



**HAL**  
open science

## Estimation of the conditional tail index using a smoothed local Hill estimator

Laurent Gardes, Gilles Stupfler

► **To cite this version:**

Laurent Gardes, Gilles Stupfler. Estimation of the conditional tail index using a smoothed local Hill estimator. *Extremes*, 2014, 17 (1), pp.45-75. hal-00739454v2

**HAL Id: hal-00739454**

**<https://hal.science/hal-00739454v2>**

Submitted on 1 Mar 2013

**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.

# Estimation of the conditional tail index using a smoothed local Hill estimator

Laurent Gardes<sup>(1)</sup> & Gilles Stupfler<sup>(2)</sup>

<sup>(1)</sup> Université de Strasbourg & CNRS, IRMA, UMR 7501, 7 rue René Descartes,  
67084 Strasbourg Cedex, France

<sup>(2)</sup> Université d'Aix-Marseille, CERGAM, 15-19 allée Claude Forbin,  
13628 Aix-en-Provence Cedex 1, France

**Abstract.** For heavy-tailed distributions, the so-called tail index is an important parameter that controls the behavior of the tail distribution and is thus of primary interest to estimate extreme quantiles. In this paper, the estimation of the tail index is considered in the presence of a finite-dimensional random covariate. Uniform weak consistency and asymptotic normality of the proposed estimator are established and some illustrations on simulations are provided.

**AMS Subject Classifications:** 62G05, 62G20, 62G30, 62G32.

**Keywords:** Heavy-tailed distribution, random covariate, uniform consistency, pointwise asymptotic normality.

## 1 Introduction

Extreme value analysis has attracted considerable attention in many fields of application, such as hydrology, biology and finance, for instance. It focuses on the study of random variables having survival function  $\bar{F}$  of the form  $\bar{F}(x) = x^{-1/\gamma}L(x)$ , where  $\gamma > 0$  shall be referred to as the tail index and  $L$  is a slowly varying function at infinity: namely,  $L$  satisfies, for all  $\lambda > 0$ ,  $L(\lambda x)/L(x) \rightarrow 1$  as  $x$  goes to infinity. Clearly,  $\gamma$  drives the tail behavior of  $F$  and its knowledge is necessary if, for instance, we are interested in the estimation of extreme quantiles. The estimation of the tail index is thus one of the central topics in extreme value theory: this problem has been extensively studied in the literature. Recent overviews on univariate tail index estimation can be found in the monographs of Beirlant *et al.* [2] and de Haan and Ferreira [12]. The most popular semi-parametric estimator was proposed by Hill [15]. Let  $k_n \in \{2, \dots, n\}$  and  $Y_{1,n} \leq \dots \leq Y_{n,n}$  be the ordered statistics associated to the sample  $Y_1, \dots, Y_n$  (note that, from now on, this way of denoting ordered statistics will be used in this paper).

Hill's estimator is the statistic

$$H(k_n) = \frac{1}{k_n - 1} \sum_{i=1}^{k_n-1} \log \frac{Y_{n-i+1,n}}{Y_{n-k_n+1,n}}. \quad (1)$$

In practice, it is often useful to link the variable of interest  $Y$  to a covariate  $X$ . In this situation, the tail index depends on the observed value  $x$  of the covariate  $X$  and shall be referred to, in the following, as the conditional tail index. Its estimation has been addressed in the recent extreme value literature mostly in the “fixed design” case, that is, when the covariates are nonrandom. Smith [17] and Davison and Smith [8] considered a regression model while Hall and Tajvidi [13] used a semi-parametric approach to estimate the conditional tail index. Fully non parametric methods have been considered using splines (see Chavez-Demoulin and Davison [4]), local polynomials (see Davison and Ramesh [7]), a moving window approach (see Gardes and Girard [9]), or a nearest neighbor approach (see Gardes and Girard [10]), among others.

Despite the great interest in practice, less attention has been paid to the random covariate case. One can cite the works of Wang and Tsai [18], based on a maximum likelihood approach, Daouia *et al.* [5] who use a fixed number of non parametric conditional quantile estimators to estimate the conditional tail index, later generalized in Daouia *et al.* [6] to a regression context with response distributions belonging to the general max-domain of attraction and Goegebeur *et al.* [11] who study a nonparametric regression estimator.

The aim of this paper is to adapt Hill's estimator to the presence of a random covariate. Note that the uniform weak consistency of the proposed estimator is established while, in most of the aforementioned studies, the authors only considered pointwise convergence.

The rest of the paper is organised as follows. In Section 2, we define our conditional tail index estimator. The two main results (uniform weak consistency and asymptotic normality) are stated in Section 3 and a simulation study is provided in Section 4. The proofs are given in Section 5 and in the Appendix.

## 2 Estimation of the conditional tail index

Let  $(X_1, Y_1), \dots, (X_n, Y_n)$  be  $n$  independent copies of a random pair  $(X, Y) \in S \times \mathbb{R}$  where  $S$  is a subset of  $\mathbb{R}^d$ ,  $d \geq 1$ , having nonempty interior. For all  $x \in S$ , we assume that the conditional survival function of  $Y$  given  $X = x$  is heavy-tailed with tail index  $\gamma(x) > 0$ . Equivalently (see Bingham *et al.* [3]), we consider the model:

( $\mathcal{M}$ )  $X$  has a probability density function  $f$  on  $\mathbb{R}^d$  with support  $S$  and for all  $x \in S$ , the conditional survival function  $\bar{F}(\cdot|x)$  is continuous and decreasing. Moreover, the conditional quantile of  $Y$  given  $X = x$  is such that

$$\forall \alpha \in (0, 1), q(\alpha|x) := \bar{F}^{-1}(\alpha|x) = \alpha^{-\gamma(x)} \ell(\alpha^{-1}|x), \quad (2)$$

where  $\overline{F}^{-1}(\cdot|x)$  is the inverse function of the conditional survival function and  $\ell(\cdot|x)$  is a slowly varying function at infinity.

For  $i = 1, \dots, n$ , denoting by  $X_i^*$  the covariate associated with the ordered statistic  $Y_{n-i+1,n}$ , a straightforward adaptation to the random covariate case of Hill's estimator (1) is:

$$H(x, k, h) = \frac{1}{M_k(x, h) - 1} \sum_{i=1}^{k-1} \log \frac{Y_{n-i+1,n}}{Y_{n-k+1,n}} \mathbb{I}\{\|X_i^* - x\| \vee \|X_k^* - x\| \leq h\}, \quad (3)$$

if  $M_k(x, h) > 1$  and  $H(x, k, h) = 0$  otherwise. In (3),  $\mathbb{I}\{\cdot\}$  is the indicator function,  $\|\cdot\|$  is a norm on  $\mathbb{R}^d$ ,  $k = k_n \in \{2, \dots, n\}$ ,  $h = h_n$  is a nonrandom positive sequence tending to 0 at infinity and

$$M_i(x, h) = \sum_{j=1}^i \mathbb{I}\{\|X_j^* - x\| \leq h\}, \quad i = 1, \dots, n,$$

is the number of covariates among  $X_1^*, \dots, X_i^*$  which lie in the ball  $B(x, h)$  with center  $x$  and radius  $h$ . Clearly, the choice of the number  $k$  in (3) is crucial since, for most values of  $k$ , the statistic  $H(x, k, h)$  is equal to 0. The behavior of  $H(x, k, h)$  as a function of  $k$  is thus very erratic. To overcome this drawback, we propose to estimate the conditional tail index by an average on  $k$  of the statistics defined in (3):

$$\widehat{\gamma}_a(x, k_x, h) = \frac{1}{k_x - \lfloor (1-a)k_x \rfloor + 1} \sum_{l=2}^n H(x, l, h) \mathbb{I}\{\lfloor (1-a)k_x \rfloor \leq M_l(x, h) \leq k_x\}, \quad a \in [0, 1), \quad (4)$$

where  $\lfloor z \rfloor = \max\{j \in \mathbb{N} | z \geq j\}$  is the integer part of  $z$  and  $k_x$  is a positive integer belonging to the interval  $[2/(1-a), n]$ . Clearly for all  $a \in [0, 1)$ , if  $M_n(x, h) > \lfloor (1-a)k_x \rfloor$ , then  $\widehat{\gamma}_a(x, k_x, h) > 0$ . The parameter  $a$  controls the number of statistics (3) taken into account in the estimator (4). For instance, if  $a = 0$  and if  $M_n(x, h) > k_x$ , only one statistic having the form  $H(x, l, h)$  is used to compute (4).

We point out that in practice,  $k_x$  is restricted to the interval  $[2/(1-a), M_n(x, h)]$ , since  $k_x$  is the number of statistics  $Y_{n-i+1,n}$ , whose associated covariates  $X_i^*$  belong to the ball with center  $x$  and radius  $h$ , which are used to compute  $\widehat{\gamma}_a(x, k_x, h)$ : see also Section 4.

### 3 Main results

In this Section, we state the two main results of the paper: the uniform weak consistency and pointwise asymptotic normality of  $\widehat{\gamma}_a(x, k_x, h)$  on a compact subset  $\Omega$  of the interior of  $S$ . To this aim, we introduce some assumptions. The following condition specifies the regularity of the conditional tail index  $\gamma$  and of the probability density function  $f$  of the covariates.

**(A1)** The function  $\gamma$  is positive and continuous on  $S$  and the probability density function  $f$  is a positive Hölder continuous function on  $S$  with exponent  $\beta_f \in (0, 1]$ .

Note that this condition especially implies that, on the compact set  $\Omega$ , the function  $\gamma$  and the probability density function  $f$  are bounded from below and above by finite positive constants:

$$0 < \underline{\gamma} := \inf_{x \in \Omega} \gamma(x) \leq \sup_{x \in \Omega} \gamma(x) =: \overline{\gamma} < \infty \quad \text{and} \quad 0 < \underline{f} := \inf_{x \in \Omega} f(x) \leq \sup_{x \in \Omega} f(x) =: \overline{f} < \infty.$$

The next assumption controls the largest oscillation of the log-quantile function with respect to its second variable. For all  $u < v \in (0, 1)$ , let

$$\omega(u, v, x, h) = \sup_{\alpha \in [u, v]} \sup_{\|x' - x\| \leq h} |\log q(\alpha|x) - \log q(\alpha|x')|.$$

We assume that

**(A2)** There exists  $\delta > 0$  such that

$$\lim_{n \rightarrow \infty} \sup_{x \in \Omega} \omega(n^{-(1+\delta)}, 1 - n^{-(1+\delta)}, x, h) = 0.$$

Now, in order to deal with the slowly varying function in (2), we assume that

**(A3)** For all  $x \in S$  and  $t \geq 1$ ,

$$\ell(t|x) = c(x) \exp\left(\int_1^t \frac{\Delta(u|x)}{u} du\right),$$

where  $c(x) > 0$  and  $\Delta(\cdot|x)$  is an ultimately monotonic function converging to 0 at infinity.

Note that **(A3)** implies in particular that for all  $x \in S$ ,  $\ell(\cdot|x)$  is a normalised slowly varying function (see Bingham *et al.* [3]). We also introduce the notation

$$\bar{\Delta}_x(z) := \sup_{u \in [z^{-1}, \infty)} |\Delta(u|x)|.$$

We can now state the uniform weak consistency of our estimator.

**Theorem 1.** *Under model  $(\mathcal{M})$ , assume that **(A1)**, **(A2)** and **(A3)** hold. If  $nh^d/\log n \rightarrow \infty$ ,*

$$\inf_{x \in \Omega} \min \left\{ \frac{k_x}{\log n}, \frac{nh^d}{k_x \log(nh^d)} \right\} \rightarrow \infty, \quad \limsup_{t \rightarrow 0} \sup_{x \in \Omega} \bar{\Delta}_x(t) = 0,$$

*and if there exists a finite positive constant  $K_1$  such that*

$$\sup_{x \in \Omega} \sup_{\|x' - x\| \leq h} |k_x - k_{x'}| \leq K_1,$$

*then, if  $a \in (0, 1)$ , it holds that, as  $n$  goes to infinity,*

$$\sup_{x \in \Omega} |\widehat{\gamma}_a(x, k_x, h) - \gamma(x)| \xrightarrow{\mathbb{P}} 0.$$

It is straightforward that under condition **(A1)**, the number  $M_n(x, h)$  of covariates lying in the ball  $B(x, h)$  is such that

$$\frac{1}{n} M_n(x, h) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{\|X_i - x\| \leq h\} = \mathbb{P}(\|X - x\| \leq h)(1 + o_{\mathbb{P}}(1)) = \mathcal{V}h^d f(x)(1 + o_{\mathbb{P}}(1)), \quad (5)$$

where  $\mathcal{V}$  is the volume of the unit ball of  $\mathbb{R}^d$  (see Lemma 3 for a uniform result). Thus, since  $f$  is bounded from below and above by finite positive constants, condition  $nh^d/\log n \rightarrow \infty$  implies that for all  $x \in \Omega$ ,  $M_n(x, h)$  goes to infinity in probability. Furthermore, condition

$$\inf_{x \in \Omega} \min \left\{ \frac{k_x}{\log n}, \frac{nh^d}{k_x \log(nh^d)} \right\} \rightarrow \infty$$

implies that for all  $x \in \Omega$ ,  $\lfloor (1-a)k_x \rfloor = (1-a)k_x(1+o(1)) \rightarrow \infty$  and that, with arbitrary large probability, we have  $k_x < M_n(x, h)$  for  $n$  sufficiently large. Hence, for  $n$  large enough,  $\widehat{\gamma}_a(x, k_x, h) > 0$  for all  $a \in [0, 1)$  and  $x \in \Omega$ .

