



**HAL**  
open science

# The stochastic approximation method for recursive kernel estimation of the conditional extreme value index

Fatma Ben Khadher, Yousri Slaoui

► **To cite this version:**

Fatma Ben Khadher, Yousri Slaoui. The stochastic approximation method for recursive kernel estimation of the conditional extreme value index. 2021. hal-03145485

**HAL Id: hal-03145485**

**<https://hal.science/hal-03145485>**

Preprint submitted on 18 Feb 2021

**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.

# The stochastic approximation method for recursive kernel estimation of the conditional extreme value index

Fatma Ben Khadher<sup>1\*</sup> and Yousri Slaoui<sup>2†</sup>

<sup>1</sup>Univ. Monastir, Laboratoire analyse, géométrie et applications, FSM, Tunisie

<sup>2</sup>Univ. Poitiers, Lab. Math. et Appl., Futuroscope Chasseneuil, France

*Abstract:* In this research paper, our central focus is upon applying the stochastic approximation method to define a class of recursive kernel estimator of the conditional extreme value index. We investigate the properties of the proposed recursive estimator and compare them to those pertaining to Hill's non-recursive kernel estimator. We attempt to demonstrate that using some optimal parameters, the proposed recursive estimator defined by the stochastic approximation algorithm proves to be very competitive to Hill's non-recursive kernel estimator. Finally, the theoretical results are explored through simulation experiments and illustrated using real dataset about Malaria in Senegalese children.

Keywords: Stochastic approximation algorithm; tail index; extreme value; Pareto distribution.

## 1 Introduction

The extreme value theory is a branch of statistics which studies the asymptotic distributions of extreme values. It can be a maximum or a minimum of a set of random variables. This theory was developed by [Emil Julius Gumbel \(1958\)](#). It is widely applied in many research areas like climate changes, environmental risks, insurance and financial banking (see [Beirlant et al. \(2004\)](#) for a list of interesting examples). Estimation of the tail index, associated with a random variable  $Y$ , is one of the main problems in the area of extreme value theory. Therefore, a lot of research, aiming to estimate this parameter, has been performed during last decades (see for example [Embrechts et al. \(1997\)](#), [Beirlant et al. \(2004\)](#), [De Haan and Ferreira \(2006\)](#), [Reiss and Thomas \(2007\)](#), [Gardes and Girard \(2008\)](#), [Gardes et al. \(2010\)](#) and [Stupfler \(2013\)](#)). We denote by  $\gamma$  the tail index which characterizes the distribution tail heaviness of  $Y$ . It can be estimated parametrically using the [Hill \(1975\)](#) estimator and nonparametrically using a kernel version of the Hill's estimator proposed by [Goegebeur et al. \(2014\)](#). Improved approaches have recently appeared in literature. Among them, we state [Brilhante et al. \(2013\)](#) to defined a moment of order  $p$ ; [Beran et al. \(2014\)](#) who proposed a harmonic moment tail index estimator; [Paulauskas and Vaičiulis \(2013, 2017\)](#) who elaborated parametric families of functions of the order statistics.

Recently, recursive estimation has drawn the attention and whetted the interest of multiple researchers. Recursivity means that the estimator calculated from the first  $n$  observations, say

---

\*e-mail adress: fatma.b.khadher@gmail.com

†e-mail adress: Yousri.Slaoui@math.univ-poitiers.fr

$\theta_n$ , is a function of  $\theta_{n-1}$ . More precisely, we can easily update the estimator value with each additional observation specially in large sample sizes. The basic objective of the present work lies in introducing a recursive kernel estimator of the conditional extreme value index defined by applying a stochastic approximation algorithm. To the best of our knowledge, this tail index estimator construction was not previously considered in literature and it aims to improve the estimation accuracy.

The application of the stochastic approximation algorithm was introduced first by Révész (1973, 1977) and then extended by Mokkadem et al. (2009b) to estimate a regression function. Using this algorithm, Tsybakov (1990) approximated the mode of a probability density, Mokkadem et al. (2009a) estimated a multivariate probability density and recently Slaoui (2016) estimated a hazard function. It turns out that this estimator depends on two important parameters, which are the bandwidth and the stepsize of the stochastic algorithm. By making an adequate choice of the two parameters, the proposed recursive estimator can be very competitive to Hill's non-recursive kernel estimator in terms of estimation error and much better in terms of computational costs.

The structure of this paper is organized as follows. In Section 2, we identify our estimator and we set forward its asymptotic properties. Simulation experiments and investigation of real dataset are depicted in Section 3. Finally, the last section wraps up the conclusion and provides new perspectives for future works.

## 2 Construction of the estimator and asymptotic properties

### 2.1 The proposed estimator

Let  $(X_i, Y_i)$ ,  $i = 1, \dots, n$ , be independent realizations of the random vectors  $(X, Y) \in \mathbb{R}^d \times \mathbb{R}_0^+$ , where  $X$  is a  $d$ -dimensional covariate with joint density function  $g$ ,  $d \geq 1$ . The probability density function of  $Y$  given  $X = x$  is defined as  $f(y|x) = \mathbb{P}(Y = y|X = x)$  and the conditional survival function of  $Y$  given  $X = x$  is denoted by  $\bar{F}(y|x) = \mathbb{P}(Y > y|X = x)$ . In this paper, we are basically interested in heavy tails. More precisely, we assume that the conditional survival function of  $Y$  given  $X = x$  satisfies

$$(C1): \bar{F}(y|x) = y^{-\frac{1}{\gamma(x)}} l(y|x),$$

where  $\gamma(\cdot)$  is an unknown positive function of the covariate  $x$  called the tail function and for a fixed  $x$ ,  $l(\cdot|x)$  is a function that varies slowly at infinity, i.e for all  $\lambda > 0$ ,

$$\lim_{y \rightarrow \infty} \frac{l(\lambda y|x)}{l(y|x)} = 1.$$

Condition (C1) means that the conditional distribution of  $Y$  given  $X = x$  is in the Frechet maximum domain of attraction. The tail function  $\gamma(x)$  is the conditional extreme value index function which needs to be adequately estimated from the available data.

$$(C2): l(\cdot|x) \text{ is normalized.}$$

The Karamata representation (Theorem 1.3.1 given in Bingham et al. (1987)) of the slowly-varying function,  $l(\cdot|x)$ , can be written as

$$l(y|x) = c(x) \exp \left( \int_1^y \frac{\varepsilon(z|x)}{z} dz \right),$$

where  $c(\cdot)$  is a positive function and  $\varepsilon(z|x) \rightarrow 0$  as  $z \rightarrow \infty$ . Thus,  $l(\cdot|x)$  is differentiable and

the function  $\varepsilon(\cdot|x)$  is given by  $\varepsilon(z|x) = z \frac{l'(z|x)}{l(z|x)}$ .

**(C3):** There exists a strictly negative function  $\rho(\cdot)$ , a strictly positive function  $\gamma(\cdot)$  and a rate function  $b(\cdot|x)$ ,  $b(y|x) \rightarrow 0$  as  $y \rightarrow \infty$ , of constant sign for large values of  $y$  such that for all  $v > 0$

$$\lim_{y \rightarrow \infty} \frac{\frac{\overline{F}(vy|x)}{\overline{F}(y|x)} - v^{-\frac{1}{\gamma(x)}}}{b(y|x)} = v^{-\frac{1}{\gamma(x)}} \frac{v^{\frac{\rho(x)}{\gamma(x)}} - 1}{\rho(x)\gamma(x)}.$$

Additional conditions are needed for ensuring the asymptotic properties of the estimators. Let  $d(x, y)$  denote the Euclidean distance between  $x$  and  $y$ , for all  $x, y \in \mathbb{R}^d$ .

**(C4):** There exists  $c_g \in \mathbb{R}$  such that for all  $x, y \in \mathbb{R}^d$ ,

$$g(x) - g(y) = c_g d(x, y).$$

**(C5):** There exists  $c_{\overline{F}} \in \mathbb{R}$  such that for all  $x, y, z \in \mathbb{R}^d$ ,

$$\frac{\ln \overline{F}(y|x)}{\ln \overline{F}(y|z)} - 1 = c_{\overline{F}} d(x, z)$$

and  $\gamma(x)c_{\overline{F}} < 0$ .

Moreover, we impose a condition on the kernel function  $K$ .

**(C6):**  $K$  is a bounded density function on  $\mathbb{R}^d$ , with support  $\Omega$  included in the unit hypersphere of  $\mathbb{R}^d$  and satisfying  $\int_{\mathbb{R}^d} K(z) dz = 1$ .

Our idea rests upon to construct a recursive estimator for the conditional tail index  $\gamma(x)$ . It will be presented as a ratio of two estimators  $a_n(x)$  and  $b_n(x)$ . The denominator  $b_n(x)$  is an estimator of the function  $b(x) = g(x) \overline{F}(t_n|x)$ , where  $(t_n)$  is a nonrandom threshold sequence tending to  $\infty$  as  $n \rightarrow \infty$ . The nominator  $a_n(x)$  is an estimator of the function  $a(x) = \gamma(x) \overline{F}(t_n|x) C_x g(x)$ , where

$$C_x = 1 + \frac{b(t_n|x)}{\gamma(x)\rho(x)} \left[ \frac{1}{1 - \rho(x)} - 1 + r_{n,x} \right]$$

and  $(r_{n,x})$  is a non-random sequence, tending to 0 as  $n \rightarrow \infty$ , defined as

$$r_{n,x} = \frac{\rho(x)}{\gamma^2(x)} \int_1^\infty z^{-\frac{1}{\gamma(x)}-1} \left( \gamma^2(x) \frac{z^{\frac{1}{\gamma(x)}} \overline{F}(t_n z|x)}{\overline{F}(t_n|x)} - 1 - \frac{z^{\frac{\rho(x)}{\gamma(x)}} - 1}{\rho(x)\gamma(x)} \right) dz.$$

### Construction of a recursive estimator of the function $a(x)$ :

Let us introduce the stochastic algorithm to estimate the function  $a(\cdot)$  at a point  $x$ . It is based on searching the zero of the function  $f_1 : y \mapsto a(x) - y$ . Following Robbins-Monro's procedure, this algorithm is defined as below

- (i)  $a_0(x) \in \mathbb{R}$ ,
- (ii)  $\forall n \geq 1$ , we set  $a_n(x) = a_{n-1}(x) + \gamma_n \mathcal{Z}_n(x)$ , where the stepsize  $(\gamma_n)$  is a sequence of positive real numbers that goes to zero and  $\mathcal{Z}_n(x)$  is an observation of the function  $f_1$  at the point  $a_{n-1}(x)$ .

To construct  $\mathcal{Z}_n(x)$ , we follow the approach of Révész (1973, 1977), Tsybakov (1990) and Slaoui (2013, 2014a,b, 2018) which are based on the classical property of stochastic algorithms (which is  $\mathbb{E}[\mathcal{Z}_n(x)|\mathbb{F}_{n-1}] = 0$ , where  $\mathbb{F}_{n-1}$  stands for the  $\sigma$ -algebra of the events occurring at the time  $n - 1$ ). In addition, we introduce a kernel  $K$  (which is a function satisfying  $\int_{\mathbb{R}^d} K(z)dz = 1$ ), and a bandwidth  $(h_n)$  (which is a sequence of positive real numbers that goes to zero when  $n \rightarrow \infty$ ), and set  $\mathcal{Z}_n(x) = K_{h_n}(x - X_n) [\ln Y_n - \ln t_n] \mathbb{1}_{\{Y_n > t_n\}} - a_{n-1}(x)$ , with  $K_h(x) := h^{-d}K(h^{-1}x)$ . The stochastic approximation algorithm introduced in Mokkadem et al. (2009a) which estimates recursively the function  $a$  at the point  $x$  is defined as follows:

$$a_n(x) = (1 - \gamma_n)a_{n-1}(x) + \gamma_n K_{h_n}(x - X_n) [\ln Y_n - \ln t_n] \mathbb{1}_{\{Y_n > t_n\}}. \quad (2.1)$$

Considering  $a_0(x) = 0$ , the estimator  $a_n$  defined in (2.1) can be rewritten as

$$a_n(x) = \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k K_{h_k}(x - X_k) [\ln Y_k - \ln t_k] \mathbb{1}_{\{Y_k > t_k\}}, \quad (2.2)$$

with

$$\Pi_n = \prod_{k=1}^n (1 - \gamma_k). \quad (2.3)$$

### Construction of a recursive estimator of the function $b(x)$ :

We apply the stochastic algorithm to estimate the function  $b(\cdot)$  at a point  $x$ . It is based on searching the zero of the function  $f_2 : y \mapsto b(x) - y$ . Following Robbins-Monro's procedure, this algorithm is defined as below

- (i)  $b_0(x) \in \mathbb{R}$ ,
- (ii)  $\forall n \geq 1$ , we set  $b_n(x) = b_{n-1}(x) + \beta_n \mathcal{T}_n(x)$ , where the stepsize  $(\beta_n)$  is a sequence of positive real numbers that goes to zero and  $\mathcal{T}_n(x)$  is an observation of the function  $f_2$  at the point  $b_{n-1}(x)$ .

