau(x)} \left( \ln \frac{p}{k_x} \right)} \frac{1}{k_x} \sum_{i=1}^{k_x} \ln T_i^*(p) \\ &\quad + \frac{\theta(x)}{2} \frac{\tau(x) - 1}{\mu_{\tau(x)} \left( \ln \frac{p}{k_x} \right)} \frac{1}{k_x} \sum_{i=1}^{k_x} \left( \ln T_{p-k_x,p} + \ln \tilde{T}_i(p) \right)^{\tau(x)-2} (\ln T_i^*(p))^2 \\ &=: \tilde{\theta}_{11}(k_x; x) + \tilde{\theta}_{12}(k_x; x)\end{aligned}$$

where  $\ln \tilde{T}_i(p)$  is a random value between 0 and  $\ln T_i^*(p)$ . The cases  $\tau(x) = 1$  and  $\tau(x) \neq 1$  can now be studied separately. If  $\tau(x) = 1$ , we have that  $\tilde{\theta}_{11}(k_x; x) = \theta(x) \frac{1}{k_x} \sum_{i=1}^{k_x} \ln T_i^*(p)$  and  $\tilde{\theta}_{12}(k_x; x) = 0$ , and thus for any  $t > 0$

$$\begin{aligned}\sup_{p \in I_x} \mathbb{P} \left( \left| \tilde{\theta}_1(k_x; x) - \theta(x) \right| > t \right) &= \sup_{p \in I_x} \mathbb{P} \left( \left| \theta(x) \frac{1}{k_x} \sum_{i=1}^{k_x} \ln T_i^*(p) - \theta(x) \right| > t \right) \\ &= \sup_{p \in I_x} \mathbb{P} \left( \left| \theta(x) \frac{1}{k_x} \sum_{i=1}^{k_x} \ln T_{k_x-i+1,k_x} - \theta(x) \right| > t \right) \\ &= \mathbb{P} \left( \left| \theta(x) \frac{1}{k_x} \sum_{i=1}^{k_x} \ln T_i - \theta(x) \right| > t \right) \\ &\rightarrow 0,\end{aligned}$$

by the law of large numbers. Otherwise, if  $\tau(x) < 1$ , by combining Lemma 6 in Stupfler (2013) with our Lemmas 1 and 3, we deduce that

$$\sup_{p \in I_x} \mathbb{P} \left( \left| \tilde{\theta}_{11}(k_x; x) - \theta(x) \right| > t \right) \rightarrow 0,$$

while concerning  $\tilde{\theta}_{12}(k_x; x)$ ,

$$\left| \tilde{\theta}_{12}(k_x; x) \right| \leq \frac{\theta(x)}{2} (\ln T_{p-k_x, p})^{-1} \frac{(\ln T_{p-k_x, p})^{\tau(x)-1} \left( \ln \frac{p}{k_x} \right)^{\tau(x)-1}}{\left( \ln \frac{p}{k_x} \right)^{\tau(x)-1} \mu_{\tau(x)} \left( \ln \frac{p}{k_x} \right)} \frac{1}{k_x} \sum_{i=1}^{k_x} (\ln T_i^*(p))^2.$$

Using again the law of large numbers combining with the convergence  $\sup_{p \in I_x} \mathbb{P} \left( (\ln T_{p-k_x, p})^{-1} > t \right) \rightarrow 0$  and our Lemma 3, we deduce that

$$\sup_{p \in I_x} \mathbb{P} \left( \left| \tilde{\theta}_{12}(k_x; x) \right| > t \right) \rightarrow 0.$$

This leads also for  $\tau(x) < 1$  to

$$\sup_{p \in I_x} \mathbb{P} \left( \left| \tilde{\theta}_1(k_x; x) - \theta(x) \right| > t \right) \rightarrow 0. \quad (3)$$