We now wish to state the pointwise asymptotic normality of the estimator at a point  $x \in S$ . To this aim, the following assumption is required:

**(A4)** For all  $x \in S$ , the function  $|\Delta(\cdot|x)|$  is regularly varying with index  $\rho(x) < 0$  i.e., for all  $\lambda > 0$ ,

$$\lim_{t \rightarrow \infty} \frac{|\Delta(\lambda t|x)|}{|\Delta(t|x)|} = \lambda^{\rho(x)}.$$

Note that conditions **(A3)** and **(A4)** entail that

$$\lim_{t \rightarrow \infty} \frac{\log \ell(\lambda t|x) - \log \ell(t|x)}{\Delta(t|x)} = \frac{\lambda^{\rho(x)} - 1}{\rho(x)}, \quad (6)$$

which is the standard second-order condition classically used to prove the asymptotic normality of tail index estimators. The asymptotic normality of our estimator is obtained conditionally to the event  $\{M_n(x, h) = m_x\}$ . Note that for instance, under **(A1)** and from (5), a typical sequence  $(m_x)$  in this case is  $m_x = \mathcal{V}f(x)nh^d$ .

**Theorem 2.** *Under model  $(\mathcal{M})$ , assume that **(A1)**, **(A3)** and **(A4)** hold. If, as  $n$  goes to infinity,  $k_x \rightarrow \infty$ ,  $k_x/m_x \rightarrow 0$ ,  $k_x^{1/2}\omega(m_x^{-1-\delta}, 1 - m_x^{-1-\delta}, x, h) \rightarrow 0$  and  $k_x^{1/2}\Delta(m_x/k_x|x) \rightarrow \xi(x) \in \mathbb{R}$ , then for  $a \in [0, 1)$  and conditionally to the event  $\{M_n(x, h) = m_x\}$  one has*

$$k_x^{1/2} \left( \widehat{\gamma}_a(x, k_x, h) - \gamma(x) - \frac{\Delta(m_x/k_x|x)}{1 - \rho(x)} \mathcal{AB}(a, x) \right) \xrightarrow{d} \mathcal{N}(0, \gamma^2(x) \mathcal{AV}(a)),$$

where if  $a \in (0, 1)$ ,

$$\mathcal{AB}(a, x) = \frac{1 - (1-a)^{1-\rho(x)}}{a(1-\rho(x))} \text{ and } \mathcal{AV}(a) = \frac{2(a + (1-a)\log(1-a))}{a^2},$$

and if  $a = 0$ ,  $\mathcal{AB}(0, x) = 1$  and  $\mathcal{AV}(0) = 1$ .

As expected, the asymptotic bias is a decreasing function of  $a$  while the asymptotic variance is increasing. For  $a = 0$ , we find back the asymptotic bias and variance of Hill's estimator.

## 4 Simulation study

To assess the finite-sample performance of the proposed conditional tail-index estimator, some simulation experiments were carried out using the following model: the conditional distribution function of  $Y$  given  $X = x$  is given by

$$\forall y > 0, \overline{F}(y|x) = \left(1 + y^{-\rho/\gamma(x)}\right)^{-1/\rho},$$

where  $X$  is uniformly distributed on  $S = [0, 1]$ . The negative second-order parameter  $\rho$  is chosen to be independent of  $x$  and its value is picked in the set  $\{-1.2, -1, -0.8\}$ . Recall that the smaller is  $|\rho|$ ,

the slower is the convergence (6) and therefore, the more difficult is the estimation. As far as the conditional tail-index  $\gamma$  is concerned, two situations are considered:

$$\begin{aligned}\gamma_1(x) &= \frac{1}{3} + \frac{1}{8} \sin(2\pi x) \\ \text{and } \gamma_2(x) &= \frac{1}{4} \left\{ 1 + \exp(-60(x - 1/4)^2) \mathbb{I}\{3x \in [0, 1]\} + \exp(-5/12) \mathbb{I}\{3x \in (1, 2]\} \right. \\ &\quad \left. + (5 - 6x) (\exp(-5/12) \mathbb{I}\{3x \in (2, 5/2]\} - \mathbb{I}\{3x \in (5/2, 3]\}) \right\}.\end{aligned}$$

Note that  $\gamma_1$  is infinitely differentiable and  $\gamma_2$  is continuous but not differentiable at  $x \in \{1/3, 2/3, 5/6\}$ . The aim of this simulation study is to estimate the conditional tail-index on a grid of points  $\{x_1, \dots, x_M\}$  of  $[0, 1]$ . A small preliminary practical investigation leads to take  $a = 3/7$  which provides reasonable performances in a large range of situations. This leaves two parameters to be chosen: the bandwidth  $h$  and the number of upper order statistics  $k_x$ . Our selection procedure for these parameters goes as follows.

- 1) We choose a grid  $\{h_1, \dots, h_P\}$  of possible values of  $h$ . In what follows, we let  $\widehat{\gamma}_{i,j}(k) := \widehat{\gamma}_{3/7}(x_i, k, h_j)$ . For each  $i \in \{1, \dots, M\}$ ,  $j \in \{1, \dots, P\}$  and  $k \in \{q_{i,j} + 4, \dots, M_n(x_i, h_j) - q_{i,j}\}$ , where  $q_{i,j} \in \mathbb{N} \setminus \{0\}$ , we introduce the set  $E_{i,j,k} = \{\widehat{\gamma}_{i,j}(l), l \in \{k - q_{i,j}, \dots, k + q_{i,j}\}\}$ . For two fixed indices  $i$  and  $j$ , our aim is to select the number of upper order statistics  $k_{i,j}$  in a region of stability for  $\widehat{\gamma}_{i,j}$ . To do that, we compute the variance of the set  $E_{i,j,k}$  for every possible value of  $k$ . We then record the number  $K_{i,j}$  for which this variance is minimal. More precisely,

$$K_{i,j} = \arg \min_k \frac{1}{2q_{i,j} + 1} \sum_{l=k-q_{i,j}}^{k+q_{i,j}} \left( \widehat{\gamma}_{i,j}(l) - \overline{\widehat{\gamma}}_{i,j}(k) \right)^2 \quad \text{with } \overline{\widehat{\gamma}}_{i,j}(k) = \frac{1}{2q_{i,j} + 1} \sum_{l=k-q_{i,j}}^{k+q_{i,j}} \widehat{\gamma}_{i,j}(l).$$

Hence, for a given point  $x_i$  and a given bandwidth  $h_j$ , the selected number of upper order statistics  $k_{i,j}$  is picked in the set  $\{K_{i,j} - q_{i,j}, \dots, K_{i,j} + q_{i,j}\}$ . We propose to record the value  $k_{i,j}$  such that  $\widehat{\gamma}_{i,j}(k_{i,j})$  is the median of the set  $E_{i,j,K_{i,j}}$ . For the sake of simplicity, the estimate  $\widehat{\gamma}_{i,j}(k_{i,j})$  will be denoted by  $\widetilde{\gamma}_{i,j}$ .

- 2) We now want to select a bandwidth that does not depend on  $x$  and which is such that the estimation carried out for bandwidths in its neighborhood does not show a large variance. To achieve that, we let  $q'$  be a positive integer such that  $2q' + 1 < P$  and we compute for each  $j \in \{q' + 1, \dots, P - q'\}$  the stability criterion

$$\overline{\sigma}(j) = \frac{1}{M} \sum_{i=1}^M \sigma_i(j),$$

where, for  $i \in \{1, \dots, M\}$ ,

$$\sigma_i(j) = \left[ \frac{1}{2q' + 1} \sum_{l=j-q'}^{j+q'} (\widetilde{\gamma}_{i,l} - \overline{\widetilde{\gamma}}_{i,\cdot}(j))^2 \right]^{1/2} \quad \text{with } \overline{\widetilde{\gamma}}_{i,\cdot}(j) = \frac{1}{2q' + 1} \sum_{l=j-q'}^{j+q'} \widetilde{\gamma}_{i,l}.$$

We next record the integer  $J$  such that  $\bar{\sigma}(J)$  is the first local minimum of the application  $j \mapsto \bar{\sigma}(j)$  which is less than the average of the  $\bar{\sigma}(j)$ , see Figure 1. In other words,  $J = q' + 1$  if  $\bar{\sigma}(\cdot)$  is increasing,  $J = P - q'$  if  $\bar{\sigma}(\cdot)$  is decreasing and

$$J = \min \left\{ j \text{ such that } \bar{\sigma}(j) \leq \bar{\sigma}(j-1) \wedge \bar{\sigma}(j+1) \text{ and } \bar{\sigma}(j) \leq \frac{1}{P-2q'} \sum_{l=q'+1}^{P-q'} \bar{\sigma}(l) \right\} \quad (7)$$

otherwise, where for convenience we extend  $\bar{\sigma}$  by setting  $\bar{\sigma}(q') := \bar{\sigma}(q' + 1)$  and  $\bar{\sigma}(P - q') := \bar{\sigma}(P - q' + 1)$ . The selected bandwidth is then  $h^* = h_J$ .

To summarize, the bandwidth and the number of upper order statistics are selected in order to satisfy a stability criterion. The selected bandwidth is independent of  $x$  and is given by  $h^* = h_J$  where  $J$  is defined in (7). The selected number of upper order statistics is given, for  $x = x_i$ , by  $k_{x_i}^* = k_{i,J}$ .

This estimation procedure is carried out on  $N = 100$  independent samples of size  $n = 1000$ . The conditional tail-index is estimated on a grid of  $M = 35$  evenly spaced points in  $[0, 1]$ . Regarding the selection procedure,  $P = 100$  values of  $h$  ranging from 0.025 to 0.25 are tested. The parameter  $q_{i,j}$  is chosen so that  $2q_{i,j} + 1$  is approximately equal to 5% of  $M_n(x_i, h_j)$  and  $q'$  is set to 3.

To have an idea of how our estimator behaves compared to other estimators in the conditional tail-index estimation literature, it is compared to:

- The estimator  $\tilde{\gamma}_D(x) := \hat{\gamma}_n^H(x)$  of Daouia *et al.* [5]:

$$\tilde{\gamma}_D(x) = \frac{\sum_{j=1}^9 [\log \hat{q}_n(\alpha_n/j|x) - \log \hat{q}_n(\alpha_n|x)]}{\sum_{j=1}^9 \log j}$$

where  $\hat{q}_n(\alpha|x) = \inf\{t \in \mathbb{R}, \hat{F}_n(t|x) \leq \alpha\}$  is the generalized inverse of the kernel estimator of the conditional survival function

$$\hat{F}_n(y|x) = \frac{\sum_{i=1}^n K_h(x - X_i) \mathbb{I}\{Y_i > y\}}{\sum_{i=1}^n K_h(x - X_i)}.$$

Here  $K_h(x) = h^{-1}K(x/h)$  where

$$K(x) = \frac{15}{16}(1 - x^2)^2 \mathbb{I}[-1, 1](x)$$

is the bi-quadratic kernel function,  $h := h_n$  is a positive sequence tending to 0 and  $\alpha_n = 0.3$ . This estimator is computed using the data-driven procedure described in Daouia *et al.* [5].