Based on the same previously used approach, we consider  $\mathcal{T}_n(x) = K_{h_n}(x - X_n) \mathbb{1}_{\{Y_n > t_n\}} - b_{n-1}(x)$ , with the same bandwidth  $(h_n)$  and kernel function  $K_h$  previously defined. Then, the stochastic approximation algorithm to estimate recursively the function  $b$  at the point  $x$  is defined as follows:

$$b_n(x) = (1 - \beta_n)b_{n-1}(x) + \beta_n K_{h_n}(x - X_n) \mathbb{1}_{\{Y_n > t_n\}}. \quad (2.4)$$

Considering  $b_0(x) = 0$ , the estimator  $b_n$  defined by (2.4) can be rewritten as

$$b_n(x) = Q_n \sum_{k=1}^n Q_k^{-1} \beta_k K_{h_k}(x - X_k) \mathbb{1}_{\{Y_k > t_k\}}, \quad (2.5)$$

with

$$Q_n = \prod_{k=1}^n (1 - \beta_k). \quad (2.6)$$

Then, our proposed recursive estimator for the conditional tail index  $\gamma(x)$  is defined as:

$$\widehat{\gamma}_n^H(x) := \frac{a_n(x)}{b_n(x)} = \frac{\Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k K_{h_k}(x - X_k) [\ln Y_k - \ln t_n] \mathbb{1}_{\{Y_k > t_n\}}}{Q_n \sum_{k=1}^n Q_k^{-1} \beta_k K_{h_k}(x - X_k) \mathbb{1}_{\{Y_k > t_n\}}}. \quad (2.7)$$

The second objective of our paper is to study the properties of the recursive estimator defined by (2.7) and to compare them with the kernel version of Hill's estimator of the conditional extreme value index proposed by Goegebeur et al. (2014), which is defined as

$$\widetilde{\gamma}_n^H(x) = \frac{\frac{1}{n} \sum_{i=1}^n K_{h_i}(x - X_i) [\ln Y_i - \ln t_n] \mathbb{1}_{\{Y_i > t_n\}}}{\frac{1}{n} \sum_{i=1}^n K_{h_i}(x - X_i) \mathbb{1}_{\{Y_i > t_n\}}}. \quad (2.8)$$

The asymptotic properties of  $\widehat{\gamma}_n^H$  are investigated in the next subsection.

## 2.2 Asymptotic results

In order to obtain the bias and the variance of the recursive estimator  $\widehat{\gamma}_n^H$  defined by (2.7), we first calculate those of the recursive estimator  $a_n$  defined by (2.2). Then, we calculate those of the recursive estimator  $b_n$  defined by (2.5).

Throughout this paper, stepsizes and bandwidths are considered to belong to the following regularly varying sequences class.

**Definition 1** Let  $u \in \mathbb{R}$  and  $(u_n)_{n \geq 1}$  be a nonrandom positive sequence. We say that  $u_n \in \mathcal{GS}(u)$  if

$$\lim_{n \rightarrow \infty} n \left[ 1 - \frac{u_{n-1}}{u_n} \right] = u.$$

This condition was introduced by Galambos and Seneta (1973). The acronym  $\mathcal{GS}$  stands for (Galambos and Seneta). Typical sequences in  $\mathcal{GS}(u)$  are, for  $b \in \mathbb{R}$ ,  $n^u (\log n)^b$ ,  $n^u (\log \log n)^b$  and so on.

Finally, we impose the following additional conditions:

(C7):

- i)  $\gamma_n \in \mathcal{GS}(-\alpha)$  with  $\alpha \in (1/2, 1]$ .
- ii)  $h_n \in \mathcal{GS}(-p)$  with  $p \in (0, \alpha/d)$ .
- iii)  $\lim_{n \rightarrow \infty} n\gamma_n \in (\min(p, (\alpha - pd)/2), \infty]$ .
- iv)  $\beta_n \in \mathcal{GS}(-b)$  with  $b \in (1/2, 1]$ .
- v)  $\lim_{n \rightarrow \infty} n\beta_n \in (\min(p, (b - pd)/2), \infty]$ .
- vi)  $nh_n^{d+2} \ln^2 t_n \xrightarrow[n \rightarrow \infty]{} \infty$ .

The following notations will be often used in this paper:

$$\varepsilon = \lim_{n \rightarrow \infty} (n\gamma_n)^{-1}. \quad (2.9)$$

$$\varepsilon_1 = \lim_{n \rightarrow \infty} (n\beta_n)^{-1}. \quad (2.10)$$

$$c_{\overline{F}}' = c_{\overline{F}} \|z\|_2 \text{ such as } z \in \mathbf{B}_{\mathbb{R}^d}^*(0, 1) = \{x \in \mathbb{R}^d; 0 < \|x\|_2 \leq 1\}.$$

$$\begin{aligned}
C &= \left( -\frac{1}{\gamma(x)} + o(1) \right) c_{\overline{F}} \|u\|_2 \text{ for all } u \in \Omega. \\
\tilde{m}_n(x) &= \mathbb{E} \left[ K_{h_n}(x - X_n) (\ln Y_n - \ln t_n) \mathbb{1}_{\{Y_n > t_n\}} \right]. \\
m_n(x) &= \mathbb{E} \left[ (\ln Y_n - \ln t_n) \mathbb{1}_{\{Y_n > t_n\}} | X_n = x \right].
\end{aligned}$$

Since we are interested in the asymptotic behavior of the estimator  $\widehat{\gamma}_n^H$ , we shall start by giving the asymptotic behavior of the estimator  $a_n$ .

**Theorem 1** (*Bias and variance of the estimator  $a_n$* )

Let Assumptions **(C1)** – **(C7)** hold.

1. If  $p \in (0, \alpha / (d + 2)]$ , then

$$\mathbb{E}(a_n(x)) = a(x) - \frac{C}{(1 - p\varepsilon)C_x} a(x) h_n \ln t_n + o(h_n \ln t_n). \quad (2.11)$$

If  $p \in (\alpha / (d + 2), 1/d)$ , then

$$\mathbb{E}(a_n(x)) = o(\sqrt{\gamma_n h_n^{-d}}). \quad (2.12)$$

2. If  $p \in (0, \alpha / (d + 2))$ , then

$$\mathbb{V}ar(a_n(x)) = o(h_n^2 \ln^2 t_n). \quad (2.13)$$

If  $p \in [\alpha / (d + 2), 1/d)$ , then

$$\mathbb{V}ar(a_n(x)) = \frac{1}{C_x} \frac{6}{2 - (\alpha - pd)\varepsilon} \|K^2\|_1 m_n(x) g(x) \gamma(x) \gamma_n h_n^{-d} + o(\gamma_n h_n^{-d}). \quad (2.14)$$

Departing from the above, we infer that the bias and the variance of the estimator  $a_n$  heavily depend on the choice of the stepsize ( $\gamma_n$ ). We consider two classical choices of ( $\gamma_n$ ). The first one rests on the minimization of the variance, while the second one relies on the Mean Integrated Squared Error (*MISE*).

**Choices of ( $\gamma_n$ ) minimizing the variance of the estimator  $a_n$ :**

As mentioned in [Mokkadem et al. \(2009a\)](#), by considering the point of view of estimation by confidence intervals, it is recommended to minimize the variance of the proposed estimator for confidence interval estimation (see also [Hall \(1992\)](#)).

**Corollary 1** *Let the assumptions of Theorem 1 hold. To minimize the asymptotic variance of the estimator  $a_n$ ,  $\alpha$  must be chosen equal to 1, ( $\gamma_n$ ) must satisfy  $\lim_{n \rightarrow \infty} n\gamma_n = 1 - pd$ , and we then have*

$$\mathbb{V}ar(a_n(x)) = \frac{6}{C_x} \frac{(1 - pd)}{nh_n^d} \|K^2\|_1 m_n(x) g(x) + o\left(\frac{1}{nh_n^d}\right).$$

The proof of Corollary 1 follows immediately from (2.14).

**Choices of  $(\gamma_n)$  minimizing the MISE of the estimator  $a_n$ :**

First, we have

$$MISE(a_n(x)) = \int_{\mathbb{R}^d} MSE(a_n(x))dx = \int_{\mathbb{R}^d} [(\mathbb{E}(a_n(x)) - a(x))^2 + \text{Var}(a_n(x))] dx.$$

The following proposition provides the MISE of the estimator  $a_n$ .

**Proposition 1** *Let Assumptions (C1)-(C7) hold.*

1. If  $p \in (0, \alpha / (d + 2))$ ,

$$MISE(a_n(x)) = \frac{C^2}{(1 - p\varepsilon)^2} h_n^2 \ln^2 t_n \int_{\mathbb{R}^d} \overline{F}^2(t_n|x) g^2(x) \gamma^2(x) dx + o(h_n^2 \ln^2 t_n).$$

2. If  $p = \alpha / (d + 2)$ ,

$$\begin{aligned} MISE(a_n(x)) &= \frac{C^2}{(1 - p\varepsilon)^2} h_n^2 \ln^2 t_n \int_{\mathbb{R}^d} \overline{F}^2(t_n|x) g^2(x) \gamma^2(x) dx \\ &+ \frac{6}{2 - (\alpha - pd)\varepsilon} \frac{\|K^2\|_1 m_n(x) g(x) \gamma(x)}{C_x} \gamma_n h_n^{-d} + o(h_n^2 \ln^2 t_n + \gamma_n h_n^{-d}). \end{aligned}$$

3. If  $p \in (\alpha / (d + 2), 1/d)$ ,

$$MISE(a_n(x)) = \frac{6}{2 - (\alpha - pd)\varepsilon} \frac{\|K^2\|_1 m_n(x) g(x) \gamma(x)}{C_x} \gamma_n h_n^{-d} + o(\gamma_n h_n^{-d}).$$

The following corollary is an immediate consequence of the second part of Proposition 1.

**Corollary 2** *Let Assumptions (C1)-(C7) hold. To minimize the MISE of the estimator  $a_n$ , the stepsize  $(\gamma_n)$  must be chosen in  $\mathcal{GS}(-1)$  such that  $\lim_{n \rightarrow \infty} n\gamma_n = 1$  and the bandwidth  $(h_n)$  must equal*

$$h_n = \left( \left[ \frac{3d(d+1)}{2(d+2)} \frac{\|K^2\|_1}{C^2 \ln^2(t_n)} \frac{\int_{\mathbb{R}^d} \overline{F}(t_n|x) g(x) \gamma^2(x) dx}{\int_{\mathbb{R}^d} \overline{F}^2(t_n|x) g^2(x) \gamma^2(x) dx} \right]^{\frac{1}{d+2}} \gamma_n^{\frac{1}{d+2}} \right).$$

Now, we treat the asymptotic behavior of the estimator  $b_n$ , in order to deduce the one of the estimator  $\widehat{\gamma}_n^H$ .

**Theorem 2** *(Bias and variance of the estimator  $b_n$ )*

*Let Assumptions (C1)-(C7) hold.*

1. If  $p \in (0, b / (d + 2)]$ , then

$$\mathbb{E}(b_n(x)) = b(x) + \frac{C}{1 - p\varepsilon_1} b(x) h_n \ln t_n + o(h_n \ln t_n). \quad (2.15)$$

If  $p \in (b / (d + 2), 1/d)$ , then

$$\mathbb{E}(b_n(x)) = b(x) - o(\sqrt{\beta_n h_n^{-d}}). \quad (2.16)$$



2. If  $p \in (0, b/(d+2))$ , then

$$\text{Var}(b_n(x)) = o(h_n^2 \ln^2 t_n). \quad (2.17)$$

If  $p \in [b/(d+2), 1/d)$ , then

$$\text{Var}(b_n(x)) = \frac{\beta_n}{h_n^d} \frac{1}{2 - (b - pd)\varepsilon_1} \|K\|_2^2 g(x) \bar{F}(t_n|x) + o\left(\frac{\beta_n}{h_n^d}\right). \quad (2.18)$$

The bias and the variance of the estimator  $b_n$  defined by the stochastic approximation algorithm (2.4), then heavily depend on the choice of the stepsize  $(\beta_n)$ . For an adequate choice, we consider two possible choices of  $(\beta_n)$ . The first one is based on the minimization of the variance, while the second one is grounded on the minimization of *MISE*.