Concerning now  $\tilde{\theta}_2(k_x; x)$ , we have to use assumption  $A(\rho(x))$  which ensures that

$$\begin{aligned} \tilde{\theta}_2(k_x; x) &= \frac{1}{\mu_{\tau(x)} \left( \ln \frac{p}{k_x} \right)} \\ &\cdot \frac{1}{k_x} \sum_{i=1}^{k_x} \ln \frac{\ell \left( \exp \left( D_{\tau(x)} \left( \ln T_{p-k_x, p} + \ln T_i^*(p) \right) - D_{\tau(x)} \left( \ln T_{p-k_x, p} \right) \right) \exp \left( D_{\tau(x)} \left( \ln T_{p-k_x, p} \right) \right); x}{\ell \left( \exp \left( D_{\tau(x)} \left( \ln T_{p-k_x, p} \right) \right); x \right)} \\ &= \frac{b \left( \exp \left( D_{\tau(x)} \left( \ln T_{p-k_x, p} \right) \right); x \right)}{\mu_{\tau(x)} \left( \ln \frac{p}{k_x} \right)} \\ &\cdot \frac{1}{k_x} \sum_{i=1}^{k_x} D_{\rho(x)} \left( \exp \left( D_{\tau(x)} \left( \ln \left( T_{p-k_x, p} T_i^*(p) \right) \right) - D_{\tau(x)} \left( \ln \left( T_{p-k_x, p} \right) \right) \right) (1 + \delta_n) \end{aligned}$$

where  $\delta_n \xrightarrow{\mathbb{P}} 0$  uniformly in  $i$  and  $p$ . An application of the mean value theorem, shows that

$$\begin{aligned} &D_{\rho(x)} \left( \exp \left( D_{\tau(x)} \left( \ln \left( T_{p-k_x, p} T_i^*(p) \right) \right) - D_{\tau(x)} \left( \ln \left( T_{p-k_x, p} \right) \right) \right) \right) \\ &= \left[ \exp \left( D_{\tau(x)} \left( \ln \tilde{T}_i(p) + \ln T_{p-k_x, p} \right) - D_{\tau(x)} \left( \ln T_{p-k_x, p} \right) \right) \right]^{\rho(x)} \left( \ln \tilde{T}_i(p) + \ln T_{p-k_x, p} \right)^{\tau(x)-1} \ln T_i^*(p), \end{aligned}$$

where  $\ln \tilde{T}_i(p)$  is a random value between 0 and  $\ln T_i^*(p)$ . Since

$$\left[ \exp \left( D_{\tau(x)} \left( \ln \tilde{T}_i(p) + \ln T_{p-k_x, p} \right) - D_{\tau(x)} \left( \ln T_{p-k_x, p} \right) \right) \right]^{\rho(x)} \leq 1,$$

it follows that

$$\left| \tilde{\theta}_2(k_x; x) \right| \leq \left| \left( 1 + \delta_n \right) \frac{(\ln T_{p-k_x, p})^{\tau(x)-1} \left( \ln \frac{p}{k_x} \right)^{\tau(x)-1}}{\left( \ln \frac{p}{k_x} \right)^{\tau(x)-1} \mu_{\tau(x)} \left( \ln \frac{p}{k_x} \right)} b \left( \exp \left( D_{\tau(x)} \left( \ln T_{p-k_x, p} \right) \right); x \right) \frac{1}{k_x} \sum_{i=1}^{k_x} \ln T_i^*(p) \right|.$$

Clearly,

$$\sup_{p \in I_x} \mathbb{P} (|(1 + \delta_n) - 1| > t) \longrightarrow 0$$

and

$$\sup_{p \in I_x} \mathbb{P} (|b(\exp(D_{\tau(x)}(\ln T_{p-k_x,p})); x)| > t) \longrightarrow 0,$$

(observe that  $b(\exp(D_{\tau(x)}(\ln y)); x)$  is regularly varying at infinity, and apply Lemma 6 of Stupfler, 2013), from which we deduce that

$$\sup_{p \in I_x} \mathbb{P} (|\tilde{\theta}_2(k_x; x)| > t) \longrightarrow 0$$

according to our Lemma 3. Finally, coming back to  $R_p(x)$ , we have

$$|R_p(x)| \leq \frac{2\omega(T_{p-k_x,p}, T_{p,p}, x, h) \mu_{\tau(x)} \left( \ln \frac{n_x}{k_x} \right)}{\mu_{\tau(x)} \left( \ln \frac{n_x}{k_x} \right) \mu_{\tau(x)} \left( \ln \frac{p}{k_x} \right)}. \quad (4)$$