- The estimator  $\tilde{\gamma}_G(x) := \hat{\gamma}_n^{(2)}(x, 0, \mathbb{K}, \mathbb{K})$  of Goegebeur *et al.* [11]:  $\tilde{\gamma}_G(x) := T_n^{(1,1)}(x)/T_n^{(1,0)}(x)$



with

$$\forall s \geq 1, \forall t \geq 0, T_n^{(s,t)}(x) = \frac{\sum_{i=1}^n K_h^s(x - X_i)(\log Y_i - \log \omega_{n,x})_+^t \mathbb{I}\{Y_i > \omega_{n,x}\}}{\sum_{i=1}^n K_h^s(x - X_i) \mathbb{I}\{Y_i > \omega_{n,x}\}}.$$

Here  $K_h(u) = h^{-1}K(u/h)$  where  $K$  is once again the bi-quadratic kernel function,  $h := h_n$  is a positive sequence tending to 0 and for all  $x$ ,  $(\omega_{n,x})$  is a positive sequence tending to infinity. Note that this estimator is a kernel version of the case  $a = 0$  of our estimator; to compute  $\tilde{\gamma}$ , we shall use the data-driven method described in [11].

- The bias-corrected version  $\tilde{\gamma}_{G,BC}(x) := \hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\hat{\rho}_n(x; \mathbb{K}, \mathbb{K}, 0.5)))$  of  $\tilde{\gamma}_G(x)$ , also presented in Goegebeur *et al.* [11]:

$$\tilde{\gamma}_{G,BC}(x) = \frac{\hat{\gamma}_n^{(2)}(x, 0, \mathbb{K}, \mathbb{K})}{\hat{\rho}_n(x; \mathbb{K}, \mathbb{K}, 0.5)} + \left[ 1 - \frac{1}{\hat{\rho}_n(x; \mathbb{K}, \mathbb{K}, 0.5)} \right] \hat{\gamma}_n^{(2)}(x, 1, \mathbb{K}, \mathbb{K})$$

where

$$\hat{\gamma}_n^{(2)}(x, 1, \mathbb{K}, \mathbb{K}) = \frac{T_n^{(1,2)}(x)}{2T_n^{(1,1)}(x)}$$

and

$$\hat{\rho}_n(x; \mathbb{K}, \mathbb{K}, 0.5) = \frac{3(R_n(x; \mathbb{K}, \mathbb{K}, 0.5) - 1)}{R_n(x; \mathbb{K}, \mathbb{K}, 0.5) - 3}$$

provided  $1 \leq R_n(x; \mathbb{K}, \mathbb{K}, 0.5) < 3$  with

$$R_n(x; \mathbb{K}, \mathbb{K}, 0.5) = \frac{\left(\frac{T_n^{(1,1)}(x)}{T_n^{(1,0)}(x)}\right)^\tau - \left(\frac{T_n^{(1,2)}(x)}{2T_n^{(1,0)}(x)}\right)^{\tau/2}}{\left(\frac{T_n^{(1,2)}(x)}{2T_n^{(1,0)}(x)}\right)^{\tau/2} - \left(\frac{T_n^{(1,3)}(x)}{6T_n^{(1,0)}(x)}\right)^{\tau/3}}.$$

For this estimator, the data-driven method described in [11] is also used.

For every estimator, we compute the empirical MSEs, averaged over the  $M = 35$  evenly spaced points in  $[0, 1]$ . Numerical results are given in Table 1. This chart shows that our estimator yields performances which are similar to the estimator  $\tilde{\gamma}_D$  of Daouia *et al.* [5]. Besides, it outperforms the estimator  $\tilde{\gamma}_G$  of Goegebeur *et al.* [11] in terms of MSEs by a 2:1 ratio in every case, while being outperformed by the bias-corrected version  $\tilde{\gamma}_{G,BC}$  of this estimator. This was expected, since the bias-corrected estimator  $\tilde{\gamma}_{G,BC}$  was shown to display far better performances than the simple estimator  $\tilde{\gamma}_G$  and that our method was not originally targeted at correcting any specific bias that the Hill statistics  $H(x, l, h)$  used for its computation may possess.

We display some results in Figures 2–4: the estimations corresponding to the median, 10% and 90% quantiles of the MSE are represented. Besides, we represent in Figure 5 boxplots of the bandwidths and in Figure 6 boxplots of the ratios  $k_x^*/M_n(x, h^*)$  at  $x = 1/2$  used to compute our estimator. It can be seen that the estimator  $\hat{\gamma}_a$  generally uses a small bandwidth, which can be interpreted as an indicator of why our estimator generally mimics the shape of the function  $\gamma$  fairly well.

## 5 Proofs

For the sake of simplicity, we introduce the notation  $k_{x,a} := \lfloor (1-a)k_x \rfloor$ .

### 5.1 Proof of the uniform weak consistency

We shall prove that for all  $\varepsilon > 0$ , the probability

$$p_n := \mathbb{P} \left( \sup_{x \in \Omega} |\hat{\gamma}_a(x, k_x, h) - \gamma(x)| > \varepsilon \right),$$

converges to 0 as  $n$  goes to infinity. The proof is based on [14, Lemma 1]: the basic idea is that instead of showing the uniform consistency on the whole set  $\Omega$ , one can show uniform consistency on a sequence of “sufficiently large” subsets  $\Omega_n$  of  $\Omega$  and deal with the oscillation of the estimator.

First note that, since  $\Omega$  is a compact subset of  $\mathbb{R}^d$ , for a fixed  $\eta > 1/\beta_f$  and every  $n \in \mathbb{N} \setminus \{0\}$ , there exists a finite subset  $\Omega_n$  of  $\Omega$  with  $\text{card}(\Omega_n) = \mathcal{O}(n^c)$ ,  $c > 0$  such that for all  $x \in \Omega$ , one can find  $\chi(x) \in \Omega_n$  satisfying  $\|x - \chi(x)\| < n^{-\eta}$ . The triangular inequality yields:

$$\begin{aligned} p_n &\leq \mathbb{I} \left\{ \sup_{x \in \Omega} |\gamma(x) - \gamma(\chi(x))| > \varepsilon/3 \right\} + \mathbb{P} \left( \sup_{\omega \in \Omega_n} |\hat{\gamma}_a(\omega, k_\omega, h) - \gamma(\omega)| > \varepsilon/3 \right) \\ &\quad + \mathbb{P} \left( \sup_{x \in \Omega} |\hat{\gamma}_a(x, k_x, h) - \hat{\gamma}_a(\chi(x), k_{\chi(x)}, h)| > \varepsilon/3 \right). \end{aligned} \quad (8)$$

The proof of the uniform weak consistency of our estimator consists in showing that the three terms in the above inequality converge to 0 as  $n$  goes to infinity. This is carried out in Propositions 1, 2 and 3. Theorem 1 is thus a direct consequence of these results. We start by focusing on the convergence of the first term.

**Proposition 1.** *Under model  $(\mathcal{M})$  and **(A1)**, for  $n$  large enough,*

$$\sup_{x \in \Omega} |\gamma(x) - \gamma(\chi(x))| \leq \varepsilon/3.$$

**Proof of Proposition 1** – Recall that for all  $x \in \Omega$ ,  $\|x - \chi(x)\| < n^{-\eta} \rightarrow 0$ . Since  $\Omega$  is compact, **(A1)** entails that the function  $\gamma$  is uniformly continuous, which shows the result. ■

We are now interested in the second term, namely in the uniform convergence of our estimator on the finite subsets  $\Omega_n$  of  $\Omega$ . Some preliminary lemmas are required, whose proofs are postponed to the Appendix. The first one is a useful result of real analysis.

**Lemma 1.** *Let  $(a_1, \dots, a_n)$  and  $(b_1, \dots, b_n)$  be two  $n$ -tuples of pairwise distinct real numbers such that for all  $i \in \{1, \dots, n\}$ ,  $a_i \leq b_i$ . Let further  $a_{1,n} \leq \dots \leq a_{n,n}$  and  $b_{1,n} \leq \dots \leq b_{n,n}$  be the associated ordered ntuples. Then for all  $i \in \{1, \dots, n\}$ ,  $a_{i,n} \leq b_{i,n}$ .*

Lemma 2 is a topological result which shall be needed in several proofs: it essentially implies that for  $n$  large enough, the ball  $B(x, h)$  is contained in  $S$  for all  $x \in \Omega$ .

**Lemma 2.** *There exists  $\beta > 0$  such that for every  $x \in \Omega$ ,  $B(x, \beta) \subset S$ .*

Lemma 3 below gives an asymptotic uniform estimation of the total number of covariates  $M_n(\omega, h)$  contained in the balls with center  $\omega \in \Omega_n$  and radius  $h$ .

**Lemma 3.** *Under model  $(\mathcal{M})$ , assume that **(A1)** holds together with  $nh^d/\log n \rightarrow \infty$ . Then, as  $n$  goes to infinity,*

$$\frac{1}{nh^d} \sup_{\omega \in \Omega_n} |M_n(\omega, h) - \mathcal{V}nh^d f(\omega)| \xrightarrow{\mathbb{P}} 0.$$

Given  $M_n(x, h) \geq 1$ , for  $i = 1, \dots, M_n(x, h)$ , let  $Z_i^{(x)}$  be the response variable whose associated covariate  $W_i^{(x)}$  belongs to the ball  $B(x, h)$ . Let us also introduce the notations  $U_i^{(x)} := \overline{F}(Z_i^{(x)} | W_i^{(x)})$  for  $i = 1, \dots, M_n(x, h)$  and  $V_i = \overline{F}(Y_i | X_i)$  for  $i = 1, \dots, n$ . In the following,  $\tilde{\Omega}$  denotes a finite subset of  $\Omega$ ,  $m := (m_\omega)_{\omega \in \tilde{\Omega}}$  is a list of positive integers and  $\mathcal{B}_{\tilde{\Omega}}(m)$  is the Borel measurable set

$$\mathcal{B}_{\tilde{\Omega}}(m) := \bigcap_{\omega \in \tilde{\Omega}} \{M_n(\omega, h) = m_\omega\}.$$

The distributions of  $U_i^{(x)}$  and  $V_i$  are given in the following result.

**Lemma 4.** *Under model  $(\mathcal{M})$ , the random variables  $V_1, \dots, V_n$  are independent standard uniform random variables which are independent from  $X_1, \dots, X_n$ . Furthermore, for all  $\omega \in \tilde{\Omega}$  and conditionally to  $\mathcal{B}_{\tilde{\Omega}}(m)$ , the random variables  $U_1^{(\omega)}, \dots, U_{m_\omega}^{(\omega)}$  are independent standard uniform random variables.*

The next lemma provides a representation of our estimator in terms of independent standard exponential random variables, which is the key argument to show Proposition 2.

**Lemma 5.** *Under model  $(\mathcal{M})$  and **(A3)**, for all  $\omega \in \tilde{\Omega}$  and conditionally to  $\mathcal{B}_{\tilde{\Omega}}(m)$ , there exist independent standard exponential random variables  $E_1^{(\omega)}, \dots, E_{m_\omega}^{(\omega)}$  such that for every sequence of real-valued functions  $(a_n)$  defined on  $\Omega$  such that  $a_n(x) \rightarrow a \in (0, 1)$  uniformly in  $x \in \Omega$ , one has for  $n$  large enough, uniformly in  $\omega \in \tilde{\Omega}$ ,*

$$\begin{aligned} \left| \widehat{\gamma}_{a_n(\omega)}(\omega, k_\omega, h) - \gamma(\omega) \overline{E}_n^{(\omega)} \right| &\leq 2\omega(U_{1, m_\omega}^{(\omega)}, U_{m_\omega, m_\omega}^{(\omega)}, \omega, h) + \overline{E}_n^{(\omega)} \overline{\Delta}_\omega(U_{k_\omega, m_\omega}^{(\omega)}) \\ &\leq 2\omega(V_{1, n}, V_{n, n}, \omega, h) + \overline{E}_n^{(\omega)} \overline{\Delta}_\omega(U_{k_\omega, m_\omega}^{(\omega)}) \end{aligned}$$

where

$$\overline{E}_n^{(\omega)} := \frac{1}{k_\omega - k_{\omega, a_n(\omega)} + 1} \sum_{l=k_{\omega, a_n(\omega)}}^{k_\omega} \frac{1}{l-1} \sum_{i=1}^{l-1} E_i^{(\omega)}.$$

We are now in position to prove the uniform consistency of our estimator on the finite subsets  $\Omega_n$ .