**Choices of  $(\beta_n)$  minimizing the variance of the estimator  $b_n$ :**

As mentioned in Mokkadem et al. (2009a), it is recommended to minimize the variance of the proposed estimator for confidence interval estimation.

**Corollary 3** *Let the assumptions of Theorem 2 hold. To minimize the asymptotic variance of the estimator  $b_n$ ,  $b$  must be chosen equal to 1,  $(\beta_n)$  must satisfy  $\lim_{n \rightarrow \infty} n\beta_n = 1 - pd$ , and we then have*

$$\text{Var}(b_n(x)) = (1 - pd) \|K\|_2^2 g(x) \bar{F}(t_n|x) \frac{1}{nh_n^d} + o\left(\frac{1}{nh_n^d}\right).$$

The proof of Corollary 3 follows immediately from (2.18).

**Choices of  $(\beta_n)$  minimizing the *MISE* of the estimator  $b_n$ :**

The following proposition provides the *MISE* of the estimator  $b_n$ .

**Proposition 2** *Let Assumptions (C1)-(C7) hold.*

1. If  $p \in (0, b/(d+2))$ ,

$$\text{MISE}(b_n(x)) = \frac{C^2}{(1 - p\varepsilon_1)^2} h_n^2 \ln^2 t_n \int_{\mathbb{R}^d} \bar{F}^2(t_n|x) g^2(x) dx + o(h_n^2 \ln^2 t_n).$$

2. If  $p = b/(d+2)$ ,

$$\begin{aligned} \text{MISE}(b_n(x)) &= \frac{C^2}{(1 - p\varepsilon_1)^2} h_n^2 \ln^2 t_n \int_{\mathbb{R}^d} \bar{F}^2(t_n|x) g^2(x) dx \\ &+ \frac{1}{2 - (b - pd)\varepsilon_1} \|K\|_2^2 \frac{\beta_n}{h_n^d} \int_{\mathbb{R}^d} g(x) \bar{F}(t_n|x) dx + o\left(h_n^2 \ln^2 t_n + \frac{\beta_n}{h_n^d}\right). \end{aligned}$$

3. If  $p \in (b/(d+2), 1/d)$ ,

$$\text{MISE}(b_n(x)) = \frac{1}{2 - (b - pd)\varepsilon_1} \|K\|_2^2 \frac{\beta_n}{h_n^d} \int_{\mathbb{R}^d} g(x) \bar{F}(t_n|x) dx + o\left(\frac{\beta_n}{h_n^d}\right).$$

The following corollary is an immediate consequence of the second part of Proposition 2.

**Corollary 4** *Let Assumptions (C1)-(C7) hold. To minimize the MISE of the estimator  $b_n$ , the stepsize  $(\beta_n)$  must be chosen in  $\mathcal{GS}(-1)$  such that  $\lim_{n \rightarrow \infty} n\beta_n = 1$  and the bandwidth  $(h_n)$  must equal*

$$h_n = \left( \left[ \frac{d(d+1)}{4(d+2)} \frac{\|K\|_2^2}{C^2 \ln^2(t_n)} \frac{\int_{\mathbb{R}^d} \overline{F}(t_n|x)g(x) dx}{\int_{\mathbb{R}^d} \overline{F}^2(t_n|x)g^2(x)dx} \right]^{\frac{1}{d+2}} \beta_n^{\frac{1}{d+2}} \right).$$

Now we present the bias and the variance of  $\widehat{\gamma}_n^H$ .

**Theorem 3** *(Bias and variance of  $\widehat{\gamma}_n^H$ )*

*Let Assumptions (C1)-(C7) hold, and suppose that the stepsize  $(\beta_n) = (n^{-1})$ .*

1. *If  $p \in (0, \alpha/(d+2)]$ , then*

$$\begin{aligned} \mathbb{E}(\widehat{\gamma}_n^H(x)) - \gamma(x) &= - \left( \frac{C}{1-p\varepsilon} + \frac{C}{1-p} \right) \gamma(x) h_n \ln t_n \\ &\quad + o(h_n \ln t_n). \end{aligned} \quad (2.19)$$

*If  $p \in (\alpha/(d+2), 1/d)$ , then*

$$\mathbb{E}(\widehat{\gamma}_n^H(x)) - \gamma(x) = o\left(\sqrt{\frac{\gamma_n}{h_n^d}}\right). \quad (2.20)$$

2. *If  $p \in (0, \alpha/(d+2))$ , then*

$$\text{Var}(\widehat{\gamma}_n^H(x)) = o(h_n^2 \ln^2 t_n). \quad (2.21)$$

*If  $p \in [\alpha/(d+2), 1/d)$ , then*

$$\text{Var}(\widehat{\gamma}_n^H(x)) = \frac{1}{b^2(x)} \frac{1}{C_x} \frac{6}{2 - (\alpha - pd)\varepsilon} \|K^2\|_1 m_n(x) g(x) \gamma(x) \frac{\gamma_n}{h_n^d} + o\left(\frac{\gamma_n}{h_n^d}\right). \quad (2.22)$$

Clearly, the bias and the variance of the estimator  $\widehat{\gamma}_n^H$  depend on the choice of the two stepsizes  $(\gamma_n)$  and  $(\beta_n)$ .

Let us state now the following Theorem, which gives the weak convergence rate of the proposed recursive estimator  $\widehat{\gamma}_n^H$  defined in 2.7 in the special case of  $(\beta_n) = (n^{-1})$ .

**Theorem 4** *Let Assumptions (C1)-(C7) hold, and suppose that  $(\beta_n) = (n^{-1})$ .*

1. *If there exists  $r > 0$  such that  $\gamma_n^{-1} h_n^{d+2} \ln^2 t_n \xrightarrow[n \rightarrow \infty]{} r$  then*

$$\sqrt{\gamma_n^{-1} h_n^d} (\widehat{\gamma}_n^H(x) - \gamma(x)) \xrightarrow{\mathcal{D}} \mathcal{N}(\sqrt{r}\mathcal{B}(x), \mathcal{V}ar(x)),$$

where

$$\begin{aligned} \mathcal{B}(x) &= - \left( \frac{C}{1-p\varepsilon} + \frac{C}{1-p} \right) \gamma(x), \\ \mathcal{V}ar(x) &= \frac{1}{b^2(x)} \frac{1}{C_x} \frac{6}{2 - (\alpha - pd)\varepsilon} \|K^2\|_1 m_n(x) g(x) \gamma(x). \end{aligned}$$

2. If  $\gamma_n^{-1} h_n^{d+2} \ln^2 t_n \xrightarrow[n \rightarrow \infty]{} \infty$ , then

$$\frac{1}{h_n \ln t_n} (\widehat{\gamma}_n^H(x) - \gamma(x)) \xrightarrow{\mathbb{P}} \mathcal{B}(x),$$

where  $\xrightarrow{\mathcal{D}}$  denotes the convergence in distribution,  $\mathcal{N}$  the gaussian-distribution and  $\xrightarrow{\mathbb{P}}$  the convergence in probability.

We can consider the case when the stepsize  $(\beta_n)$  is chosen to minimise the variance of the estimator  $b_n$ . Similarly, we can obtain the weak convergence rate of the estimator  $\widehat{\gamma}_n^H$  defined in 2.7.

### 3 Simulation study

The target of our applications is to compare the performance of the proposed recursive kernel estimator of the conditional extreme value index given in (2.7) to that of Hill's non-recursive estimator defined in (2.8) using the "Leave One Out" cross-validation bandwidth selection.

#### 3.1 The study design

We use the following simulation design: we consider the unidimensional case  $d = 1$  and we simulate  $N = 500$  samples of size  $n$  ( $n = 50, 250$ ) of independent replicates  $(X_i, Y_i)$  where  $X$  is uniformly distributed on  $[0, 1]$  and the conditional distribution of  $Y_i$  given  $X_i = x$  is Pareto with parameter  $\gamma(x) = 0.5(0.1 + \sin(\pi x) \times (1.1 - 0.5 \exp(-64(x - 0.5)^2)))$  (this function was proposed by Daouia et al. (2011)), it was also used in Goegebeur et al. (2014) and in Ndao et al. (2016). The pattern of  $\gamma$  is given in Figure 1.

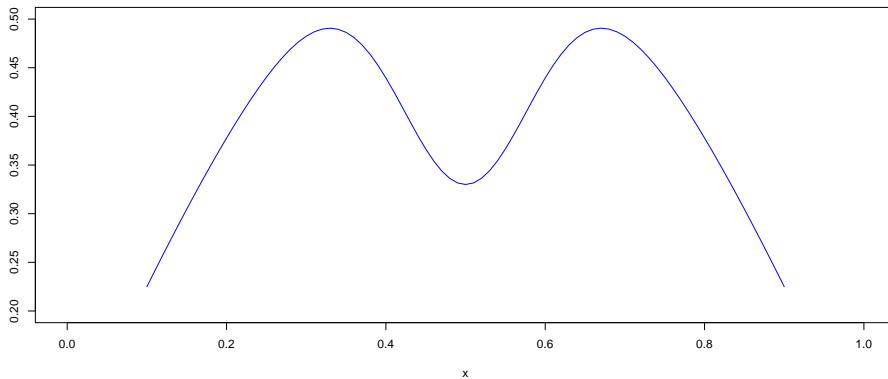


Figure 1: Pattern of  $\gamma(\cdot)$  on  $[0, 1]$

For each of the  $N$  simulated samples, we estimate  $\gamma(\cdot)$  at  $x = (0.1, 0.2, 0.3, \dots, 0.8, 0.9)$  using the estimator (2.7) with a biquadratic kernel  $K(x) = \frac{15}{16}(1 - x^2)^2 \mathbf{1}_{[-1,1]}$ . As mentioned in previous papers (see Slaoui (2014a,b)), there is no big influence on the choice of the kernel  $K$  in our setup when the observations are not contaminated. However, investigating deconvolution problem of such estimator is out of the scope of our area of research (see Slaoui (2019b)).

In order to calculate our estimator, we need to choose the bandwidth  $h$  and the threshold  $t_n$ . We take  $t_n$  to be the  $(n - k)$  th order statistic  $Y_{(n-k)}$  as is usual in extreme value statistics. Moreover, we propose an algorithm for choosing  $(h, k)$ . This algorithm adapted from [Goegebeur et al. \(2014\)](#), was considered recently by [Ndao et al. \(2016\)](#). The purpose is then to select the bandwidth  $h$  using the following cross-validation criterion

$$h_{cv} = \arg \min_{h \in \mathcal{H}} \sum_{i=1}^n \sum_{j=1}^n \left( \mathbf{1}_{\{Y_i \leq Y_j\}} - \widehat{F}_{n,-i}(Y_j | X_i) \right)^2,$$

where  $\mathcal{H} = \{h_n = cn^{-1}; n \geq 1 \text{ and } (c, v) \in \{0.1, 0.2, \dots, 0.9\}\}$  is a grid of values for  $h$  and

$$\widehat{F}_{n,-i}(y|x) := \frac{\sum_{j=1, j \neq i}^n K_h(x - X_j) \mathbf{1}_{\{Y_j \leq y\}}}{\sum_{j=1, j \neq i}^n K_h(x - X_j)}.$$

This criterion was introduced in [Yao \(1999\)](#), implemented by [Gannoun et al. \(2002\)](#), and established in an extreme value context by [Daouia et al. \(2011, 2013\)](#), [Goegebeur et al. \(2014\)](#) and [Ndao et al. \(2016\)](#). Using this bandwidth selection, we consider the following procedure to determine the number of threshold excesses  $k$ . This procedure rests on considering for each point  $x$ , the following steps:

Step 1: we compute the estimates for  $\widehat{\gamma}_{n, Y_{(n-k)}}^H(x)$  with  $k = 1, \dots, n - 1$ .