Since  $\omega(u, v, x, h)$  is a decreasing function in  $u$  and an increasing function in  $v$ , it is clear that for all  $t > 0$ ,

$$\left\{ \left| \frac{2\omega \left( \frac{n_x}{(1+\delta)k_x}, n_x^{1+\delta}, x, h \right)}{\mu_{\tau(x)} \left( \ln \frac{n_x}{k_x} \right)} \right| \leq t \right\} \cap \left\{ T_{p-k_x,p} \geq \frac{n_x}{(1+\delta)k_x} \right\} \cap \left\{ T_{p,p} \leq n_x^{1+\delta} \right\} \subseteq \left\{ \left| \frac{2\omega(T_{p-k_x,p}, T_{p,p}, x, h)}{\mu_{\tau(x)} \left( \ln \frac{n_x}{k_x} \right)} \right| \leq t \right\}.$$

By considering the complementary event, we have

$$\left\{ \left| \frac{2\omega(T_{p-k_x,p}, T_{p,p}, x, h)}{\mu_{\tau(x)} \left( \ln \frac{n_x}{k_x} \right)} \right| > t \right\} \subseteq \left\{ \left| \frac{2\omega \left( \frac{n_x}{(1+\delta)k_x}, n_x^{1+\delta}, x, h \right)}{\mu_{\tau(x)} \left( \ln \frac{n_x}{k_x} \right)} \right| > t \right\} \cup \left\{ T_{p-k_x,p} < \frac{n_x}{(1+\delta)k_x} \right\} \cup \left\{ T_{p,p} > n_x^{1+\delta} \right\}.$$

Taking  $n_x$  sufficiently large, under the assumption of Theorem 1, we have

$$\begin{aligned} \sup_{p \in I_x} \mathbb{P} \left( \left| \frac{2\omega(T_{p-k_x,p}, T_{p,p}, x, h)}{\mu_{\tau(x)} \left( \ln \frac{n_x}{k_x} \right)} \right| > t \right) &\leq \sup_{p \in I_x} \mathbb{P} \left( T_{p-k_x,p} < \frac{n_x}{(1+\delta)k_x} \right) + \sup_{p \in I_x} \mathbb{P} \left( T_{p,p} > n_x^{1+\delta} \right) \\ &\longrightarrow 0, \end{aligned}$$

by Lemma 6 in Stupfler (2013) and using the properties of the largest order statistic  $T_{p,p}$ . This ensures then under our Lemma 2 that

$$\sup_{p \in I_x} \mathbb{P} (|R_p(x)| > t) \longrightarrow 0.$$

Combining the above results, Theorem 1 follows. ■

Now we establish the asymptotic normality of  $\hat{\theta}(k_x; x)$ , when properly normalised.



**Theorem 2** Assume that  $\bar{F}(\cdot; x)$  satisfies (1) and that  $A(\rho(x))$  holds. If  $n_x \rightarrow \infty$ ,  $k_x \rightarrow \infty$  and  $\frac{k_x}{n_x} \rightarrow 0$  in such a way that for some  $\delta > 0$ ,

$$\frac{\sqrt{k_x}}{\mu_{\tau(x)}\left(\ln \frac{n_x}{k_x}\right)} \omega\left(\frac{n_x}{(1+\delta)k_x}, n_x^{1+\delta}, x, h\right) \rightarrow 0,$$

and if additionally

$$\sqrt{k_x} b\left(\exp\left(D_{\tau(x)}\left(\ln \frac{n_x}{k_x}\right)\right); x\right) \rightarrow \lambda \in \mathbb{R}$$

and for  $\tau(x) < 1$

$$\frac{\sqrt{k_x}}{\ln \frac{n_x}{k_x}} \rightarrow 0$$

then

$$\sqrt{k_x} \left(\hat{\theta}(k_x; x) - \theta(x)\right) \xrightarrow{D} \mathcal{N}\left(\frac{\lambda}{1-\rho(x)} \mathbb{1}_{\{\tau(x)=1\}} + \lambda \mathbb{1}_{\{\tau(x)<1\}}, \theta^2(x)\right).$$