**Proposition 2.** *Under model  $(\mathcal{M})$ , assume that **(A1)**, **(A2)** and **(A3)** hold. If  $nh^d/\log n \rightarrow \infty$ ,*

$$\inf_{x \in \Omega} \min \left\{ \frac{k_x}{\log n}, \frac{nh^d}{k_x \log(nh^d)} \right\} \rightarrow \infty \text{ and } \limsup_{t \rightarrow 0} \sup_{x \in \Omega} \overline{\Delta}_x(t) = 0,$$

*then, for every sequence of real-valued functions  $(a_n)$  defined on  $\Omega$  such that  $a_n(x) \rightarrow a \in (0, 1)$  uniformly in  $x \in \Omega$  as  $n$  goes to infinity,*

$$\sup_{\omega \in \Omega_n} |\widehat{\gamma}_{a_n(\omega)}(\omega, k_\omega, h) - \gamma(\omega)| \xrightarrow{\mathbb{P}} 0.$$

Note that to show the convergence to zero of the second term in (8), it is obviously sufficient to use Proposition 2 with the constant sequence  $a_n = a$  for all  $n \geq 1$ . Proposition 2 also handles the case when  $(a_n)$  is an arbitrary sequence of real-valued functions on  $\Omega$  uniformly converging to  $a$ , which shall be useful to establish Proposition 3.

**Proof of Proposition 2** – Let  $m = (m_\omega)_{\omega \in \Omega_n}$  be a list of positive integers such that

$$\forall \omega \in \Omega_n, \frac{m_\omega}{f(\omega)nh^d} \in \left[ \frac{\mathcal{V}}{2}, \frac{3\mathcal{V}}{2} \right], \quad (9)$$

and let  $L_n$  be the set of all possible lists satisfying (9). From Lemma 3, it is clear that  $\mathbb{P}(\mathcal{A}_n) \rightarrow 1$  as  $n$  goes to infinity, where

$$\mathcal{A}_n := \bigcup_{m \in L_n} \mathcal{B}_{\Omega_n}(m)$$

is the disjoint union of the  $\mathcal{B}_{\Omega_n}(m)$  for  $m \in L_n$ . Let  $\varepsilon > 0$ . Remarking that

$$\begin{aligned} \mathbb{P} \left( \sup_{\omega \in \Omega_n} |\widehat{\gamma}_{a_n(\omega)}(\omega, k_\omega, h) - \gamma(\omega)| > \varepsilon \right) &\leq \mathbb{P}(\mathcal{A}_n^C) + \sum_{m \in L_n} \mathbb{P} \left( \sup_{\omega \in \Omega_n} |\widehat{\gamma}_{a_n(\omega)}(\omega, k_\omega, h) - \gamma(\omega)| > \varepsilon; \mathcal{B}_{\Omega_n}(m) \right) \\ &\leq \mathbb{P}(\mathcal{A}_n^C) + \sup_{m \in L_n} \mathbb{P} \left( \sup_{\omega \in \Omega_n} |\widehat{\gamma}_{a_n(\omega)}(\omega, k_\omega, h) - \gamma(\omega)| > \varepsilon \middle| \mathcal{B}_{\Omega_n}(m) \right), \end{aligned}$$

where  $\mathcal{A}_n^C$  is the complement of  $\mathcal{A}_n$ , it is sufficient to prove that as  $n$  goes to infinity,

$$\sup_{m \in L_n} T(m) := \sup_{m \in L_n} \mathbb{P} \left( \sup_{\omega \in \Omega_n} |\widehat{\gamma}_{a_n(\omega)}(\omega, k_\omega, h) - \gamma(\omega)| > \varepsilon \middle| \mathcal{B}_{\Omega_n}(m) \right) \rightarrow 0. \quad (10)$$

Let  $m \in L_n$ . Remarking that

$$T(m) \leq \mathbb{P} \left( \sup_{\omega \in \Omega_n} |\widehat{\gamma}_{a_n(\omega)}(\omega, k_\omega, h) - \gamma(\omega) \overline{E}_n^{(\omega)}| > \frac{\varepsilon}{2} \middle| \mathcal{B}_{\Omega_n}(m) \right) + \mathbb{P} \left( \sup_{\omega \in \Omega_n} |\gamma(\omega)(\overline{E}_n^{(\omega)} - 1)| > \frac{\varepsilon}{2} \middle| \mathcal{B}_{\Omega_n}(m) \right),$$

we have from Lemmas 4 and 5 that

$$\begin{aligned} T(m) &\leq \mathbb{P} \left( \sup_{\omega \in \Omega_n} \omega(V_{1,n}, V_{n,n}, \omega, h) > \frac{\varepsilon}{8} \right) + \mathbb{P} \left( \sup_{\omega \in \Omega_n} \overline{E}_n^{(\omega)} \overline{\Delta}_\omega(U_{k_\omega, m_\omega}^{(\omega)}) > \frac{\varepsilon}{4} \middle| \mathcal{B}_{\Omega_n}(m) \right) \\ &\quad + \mathbb{P} \left( \sup_{\omega \in \Omega_n} |\gamma(\omega)(\overline{E}_n^{(\omega)} - 1)| > \frac{\varepsilon}{2} \middle| \mathcal{B}_{\Omega_n}(m) \right) \\ &\leq \mathbb{P} \left( \sup_{\omega \in \Omega_n} \omega(V_{1,n}, V_{n,n}, \omega, h) > \frac{\varepsilon}{8} \right) + \text{card}(\Omega_n) \left\{ \sup_{\omega \in \Omega_n} \mathbb{P} \left( |\gamma(\omega)(\overline{E}_n^{(\omega)} - 1)| > \frac{\varepsilon}{2} \middle| \mathcal{B}_{\Omega_n}(m) \right) \right. \\ &\quad \left. + \sup_{\omega \in \Omega_n} \mathbb{P} \left( \overline{E}_n^{(\omega)} \overline{\Delta}_\omega(U_{k_\omega, m_\omega}^{(\omega)}) > \frac{\varepsilon}{4} \middle| \mathcal{B}_{\Omega_n}(m) \right) \right\} =: T_1(m) + \text{card}(\Omega_n)(T_2(m) + T_3(m)). \end{aligned}$$

First, let us consider the term  $T_1(m)$ . Under condition **(A2)**, for  $n$  large enough and uniformly in  $m$ ,

$$T_1(m) \leq \mathbb{P}(V_{1,n} < n^{-(1+\delta)}) + \mathbb{P}(V_{n,n} > 1 - n^{-(1+\delta)}) = 2(1 - (1 - n^{-(1+\delta)})^n) \rightarrow 0. \quad (11)$$

Regarding  $T_2(m)$ , it is easy to see that for  $n$  large enough

$$T_2(m) \leq \sup_{\omega \in \Omega_n} \sum_{l=k_\omega, a_n(\omega)}^{k_\omega} \mathbb{P} \left( \left| \frac{1}{l-1} \sum_{i=1}^{l-1} \gamma(\omega)(E_i^{(\omega)} - 1) \right| > \frac{\varepsilon}{2} \middle| \mathcal{B}_{\Omega_n}(m) \right).$$

From Lemma 4, using a classical Chernoff bound for independent standard random exponential variables together with **(A1)**, there exists a positive constant  $C_\varepsilon$  such that, for  $n$  large enough,

$$T_2(m) \leq 2 \sup_{\omega \in \Omega_n} \sum_{l=k_\omega, a_n(\omega)}^{k_\omega} \exp(-C_\varepsilon(l-1)) \leq 2 \exp\left(-\frac{C_\varepsilon}{2} \inf_{\omega \in \Omega_n} (k_{\omega, a_n(\omega)} - 1)\right).$$

Finally, using the fact that  $\text{card}(\Omega_n) = \mathcal{O}(n^c)$ ,  $k_{x, a_n(x)}/k_x \rightarrow 1 - a$  and  $k_x/\log(n) \rightarrow \infty$  uniformly in  $x \in \Omega$ , one has, for  $n$  sufficiently large, uniformly in  $m$ ,

$$\text{card}(\Omega_n) T_2(m) \leq 2 \exp\left(-\frac{C_\varepsilon}{4} \inf_{\omega \in \Omega_n} (k_{\omega, a_n(\omega)} - 1)\right) \rightarrow 0. \quad (12)$$

We now focus on  $T_3(m)$ . Let us define

$$\varepsilon_n^2 = \sup_{x \in \Omega} \frac{k_x \log(nh^d)}{nh^d}.$$

Clearly,  $\varepsilon_n \rightarrow 0$  as  $n$  goes to infinity and

$$T_3(m) \leq \sup_{\omega \in \Omega_n} \mathbb{P}\left(U_{k_\omega, m_\omega}^{(\omega)} > \varepsilon_n \mid \mathcal{B}_{\Omega_n}(m)\right) + \sup_{\omega \in \Omega_n} \mathbb{P}\left(\overline{E}_n^{(\omega)} \sup_{x \in \Omega} \overline{\Delta}_x(\varepsilon_n) > \frac{\varepsilon}{4} \mid \mathcal{B}_{\Omega_n}(m)\right).$$

Using Lemma 4, we have:

$$\begin{aligned} \mathbb{P}\left(U_{k_\omega, m_\omega}^{(\omega)} > \varepsilon_n \mid \mathcal{B}_{\Omega_n}(m)\right) &= \frac{m_\omega!}{(k_\omega - 1)!(m_\omega - k_\omega)!} \int_{\varepsilon_n}^1 x^{k_\omega - 1} (1-x)^{m_\omega - k_\omega} dx \\ &\leq m_\omega^{k_\omega} (1 - \varepsilon_n)^{m_\omega - k_\omega}. \end{aligned}$$

Remarking that  $\log(1 - \varepsilon_n) < -\varepsilon_n/2$  for  $n$  large enough, one has, for all  $m$  and  $\omega$ ,

$$\mathbb{P}\left(U_{k_\omega, m_\omega}^{(\omega)} > \varepsilon_n \mid \mathcal{B}_{\Omega_n}(m)\right) \leq \exp\left(-m_\omega \varepsilon_n \left(\frac{m_\omega - k_\omega}{2m_\omega} - \frac{k_\omega \log m_\omega}{m_\omega \varepsilon_n}\right)\right).$$

Furthermore, under **(A1)**, since  $m$  satisfies (9), we have:

$$\frac{k_\omega}{m_\omega} \leq \frac{2}{f\mathcal{V}} \frac{\varepsilon_n^2}{\log(nh^d)} \text{ and } \log(m_\omega) \leq \log\left(\frac{3f\mathcal{V}}{2} nh^d\right) \leq \frac{3}{2} \log(nh^d),$$

for all  $m$  and  $\omega$ . Thus, for  $n$  sufficiently large, uniformly in  $m$  and  $\omega$ ,

$$\mathbb{P}\left(U_{k_\omega, m_\omega}^{(\omega)} > \varepsilon_n \mid \mathcal{B}_{\Omega_n}(m)\right) \leq \exp\left(-\frac{1}{4} \inf_{\omega \in \Omega_n} m_\omega \varepsilon_n\right) \leq \exp\left(-\frac{f\mathcal{V}}{8} nh^d \varepsilon_n\right).$$

Furthermore, since  $\log n/(nh^d \varepsilon_n) \rightarrow 0$  and  $\text{card}(\Omega_n) = \mathcal{O}(n^c)$ , it is straightforward that for  $n$  sufficiently large, uniformly in  $m$ ,

$$\text{card}(\Omega_n) \sup_{\omega \in \Omega_n} \mathbb{P}\left(U_{k_\omega, m_\omega}^{(\omega)} > \varepsilon_n \mid \mathcal{B}_{\Omega_n}(m)\right) \leq \exp\left(-\frac{f\mathcal{V}}{16} nh^d \varepsilon_n\right). \quad (13)$$

Next, since  $\varepsilon_n \rightarrow 0$  and that, by assumption, for  $n$  large enough,

$$\sup_{x \in \Omega} \overline{\Delta}_x(\varepsilon_n) \leq \frac{\varepsilon}{8},$$

one has, under **(A1)**:

$$\begin{aligned} \sup_{\omega \in \Omega_n} \mathbb{P}\left(\overline{E}_n^{(\omega)} \sup_{x \in \Omega} \overline{\Delta}_x(\varepsilon_n) > \frac{\varepsilon}{4} \mid \mathcal{B}_{\Omega_n}(m)\right) &\leq \sup_{\omega \in \Omega_n} \mathbb{P}\left(\sup_{x \in \Omega} \overline{\Delta}_x(\varepsilon_n) |\overline{E}_n^{(\omega)} - 1| > \frac{\varepsilon}{8} \mid \mathcal{B}_{\Omega_n}(m)\right) \\ &\leq \sup_{\omega \in \Omega_n} \mathbb{P}\left(\gamma(\omega) |\overline{E}_n^{(\omega)} - 1| > \gamma \mid \mathcal{B}_{\Omega_n}(m)\right). \end{aligned}$$

The right hand-side of the above inequality is similar to  $T_2(m)$  and thus (12) and (13) lead to

$$\text{card}(\Omega_n)T_3(m) \leq \exp\left(-\frac{f\mathcal{V}}{16}nh^d\varepsilon_n\right) + 2\exp\left(-\frac{C'_\varepsilon}{4}\inf_{\omega \in \Omega_n}(k_{\omega, a_n(\omega)} - 1)\right) \rightarrow 0, \quad (14)$$

for  $n$  large enough, uniformly in  $m$ , where  $C'_\varepsilon$  is a positive constant. We then easily obtain (10) using (11), (12) and (14) and the proof is complete.  $\blacksquare$

The oscillation of the function  $x \mapsto \widehat{\gamma}_a(x, k_x, h)$  is studied in Proposition 3. The proof of this result requires to control the random variable

$$C_h(x, r) = \sum_{i=1}^n \mathbb{I}\{h - r \leq \|X_i - x\| \leq h + r\},$$

which is the total number of covariates in the annulus with center  $x$ , inner radius  $h - r$  and outer radius  $h + r$ . Lemma 6 below essentially states that this number is asymptotically bounded with arbitrarily large probability.