Step 2: we constitute several successive "blocks" of the estimates  $\widehat{\gamma}_{n, Y_{(n-k)}}^H(x)$  (one block for  $k \in \{1, \dots, 15\}$ , a second block for  $k \in \{16, \dots, 30\}$  and so on).

Step 3: we calculate the standard deviation of the estimate in each block.

Step 4: we determine the  $k$ -value (denoted by  $k_1$ ) from the block with minimal standard deviation (in particular, we take the median of the  $k$ -values in that block).

Finally, we estimate  $\gamma(x)$  by using the estimator  $\widehat{\gamma}_n^H$  (2.7) by taking  $(h, k) = (h_{cv}, k_1)$

## 3.2 Results

For each configuration of the simulation design parameters (sample size  $n$ , stepsize parameters  $(\gamma_n, \beta_n)$  and covariate value  $x$ ), we calculate the average *IAE* (Integrated Absolute Error), the average *ISE* (Integrated Squared Error) and  $L_\infty$  of the estimators over  $N = 500$  trials;  $\overline{IAE} = \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}} |\widehat{g}^{[i]}(x) - \gamma(x)| dx$ ,  $\overline{ISE} = \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}} (\widehat{g}^{[i]}(x) - \gamma(x))^2 dx$  and  $L_\infty = \max_{i=1, \dots, N} \int_{\mathbb{R}} |\widehat{g}^{[i]}(x) - \gamma(x)| dx$ , where  $\widehat{g}^{[i]}$  corresponds to the estimator computed from the  $i$ th sample. In order to investigate the comparison estimators, we consider the stepsizes  $(\gamma_n, \beta_n)$  equal to  $(n^{-1}, n^{-1})$ ,  $((2/3)n^{-1}, n^{-1})$ ,  $(n^{-1}, (2/3)n^{-1})$  and  $((2/3)n^{-1}, (2/3)n^{-1})$  respectively. These four choices of parameters of the recursive estimator are referred to as *R1*, *R2*, *R3* and *R4* respectively. Results are highlighted in [Table 1](#). We point out that the major merit of our proposed estimator lies in its update aspect. Indeed, when new sample points are available, it requires less computational cost than non-recursive estimator. Moreover, [Table 1](#) reveals that our proposed recursive estimator can provide better results in some specific situations that are very close in general to the reference values, which proves the effectiveness of our proposed recursive estimator in terms of the estimation error. [Figure 2](#) discloses that all the considered estimators yield good results since the values of  $\gamma$  at each point  $x \in \{0.1, 0.2, \dots, 0.9\}$  are very close to the median.

$n$	$\gamma(0.1) = 0.225$					$\gamma(0.2) = 0.3777$					$\gamma(0.3) = 0.4824$				
	$NR$	$R1$	$R2$	$R3$	$R4$	$NR$	$R1$	$R2$	$R3$	$R4$	$NR$	$R1$	$R2$	$R3$	$R4$
50	<b>0.2095</b>	0.1829	0.1589	0.2089	0.1814	0.4012	<b>0.3901</b>	0.3446	0.4057	0.3583	0.3566	0.3647	0.3324	<b>0.3843</b>	0.3503
	<b>0.0224</b>	0.0238	0.0335	0.023	0.0227	<b>0.023</b>	0.0239	0.0326	0.0301	0.0249	<b>0.0223</b>	0.0241	0.0344	0.0306	0.0259
	<b>0.0977</b>	0.1003	0.1087	0.1012	0.0994	<b>0.0995</b>	0.1006	0.1083	0.1083	0.1015	<b>0.0982</b>	0.1015	0.1082	0.1105	0.1036
	0.4173	0.4028	0.6902	<b>0.3668</b>	0.4225	0.41	<b>0.3877</b>	0.5269	0.5114	0.4687	<b>0.3381</b>	0.4909	0.7606	0.6507	0.46
250	$NR$	$R1$	$R2$	$R3$	$R4$	$NR$	$R1$	$R2$	$R3$	$R4$	$NR$	$R1$	$R2$	$R3$	$R4$
	0.3189	0.3166	<b>0.2971</b>	0.3312	0.3108	0.3355	0.3115	<b>0.3391</b>	0.3	0.3253	<b>0.4204</b>	0.3709	0.3857	0.3744	0.3893
	<b>0.0036</b>	0.0036	0.0042	0.0042	0.0038	<b>0.0037</b>	0.0038	0.0043	0.0042	0.004	<b>0.0038</b>	0.0039	0.0044	0.0044	0.004
	<b>0.0422</b>	0.0425	0.0457	0.0453	0.0438	<b>0.0427</b>	0.0435	0.0465	0.0457	0.0447	<b>0.0433</b>	0.0436	0.0466	0.0466	0.045
	<b>0.088</b>	0.0883	0.1025	0.1016	0.0957	0.0995	0.1039	<b>0.0926</b>	0.1069	0.1237	0.1265	0.1253	<b>0.1156</b>	0.1352	0.1177
$n$	$\gamma(0.4) = 0.4395$					$\gamma(0.5) = 0.33$					$\gamma(0.6) = 0.4395$				
	$NR$	$R1$	$R2$	$R3$	$R4$	$NR$	$R1$	$R2$	$R3$	$R4$	$NR$	$R1$	$R2$	$R3$	$R4$
50	0.3936	0.3723	0.357	<b>0.3983</b>	0.3819	0.4181	<b>0.4306</b>	0.3539	0.5089	0.4183	0.3578	0.3581	0.3029	<b>0.4345</b>	0.3676
	<b>0.0248</b>	0.0259	0.0352	0.0335	0.0282	0.0245	<b>0.0239</b>	0.0314	0.031	0.026	<b>0.0243</b>	0.0249	0.0336	0.031	0.0265
	<b>0.1011</b>	0.1029	0.1102	0.1124	0.1057	0.1021	<b>0.1011</b>	0.1071	0.1095	0.1029	<b>0.1017</b>	0.1018	0.1088	0.1091	0.1035
	<b>0.4756</b>	0.4909	0.7606	0.6464	0.5454	<b>0.3842</b>	0.3877	0.5269	0.5114	0.4114	<b>0.4756</b>	0.4909	0.7606	0.5158	0.5454
250	$NR$	$R1$	$R2$	$R3$	$R4$	$NR$	$R1$	$R2$	$R3$	$R4$	$NR$	$R1$	$R2$	$R3$	$R4$
	<b>0.4269</b>	0.5024	0.4624	0.523	0.4814	0.382	0.359	<b>0.346</b>	0.3736	0.3605	0.3848	0.3862	<b>0.4339</b>	0.382	0.4292
	<b>0.0036</b>	0.0037	0.0043	0.0043	0.004	<b>0.0038</b>	0.0039	0.0045	0.0043	0.004	<b>0.0037</b>	0.0038	0.0044	0.0044	0.004
	<b>0.0425</b>	0.0428	0.0462	0.0458	0.0449	<b>0.0432</b>	0.0435	0.0472	0.046	0.0448	<b>0.0425</b>	0.0431	0.0463	0.046	0.0444
	0.0952	0.0895	<b>0.0941</b>	0.1173	0.1039	<b>0.0925</b>	0.1099	0.1283	0.1006	0.1125	0.098	0.0894	0.1218	0.1216	<b>0.0968</b>
$n$	$\gamma(0.7) = 0.4824$					$\gamma(0.8) = 0.3777$					$\gamma(0.9) = 0.2250$				
	$NR$	$R1$	$R2$	$R3$	$R4$	$NR$	$R1$	$R2$	$R3$	$R4$	$NR$	$R1$	$R2$	$R3$	$R4$
50	0.4085	0.4032	0.3457	<b>0.4837</b>	0.4147	0.3496	0.3495	0.3223	<b>0.3680</b>	0.3394	0.5872	0.5987	0.4448	<b>0.3635</b>	0.4021
	<b>0.0239</b>	0.0257	0.0348	0.0291	0.0255	<b>0.0217</b>	0.0225	0.0305	0.0285	0.0257	<b>0.0205</b>	0.0231	0.0318	0.027	0.0234
	<b>0.0999</b>	0.1024	0.1097	0.1081	0.1032	<b>0.0982</b>	0.0995	0.1062	0.1079	0.1034	<b>0.0964</b>	0.0987	0.1078	0.105	0.1005
	<b>0.4801</b>	0.4909	0.7606	0.5158	0.5192	<b>0.3284</b>	0.3396	0.5352	0.3997	0.39	<b>0.3037</b>	0.4909	0.7606	0.5114	0.3913
250	$NR$	$R1$	$R2$	$R3$	$R4$	$NR$	$R1$	$R2$	$R3$	$R4$	$NR$	$R1$	$R2$	$R3$	$R4$
	0.4431	<b>0.4751</b>	0.5599	0.4351	0.5128	<b>0.3708</b>	0.3571	0.405	0.3071	0.3483	0.2428	0.2439	<b>0.2223</b>	0.2372	0.2162
	<b>0.0038</b>	0.0039	0.0045	0.0044	0.004	<b>0.0037</b>	0.0037	0.0043	0.0041	0.0039	<b>0.0034</b>	0.0036	0.0041	0.0041	0.0037
	<b>0.0433</b>	0.0436	0.0469	0.0462	0.0449	<b>0.0426</b>	0.043	0.0464	0.0453	0.0441	<b>0.0414</b>	0.0421	0.0455	0.0447	0.0433
	<b>0.092</b>	0.1177	0.1152	0.12	0.0994	<b>0.0959</b>	0.1482	0.0986	0.1484	0.0925	<b>0.0825</b>	0.0888	0.0922	0.0899	

Table 1: Simulation results for  $\gamma(x)$ .  $N = 500$  trials are considered. For each configuration of the simulated parameters  $(n, \gamma_n, \beta_n, x)$ , the first line indicates the value of each estimator in different points  $x$ . Lines two, three and four provide the averages  $IAEs$ ,  $ISEs$  and  $L_\infty$  respectively of the five different kernel density estimators.  $NR$  corresponds to the non-recursive estimator and  $R1$ ,  $R2$ ,  $R3$  and  $R4$  correspond to the proposed recursive estimators when  $(\gamma_n, \beta_n)$  takes  $(n^{-1}, n^{-1})$ ,  $((2/3)n^{-1}, n^{-1})$ ,  $(n^{-1}, (2/3)n^{-1})$  and  $((2/3)n^{-1}, (2/3)n^{-1})$  respectively.

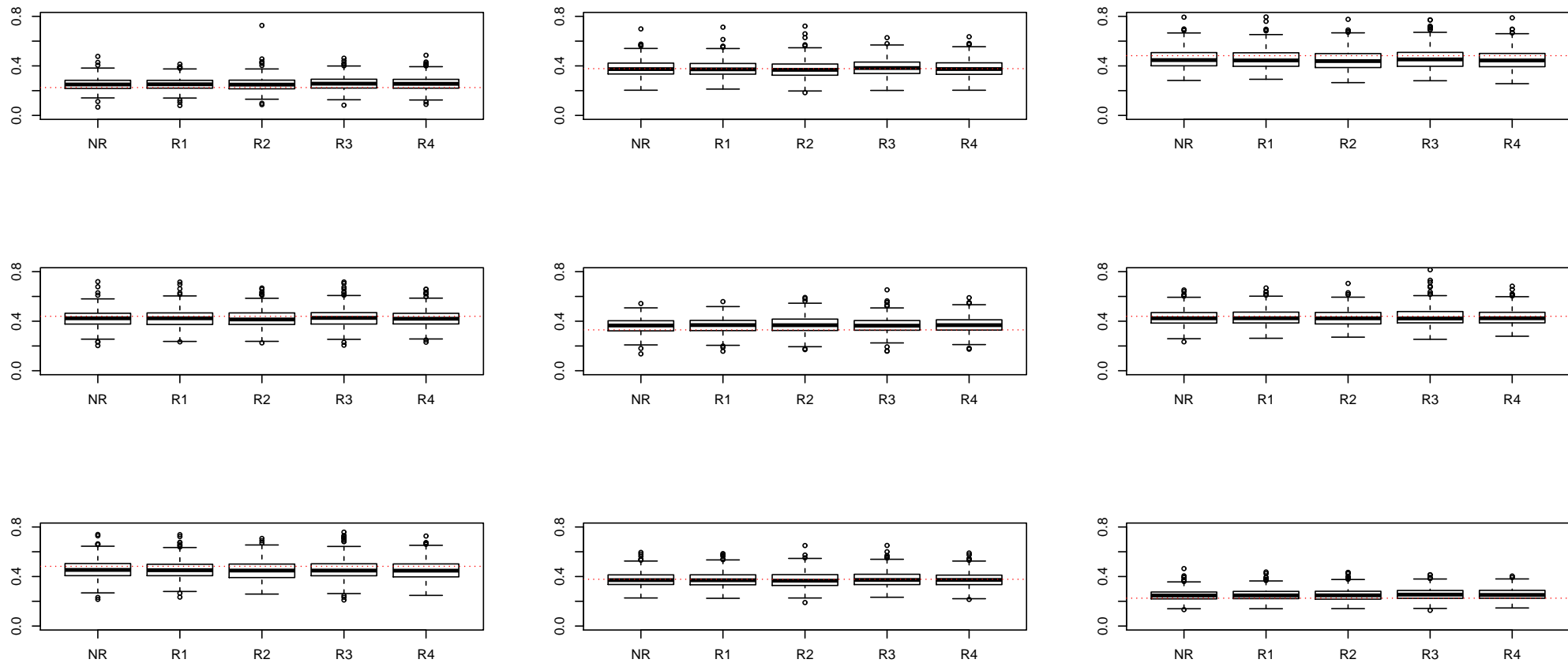


Figure 2: Boxplots of the  $N = 500$  estimates of our five considered estimators ( $NR$ ,  $R1$ ,  $R2$ ,  $R3$ ,  $R4$ ) in points  $x = 0.1, 0.2, 0.3$  (1st line),  $x = 0.4, 0.5, 0.6$  (2nd line) and  $x = 0.7, 0.8, 0.9$  (3rd line), and dashed lines represent the values of reference estimator  $\gamma(\cdot)$  in each point  $x$  as mentioned above.