**Proof:** Given  $N_{n,x,h} = p \geq 1$ , the distribution of  $\sqrt{k_x}(\hat{\theta}(k_x; x) - \theta(x))$  is the same as that of  $\sqrt{k_x}(\check{\theta}(k_x; x) - \theta(x))$ . Thus according to Lemma 5 in Stupfler (2013), it is sufficient to prove that the latter has the same distribution as a triangular array of the form

$$D_n + \phi_{np}$$

where  $D_n \xrightarrow{D} \mathcal{N}\left(\frac{\lambda}{1-\rho(x)} \mathbb{1}_{\{\tau(x)=1\}} + \lambda \mathbb{1}_{\{\tau(x)<1\}}, \theta^2(x)\right)$  and  $\sup_{p \in I_x} \mathbb{P}(|\phi_{np}| > t) \rightarrow 0$  for all  $t > 0$ , as  $n_x \rightarrow \infty$ . We can use the same decomposition of  $\check{\theta}(k_x; x)$  as in the proof of Theorem 1, that is in terms of  $\tilde{\theta}_{11}(k_x; x)$ ,  $\tilde{\theta}_{12}(k_x; x)$ ,  $\tilde{\theta}_2(k_x; x)$  and  $R_p(x)$ . Expanding further on the term  $\tilde{\theta}_{11}(k_x; x)$  gives

$$\begin{aligned} \tilde{\theta}_{11}(k_x; x) &\stackrel{D}{=} \theta(x) \frac{1}{k_x} \sum_{i=1}^{k_x} \ln T_i + \theta(x) \left[ \frac{(\ln T_{p-k_x, p})^{\tau(x)-1} \left(\ln \frac{p}{k_x}\right)^{\tau(x)-1}}{\left(\ln \frac{p}{k_x}\right)^{\tau(x)-1} \mu_{\tau(x)}\left(\ln \frac{p}{k_x}\right)} - 1 \right] \frac{1}{k_x} \sum_{i=1}^{k_x} \ln T_i \\ &=: \tilde{\theta}_{111}(k_x; x) + \tilde{\theta}_{112}(k_x; x). \end{aligned}$$

The first term  $\tilde{\theta}_{111}(k_x; x)$  can be dealt with directly with the central limit theorem

$$\sqrt{k_x} \left(\tilde{\theta}_{111}(k_x; x) - \theta(x)\right) \xrightarrow{D} \mathcal{N}(0, \theta^2(x)).$$

Note that  $\tilde{\theta}_{112}(k_x; x) = 0$  if  $\tau(x) = 1$ , so we only need to consider the case  $\tau(x) < 1$ . For  $\tilde{\theta}_{112}(k_x; x)$ , we have thus to show that for all  $t > 0$

$$\sup_{p \in I_x} \mathbb{P} \left( \left| \sqrt{k_x} \left| \left( \frac{\ln T_{p-k_x, p}}{\ln p/k_x} \right)^{\tau(x)-1} - 1 \right| > t \right) \longrightarrow 0.$$

From the mean value theorem we get

$$\begin{aligned} & \sup_{p \in I_x} \mathbb{P} \left( \left| \sqrt{k_x} \left| \left( \frac{\ln T_{p-k_x, p}}{\ln p/k_x} \right)^{\tau(x)-1} - 1 \right| > t \right) \\ & \leq \sup_{p \in I_x} \mathbb{P} \left( \left( \left| 1 - \frac{\ln \left( \frac{k_x}{p} T_{p-k_x, p} \right)}{\ln(p/k_x)} \right| \right)^{\tau(x)-2} \frac{\sqrt{k_x}}{\ln[(1 - n_x^{-1/4})n_x/k_x]} \left| \ln \left( \frac{k_x}{p} T_{p-k_x, p} \right) \right| > t \right). \end{aligned}$$