**Lemma 6.** *Under model  $(\mathcal{M})$ , assume that **(A1)** holds together with  $nh^d \rightarrow \infty$ . Then, for every arbitrary integer  $K_2 > c/(\eta\beta_f - 1)$ ,  $\mathbb{P}(\mathcal{A}_{n, K_2}) \rightarrow 0$  as  $n$  goes to infinity, where*

$$\mathcal{A}_{n, K_2} = \left\{ \sup_{\omega \in \Omega_n} C_h(\omega, n^{-\eta}) \geq K_2 \right\}.$$

We can now state and prove Proposition 3: the oscillation of the function  $x \mapsto \widehat{\gamma}_a(x, k_x, h)$  converges uniformly to 0 in probability.

**Proposition 3.** *Under model  $(\mathcal{M})$ , assume that **(A1)**, **(A2)** and **(A3)** hold. If  $nh^d/\log n \rightarrow \infty$ ,*

$$\inf_{x \in \Omega} \min \left\{ \frac{k_x}{\log n}, \frac{nh^d}{k_x \log(nh^d)} \right\} \rightarrow \infty, \quad \limsup_{t \rightarrow 0} \sup_{x \in \Omega} \overline{\Delta}_x(t) = 0,$$

and if there exists a finite positive constant  $K_1$  such that

$$\sup_{x \in \Omega} \sup_{\|x' - x\| \leq h} |k_x - k_{x'}| \leq K_1,$$

then, as  $n$  goes to infinity, if  $a \in (0, 1)$ ,

$$\sup_{x \in \Omega} |\widehat{\gamma}_a(x, k_x, h) - \widehat{\gamma}_a(\chi(x), k_{\chi(x)}, h)| \xrightarrow{\mathbb{P}} 0.$$

**Proof of Proposition 3** – From Lemma 6, it is enough to show that for all  $\varepsilon > 0$  and for a fixed integer  $K_2 > c/(\eta\beta_f - 1)$ ,

$$\mathbb{P}\left(\sup_{x \in \Omega} |\widehat{\gamma}_a(x, k_x, h) - \widehat{\gamma}_a(\chi(x), k_{\chi(x)}, h)| > \varepsilon \mid \mathcal{A}_{n, K_2}^C\right) \rightarrow 0.$$

For  $(k, l) \in \{2, \dots, n\}^2$  and  $i \in \{1, \dots, n - 1\}$ , let

$$r_{i, l}(x, k, h) = \frac{\mathbb{I}\{\|X_i^* - x\| \leq h; M_l(x, h) \geq k\}}{M_l(x, h) - 1},$$

if  $M_l(x, h) > 1$  and 0 elsewhere, and, for  $a \in (0, 1)$  and  $k_a = \lfloor (1 - a)k \rfloor$ ,

$$s_{l, a}(x, k, h) = \frac{\mathbb{I}\{\|X_l^* - x\| \leq h; M_l(x, h) \leq k\}}{k - k_a + 1}.$$

Clearly

$$\widehat{\gamma}_a(x, k_x, h) = \sum_{l=2}^n \sum_{i=1}^{l-1} \log \frac{Y_{n-i+1, n}}{Y_{n-l+1, n}} r_{i, l}(x, k_{x, a}, h) s_{l, a}(x, k_x, h),$$

and thus  $|\widehat{\gamma}_a(x, k_x, h) - \widehat{\gamma}_a(\chi(x), k_{\chi(x)}, h)| \leq S_{n,1}(x) + S_{n,2}(x)$ , where

$$S_{n,1}(x) := \sum_{l=2}^n \sum_{i=1}^{l-1} \log \frac{Y_{n-i+1, n}}{Y_{n-l+1, n}} |r_{i, l}(x, k_{x, a}, h) - r_{i, l}(\chi(x), k_{\chi(x), a}, h)| s_{l, a}(x, k_x, h),$$

and  $S_{n,2}(x) := \sum_{l=2}^n \sum_{i=1}^{l-1} \log \frac{Y_{n-i+1, n}}{Y_{n-l+1, n}} |s_{l, a}(x, k_x, h) - s_{l, a}(\chi(x), k_{\chi(x)}, h)| r_{i, l}(\chi(x), k_{\chi(x), a}, h).$

The idea of the rest of the proof is quite simple. We will show that on the event  $\mathcal{A}_{n, K_2}^C$ , there exist two sequences of real-valued functions  $(a_n^-)$  and  $(a_n^+)$  on  $\Omega$  uniformly tending to  $a$ , four sequences  $(\alpha_{1, n}^-)$ ,  $(\alpha_{1, n}^+)$ ,  $(\alpha_{2, n}^-)$  and  $(\alpha_{2, n}^+)$  tending to 1 and a positive constant  $K_3$  such that, for all  $x \in \Omega$

$$S_{n,1}(x) \leq 2 \left( \alpha_{1, n}^+ \widehat{\gamma}_{a_n^+}(\chi(x))(\chi(x), k_{\chi(x)} + K_3, h^+) - \alpha_{1, n}^- \widehat{\gamma}_{a_n^-}(\chi(x))(\chi(x), k_{\chi(x)} - K_3, h^-) \right), \quad (15)$$

and

$$S_{n,2}(x) \leq 2 \left( \alpha_{2, n}^+ \widehat{\gamma}_{a_n^+}(\chi(x))(\chi(x), k_{\chi(x)} + K_3, h^+) - \alpha_{2, n}^- \widehat{\gamma}_{a_n^-}(\chi(x))(\chi(x), k_{\chi(x)} - K_3, h^-) \right) \quad (16)$$

where  $h^\pm := h \pm n^{-\eta}$ . Since  $\inf_{x \in \Omega} k_x \rightarrow \infty$ ,  $h^\pm = h(1 + o(1))$  and the function  $\gamma$  is bounded from below and above by positive constants, a direct use of Proposition 2 shall then lead to

$$\sup_{x \in \Omega} S_{n,1}(x) \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad \sup_{x \in \Omega} S_{n,2}(x) \xrightarrow{\mathbb{P}} 0,$$

which will then conclude the proof of Proposition 3. To obtain (15) and (16), the following straightforward results will be useful. For all  $(x, x') \in \Omega^2$  such that  $\|x - x'\| \leq n^{-\eta}$  and for all  $i \in \{1, \dots, n-1\}$ ,

$$|\mathbb{I}\{\|X_i^* - x\| \leq h\} - \mathbb{I}\{\|X_i^* - x'\| \leq h\}| \leq \mathbb{I}\{h^- \leq \|X_i^* - x'\| \leq h^+\}. \quad (17)$$

Furthermore, from the inequalities

$$|M_l(x, h) - M_l(x', h)| \leq C_h(x', n^{-\eta}) \quad \text{and} \quad |M_l(x', h) - M_l(x', h^\pm)| \leq C_h(x', n^{-\eta}),$$

the triangular inequality yields, for all  $l \in \{2, \dots, n\}$ , on  $\mathcal{A}_{n, K_2}^C$ ,

$$|M_l(x, h) - M_l(x', h^\pm)| \leq 2C_h(x', n^{-\eta}) \leq 2K_2. \quad (18)$$

Especially, if  $M_l(x, h) > 1$  and on  $\mathcal{A}_{n, K_2}^C$ ,

$$\frac{M_l(x', h^+) - 1}{M_l(x, h) - 1} \leq 1 + \frac{2K_2}{M_l(x, h) - 1} \quad \text{and} \quad \frac{M_l(x', h^-) - 1}{M_l(x, h) - 1} \geq 1 - \frac{2K_2}{M_l(x, h) - 1}. \quad (19)$$

Let us first focus on the term  $S_{n,1}(x)$ . It is easy to see that

$$D_{i, l}^{(r)}(x, a, h) := |r_{i, l}(x, k_{x, a}, h) - r_{i, l}(\chi(x), k_{\chi(x), a}, h)| \leq T_{n,1}^{(r)}(x) + T_{n,2}^{(r)}(x) + T_{n,3}^{(r)}(x),$$

where

$$T_{n,1}^{(r)}(x) = \frac{\mathbb{I}\{\|X_i^* - x\| \leq h\} - \mathbb{I}\{\|X_i^* - \chi(x)\| \leq h\} \mathbb{I}\{M_l(\chi(x), h) \geq k_{\chi(x),a}\}}{M_l(x, h) - 1},$$

if  $M_l(\chi(x), h) \geq k_{\chi(x),a}$  and 0 otherwise,

$$T_{n,2}^{(r)}(x) = \frac{|\mathbb{I}\{M_l(x, h) \geq k_{x,a}\} - \mathbb{I}\{M_l(\chi(x), h) \geq k_{\chi(x),a}\}| \mathbb{I}\{\|X_i^* - x\| \leq h\}}{M_l(x, h) - 1},$$

if  $M_l(x, h) \geq k_{x,a}$  or  $M_l(\chi(x), h) \geq k_{\chi(x),a}$  and 0 otherwise and

$$T_{n,3}^{(r)}(x) = \frac{|M_l(\chi(x), h) - M_l(x, h)| \mathbb{I}\{\|X_i^* - \chi(x)\| \leq h\} \mathbb{I}\{M_l(\chi(x), h) \geq k_{\chi(x),a}\}}{(M_l(x, h) - 1)(M_l(\chi(x), h) - 1)},$$

if  $M_l(\chi(x), h) \geq k_{\chi(x),a}$  and 0 otherwise. Note that for  $n$  large enough, since  $\inf_{x \in \Omega} k_{x,a} \rightarrow \infty$  and (18) holds, if  $M_l(\chi(x), h) \geq k_{\chi(x),a}$  then  $M_l(x, h) > 1$  and thus the terms  $T_{n,i}^{(r)}(x)$ ,  $i = 1, 2, 3$  are asymptotically well defined. We now study separately these three terms. For  $u \in \mathbb{R}$ , let us introduce the quantities

$$\begin{aligned} \xi^+(u) &= \sup_{x \in \Omega} \left( 1 + \frac{2K_2}{k_{x,a} - u - 1} \right), \quad \xi^-(u) = \inf_{x \in \Omega} \left( 1 - \frac{2K_2}{k_{x,a} - u - 1} \right), \\ \zeta^+(u) &= \sup_{x \in \Omega} \frac{k_{\chi(x)} - k_{\chi(x),a} + 1 + u}{k_x - k_{x,a} + 1} \quad \text{and} \quad \zeta^-(u) = \inf_{x \in \Omega} \frac{k_{\chi(x)} - k_{\chi(x),a} + 1 - u}{k_x - k_{x,a} + 1}. \end{aligned}$$