	ISE	IAE
Recursive estimator	<b>0.00639654</b>	<b>0.0599912</b>
Non-recursive estimator	0.009305902	0.07006466

Table 2: The comparison between errors of the non-recursive estimator (2.8) and the proposed recursive estimator (2.7).

### 3.3 Real dataset

We considered a Malaria dataset of 176 families in Senegal, totalizing 505 children between 2 and 19 years old, living in two villages of Niakhar (Toucar and Diohine). The number of observations was 6986. We measured *Plasmodium falciparum* Parasite Load (PL) from thick blood smears obtained by finger-prick during two different seasons and regularly over a three-year observation period (2001-2003). The number of measurements per child ranged from 1 to 15. We refer readers to consult Milet et al. (2010) for more details about data. These data were used also in Slaoui and Nuel (2014c) in a parametric context and more recently in Slaoui (2019a) in a non-parametric context. In the real example, the proposed recursive estimator (2.7) was compared to Hill’s non-recursive kernel estimator (2.8) proposed by Goegebeur et al. (2014). Therefore, for any considered estimator  $\hat{\gamma}$  of the index function  $\gamma$ , we propose to compute *IAE* and *ISE* defined as:

$$IAE(\hat{\gamma}) = \int_{\mathbb{R}} |\hat{\gamma}(x) - \gamma(x)| dx$$

and

$$ISE(\hat{\gamma}) = \int_{\mathbb{R}} (\hat{\gamma}(x) - \gamma(x))^2 dx.$$

Departing from Table 2 and Figure 3, we infer that the *IAE* and the *ISE* of the proposed recursive estimator are smaller than those of the non-recursive estimator set forward by Goegebeur et al. (2014). Thus, demonstrating the effectiveness of our considered estimator.

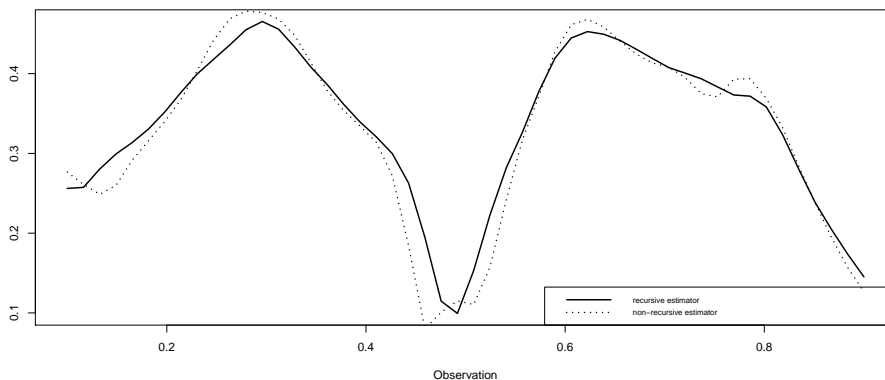


Figure 3: Qualitative comparison between the non-recursive estimator (2.8) and the proposed recursive one (2.7).

## 4 Conclusion

In this research paper, we tackled the estimation of the conditional extreme value index  $\gamma(x)$  of a heavy-tailed distribution when some random covariate information is available. We elaborate recursive kernel estimator of the extreme value index function based on the stochastic approximation algorithm. The proposed estimator asymptotically follows normal distribution. We subsequently compared the proposed estimator to Hill's non-recursive extreme value index estimator. We demonstrated that using some particular stepsizes and a specific bandwidth selection through a cross-validation procedure, the proposed recursive estimator could be very competitive to the non-recursive version. Moreover, we highlighted that the proposed estimator is much better in terms of computational costs. Numerical results illustrate the effectiveness of our recursive approach. To this extent, we would state that although our work is an extension to a wealthy historical background, it may be taken further, extended and built upon since it offers different perspectives and opens new horizons for future research. We can extend our recursive extreme value index estimator to the case of censored data. We can also propose a new estimator of the conditional extreme quantile using our recursive estimator defined by (2.7) and compare it to the classical Weissman estimator. Another direction is to investigate the almost sure convergence and the large and moderate deviation principles of the proposed estimator, which requires non trivial mathematics. This would go well beyond the scope of the present paper.

## Funding

This work benefited from the financial support of the GDR 3477 GeoSto. It is supported also by LR18E16: Analyse, Géométrie et Applications, University of Monastir (Tunisia).

## Acknowledgments

This work was performed while the first author was at the Laboratory of Mathematics and Applications of Poitiers as a visiting scientist, partly funded by the University of Monastir. The authors would like to thank the Editor-in-Chief Prof. Narayanaswamy Balakrishnan of Journal of Communications in Statistics - Theory and Methods and the reviewers for their helpful comments, which helped us focus on improving the original version of the paper.

## A Proofs

We introduce the following Lemmas that will enable us to obtain the asymptotic expansion of  $a_n$ .

**Lemma 1** *Let assumption (C3) holds. Then, for  $t_n \rightarrow \infty$  as  $n \rightarrow \infty$  we have*

$$m_n(x) = \gamma(x)\overline{F}(t_n|x)C_x.$$

The proof of Lemma 1 is presented in Goegebeur et al. (2014).

**Lemma 2** *Let assumptions (C1)-(C6) hold. Then, for all  $x \in \mathbb{R}^d$  such that  $g(x) > 0$  we have*



for  $t_n \xrightarrow[n \rightarrow \infty]{} \infty$  and  $h_n \xrightarrow[n \rightarrow \infty]{} 0$  with  $h_n \ln t_n \xrightarrow[n \rightarrow \infty]{} 0$

$$\tilde{m}_n(x) = m_n(x)g(x) \left\{ 1 + \frac{C}{C_x} h_n \ln t_n + o(h_n \ln t_n) \right\}. \quad (\text{A.1})$$

### Proof of Lemma 2

$$\begin{aligned} \tilde{m}_n(x) &= \mathbb{E} [K_{h_n}(x - X_n)(\ln Y_n - \ln t_n)\mathbb{1}_{\{Y > t_n\}}] \\ &= \mathbb{E} [K_{h_n}(x - X_n)\mathbb{E}((\ln Y_n - \ln t_n)\mathbb{1}_{\{Y > t_n\}}|X_n)] \\ &= \mathbb{E} [K_{h_n}(x - X_n)m_n(X_n)] \\ &= \int_{\mathbb{R}^d} K(z)m_n(x - h_n z)g(x - h_n z)dz. \end{aligned}$$

An integration by parts then yields

$$\tilde{m}_n(x) = \int_{\Omega} K(z) \int_{t_n}^{\infty} \frac{\overline{F}(y|x - h_n z)}{y} dy g(x - h_n z) dz,$$

and consequently,

$$\begin{aligned} \tilde{m}_n(x) - m_n(x)g(x) &= \int_{\Omega} K(z) \int_{t_n}^{\infty} \frac{\overline{F}(y|x - h_n z)}{y} dy g(x - h_n z) dz - g(x) \int_{t_n}^{\infty} \frac{\overline{F}(y|x)}{y} dy \\ &\stackrel{(\text{C6})}{=} \int_{\Omega} K(z) \int_{t_n}^{\infty} \frac{\overline{F}(y|x)}{y} \frac{\overline{F}(y|x - h_n z)}{\overline{F}(y|x)} dy (g(x - h_n z) - g(x) + g(x)) dz \\ &\quad - \int_{\Omega} K(z) \int_{t_n}^{\infty} \frac{\overline{F}(y|x)}{y} dy (g(x) - g(x - h_n z) + g(x - h_n z)) dz \\ &= \int_{\Omega} K(z) \int_{t_n}^{\infty} \frac{\overline{F}(y|x)}{y} \left( \frac{\overline{F}(y|x - h_n z)}{\overline{F}(y|x)} - 1 \right) dy g(x) dz \\ &\quad + \int_{\Omega} K(z) \int_{t_n}^{\infty} \frac{\overline{F}(y|x)}{y} dy (g(x - h_n z) - g(x)) dz \\ &\quad + \int_{\Omega} K(z) \int_{t_n}^{\infty} \frac{\overline{F}(y|x)}{y} \left( \frac{\overline{F}(y|x - h_n z)}{\overline{F}(y|x)} - 1 \right) dy (g(x - h_n z) - g(x)) dz \\ &=: \mathbb{I}_4 + \mathbb{I}_5 + \mathbb{I}_6. \end{aligned} \quad (\text{A.2})$$

The expression of  $\mathbb{I}_4$  can be written as follows

$$\mathbb{I}_4 = \int_{\Omega} K(z) \tilde{\mathbb{I}}_4 g(x) dz,$$

with

$$\tilde{\mathbb{I}}_4 = \int_{t_n}^{\infty} \frac{\overline{F}(y|x)}{y} \left( \frac{\overline{F}(y|x - h_n z)}{\overline{F}(y|x)} - 1 \right) dy,$$

and grounded on the fact that

$$\frac{\overline{F}(y|x - h_n z)}{\overline{F}(y|x)} = \exp \left[ \ln \overline{F}(y|x) \left( \frac{\ln \overline{F}(y|x - h_n z)}{\ln \overline{F}(y|x)} - 1 \right) \right], \quad (\text{A.3})$$

under **(C5)**, we readily obtain

$$\frac{\ln \bar{F}(y|x - h_n z)}{\ln \bar{F}(y|x)} - 1 = c'_F \sqrt{\sum_{i=1}^d (h_n z_i)^2} = c'_F h_n.$$

Moreover, using the following property:  $\frac{\ln l(y|x)}{\ln y} \rightarrow 0$  as  $y \rightarrow \infty$ , ensures that

$$\begin{aligned} \ln \bar{F}(y|x) \left( \frac{\ln \bar{F}(y|x - h_n z)}{\ln \bar{F}(y|x)} - 1 \right) &= \ln \bar{F}(y|x) c'_F h_n \\ &= \left( -\frac{1}{\gamma(x)} \ln y + \ln l(y|x) \right) c'_F h_n \\ &= \left( -\frac{1}{\gamma(x)} + o(1) \right) c'_F h_n \ln y \\ &= C h_n \ln y, \end{aligned}$$

for some positive constant  $C = \left( -\frac{1}{\gamma(x)} + o(1) \right) c'_F$ . Using the following Taylor's formula,  $\exp(x) - 1 = x + \frac{x^2}{2} + o(1)$ , we readily obtain:

$$\begin{aligned} \tilde{\mathbb{I}}_4 &= \int_{t_n}^{\infty} \frac{\bar{F}(y|x)}{y} \left( \exp \left[ \ln \bar{F}(y|x) \left( \frac{\ln \bar{F}(y|x - h_n z)}{\ln \bar{F}(y|x)} - 1 \right) \right] - 1 \right) dy \\ &= \int_{t_n}^{\infty} \frac{\bar{F}(y|x)}{y} (\exp(C h_n \ln y) - 1) dy \\ &= C h_n \int_{t_n}^{\infty} \frac{\bar{F}(y|x)}{y} \ln y dy + \frac{C^2 h_n^2}{2} \int_{t_n}^{\infty} \frac{\bar{F}(y|x)}{y} \ln^2 y dy + o(1) \\ &=: \tilde{\mathbb{I}}_{4,1} + \tilde{\mathbb{I}}_{4,2} + \tilde{\mathbb{I}}_{4,3}. \end{aligned}$$