Taylor's theorem gives now

$$\sup_{p \in I_x} \mathbb{P} \left( \left| \ln \left( \frac{k_x}{p} T_{p-k_x, p} \right) \right| > t \right) \leq \sup_{p \in I_x} \mathbb{P} \left( \frac{\left| \frac{k_x}{p} T_{p-k_x, p} - 1 \right|}{\left| 1 - \frac{k_x}{p} T_{p-k_x, p} - 1 \right|} > t \right) = \sup_{p \in I_x} \mathbb{P} \left( \left| \frac{k_x}{p} T_{p-k_x, p} - 1 \right| > \frac{t}{1+t} \right),$$

which tends to zero by Lemma 6 in Stupfler (2013), and, with  $a > 1$ ,

$$\begin{aligned} & \sup_{p \in I_x} \mathbb{P} \left( \left( \left| 1 - \frac{\ln \left( \frac{k_x}{p} T_{p-k_x, p} \right)}{\ln(p/k_x)} \right| \right)^{\tau(x)-2} - 1 > t \right) \\ & \leq \sup_{p \in I_x} \mathbb{P} \left( \left( \left| 1 - \frac{\ln T_{p-k_x, p}}{\ln(p/k_x)} - 1 \right| \right)^{\tau(x)-3} > a \right) + \sup_{p \in I_x} \mathbb{P} \left( \left| \frac{\ln T_{p-k_x, p}}{\ln(p/k_x)} - 1 \right| > \frac{t}{2a} \right) \\ & = \sup_{p \in I_x} \mathbb{P} \left( \left| \frac{\ln T_{p-k_x, p}}{\ln(p/k_x)} - 1 \right| > 1 - a^{\frac{1}{\tau(x)-3}} \right) + \sup_{p \in I_x} \mathbb{P} \left( \left| \frac{\ln T_{p-k_x, p}}{\ln(p/k_x)} - 1 \right| > \frac{t}{2a} \right) \\ & \rightarrow 0. \end{aligned}$$

Concerning now the term  $\tilde{\theta}_{12}(k_x; x)$  (which only needs to be considered in case  $\tau(x) < 1$ ), remark that

$$\left| \sqrt{k_x} \tilde{\theta}_{12}(k_x; x) \right| \leq \left| \frac{\theta(x)}{2} \frac{\sqrt{k_x}}{\ln \frac{n_x}{k_x}} \frac{\ln \frac{n_x}{k_x}}{\ln T_{p-k_x, p}} \frac{(\ln T_{p-k_x, p})^{\tau(x)-1}}{\left( \ln \frac{p}{k_x} \right)^{\tau(x)-1}} \frac{\left( \ln \frac{p}{k_x} \right)^{\tau(x)-1}}{\mu_{\tau(x)} \left( \ln \frac{p}{k_x} \right)} \frac{1}{k_x} \sum_{i=1}^{k_x} (\ln T_i^*(p))^2 \right|.$$

Combining again Lemma 6 in Stupfler (2013) with our Lemmas 1 and 3 together with our assumptions, we infer that

$$\sup_{p \in I_x} \mathbb{P} \left( \left| \sqrt{k_x} \tilde{\theta}_{12}(k_x; x) \right| > t \right) \longrightarrow 0.$$

For  $\tilde{\theta}_2(k_x; x)$ , we need also to distinguish between the two cases  $\tau(x) = 1$  and  $\tau(x) < 1$ . We first consider the case  $\tau(x) = 1$ , where we use the fact that  $b(\cdot; x)$  is regularly varying at infinity combining with Lemma 6 in Stupfler (2013) and the law of large numbers according to which

$$\sup_{p \in I_x} \mathbb{P} \left( \left| \frac{1}{k_x} \sum_{i=1}^{k_x} \frac{(T_i^*(p))^{\rho(x)} - 1}{\rho(x)} - \frac{1}{1 - \rho(x)} \right| > t \right) = \mathbb{P} \left( \left| \frac{1}{k_x} \sum_{i=1}^{k_x} \frac{T_i^{\rho(x)} - 1}{\rho(x)} - \frac{1}{1 - \rho(x)} \right| > t \right) \rightarrow 0.$$