Clearly, for all  $u \in \mathbb{R}$ ,  $\xi^\pm(u)$  and  $\zeta^\pm(u)$  converge to one as  $n$  goes to infinity. From (17), (18) and (19), since for all  $l \in \{2, \dots, n\}$ ,  $i \in \{1, \dots, n-1\}$  and  $x \in \Omega$ ,  $r_{i,l}(x, \cdot, h)$  is a decreasing function, one has

$$\begin{aligned} T_{n,1}^{(r)}(x) &\leq \frac{M_l(\chi(x), h^+) - 1}{M_l(x, h) - 1} r_{i,l}(\chi(x), k_{\chi(x),a}, h^+) - \frac{M_l(\chi(x), h^-) - 1}{M_l(x, h) - 1} r_{i,l}(\chi(x), k_{\chi(x),a}, h^-) \\ &\leq \xi^+(2K_2) r_{i,l}(\chi(x), k_{\chi(x),a} + K_4, h^+) - \xi^-(2K_2) r_{i,l}(\chi(x), k_{\chi(x),a} - K_4, h^-), \end{aligned} \quad (20)$$

where  $K_4 = (a-1)K_1 - 2K_2 - 1$ . Similarly, since  $|k_{x,a} - k_{\chi(x),a}| \leq (1-a)K_1 + 1$  uniformly in  $x \in \Omega$ , noting that

$$|\mathbb{I}\{M_l(x, h) \geq k_{x,a}\} - \mathbb{I}\{M_l(\chi(x), h) \geq k_{\chi(x),a}\}| \leq \mathbb{I}\{k_{\chi(x),a} + K_4 \leq M_l(\chi(x), h) < k_{\chi(x),a} - K_4\}$$

yields

$$\begin{aligned} T_{n,2}^{(r)}(x) &\leq \frac{M_l(\chi(x), h^+) - 1}{M_l(x, h) - 1} r_{i,l}(\chi(x), k_{\chi(x),a} + K_4, h^+) - \frac{M_l(\chi(x), h^-) - 1}{M_l(x, h) - 1} r_{i,l}(\chi(x), k_{\chi(x),a} - K_4, h^-) \\ &\leq \xi^+(2K_2 - K_4) r_{i,l}(\chi(x), k_{\chi(x),a} + K_4, h^+) - \xi^-(2K_2 + K_4) r_{i,l}(\chi(x), k_{\chi(x),a} - K_4, h^-). \end{aligned} \quad (21)$$

Clearly

$$T_{n,3}^{(r)}(x) \leq \frac{K_2 \xi^+(2K_2)}{k_{\chi(x),a} - 1} r_{i,l}(\chi(x), k_{\chi(x),a} + K_4, h^+), \quad (22)$$

and  $K_2 \xi^+(2K_2)/(k_{\chi(x),a} - 1) \rightarrow 0$  uniformly in  $x \in \Omega$ . Furthermore, using once again (17), (18) and (19), letting  $K_3 = K_1 + 2K_2$  and  $K_5 = K_3 - K_4$ , one has

$$\zeta^-(K_5) s_{l,a_n^-(\chi(x))}(\chi(x), k_{\chi(x)} - K_3, h^-) \leq s_{l,a}(x, k_x, h) \leq \zeta^+(K_5) s_{l,a_n^+(\chi(x))}(\chi(x), k_{\chi(x)} + K_3, h^+), \quad (23)$$



where the sequences of functions  $(a_n^+)$  and  $(a_n^-)$  are given by

$$\forall x \in \Omega, a_n^\pm(x) = 1 - \frac{k_{x,a} \pm K_4}{k_x \pm K_3}.$$

Collecting (20) to (23) it is easy to construct two sequences  $(\alpha_{1,n}^-)$  and  $(\alpha_{1,n}^+)$  tending to 1 such that

$$\begin{aligned} D_{i,l}^{(r)}(x, a, h) s_{l,a}(x, k_x, h) &\leq 2 \left( \alpha_{1,n}^+ r_{i,l}(\chi(x), k_{\chi(x),a} + K_4, h^+) s_{l,a_n^+(\chi(x))}(\chi(x), k_{\chi(x)} + K_3, h^+) \right. \\ &\quad \left. - \alpha_{1,n}^- r_{i,l}(\chi(x), k_{\chi(x),a} - K_4, h^-) s_{l,a_n^-(\chi(x))}(\chi(x), k_{\chi(x)} - K_3, h^-) \right), \end{aligned}$$

which concludes the proof of (15). We now turn to  $S_{n,2}(x)$ . We first start from the decomposition

$$D_{l,a}^{(s)}(x, k_x, h) := |s_{l,a}(x, k_x, h) - s_{l,a}(\chi(x), k_{\chi(x)}, h)| \leq T_{n,1}^{(s)}(x) + T_{n,2}^{(s)}(x) + T_{n,3}^{(s)}(x),$$

where

$$\begin{aligned} T_{n,1}^{(s)}(x) &= \frac{|\mathbb{I}\{M_l(x, h) \leq k_x\} - \mathbb{I}\{M_l(\chi(x), h) \leq k_{\chi(x)}\}| \mathbb{I}\{\|X_l^* - \chi(x)\| \leq h\}}{k_{\chi(x)} - k_{\chi(x),a} + 1}, \\ T_{n,2}^{(s)}(x) &= \frac{|\mathbb{I}\{\|X_l^* - x\| \leq h\} - \mathbb{I}\{\|X_l^* - \chi(x)\| \leq h\}| \mathbb{I}\{M_l(x, h) \leq k_x\}}{k_{\chi(x)} - k_{\chi(x),a} + 1}, \end{aligned}$$

and

$$T_{n,3}^{(s)}(x) = \left| \frac{1}{k_x - k_{x,a} + 1} - \frac{1}{k_{\chi(x)} - k_{\chi(x),a} + 1} \right| \mathbb{I}\{\|X_l^* - x\| \leq h\} \mathbb{I}\{M_l(x, h) \leq k_x\}.$$

A conjoint use of (17), (18) and (19) leads to

$$\begin{aligned} T_{n,1}^{(s)}(x) &\leq \frac{\mathbb{I}\{\|X_l^* - \chi(x)\| \leq h^+\} \mathbb{I}\{M_l(\chi(x), h^+) \leq k_{\chi(x)} + K_3\}}{k_{\chi(x)} - k_{\chi(x),a} + 1} \\ &\quad - \frac{\mathbb{I}\{\|X_l^* - \chi(x)\| \leq h^-\} \mathbb{I}\{M_l(\chi(x), h^-) \leq k_{\chi(x)} - K_3\}}{k_{\chi(x)} - k_{\chi(x),a} + 1} \\ &\leq \frac{\zeta^+(K_5)}{\zeta^-(0)} s_{l,a_n^+(\chi(x))}(\chi(x), k_{\chi(x)} + K_3, h^+) - \frac{\zeta^-(K_5)}{\zeta^+(0)} s_{l,a_n^-(\chi(x))}(\chi(x), k_{\chi(x)} - K_3, h^-). \end{aligned} \quad (24)$$

Similarly,

$$\begin{aligned} T_{n,2}^{(s)}(x) &\leq \frac{\mathbb{I}\{h^- \leq \|X_l^* - \chi(x)\| \leq h^+\} \mathbb{I}\{M_l(x, h) \leq k_x\}}{k_{\chi(x)} - k_{\chi(x),a} + 1} \\ &\leq \frac{\zeta^+(K_5)}{\zeta^-(0)} s_{l,a_n^+(\chi(x))}(\chi(x), k_{\chi(x)} + K_3, h^+) - \frac{\zeta^-(K_5)}{\zeta^+(0)} s_{l,a_n^-(\chi(x))}(\chi(x), k_{\chi(x)} - K_3, h^-). \end{aligned} \quad (25)$$

Next, (23) yields

$$\begin{aligned} T_{n,3}^{(s)}(x) &\leq [(\zeta^+(0) - 1) \vee (1 - \zeta^-(0))] \frac{k_x - k_{x,a} + 1}{k_{\chi(x)} - k_{\chi(x),a} + 1} s_{l,a}(x, k_x, h) \\ &\leq [(\zeta^+(0) - 1) \vee (1 - \zeta^-(0))] \frac{\zeta^+(K_5)}{\zeta^-(0)} s_{l,a_n^+(\chi(x))}(\chi(x), k_{\chi(x)} + K_3, h^+). \end{aligned} \quad (26)$$

Remarking that

$$\xi^-(0) r_{i,l}(\chi(x), k_{\chi(x),a} - K_4, h^-) \leq r_{i,l}(\chi(x), k_{\chi(x),a}, h) \leq \xi^+(0) r_{i,l}(\chi(x), k_{\chi(x),a} + K_4, h^+), \quad (27)$$

and collecting (24) to (27), one can find sequences  $(\alpha_{2,n}^-)$  and  $(\alpha_{2,n}^+)$  tending to 1 such that

$$\begin{aligned} D_{l,a}^{(s)}(x, k_x, h) r_{i,l}(\chi(x), k_{\chi(x),a}, h) &\leq 2 \left( \alpha_{2,n}^+ r_{i,l}(\chi(x), k_{\chi(x),a} + K_4, h^+) s_{l,a_n^+(\chi(x))}(\chi(x), k_{\chi(x)} + K_3, h^+) \right. \\ &\quad \left. - \alpha_{2,n}^- r_{i,l}(\chi(x), k_{\chi(x),a} - K_4, h^-) s_{l,a_n^-(\chi(x))}(\chi(x), k_{\chi(x)} - K_3, h^-) \right), \end{aligned}$$

which entails (16) and thus concludes the proof.  $\blacksquare$

## 5.2 Proof of the asymptotic normality

The following lemma provides a representation of the estimator  $\hat{\gamma}_a(x, k_x, h)$ .

**Lemma 7.** *Under model  $(\mathcal{M})$ , assume that **(A1)**, **(A3)** and **(A4)** hold,  $m_x \rightarrow \infty$ ,  $k_x/m_x \rightarrow 0$ ,  $k_x^{1/2}\omega(m_x^{-1-\delta}, 1 - m_x^{-1-\delta}, x, h) \rightarrow 0$  and  $k_x^{1/2}\Delta(m_x/k_x|x) \rightarrow \xi(x) \in \mathbb{R}$  as  $n$  goes to infinity. Under  $\{M_n(x, h) = m_x\}$ , there exist independent standard exponential variables  $\tilde{E}_1^{(x)}, \dots, \tilde{E}_{k_x}^{(x)}$  such that for all  $a \in [0, 1)$ ,*

$$\hat{\gamma}_a(x, k_x, h) = \gamma(x) + \frac{\Delta(m_x/k_x|x)}{1 - \rho(x)} \mathcal{AB}(a, x) + S_n(x) + o_{\mathbb{P}}(k_x^{-1/2}),$$

where

$$S_n(x) = \frac{1}{k_x - k_{x,a} + 1} \sum_{l=k_{x,a}}^{k_x} \frac{1}{l-1} \sum_{i=1}^{l-1} \left( \gamma(x) + \Delta(m_x/k_x|x) \left( \frac{i}{k_x} \right)^{-\rho(x)} \right) (\tilde{E}_i^{(x)} - 1).$$

In view of the previous lemma, to obtain the asymptotic normality of our estimator, we shall show that  $k_x^{1/2}S_n(x) \xrightarrow{d} \mathcal{N}(0, \gamma^2(x)\mathcal{AV}(a))$ . Since  $k_x^{1/2}\Delta(m_x/k_x|x) \rightarrow \xi(x) \in \mathbb{R}$ , we get  $\Delta(m_x/k_x|x) \rightarrow 0$  as  $n$  goes to infinity. Therefore

$$S_n(x) = \frac{\gamma(x)(1 + o(1))}{k_x - k_{x,a} + 1} \sum_{i=1}^{k_x-1} \left( \sum_{l=(k_{x,a}-1) \vee i}^{k_x-1} \frac{1}{l} \right) (\tilde{E}_i^{(x)} - 1),$$

which makes it enough to prove that for all  $a \in [0, 1)$ ,

$$W_{n,a}(x) := \frac{k_x^{1/2}}{(\mathcal{AV}(a))^{1/2}(k_x - k_{x,a} + 1)} \sum_{i=1}^{k_x-1} \left( \sum_{l=(k_{x,a}-1) \vee i}^{k_x-1} \frac{1}{l} \right) (\tilde{E}_i^{(x)} - 1) \xrightarrow{d} \mathcal{N}(0, 1). \quad (28)$$

Let us first show that  $\text{Var}(W_{n,a}(x)) \rightarrow 1$  as  $n$  goes to infinity. In the simple case  $a = 0$ ,

$$\text{Var}(W_{n,0}(x)) = \frac{k_x}{\mathcal{AV}(0)} \sum_{i=1}^{k_x-1} \frac{1}{(k_x - 1)^2} \rightarrow 1,$$

since  $\mathcal{AV}(0) = 1$ . Now, if  $a \in (0, 1)$ ,

$$\text{Var}(W_{n,a}(x)) = \frac{k_x}{\mathcal{AV}(a)} \left( \frac{\sigma_{n,a}(x)}{k_x - k_{x,a} + 1} \right)^2,$$

where

$$\sigma_{n,a}^2(x) := \sum_{i=1}^{k_x-1} \left( \sum_{l=(k_{x,a}-1) \vee i}^{k_x-1} \frac{1}{l} \right)^2 = \sum_{i=1}^{k_x-1} \sum_{l, l'=(k_{x,a}-1) \vee i}^{k_x-1} \frac{1}{ll'} = \sum_{l, l'=k_{x,a}-1}^{k_x-1} \frac{l \wedge l'}{ll'},$$

by changing the order of summation. Hence, by breaking the second sum into two parts,

$$\sigma_{n,a}^2(x) = \sum_{l=k_{x,a}-1}^{k_x-1} \frac{l - k_{x,a} + 2}{l} + \sum_{l=k_{x,a}-1}^{k_x-1} \sum_{l'=l+1}^{k_x-1} \frac{1}{l'} = \sum_{l=k_{x,a}-1}^{k_x-1} \frac{2(l - k_{x,a}) + 3}{l},$$

where the last equality was obtained by changing the order of summation in the second term. Some straightforward computations lead to

$$\sigma_{n,a}^2(x) = 2k_x \left( \frac{k_x - k_{x,a} + 1}{k_x} - \frac{k_{x,a} - 3/2}{k_x} \sum_{l=k_{x,a}-1}^{k_x-1} \frac{1}{l} \right).$$