Concerning  $\tilde{\mathbb{I}}_{4,1}$ , by integration by parts and using **(C1)** and **(C2)**, we get

$$\begin{aligned} \tilde{\mathbb{I}}_{4,1} &= -C h_n \bar{F}(t_n|x) (\ln t_n - 1) - C h_n \int_{t_n}^{\infty} \frac{\partial \bar{F}(y|x)}{\partial y} (\ln y - 1) dy + C h_n \int_{t_n}^{\infty} \frac{\bar{F}(y|x)}{y} (\ln y - 1) dy \\ &= -C h_n \bar{F}(t_n|x) (\ln t_n - 1) + \left( 1 + \frac{1}{\gamma(x)} + o(1) \right) \left( C h_n \int_{t_n}^{\infty} \frac{\bar{F}(y|x)}{y} \ln y dy \right) \\ &\quad - \left( C h_n \left( 1 + \frac{1}{\gamma(x)} \right) + o(h_n) \right) \int_{t_n}^{\infty} \frac{\bar{F}(y|x)}{y} dy \\ &= -C h_n \bar{F}(t_n|x) (\ln t_n - 1) + \left( 1 + \frac{1}{\gamma(x)} + o(1) \right) \tilde{\mathbb{I}}_{4,1} - \left( C h_n \left( 1 + \frac{1}{\gamma(x)} \right) + o(h_n) \right) m_n(x). \end{aligned}$$

Hence, it follows that

$$\tilde{\mathbb{I}}_{4,1} = \frac{C h_n \bar{F}(t_n|x) (\ln t_n - 1)}{\left( \frac{1}{\gamma(x)} + o(1) \right)} + \frac{C h_n \left( 1 + \frac{1}{\gamma(x)} + o(1) \right) m_n(x)}{\left( \frac{1}{\gamma(x)} + o(1) \right)}$$

$$= \frac{Ch_n \overline{F}(t_n|x) \gamma(x) \ln t_n C_x}{(1+o(1)) C_x} - \frac{Ch_n \overline{F}(t_n|x) \gamma(x) C_x}{(1+o(1)) C_x} + \frac{Ch_n (1+\gamma(x)) (1+o(1)) m_n(x)}{1+o(1)},$$

with

$$C_x = 1 + \frac{b(t_n|x)}{\gamma(x)\rho(x)} \left[ \frac{1}{1-\rho(x)} - 1 \right] (1+o(1));$$

which can be written as

$$\begin{aligned} \tilde{\mathbb{I}}_{4,1} &= \frac{C}{C_x} h_n \ln t_n m_n(x) (1+o(1)) - \frac{C}{C_x} h_n m_n(x) (1+o(1)) \\ &\quad + C (1+\gamma(x)) h_n m_n(x) (1+o(1)) \\ &= \frac{C}{C_x} h_n \ln t_n m_n(x) (1+o(1)). \end{aligned} \tag{A.4}$$

Concerning  $\tilde{\mathbb{I}}_{4,2}$ , integration by parts and using **(C1)** and **(C2)**, we get

$$\begin{aligned} \tilde{\mathbb{I}}_{4,2} &= -\frac{C^2 h_n^2}{2} \overline{F}(t_n|x) (\ln^2 t_n - 2 \ln t_n + 2) - \frac{C^2 h_n^2}{2} \int_{t_n}^{\infty} \frac{\partial \overline{F}(y|x)}{\partial y} (\ln^2 y - 2 \ln y + 2) dy \\ &\quad + \frac{C^2 h_n^2}{2} \int_{t_n}^{\infty} \frac{\overline{F}(y|x)}{y} (\ln^2 y - 2 \ln y + 2) dy \\ &= -\frac{C^2 h_n^2}{2} \overline{F}(t_n|x) (\ln^2 t_n - 2 \ln t_n + 2) \\ &\quad + \frac{C^2 h_n^2}{2} \left( \frac{1}{\gamma(x)} + o(1) \right) \int_{t_n}^{\infty} \frac{\overline{F}(y|x)}{y} (\ln^2 t_n - 2 \ln t_n + 2) dy \\ &\quad + \frac{C^2 h_n^2}{2} \int_{t_n}^{\infty} \frac{\overline{F}(y|x)}{y} \ln^2 y dy - C^2 h_n^2 \int_{t_n}^{\infty} \frac{\overline{F}(y|x)}{y} \ln y dy + C^2 h_n^2 \int_{t_n}^{\infty} \frac{\overline{F}(y|x)}{y} dy \\ &= -\frac{C^2 h_n^2}{2} \overline{F}(t_n|x) \ln^2 t_n + C^2 h_n^2 \overline{F}(y|x) \ln t_n - C^2 h^2 \overline{F}(y|x) + \left( 1 + \frac{1}{\gamma(x)} + o(1) \right) \tilde{\mathbb{I}}_{4,2} \\ &\quad - Ch_n \left( 1 + \frac{1}{\gamma(x)} + o(1) \right) \tilde{\mathbb{I}}_{4,1} + C^2 h_n^2 \left( 1 + \frac{1}{\gamma(x)} + o(1) \right) m_n(x). \end{aligned}$$

Then, it comes that

$$\begin{aligned} \tilde{\mathbb{I}}_{4,2} &= \frac{C^2}{2C_x} h_n^2 \gamma(x) \overline{F}(t_n|x) C_x \ln^2 t_n (1+o(1)) - \frac{C^2}{C_x} h_n^2 \gamma(x) \overline{F}(t_n|x) C_x \ln t_n (1+o(1)) \\ &\quad + C^2 h_n^2 \gamma(x) \overline{F}(t_n|x) (1+o(1)) + Ch_n (1+\gamma(x)) (1+o(1)) \tilde{\mathbb{I}}_{4,1} \\ &\quad - C^2 h_n^2 (1+\gamma(x)) (1+o(1)) m_n(x) \\ &= \frac{C^2}{2C_x} h_n^2 \ln^2 t_n m_n(x) (1+o(1)). \end{aligned} \tag{A.5}$$

Moreover, note that

$$\tilde{\mathbb{I}}_{4,3} = o(1). \tag{A.6}$$

Now, the combination of **(A.4)**, **(A.5)** **(A.6)**, ensures that

$$\tilde{\mathbb{I}}_4 = \frac{C}{C_x} h_n \ln t_n m_n(x) (1+o(1)).$$

Furthermore, we have

$$\mathbb{I}_4 = \int_{\Omega} K(z) \tilde{\mathbb{I}}_4 g(x) dz = \frac{C}{C_x} h_n \ln t_n m_n(x) g(x) (1 + o(1)). \quad (\text{A.7})$$

Concerning  $\mathbb{I}_5$ , we have under the assumption **(C4)** and **(C6)**

$$\begin{aligned} \mathbb{I}_5 &= \int_{\Omega} K(z) \int_{t_n}^{\infty} \frac{\bar{F}(y|x)}{y} dy (g(x - h_n z) - g(x)) dz \\ &= c_g m_n(x) \int_{\Omega} K(z) d(x - h_n z, x) dz \\ &= o(1). \end{aligned} \quad (\text{A.8})$$

Moreover, under the assumption **(C4)**-**(C6)**, we have, for  $t_n$  sufficiently large,

$$\begin{aligned} \mathbb{I}_6 &= \int_{\Omega} K(z) \int_{t_n}^{\infty} \frac{\bar{F}(y|x)}{y} \left( \frac{\bar{F}(y|x - h_n z)}{\bar{F}(y|x)} - 1 \right) dy (g(x - h_n z) - g(x)) dz \\ &= o(1). \end{aligned} \quad (\text{A.9})$$

The combination of **(A.2)**, **(A.7)**, **(A.8)** and **(A.9)**, ensures that

$$\tilde{m}_n(x) - m_n(x)g(x) = \frac{C}{C_x} h_n \ln t_n m_n(x)g(x) (1 + o(1)) = m_n(x)g(x) \left\{ \frac{C}{C_x} h_n \ln t_n (1 + o(1)) \right\},$$

which gives **(A.1)**.

**Lemma 3** *Let assumptions **(C1)** and **(C4)**-**(C6)** hold. Then, for all  $x \in \mathbb{R}^d$  such that  $g(x) > 0$ , we have for  $t_n \xrightarrow{n \rightarrow \infty} \infty$  and  $h_n \xrightarrow{n \rightarrow \infty} 0$  with  $h_n \ln t_n \xrightarrow{n \rightarrow \infty} 0$*

$$\mathbb{E} [K_{h_n}(x - X_n) \mathbb{1}_{\{Y_n > t_n\}}] = g(x) \bar{F}(t_n|x) (1 + Ch_n \ln t_n + o(h_n \ln t_n)).$$

**Proof of Lemma 3** Since  $(X_i, Y_i)$ ,  $i = 1, \dots, n$  are independent and identically distributed, we have under the assumption **(C6)**

$$\begin{aligned} \mathbb{E} [K_{h_n}(x - X_n) \mathbb{1}_{\{Y_n > t_n\}}] &= \int_{\mathbb{R}^d} \int_{\mathbb{R}} \frac{1}{h_n^d} K\left(\frac{x-t}{h_n}\right) \mathbb{1}_{\{y > t_n\}} f(y|t) g(t) dt dy \\ &= \int_{\mathbb{R}^d} \frac{1}{h_n^d} K\left(\frac{x-t}{h_n}\right) \bar{F}(t_n|t) g(t) dt \\ &= \int_{\Omega} K(u) \bar{F}(t_n|x - uh_n) g(x - uh_n) du. \end{aligned}$$

Now, we consider

$$\begin{aligned} \mathbb{E} [K_{h_n}(x - X_n) \mathbb{1}_{\{Y_n > t_n\}}] - \bar{F}(t_n|x)g(x) &= \bar{F}(t_n|x) \int_{\Omega} K(u) \left( \frac{\bar{F}(t_n|x - uh_n)}{\bar{F}(t_n|x)} - 1 \right) g(x - h_n u) du \\ &\quad + \bar{F}(t_n|x) \int_{\Omega} K(u) (g(x - h_n u) - g(x)) du \\ &=: \tilde{\mathbb{J}}_1 + \tilde{\mathbb{J}}_2. \end{aligned}$$

Concerning  $\tilde{\mathbb{J}}_1$ , under the assumption **(C5)** and using the equation **(A.3)**, we have

$$\begin{aligned}\frac{\overline{F}(t_n|x - uh_n)}{\overline{F}(t_n|x)} &= \exp [\ln (\overline{F}(t_n|x)) c_{\overline{F}}\|u\|_2 h_n] \\ &= \exp \left[ \ln t_n \left( -\frac{1}{\gamma(x)} + o(1) \right) c_{\overline{F}}\|u\|_2 h_n \right] \\ &= \exp [\ln t_n C h_n].\end{aligned}$$

Moreover, since  $g(x) > 0$ , the application of Taylor's formula, ensures that

$$\begin{aligned}\tilde{\mathbb{J}}_1 &= g(x) \overline{F}(t_n|x) \int_{\Omega} K(u) [C h_n \ln t_n + o(h_n \ln t_n)] \frac{g(x - h_n u)}{g(x)} du \\ &= C g(x) \overline{F}(t_n|x) h_n \ln t_n + o(\overline{F}(t_n|x) h_n \ln t_n).\end{aligned}$$

Under **(C5)**, we have

$$\tilde{\mathbb{J}}_2 = \overline{F}(t_n|x) c_g \int_{\Omega} \|u\|_2 h_n K(u) du = o(g(x) \overline{F}(t_n|x) h_n \ln t_n).$$

Then, we get

$$\mathbb{E} [K_{h_n}(x - X_n) \mathbf{1}_{\{Y_n > t_n\}}] = g(x) \overline{F}(t_n|x) + C g(x) \overline{F}(t_n|x) h_n \ln t_n + o(\overline{F}(t_n|x) h_n \ln t_n).$$

We state now the following technical lemma, which is proved in [Mokkadem et al. \(2009a\)](#), and which will be used throughout the demonstrations.