The convergence

$$\sup_{p \in I_x} \mathbb{P} \left( \left| \sqrt{k_x} \tilde{\theta}_2(k_x; x) - \frac{\lambda}{1 - \rho(x)} \right| > t \right) \rightarrow 0$$

then follows from our assumptions and our Lemma 3. In the case where  $\tau(x) < 1$ , using the same arguments as in the proof of Theorem 1, we have the following decomposition

$$\tilde{\theta}_2(k_x; x) =: \tilde{\theta}_{21}(k_x; x) + \tilde{\theta}_{22}(k_x; x) + \tilde{\theta}_{23}(k_x; x),$$

where

$$\begin{aligned} \tilde{\theta}_{21}(k_x; x) &:= (1 + \delta_n) b \left( \exp \left( D_{\tau(x)} \left( \ln T_{p-k_x, p} \right) \right); x \right) \frac{(\ln T_{p-k_x, p})^{\tau(x)-1}}{\mu_{\tau(x)} \left( \ln \frac{p}{k_x} \right)} \frac{1}{k_x} \sum_{i=1}^{k_x} \ln T_i^*(p) \\ \tilde{\theta}_{22}(k_x; x) &:= (1 + \delta_n) \frac{b \left( \exp \left( D_{\tau(x)} \left( \ln T_{p-k_x, p} \right) \right); x \right)}{\mu_{\tau(x)} \left( \ln \frac{p}{k_x} \right)} \frac{1}{k_x} \sum_{i=1}^{k_x} \ln T_i^*(p) \\ &\quad \cdot e^{\rho(x) [D_{\tau(x)} (\ln \tilde{T}_i(p) + \ln T_{p-k_x, p}) - D_{\tau(x)} (\ln T_{p-k_x, p})]} \left\{ \left( \ln T_{p-k_x, p} + \ln \tilde{T}_i(p) \right)^{\tau(x)-1} - (\ln T_{p-k_x, p})^{\tau(x)-1} \right\} \\ \tilde{\theta}_{23}(k_x; x) &:= (1 + \delta_n) b \left( \exp \left( D_{\tau(x)} \left( \ln T_{p-k_x, p} \right) \right); x \right) \frac{(\ln T_{p-k_x, p})^{\tau(x)-1}}{\mu_{\tau(x)} \left( \ln \frac{p}{k_x} \right)} \\ &\quad \cdot \frac{1}{k_x} \sum_{i=1}^{k_x} \ln T_i^*(p) \left\{ e^{\rho(x) [D_{\tau(x)} (\ln \tilde{T}_i(p) + \ln T_{p-k_x, p}) - D_{\tau(x)} (\ln T_{p-k_x, p})]} - 1 \right\}. \end{aligned}$$

Using the regularly varying property of  $b(\cdot; x)$ , the law of large numbers, our Lemmas 1-3 and our assumptions, combining with the mean value theorem for  $\tilde{\theta}_{22}(k_x; x)$  and  $\tilde{\theta}_{23}(k_x; x)$ , we deduce that

$$\begin{aligned} \sup_{p \in I_x} \mathbb{P} \left( \left| \sqrt{k_x} \tilde{\theta}_{21}(k_x; x) - \lambda \right| > t \right) &\rightarrow 0, \\ \sup_{p \in I_x} \mathbb{P} \left( \left| \sqrt{k_x} \tilde{\theta}_{22}(k_x; x) \right| > t \right) &\rightarrow 0, \\ \sup_{p \in I_x} \mathbb{P} \left( \left| \sqrt{k_x} \tilde{\theta}_{23}(k_x; x) \right| > t \right) &\rightarrow 0. \end{aligned}$$

For what concerns the remainder term  $R_p(x)$ , using the same arguments as in the proof of Theorem 1, we get for all  $t > 0$ , that

$$\left\{ \left| \sqrt{k_x} \frac{2\omega(T_{p-k_x,p}, T_{p,p}, x, h)}{\mu_{\tau(x)}\left(\ln \frac{n_x}{k_x}\right)} \right| > t \right\} \subseteq \left\{ \left| \sqrt{k_x} \frac{2\omega\left(\frac{n_x}{(1+\delta)k_x}, n_x^{1+\delta}, x, h\right)}{\mu_{\tau(x)}\left(\ln \frac{n_x}{k_x}\right)} \right| > t \right\} \cup \left\{ T_{p-k_x,p} < \frac{n_x}{(1+\delta)k_x} \right\} \\ \cup \left\{ T_{p,p} > n_x^{1+\delta} \right\}.$$