Recall that for every nonnegative monotonic and continuous function  $\varphi$  defined on  $(0, 1)$ , we have for all  $N, N' \in \mathbb{N} \setminus \{0\}$  with  $N' \leq N$ ,

$$\left| \frac{1}{N} \sum_{i=N'}^N \varphi\left(\frac{i}{N}\right) - \int_{N'/N}^1 \varphi(t) dt \right| \leq \frac{1}{N} \left( \varphi\left(\frac{N'}{N}\right) \vee \varphi(1) \right). \quad (29)$$

Applying (29) with  $\varphi(t) = 1/t$ , since  $k_x \rightarrow \infty$  and  $k_{x,a}/k_x \rightarrow 1 - a$ , one has

$$\sigma_{n,a}^2(x) = 2k_x(a + (1 - a) \log(1 - a))(1 + o(1)),$$

and thus  $\text{Var}(W_{n,a}(x)) = 1 + o(1)$ . It now only remains to show that for  $a \in [0, 1)$ ,

$$T_{n,a}(x) := \frac{k_x^{3/2}}{(k_x - k_{x,a} + 1)^3} \sum_{i=1}^{k_x-1} \left( \sum_{l=(k_{x,a}-1) \vee i}^{k_x-1} \frac{1}{l} \right)^3 \rightarrow 0,$$

and convergence (28) will be obtained by using Lyapounov's central limit theorem. First, if  $a = 0$ ,

$$T_{n,0}(x) = k_x^{3/2} \sum_{i=1}^{k_x-1} \frac{1}{(k_x - 1)^3} = k_x^{-1/2}(1 + o(1)) = o(1).$$

Finally, if  $a \in (0, 1)$ , since from (29) with  $\varphi(t) = 1/t$

$$\sum_{l=(k_{x,a}-1) \vee i}^{k_x-1} \frac{1}{l} \leq \sum_{l=k_{x,a}-1}^{k_x-1} \frac{1}{l} = -\log(1 - a)(1 + o(1)),$$

we have

$$T_{n,a}(x) \leq \frac{k_x^{3/2}(k_x - 1)[- \log(1 - a)]^3}{(k_x - k_{x,a} + 1)^3}(1 + o(1)) = k_x^{-1/2} \left[ \frac{- \log(1 - a)}{a} \right]^3 (1 + o(1)) \rightarrow 0,$$

and the proof is complete. ■

## References

- [1] Beirlant, J., Dierckx, G., Guillou, A., Starica, C. (2002). On exponential representations of logspacings of extreme order statistics, *Extremes*, **5(2)**, 157–180.
- [2] Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J. (2004). *Statistics of extremes - Theory and applications*. Wiley Series in Probability and Statistics.
- [3] Bingham, N.H., Goldie, C.M., Teugels, J.L. (1987). *Regular Variation*, Cambridge, U.K.: Cambridge University Press.
- [4] Chavez-Demoulin, V., Davison, A.C. (2005). Generalized additive modelling of sample extremes, *Journal of the Royal Statistical Society, series C*, **54**, 207–222.
- [5] Daouia, A., Gardes, L., Girard, S., Lekina, A. (2011). Kernel estimators of extreme level curves, *Test*, **20(2)**, 311–333.

- [6] Daouia, A., Gardes, L., Girard, S. (2013). On kernel smoothing for extremal quantile regression, *Bernoulli*, to appear.
- [7] Davison, A.C., Ramesh, N.I. (2000). Local likelihood smoothing of sample extremes, *Journal of the Royal Statistical Society, series B*, **62**, 191–208.
- [8] Davison, A.C., Smith, R.L. (1990). Models for exceedances over high thresholds, *Journal of the Royal Statistical Society, series B*, **52**, 393–442.
- [9] Gardes, L., Girard, S. (2008). A moving window approach for nonparametric estimation of the conditional tail index, *Journal of Multivariate Analysis*, **99**, 2368–2388.
- [10] Gardes, L., Girard, S. (2010). Conditional extremes from heavy-tailed distributions: an application to the estimation of extreme rainfall return levels, *Extremes*, **13**, 177–204.
- [11] Goegebeur, Y., Guillou, A., Schorgen, A. (2013). Nonparametric regression estimation of conditional tails – the random covariate case, *Statistics*, to appear.
- [12] de Haan, L., Ferreira, A. (2006). *Extreme value theory: An introduction*. Springer.
- [13] Hall, P., Tajvidi, N. (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data, *Statistical Science*, **15**, 153–167.
- [14] Härdle, W., Marron, J.S. (1985). Optimal bandwidth selection in nonparametric regression function estimation, *Annals of Statistics*, **13(4)**, 1465–1481.
- [15] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *Annals of Statistics*, **3**, 1163–1174.
- [16] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables, *Journal of the American Statistical Association*, **58**, 13–30.
- [17] Smith, R. L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone (with discussion), *Statistical Science*, **4**, 367–393.
- [18] Wang, H., Tsai, C.L. (2009). Tail index regression, *Journal of the American Statistical Association*, **104:487**, 1233–1240.

## Appendix

**Proof of Lemma 1** – Without loss of generality, we may assume that  $b_1 \leq \dots \leq b_n$ . Two cases are now considered: if  $a_{1,n} = a_1$ , then the two  $(n-1)$ -tuples  $(a_2, \dots, a_n)$  and  $(b_2, \dots, b_n)$  satisfy the assumptions of Lemma 1. If, on the contrary,  $a_{1,n} = a_i$  for some  $i > 1$ , then since  $a_j \leq b_{j+1}$  for all  $j \in \{1, \dots, i-1\}$  and  $a_j \leq b_j$  for all  $j \in \{i+1, \dots, n\}$ , the two  $(n-1)$ -tuples  $(a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n)$

and  $(b_2, \dots, b_n)$  satisfy the assumptions of Lemma 1. In conclusion, removing  $a_{1,n}$  and  $b_1$  from the two  $n$ -tuples  $(a_1, \dots, a_n)$  and  $(b_1, \dots, b_n)$  leads to  $(n-1)$ -tuples satisfying the assumptions of Lemma 1. Remarking that  $a_{1,n} \leq b_1$ , the conclusion of the proof is straightforward by induction on  $n$ . ■

**Proof of Lemma 2.** Let  $U$  denote the interior of  $S$  and  $\partial S = S \setminus U$  be the (topological) boundary of  $S$ . Note that  $\partial S$  is a closed set since it is the intersection of two closed sets in  $\mathbb{R}^d$ ; since  $\Omega$  is a compact set and  $\partial S$  is a closed set with  $\Omega \cap \partial S = \emptyset$ , it holds that

$$\exists \beta > 0, \quad d(\Omega, \partial S) := \inf_{x \in \Omega} \inf_{s \in \partial S} \|x - s\| = 2\beta > 0. \quad (30)$$

We shall now prove the result. Pick  $x \in \Omega$ . If one could find  $y \in B(x, \beta) \cap S^C$ , then the real number

$$t_0 = \inf\{t \in [0, 1] \mid z_t := (1-t)x + ty \notin S\}$$

would belong to  $(0, 1)$  since  $x \in U$  and  $y \in S^C$  which are both open sets. Therefore, because for every  $t \in (0, t_0)$ ,  $z_t \in S$  and there exists a nonincreasing sequence  $(t_k)$  converging to  $t_0$  such that  $(z_{t_k}) \subset S^C \subset U^C$  which is a closed set, one has

$$z_{t_0} = \lim_{t \uparrow t_0} z_t \in S \text{ and } z_{t_0} = \lim_{k \rightarrow \infty} z_{t_k} \in U^C.$$

Hence  $z_{t_0} \in \partial S$ , but  $\|x - z_{t_0}\| = t_0\|x - y\| < \beta$ , which contradicts (30): Lemma 2 is proven. ■

**Proof of Lemma 3** – We start with the following consequence of the triangular inequality:

$$\begin{aligned} |M_n(\omega, h) - \mathcal{V}nh^d f(\omega)| &\leq |n\mathbb{P}(\|X - \omega\| \leq h) - \mathcal{V}nh^d f(\omega)| + |M_n(\omega, h) - n\mathbb{P}(\|X - \omega\| \leq h)| \\ &=: T_{1,n}(\omega) + T_{2,n}(\omega). \end{aligned}$$

Note that Lemma 2 implies that one can take  $n$  large enough such that  $B(\omega, h) \subset S$  for every  $\omega \in \Omega_n$ . Then, under **(A1)**:

$$\frac{1}{nh^d} \sup_{\omega \in \Omega_n} T_{1,n}(\omega) \leq \sup_{\omega \in \Omega_n} \int_B |f(\omega + hu) - f(\omega)| du = \mathcal{O}(h^{\beta_f}) \rightarrow 0.$$

It thus remains to prove that, for all  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \frac{1}{nh^d} \sup_{\omega \in \Omega_n} T_{2,n}(\omega) > \varepsilon \right) = 0.$$

Remark that

$$\left\{ \frac{1}{nh^d} \sup_{\omega \in \Omega_n} T_{2,n}(\omega) > \varepsilon \right\} = \bigcup_{\omega \in \Omega_n} \left\{ \frac{1}{n} \left| \sum_{i=1}^n \mathbb{I}\{\|X_i - \omega\| \leq h\} - \mathbb{P}(\|X_i - \omega\| \leq h) \right| > \varepsilon h^d \right\}. \quad (31)$$

Since the  $X_i$ ,  $1 \leq i \leq n$ , are independent and identically distributed, Bernstein's inequality (see Hoeffding [16]) yields, for all  $\omega \in \Omega_n$ ,

$$\mathbb{P} \left( \frac{1}{n} \left| \sum_{i=1}^n \mathbb{I}\{\|X_i - \omega\| \leq h\} - \mathbb{P}(\|X_i - \omega\| \leq h) \right| > \varepsilon h^d \right) \leq 2 \exp \left( -\frac{\tau_n(\omega) \lambda_n(\omega)}{2(1 + \lambda_n(\omega)/3)} \right). \quad (32)$$

where we have defined

$$\tau_n(\omega) := \varepsilon n h^d \text{ and } \lambda_n(\omega) := \frac{\varepsilon h^d}{\mathbb{P}(\|X_i - \omega\| \leq h) \mathbb{P}(\|X_i - \omega\| > h)},$$

Since **(A1)** holds, there exists a positive constant  $\kappa_f$  such that, for all  $\omega \in \Omega_n$ ,

$$\mathbb{P}(\|X_i - \omega\| \leq h) \mathbb{P}(\|X_i - \omega\| > h) \leq \mathbb{P}(\|X - \omega\| \leq h) \leq \mathcal{V} h^d (f(\omega) + \kappa_f h^{\beta_f}).$$

Hence, for  $n$  large enough, under **(A1)**,

$$\sup_{\omega \in \Omega_n} \frac{1}{\lambda_n(\omega)} \leq \frac{2\mathcal{V}\bar{f}}{\varepsilon} < \infty. \quad (33)$$