**Lemma 4**

Let  $(v_n) \in \mathcal{GS}(v^*)$ ,  $(\gamma_n) \in \mathcal{GS}(-\alpha)$  and  $m > 0$  such that  $m - v^* \varepsilon > 0$  where  $\varepsilon$  is defined in **(2.9)**, and  $\Pi_n$  in **(2.3)**. Then,

$$\lim_{n \rightarrow \infty} v_n \Pi_n^m \sum_{k=1}^n \Pi_k^{-m} \frac{\gamma_k}{v_k} = \frac{1}{m - v^* \varepsilon}.$$

Moreover, for all positive sequences  $(\alpha_n)$  such that  $\lim_{n \rightarrow \infty} \alpha_n = 0$ , and all  $C \in \mathbb{R}$ ,

$$\lim_{n \rightarrow \infty} v_n \Pi_n^m \left[ \sum_{k=1}^n \Pi_k^{-m} \frac{\gamma_k}{v_k} \alpha_k + C \right] = 0.$$

**Proof of Theorem 1**

1. The application of [Lemma 2](#), ensures that

$$\mathbb{E}(a_n(x)) = \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k \tilde{m}_k(x) = \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k m_n(x) g(x) \left\{ 1 - \frac{C}{C_x} h_k \ln t_n + o(h_k \ln t_n) \right\}.$$

In the case  $p \in (0, \alpha / (d + 2)]$ , we have  $\lim_{n \rightarrow \infty} n \gamma_n > p$ ; the application of [lemma 4](#) ensures that

$$\mathbb{E}(a_n(x)) = a(x) - \frac{C}{(1 - p\varepsilon)C_x} a(x) h_n \ln t_n + o(h_n \ln t_n),$$

and **(2.11)** follows. In the case  $p \in (\alpha / (d + 2), 1/d)$ , we have  $h_n \ln t_n = o(\sqrt{\gamma_n h_n^{-d}})$ , [Lemma 4](#) ensures  $\mathbb{E}(a_n(x)) - a(x) = o(\sqrt{\gamma_n h_n^{-d}})$ , which gives **(2.12)**.

2. Now, we have

$$\begin{aligned}\mathbb{V}ar(a_n(x)) &= \Pi_n^2 \sum_{k=1}^n \Pi_k^{-2} \gamma_k^2 \left[ \mathbb{E} \left[ h_k^{-2d} K^2 \left( \frac{x - X_k}{h_k} \right) [\ln Y_k - \ln t_n]^2 \mathbb{1}_{\{Y_k > t_n\}} \right] \right. \\ &\quad \left. - \mathbb{E}^2 \left[ h_k^{-d} K \left( \frac{x - X_k}{h_k} \right) [\ln Y_k - \ln t_n] \mathbb{1}_{\{Y_k > t_n\}} \right] \right].\end{aligned}$$

The application of Theorem 1 in [Goegebeur et al. \(2014\)](#), ensures that

$$\begin{aligned}\mathbb{V}ar(a_n(x)) &= \Pi_n^2 \sum_{k=1}^n \Pi_k^{-2} \gamma_k^2 \left[ 6 \frac{\|K^2\|_1}{h_k^d} \gamma^2(x) \bar{F}(t_n|x) g(x) (1 + o(1)) \right] \\ &= \Pi_n^2 \sum_{k=1}^n \Pi_k^{-2} \gamma_k \frac{\gamma_k}{h_k^d} \left[ 6 \|K^2\|_1 \gamma^2(x) \bar{F}(t_n|x) g(x) (1 + o(1)) \right].\end{aligned}$$

In the case when  $p \in [\alpha/(d+2), 1/d)$ , we have  $\lim_{n \rightarrow \infty} n\gamma_n > \frac{\alpha - pd}{2}$ , and the application of Lemma 4 ensures that

$$\begin{aligned}\mathbb{V}ar(a_n(x)) &= \frac{6}{2 - (\alpha - pd)\varepsilon} \|K^2\|_1 \bar{F}(t_n|x) g(x) \gamma^2(x) \gamma_n h_n^{-d} \\ &\quad + \frac{6}{2 - (\alpha - pd)\varepsilon} \|K^2\|_1 \bar{F}(t_n|x) g(x) \gamma^2(x) o(\gamma_n h_n^{-d}),\end{aligned}$$

which proves (2.14). In the case when  $p \in (0, \alpha/(d+2))$ , we have  $\gamma_n h_n^{-d} = o(h_n^2 \ln^2 t_n)$ , Lemma 4 ensures that  $\mathbb{V}ar(a_n(x)) = o(h_n^2 \ln^2 t_n)$ , which yields (2.13).

### Proof of Theorem 2

1. First, the application of Lemma 3 provides

$$\mathbb{E}(b_n(x)) = Q_n \sum_{k=1}^n Q_k^{-1} \beta_k g(x) \bar{F}(t_n|x) (1 + Ch_k \ln t_n + o(h_k \ln t_n)).$$

Now, in the case when  $p \in (0, b/(d+2)]$ , we have  $\lim_{n \rightarrow \infty} n\beta_n > p$ ; the application of Lemma 4 ensures that

$$\mathbb{E}(b_n(x)) = b(x) + \frac{C}{1 - p\varepsilon_1} b(x) h_n \ln t_n + g(x) \bar{F}(t_n|x) o(h_n \ln t_n),$$

and (2.15) follows. In the case when  $p \in (b/(d+2), 1/d)$ , we have  $h_n \ln t_n = o(\sqrt{\beta_n h_n^{-d}})$ , Lemma 4 ensures  $\mathbb{E}(b_n(x)) = o(\sqrt{\beta_n h_n^{-d}})$ , which gives (2.16).

2. Now, we have

$$\begin{aligned}\mathbb{V}ar(b_n(x)) &= Q_n^2 \sum_{k=1}^n Q_k^{-2} \beta_k^2 \left[ \mathbb{E} \left[ h_k^{-2d} K^2 \left( \frac{x - X_k}{h_k} \right) \mathbb{1}_{\{Y_k > t_n\}} \right] - \mathbb{E}^2 \left[ h_k^{-d} K \left( \frac{x - X_k}{h_k} \right) \mathbb{1}_{\{Y_k > t_n\}} \right] \right]\end{aligned}$$

$$= Q_n^2 \sum_{k=1}^n Q_k^{-2} \beta_k^2 \left[ \frac{\|K\|_2^2}{h_k^d} \mathbb{E} \left[ h_k^{-d} H \left( \frac{x - X_k}{h_k} \right) \mathbb{1}_{\{Y_k > t_n\}} \right] - \mathbb{E}^2 \left[ h_k^{-d} K \left( \frac{x - X_k}{h_k} \right) \mathbb{1}_{\{Y_k > t_n\}} \right] \right],$$

with  $H(\cdot) =: \frac{K^2(\cdot)}{\|K\|_2^2}$  also satisfying assumption **(C6)**. Using Lemma 3, we get

$$\begin{aligned} \mathbb{V}ar(b_n(x)) &= Q_n^2 \sum_{k=1}^n Q_k^{-2} \beta_k^2 \left[ \frac{\|K\|_2^2}{h_k^d} [g(x) \bar{F}(t_n|x) (1 + Ch_k \ln t_n + o(h_k \ln t_n))] \right. \\ &\quad \left. - g^2(x) \bar{F}^2(t_n|x) (1 + 2Ch_k \ln t_n + o(h_k \ln t_n)) \right], \end{aligned}$$

then, we have

$$\begin{aligned} \mathbb{V}ar(b_n(x)) &= \|K\|_2^2 g(x) \bar{F}(t_n|x) Q_n^2 \sum_{k=1}^n Q_k^{-2} \frac{\beta_k^2}{h_k^d} + C \ln t_n \|K\|_2^2 g(x) \bar{F}(t_n|x) Q_n^2 \sum_{k=1}^n Q_k^{-2} \frac{\beta_k^2}{h_k^{d-1}} \\ &\quad + \|K\|_2^2 g(x) \bar{F}(t_n|x) Q_n^2 \sum_{k=1}^n Q_k^{-2} \beta_k^2 o\left(\frac{\ln t_n}{h_k^{d-1}}\right) - g^2(x) \bar{F}^2(t_n|x) Q_n^2 \sum_{k=1}^n Q_k^{-2} \beta_k^2 \\ &\quad - 2C \ln t_n g^2(x) \bar{F}^2(t_n|x) Q_n^2 \sum_{k=1}^n Q_k^{-2} \beta_k^2 h_k - g^2(x) \bar{F}^2(t_n|x) Q_n^2 \sum_{k=1}^n Q_k^{-2} \beta_k^2 o(h_k \ln t_n). \end{aligned}$$

In the case when  $p \in [b/(d+2), 1/d)$ , we have  $\lim_{n \rightarrow \infty} n\beta_n > (b - pd)/2$ , and the application of Lemma 4 gives

$$\begin{aligned} \mathbb{V}ar(b_n(x)) &= \frac{1}{2 - (b - pd)\varepsilon_1} \|K\|_2^2 g(x) \bar{F}(t_n|x) \frac{\beta_n}{h_n^d} + \frac{C}{2 - (b - p(d-1))\varepsilon_1} \|K\|_2^2 g(x) \ln t_n \bar{F}(t_n|x) \frac{\beta_n}{h_n^{d-1}} \\ &\quad + o\left(\frac{\beta_n \ln t_n}{h_n^{d-1}}\right) - \frac{1}{2 - b\varepsilon_1} g^2(x) \bar{F}^2(t_n|x) \beta_n - \frac{2C}{2 - (b+p)\varepsilon_1} g^2(x) \ln t_n \bar{F}^2(t_n|x) \beta_n h_n + o(\ln t_n \beta_n h_n), \end{aligned}$$

which proves (2.18). In the case when  $p \in (0, b/(d+2))$ , we have  $\beta_n h_n^{-d} = o(h_n^2 \ln^2 t_n)$ , Lemma 4 ensures that  $\mathbb{V}ar(b_n(x)) = o(h_n^2 \ln^2 t_n)$ , which gives (2.17).

**Proof of Theorem 3** Let us first note that, for  $x$  such that  $b_n(x) \neq 0$ , we have

$$\hat{\gamma}_n^H(x) - \gamma(x) = D_n(x) \frac{b(x)}{b_n(x)}, \quad (\text{A.10})$$

with

$$D_n(x) = \frac{1}{b(x)} (a_n(x) - a(x)) - \frac{\gamma(x)}{b(x)} (b_n(x) - b(x)). \quad (\text{A.11})$$

It follows from (A.10), that the asymptotic behavior of  $\hat{\gamma}_n^H(x) - \gamma(x)$  can be deduced from the one of  $D_n(x)$ . Then, (2.19) follows from (2.11), (2.15) and (A.10) whereas (2.20) follows from (2.12), (2.16) and (A.10). Now it follows from (A.11) that

$$\mathbb{V}ar(D_n(x)) = \frac{1}{b^2(x)} \mathbb{V}ar(a_n(x)) - \frac{2\gamma(x)}{b^2(x)} \text{Cov}(a_n(x), b_n(x)) + \frac{\gamma^2(x)}{b^2(x)} \mathbb{V}ar(b_n(x)). \quad (\text{A.12})$$

By using Lemma 4 and choosing the stepsize  $(\gamma_n) = (n^{-1})$ , computations provide

$$\text{Cov}(a_n(x), b_n(x)) = \frac{1}{n} Q_n \sum_{k=1}^n Q_k^{-1} \beta_k h_k^{-2d} (\mathcal{J}_1 - \mathcal{J}_2 \mathcal{J}_3), \quad (\text{A.13})$$

with

$$\mathcal{J}_1 = \mathbb{E} \left[ K^2 \left( \frac{x - X_k}{h_k} \right) [\ln Y_k - \ln t_n] \mathbf{1}_{\{Y_k > t_n\}} \right], \quad \mathcal{J}_2 = \tilde{m}_n(x) \quad \text{and} \quad \mathcal{J}_3 = \mathbb{E} [K_{h_k}(x - X_k) \mathbf{1}_{\{Y_k > t_n\}}].$$

Following similar steps as Lemma 2 in Goegebeur et al. (2014) and Lemma 2, we infer that

$$\mathcal{J}_1 = m_n(x) g(x) \frac{\|K\|_2^2}{h_k^d} \left( 1 - \frac{C}{C_x} h_k \ln t_n + o(h_k \ln t_n) \right),$$

$\mathcal{J}_2$  and  $\mathcal{J}_3$  are already calculated in Lemmas 1 and 2. Then, the combination of (A.11), (A.12), (2.14), (2.18) and (A.13), gives (2.22), and the combination of (A.11), (A.12), (2.13), (2.17) and (A.13), gives (2.21).