Taking now  $n_x$  sufficiently large, this implies by assumption that

$$\sup_{p \in I_x} \mathbb{P} \left( \left| \sqrt{k_x} \frac{2\omega(T_{p-k_x,p}, T_{p,p}, x, h)}{\mu_{\tau(x)}\left(\ln \frac{n_x}{k_x}\right)} \right| > t \right) \leq \sup_{p \in I_x} \mathbb{P} \left( T_{p-k_x,p} < \frac{n_x}{(1+\delta)k_x} \right) + \sup_{p \in I_x} \mathbb{P} \left( T_{p,p} > n_x^{1+\delta} \right) \\ \rightarrow 0.$$

This convergence combined with (4) and Lemma 2 ensures that

$$\sup_{p \in I_x} \mathbb{P} \left( \left| \sqrt{k_x} R_p(x) \right| > t \right) \rightarrow 0.$$

Combining all these convergences yield our Theorem 2. ■

## Appendix

In this section we introduce some lemmas which are useful for establishing the main results.

**Lemma 1** *Assume that  $n_x \rightarrow \infty$ ,  $k_x \rightarrow \infty$  such that  $\frac{k_x}{n_x} \rightarrow 0$ . If  $\tau(x) < 1$ , then there exist a constant  $C > 0$ , such that*

$$\sup_{p \in I_x} \left| \frac{\left(\ln \frac{p}{k_x}\right)^{\tau(x)-1}}{\mu_{\tau(x)}\left(\ln \frac{p}{k_x}\right)} - 1 \right| \leq C \left(\ln \frac{n_x}{k_x}\right)^{-1}.$$

**Proof:** First note that we have  $\mu_{\tau(x)}(y) = y^{\tau(x)-1} + \tilde{R}(y)$ , with

$$\tilde{R}(y) := \frac{\tau(x)-1}{2} y^{\tau(x)-2} \int_0^\infty (1+\xi)^{\tau(x)-2} u^2 e^{-u} du,$$

where  $\xi$  is a value between 0 and  $\frac{u}{y}$ . Hence  $|\tilde{R}(y)| \leq y^{\tau(x)-2}$ . Consequently

$$\left| \frac{\left(\ln \frac{p}{k_x}\right)^{\tau(x)-1}}{\mu_{\tau(x)}\left(\ln \frac{p}{k_x}\right)} - 1 \right| = \left| \frac{\tilde{R}\left(\ln \frac{p}{k_x}\right)}{\left(\ln \frac{p}{k_x}\right)^{\tau(x)-1} + \tilde{R}\left(\ln \frac{p}{k_x}\right)} \right| \leq \left(\ln \frac{p}{k_x}\right)^{-1} \left( 1 + O\left(\left(\ln \frac{p}{k_x}\right)^{-1}\right) \right)^{-1}.$$

Since

$$\sup_{p \in I_x} \left( \ln \frac{p}{k_x} \right)^{-1} \leq \left( \ln \frac{n_x \left( 1 - n_x^{-\frac{1}{4}} \right)}{k_x} \right)^{-1},$$

the result easily follows. ■

**Lemma 2** Assume that  $n_x \rightarrow \infty$ ,  $k_x \rightarrow \infty$  such that  $\frac{k_x}{n_x} \rightarrow 0$ . Then

$$\frac{\mu_{\tau(x)} \left( \ln \frac{p}{k_x} \right)}{\mu_{\tau(x)} \left( \ln \frac{n_x}{k_x} \right)} \rightarrow 1$$

uniformly in  $p \in I_x$ .

**Proof:** We start by rewriting the term  $\frac{\mu_{\tau(x)} \left( \ln \frac{p}{k_x} \right)}{\mu_{\tau(x)} \left( \ln \frac{n_x}{k_x} \right)} - 1$  as

$$\frac{\mu_{\tau(x)} \left( \ln \frac{p}{k_x} \right)}{\mu_{\tau(x)} \left( \ln \frac{n_x}{k_x} \right)} - 1 = \left( \frac{\mu_{\tau(x)} \left( \ln \frac{p}{k_x} \right)}{\left( \ln \frac{p}{k_x} \right)^{\tau(x)-1}} - 1 \right) \frac{\left( \ln \frac{p}{k_x} \right)^{\tau(x)-1}}{\mu_{\tau(x)} \left( \ln \frac{n_x}{k_x} \right)} + \frac{\left( \ln \frac{p}{k_x} \right)^{\tau(x)-1}}{\mu_{\tau(x)} \left( \ln \frac{n_x}{k_x} \right)} - 1.$$

According to Lemma 2 in Gardes *et al.* (2011),  $\mu_{\tau(x)} \left( \ln \frac{n_x}{k_x} \right) \sim \left( \ln \frac{n_x}{k_x} \right)^{\tau(x)-1}$ . Thus, using a Taylor series expansion combining with the fact that uniformly in  $p \in I_x$ ,  $\ln \frac{p}{n_x} \rightarrow 0$ , we have

$$\left| \frac{\left( \ln \frac{p}{k_x} \right)^{\tau(x)-1}}{\mu_{\tau(x)} \left( \ln \frac{n_x}{k_x} \right)} - 1 \right| \sim \left| \left( 1 + \frac{\ln \frac{p}{n_x}}{\ln \frac{n_x}{k_x}} \right)^{\tau(x)-1} - 1 \right| \rightarrow 0 \quad (5)$$

uniformly in  $p \in I_x$ . Moreover, from the proof of Lemma 1, we know that

$$\left| \frac{\mu_{\tau(x)} \left( \ln \frac{p}{k_x} \right)}{\left( \ln \frac{p}{k_x} \right)^{\tau(x)-1}} - 1 \right| = \left| \frac{\tilde{R} \left( \ln \frac{p}{k_x} \right)}{\left( \ln \frac{p}{k_x} \right)^{\tau(x)-1}} \right| \leq \left( \ln \frac{p}{k_x} \right)^{-1} \rightarrow 0 \quad (6)$$

uniformly in  $p \in I_x$ . Combining (5) and (6), our Lemma 2 follows. ■

**Lemma 3** Assume that  $I_n$  is some index set, and, for  $p \in I_n$  let  $(X_n(p))_n$  and  $(Y_n(p))_n$  be sequences of random variables. If for all  $\varepsilon > 0$  and some  $x, y \in \mathbb{R}$ ,

$$\sup_{p \in I_n} \mathbb{P}(|X_n(p) - x| > \varepsilon) \rightarrow 0$$

and

$$\sup_{p \in I_n} \mathbb{P} (|Y_n(p) - y| > \varepsilon) \longrightarrow 0$$

as  $n \rightarrow \infty$ , then

$$\sup_{p \in I_n} \mathbb{P} (|X_n(p)Y_n(p) - xy| > \varepsilon) \longrightarrow 0$$

as  $n \rightarrow \infty$ .

**Proof:** Note that for all  $p \in I_n$ ,

$$\begin{aligned} \{|X_n(p)Y_n(p) - xy| > \varepsilon\} &\subseteq \{|(X_n(p) - x)| > 1\} \cup \left\{ |(Y_n(p) - y)| > \frac{\varepsilon}{3} \right\} \\ &\cup \left\{ |y(X_n(p) - x)| > \frac{\varepsilon}{3} \right\} \cup \left\{ |x(Y_n(p) - y)| > \frac{\varepsilon}{3} \right\}. \end{aligned}$$

Lemma 3 then follows using the subadditivity property of a probability measure. ■

## Acknowledgements

This work was supported by a research grant (VKR023480) from VILLUM FONDEN and an international project for scientific cooperation (PICS-6416). The authors are grateful to the referee and the associate editor for their comments on the preliminary version of the paper.

## References

- [1] de Wet, T., Goegebeur, G., Guillou, A., Osmann, M., 2015. Kernel regression with Weibull-type tails. Submitted.
- [2] Gardes, L., Girard, S., Guillou, A., 2011. Weibull tail-distributions revisited: A new look at some tail estimators. *J. Statist. Plann. Inference* 141, 429–444.
- [3] Goegebeur, Y., Guillou, A., Schorgen, A., 2014. Nonparametric regression estimation of conditional tails - the random covariate case. *Statistics* 48, 732–755.
- [4] Stupfler, G., 2013. A moment estimator for the conditional extreme value index. *Electron. J. Stat.* 7, 2298–2343.