**Proof of Theorem 4** Let us at first assume that, if  $p \geq \alpha/(d+2)$ , then

$$\sqrt{\gamma_n^{-1} h_n^d} (\hat{\gamma}_n^H(x) - \mathbb{E}(\hat{\gamma}_n^H(x))) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \text{Var}(x)). \quad (\text{A.14})$$

In the case when  $p > \alpha/(d+2)$ , Part 1 of the theorem follows from the combination of (2.20) and (A.14). In the case when  $p = \alpha/(d+2)$ , Parts 1 and 2 of the Theorem follow from the combination of (2.19) and (A.14). In the case  $p < \alpha/(d+2)$ , (2.21) implies that

$$\frac{1}{h_n \ln t_n} (\hat{\gamma}_n^H(x) - \mathbb{E}(\hat{\gamma}_n^H(x))) \xrightarrow{\mathbb{P}} 0,$$

and the application of (2.19) gives Part 2 of Theorem. Now (A.14) is proved. Relying on (A.11), we have

$$D_n(x) - \mathbb{E}[D_n(x)] = \frac{1}{b(x)} \Pi_n \sum_{k=1}^n (\mathcal{Y}_k(x) - \mathbb{E}[\mathcal{Y}_k(x)]),$$

where

$$\mathcal{Y}_k(x) = \Pi_k^{-1} (\gamma_k \mathcal{Z}_k(x) - \gamma(x) \eta_n \eta_k^{-1} \beta_k \mathcal{W}_k(x)),$$

with  $\mathcal{Z}_k(x) = K_{h_k}(x - X_k) [\ln Y_k - \ln t_n] \mathbf{1}_{\{Y_k > t_n\}}$ ,  $\mathcal{W}_k(x) = K_{h_k}(x - X_k) \mathbf{1}_{\{Y_k > t_n\}}$  and  $\eta_n = \Pi_n^{-1} Q_n$ . Now, in the case when  $(\beta_n) = (n^{-1})$ , we have  $\eta_n = (n \Pi_n)^{-1}$  and  $\eta_k^{-1} \beta_k = \Pi_k$ . Then,

$$\mathcal{Y}_k(x) = \Pi_k^{-1} \gamma_k \mathcal{Z}_k(x) - \gamma(x) (n \Pi_n)^{-1} \mathcal{W}_k(x).$$

Set

$$T_k(x) = \mathcal{Y}_k(x) - \mathbb{E}[\mathcal{Y}_k(x)]. \quad (\text{A.15})$$

Moreover, we have

$$s_n^2 = \sum_{k=1}^n \text{Var}(T_k(x))$$



$$\begin{aligned}
&= \sum_{k=1}^n \Pi_k^{-2} \gamma_k^2 \mathbb{V}ar(\mathcal{Z}_k(x)) + \gamma^2(x) (n\Pi_n)^{-2} \sum_{k=1}^n \mathbb{V}ar(\mathcal{W}_k(x)) \\
&\quad - 2\gamma(x) (n\Pi_n)^{-1} \sum_{k=1}^n \Pi_k^{-1} \gamma_k \mathbb{C}ov(\mathcal{Z}_k(x), \mathcal{W}_k(x)) \\
&:= \Gamma_1 + \Gamma_2 + \Gamma_3.
\end{aligned}$$

In addition, classical computations and applications of Lemma 4 ensure that

$$\begin{aligned}
\Gamma_1 &= \Pi_n^{-2} \gamma(x) \left[ \frac{1}{C_x} \frac{6}{2 - (\alpha - pd)\varepsilon} \|K^2\|_1 m_n(x) g(x) \frac{\gamma_n}{h_n^d} + o\left(\frac{\gamma_n}{h_n^d}\right) \right], \\
\Gamma_2 &= \Pi_n^{-2} \gamma(x) \left[ \frac{1}{C_x} \frac{1}{1 + pd} \|K\|_2^2 m_n(x) g(x) \frac{1}{nh_n^d} + o\left(\frac{1}{nh_n^d}\right) \right], \\
\Gamma_3 &= \Pi_n^{-2} \gamma(x) \left[ \frac{2}{1 + pd\varepsilon} \|K\|_2^2 m_n(x) g(x) \frac{1}{nh_n^d} + o\left(\frac{1}{nh_n^d}\right) \right].
\end{aligned}$$

As a matter of fact, we infer that

$$s_n^2 = \frac{b^2(x) \gamma_n}{\Pi_n^2 h_n^d} [\mathcal{V}ar(x) + o(1)].$$

On the other side, we have, for all  $q > 0$ ,

$$\mathbb{E} [|\mathcal{Y}_k(x)|^{2+q}] = O\left(\frac{1}{h_k^{(1+q)d}}\right),$$

and, since  $\lim_{n \rightarrow \infty} (n\gamma_n) > (\alpha - pd)/2$ , there exists  $q > 0$  such that  $\lim_{n \rightarrow \infty} n\gamma_n > \frac{1+q}{2+q}(\alpha - pd)$ .

Applying Lemma 4, we get

$$\sum_{k=1}^n \mathbb{E} [|\mathcal{T}_k(x)|^{2+q}] = O\left(\sum_{k=1}^n \Pi_k^{-2-q} \gamma_k^{2+q} \mathbb{E} [|\mathcal{Y}_k(x)|^{2+q}]\right) = O\left(\frac{\gamma_n^{1+q}}{\Pi_n^{2+q} h_n^{(q+1)d}}\right),$$

and we thus obtain

$$\frac{1}{s_n^{2+q}} \sum_{k=1}^n \mathbb{E} [|\mathcal{T}_k(x)|^{2+q}] = \frac{1}{s_n^{2(1+q/2)}} O\left(\frac{\gamma_n^{1+q}}{\Pi_n^{2+q} h_n^{(q+1)d}}\right) = O\left(\gamma_n^{\frac{q}{2}} h_n^{-\frac{dq}{2}}\right) = o(1).$$

The convergence in (A.14) then follows from the application of Lyapounov's Theorem.

## References

- Beirlant, J., Goegebeur, Y. and Teugels, J. (2004). *Statistics of extremes-Theory and applications*. Chichester: Wiley.
- Beran, J., Schell, D. and Stehlík, M. (2014). The harmonic moment tail index estimator: asymptotic distribution and robustness. *Ann. Inst. Statist. Math.*, **66**, 193–220.

- Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*. Cambridge: Cambridge University Press.
- Brilhante, M. F., Gomes, M. I. and Pestana, D. (2013). A simple generalisation of the Hill estimator. *Comput. Statist. Data Anal.*, **57**, 518–535.
- Daouia, A., Gardes, L., Girard, S. and Lekina, A. (2011). Kernel estimators of extreme level curves. *TEST*, **20**, 311–333.
- Daouia, A., Gardes, L. and Girard, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli*, **19**, 2557–2589.
- De Haan, L. and Ferreira, A. (2006). *Extreme Value Theory-An Introduction*. New York: Springer.
- Embrechts, P., Kluppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events*. New York: Springer.
- Galambos, J. and Seneta, E. (1973). Regularly Varying Sequences. *Proc. Amer. Math. Soc.*, **41**, 110–116.
- Gannoun, A., Girard, S., Guinot, C. and Saracco, S. (2002). Reference ranges based on nonparametric quantile regression. *Stat. Med.*, **21**, 3119–3135.
- Gardes, L. and Girard, S. (2008). A moving window approach for nonparametric estimation of the conditional tail index. *J. Multivariate Anal.*, **99**, 2368–2388.
- Gardes, L., Girard, S. and Lekina, A. (2010). Functional nonparametric estimation of conditional extreme quantiles. *J. Multivariate Anal.*, **101**, 419–433.
- Goegebeur, Y., Guillou, A. and Schorgen, A. (2014). Nonparametric regression estimation of conditional tails: the random covariate case. *Statistics*, **48**, 732–755.
- Hall, P. (1992). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Ann. Statist.*, **20**, 675–694.
- Gumbel, E. J. (1958). *Statistics of Extremes*. New York: Columbia University Press.
- Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, **3**, 1163–1174.
- Milet, J., Nuel, G., Watier, L., Courtin, D., Slaoui, Y., Senghor, P., Migot-Nabias, F., Gaye, O. and Garcia, A. (2010). Genome wide linkage study, using a 250K SNP map, of Plasmodium falciparum infection and mild malaria attack in a Senegalese population. *PLoS One*, **5**, e11616.
- Mokkadem, A., Pelletier, M. and Slaoui, Y. (2009a). The stochastic approximation method for the estimation of a multivariate probability density. *J. Statist. Plann. Inference*, **139**, 2459–2478.
- Mokkadem, A., Pelletier, M. and Slaoui, Y. (2009b). Revisiting Révész’s stochastic approximation method for the estimation of a regression function. *ALEA. Lat. Am. J. Probab. Math. Stat.*, **6**, 63–114.
- Ndao, P., Diop, A. and Dupuy, J.F. (2016). Nonparametric estimation of the conditional extreme value index with random covariates and censoring. *J. Statist. Plann. Inference*, **168**, 20–37.

- Paulauskas, V. and Vaičiulis, M. (2013). On an improvement of Hill and some other estimators. *Lith. Math. J.*, **53**, 336–355.
- Paulauskas, V. and Vaičiulis, M. (2017). A class of new tail index estimators. *Ann. Inst. Statist. Math.*, **69**, 461–487.
- Quintela-del-Río, A. and Estévez-Pérez, G. (2012). *Nonparametric Kernel Distribution Function Estimation with Kerdist: An R Package for Bandwidth Choice and Applications*. *J. Statist. Softw.*, **8**, 1–21.
- Reiss, R.D. and Thomas, M. (2007). *Statistical Analysis of Extreme Values*. Basel, Switzerland: Birkhauser.
- Révész, P. (1973). Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes I. *Studia Sci. Math. Hung.* **8**, 391–398.
- Révész, P. (1977). How to apply the method of stochastic approximation in the non-parametric estimation of a regression function. *Math. Operationsforsch. Statist., Ser. Statistics*. **8**, 119–126.
- Slaoui, Y. (2013). Large and moderate principles for recursive kernel density estimators defined by stochastic approximation method. *Serdica Math. J.*, **39**, 53–82.
- Slaoui, Y. (2014a). Bandwidth selection for recursive kernel density estimators defined by stochastic approximation method. *J. Probab. Stat* **2014**, ID 739640, doi:10.1155/2014/739640.
- Slaoui, Y. (2014b). The stochastic approximation method for the estimation of a distribution function. *Math. Methods Statist.*, **23**, 306–325.
- Slaoui, Y. and Nuel, G. (2014c). Parameter estimation in a hierarchical random intercept model with censored response approach using a SEM algorithm and Gibbs sampling, *Sankhya B*, **76**, 210–233.
- Slaoui, Y. (2016). On the choice of smoothing parameters for semi-recursive nonparametric hazard estimators. *J. Stat. Theory Pract.*, **10**, 656–672.
- Slaoui, Y. (2018). Data-driven bandwidth selection for recursive kernel density estimators under double truncation, *Sankhya B*, **80**, 341–368.
- Slaoui, Y. (2019a). Automatic bandwidth selection for recursive kernel density estimators with length biased data, *Jpn. J. Stat. Data Sci.*, doi: 10.1007/s42081-019-00053-z.
- Slaoui, Y. (2019b). Data-driven deconvolution recursive kernel density estimators defined by stochastic approximation method, *Sankhya A*, doi: 10.1007/s13171-019-00182-3.
- Stupfler, G. (2013). A moment estimator for the conditional extreme-value index. *Electron. J. Stat.*, **7**, 2298–2343.
- Tsybakov, A. B. (1990). Recurrent estimation of the mode of a multidimensional distribution. *Probl. Inf. Transm.*, **8**, 119–126.
- Yao, Q. (1999). Conditional predictive regions for stochastic processes. *Technical report*, University of Kent at Canterbury.