Recalling that  $\text{card}(\Omega_n) = \mathcal{O}(n^c)$ , (31), (32) and (33) imply that there exists a positive constant  $\kappa_\varepsilon$  such that

$$\mathbb{P}\left(\frac{1}{nh^d} \sup_{\omega \in \Omega_n} T_{2,n}(\omega) > \varepsilon\right) = \mathcal{O}(n^c \exp(-\kappa_\varepsilon n h^d)) \rightarrow 0,$$

since  $nh^d/\log n \rightarrow \infty$ . ■

**Proof of Lemma 4** – Since the random pairs  $(X_i, Y_i)$ ,  $1 \leq i \leq n$ , are independent, it is straightforward to show that  $V_1, \dots, V_n$  are also independent. Furthermore, if  $t \in (0, 1)$ , one has for all  $i = 1, \dots, n$ :

$$\mathbb{P}(V_i \leq t) = \int_{\Omega} \mathbb{P}(\bar{F}(Y|s) \leq t | X = s) f(s) ds = t,$$

since

$$\mathbb{P}(\bar{F}(Y|s) \leq t | X = s) = \int_0^\infty \mathbb{I}\{\bar{F}(y|s) \leq t\} \varphi(y|s) dy = t,$$

where  $\varphi(\cdot|x)$  is the conditional probability density function of  $Y$  given  $X = x$ . Let now  $A$  be an arbitrary Borel subset of  $\Omega$ . For all  $i = 1, \dots, n$ ,

$$\mathbb{P}(V_i \leq t; X_i \in A) = \int_A \int_0^\infty \mathbb{I}\{\bar{F}(y|s) \leq t\} \varphi(y|s) dy f(s) ds = t \mathbb{P}(X_i \in A) = \mathbb{P}(V_i \leq t) \mathbb{P}(X_i \in A),$$

which concludes the first part of the proof. Furthermore, if  $(t_1, \dots, t_{m_\omega}) \in (0, 1)^{m_\omega}$ , we have

$$\mathbb{P}(\{U_i^{(\omega)} \leq t_i, i = 1, \dots, m_\omega\}; \mathcal{B}_{\bar{\Omega}}(m)) = \binom{n}{m_\omega} \mathbb{P}(\{V_i \leq t_i; X_i \in B(\omega, h), i = 1, \dots, m_\omega\}; \mathcal{B}_{\bar{\Omega}}(m)).$$

Using the first part of the proof and the fact that the event  $\mathcal{B}_{\bar{\Omega}}(m)$  belongs to the  $\sigma$ -algebra generated by  $X_1, \dots, X_n$ , one has

$$\mathbb{P}(\{U_i^{(\omega)} \leq t_i, i = 1, \dots, m_\omega\}; \mathcal{B}_{\bar{\Omega}}(m)) = t_1 \dots t_{m_\omega} \binom{n}{m_\omega} \mathbb{P}(\{X_i \in B(\omega, h), i = 1, \dots, m_\omega\}; \mathcal{B}_{\bar{\Omega}}(m)).$$

Remarking that

$$\mathbb{P}(\mathcal{B}_{\bar{\Omega}}(m)) = \binom{n}{m_\omega} \mathbb{P}(\{X_i \in B(\omega, h), i = 1, \dots, m_\omega\}; \mathcal{B}_{\bar{\Omega}}(m))$$

concludes the proof. ■

**Proof of Lemma 5** – Note that, conditionally to  $\mathcal{B}_{\tilde{\Omega}}(m)$ , for  $n$  large enough, uniformly in  $\omega \in \tilde{\Omega}$ ,

$$\widehat{\gamma}_{a_n(\omega)}(\omega, k_\omega, h) = \frac{1}{k_\omega - k_{\omega, a_n(\omega)} + 1} \sum_{l=k_{\omega, a_n(\omega)}}^{k_\omega} \frac{1}{l-1} \sum_{i=1}^{l-1} \log \left( \frac{Z_{m_\omega-i+1, m_\omega}^{(\omega)}}{Z_{m_\omega-l+1, m_\omega}^{(\omega)}} \right).$$

In this case, one has, for all  $i = 1, \dots, m_\omega$ ,

$$\log Z_i^{(\omega)} = \log q(U_i^{(\omega)} | W_i^{(\omega)}) \leq \log q(U_i^{(\omega)} | \omega) + \omega(U_{1, m_\omega}^{(\omega)}, U_{m_\omega, m_\omega}^{(\omega)}, \omega, h).$$

Lemma 1 now gives

$$\log Z_{m_\omega-i+1, m_\omega}^{(\omega)} \leq \log q(U_{i, m_\omega}^{(\omega)} | \omega) + \omega(U_{1, m_\omega}^{(\omega)}, U_{m_\omega, m_\omega}^{(\omega)}, \omega, h).$$

A similar idea yields

$$\log q(U_{i, m_\omega}^{(\omega)} | \omega) - \omega(U_{1, m_\omega}^{(\omega)}, U_{m_\omega, m_\omega}^{(\omega)}, \omega, h) \leq \log Z_{m_\omega-i+1, m_\omega}^{(\omega)},$$

and therefore

$$\left| \log Z_{m_\omega-i+1, m_\omega}^{(\omega)} - \log q(U_{i, m_\omega}^{(\omega)} | \omega) \right| \leq \omega(U_{1, m_\omega}^{(\omega)}, U_{m_\omega, m_\omega}^{(\omega)}, \omega, h).$$

This yields

$$\begin{aligned} \left| \log \left( \frac{Z_{m_\omega-i+1, m_\omega}^{(\omega)}}{Z_{m_\omega-l+1, m_\omega}^{(\omega)}} \right) - \log \left( \frac{q(U_{i, m_\omega}^{(\omega)} | \omega)}{q(U_{l, m_\omega}^{(\omega)} | \omega)} \right) \right| &\leq \left| \log Z_{m_\omega-i+1, m_\omega}^{(\omega)} - \log q(U_{i, m_\omega}^{(\omega)} | \omega) \right| \\ &+ \left| \log Z_{m_\omega-l+1, m_\omega}^{(\omega)} - \log q(U_{l, m_\omega}^{(\omega)} | \omega) \right| \\ &\leq 2\omega \left( U_{1, m_\omega}^{(\omega)}, U_{m_\omega, m_\omega}^{(\omega)}, \omega, h \right), \end{aligned}$$

which in turn implies

$$\left| \widehat{\gamma}_{a_n(\omega)}(\omega, k_\omega, h) - \frac{1}{k_\omega - k_{\omega, a_n(\omega)} + 1} \sum_{l=k_{\omega, a_n(\omega)}}^{k_\omega} \frac{1}{l-1} \sum_{i=1}^{l-1} i \log \left( \frac{q(U_{i, m_\omega}^{(\omega)} | \omega)}{q(U_{i+1, m_\omega}^{(\omega)} | \omega)} \right) \right| \leq 2\omega \left( U_{1, m_\omega}^{(\omega)}, U_{m_\omega, m_\omega}^{(\omega)}, \omega, h \right). \quad (34)$$

Note that to obtain (34), we have used the straightforward identity

$$\sum_{i=1}^{l-1} \log \left( \frac{q(U_{i, m_\omega}^{(\omega)} | \omega)}{q(U_{l, m_\omega}^{(\omega)} | \omega)} \right) = \sum_{i=1}^{l-1} i \log \left( \frac{q(U_{i, m_\omega}^{(\omega)} | \omega)}{q(U_{i+1, m_\omega}^{(\omega)} | \omega)} \right).$$

For all  $i \in \{1, \dots, m_\omega\}$ , let now  $F_i^{(\omega)} := -\log U_i^{(\omega)}$ . From Lemma 4, conditionally to  $\mathcal{B}_{\tilde{\Omega}}(m)$ ,  $F_1^{(\omega)}, \dots, F_{m_\omega}^{(\omega)}$  are independent standard exponential random variables. Condition **(A3)** then yields, for all  $i \in \{1, \dots, k_\omega - 1\}$ ,

$$i \log \left( \frac{q(U_{i, m_\omega}^{(\omega)} | \omega)}{q(U_{i+1, m_\omega}^{(\omega)} | \omega)} \right) = \gamma(\omega) i (F_{m_\omega-i+1, m_\omega}^{(\omega)} - F_{m_\omega-i, m_\omega}^{(\omega)}) + i \int_{(U_{i+1}^{(\omega)})^{-1}}^{(U_i^{(\omega)})^{-1}} \frac{\Delta(u | \omega)}{u} du.$$

Rényi's representation shows that

$$\left\{ E_i^{(\omega)} := i (F_{m_\omega-i+1, m_\omega}^{(\omega)} - F_{m_\omega-i, m_\omega}^{(\omega)}), \quad i = 1, \dots, k_\omega - 1 \right\}$$

are independent standard exponential random variables. Moreover, for all  $i \in \{1, \dots, k_\omega - 1\}$ ,

$$\left| i \log \left( \frac{q(U_{i, m_\omega}^{(\omega)} | \omega)}{q(U_{i+1, m_\omega}^{(\omega)} | \omega)} \right) - \gamma(\omega) E_i^{(\omega)} \right| \leq E_i^{(\omega)} \overline{\Delta}_\omega(U_{k_\omega, m_\omega}^{(\omega)}). \quad (35)$$

Inequalities (34) and (35) conclude the proof.  $\blacksquare$

**Proof of Lemma 6** – First remark that

$$\mathbb{P}(\mathcal{A}_{n,K_2}) \leq \sum_{\omega \in \Omega_n} \mathbb{P}(C_h(\omega, n^{-\eta}) \geq K_2).$$

Furthermore, for all  $\omega \in \Omega_n$ ,

$$\{C_h(\omega, n^{-\eta}) \geq K_2\} \subset \bigcup_{1 \leq i_1 < \dots < i_{K_2} \leq n} \bigcap_{q=1}^{K_2} \{h^- \leq \|X_{i_q} - \omega\| \leq h^+\},$$

where  $h^\pm := h \pm n^{-\eta}$ . Therefore,

$$\mathbb{P}(\mathcal{A}_{n,K_2}) \leq \text{card}(\Omega_n) \binom{n}{K_2} \sup_{\omega \in \Omega_n} (\mathbb{P}(h^- \leq \|X - \omega\| \leq h^+))^{K_2}.$$

Some straightforward calculus leads to

$$\mathbb{P}(h^- \leq \|X - \omega\| \leq h^+) = ((h^+)^d - (h^-)^d) (f(\omega)\mathcal{V} + R_{n,1}(\omega)) + R_{n,2}^+(\omega) - R_{n,2}^-(\omega),$$

where, if  $B$  is the unit ball of  $\mathbb{R}^d$ ,

$$R_{n,1}(\omega) = \int_B (f(\omega + hu) - f(\omega)) du \text{ and } R_{n,2}^\pm(\omega) = (h^\pm)^d \int_B (f(\omega + h^\pm u) - f(\omega + hu)) du.$$

Note that Lemma 2 implies that one can take  $n$  large enough such that  $B(\omega, h^+) \subset S$  for every  $\omega \in \Omega_n$ .

Under **(A1)**, and since  $\eta > 1/\beta_f \geq 1/d$  and  $nh^d \rightarrow \infty$  imply that  $n^\eta h \rightarrow \infty$ , we have, uniformly in  $\omega$ :

$$R_{n,1}(\omega) = \mathcal{O}(h^{\beta_f}) = o(1) \text{ and } R_{n,2}^\pm(\omega) = \mathcal{O}(h^d n^{-\eta\beta_f}).$$

Furthermore, remarking that, because  $n^\eta h \rightarrow \infty$ ,

$$(h^+)^d - (h^-)^d = h^d \left( \frac{2dn^{-\eta}}{h} + \mathcal{O}\left(\frac{n^{-\eta}}{h}\right) \right) = \mathcal{O}(h^{d-1}n^{-\eta}),$$

and since, under **(A1)**  $f$  is bounded, one has uniformly in  $\omega$

$$\mathbb{P}(h^- \leq \|X - \omega\| \leq h^+) = \mathcal{O}(h^{d-1}n^{-\eta}) + \mathcal{O}(h^d n^{-\eta\beta_f}) = \mathcal{O}(n^{-\eta\beta_f}),$$

because  $h \rightarrow 0$  and  $\beta_f \leq 1$ . Using the well-known equivalent

$$\binom{n}{K_2} = \frac{n^{K_2}}{K_2!} (1 + o(1)),$$

and since  $\text{card}(\Omega_n) = \mathcal{O}(n^c)$ , we thus have

$$\sup_{\omega \in \Omega_n} \mathbb{P}(C_h(\omega, n^{-\eta}) \geq K_2) = \mathcal{O}\left(n^{c+K_2(1-\eta\beta_f)}\right) = o(1)$$

and the proof is complete. ■

**Proof of Lemma 7** – From (34) (see proof of Lemma 5), conditionally to  $\{M_n(x, h) = m_x\}$ , one has

$$\left| \widehat{\gamma}_a(x, k_x, h) - \frac{1}{k_x - k_{x,a} + 1} \sum_{l=k_{x,a}}^{k_x} \frac{1}{l-1} \sum_{i=1}^{l-1} i \log \left( \frac{q(U_{i,m_x}^{(x)} | x)}{q(U_{i+1,m_x}^{(x)} | x)} \right) \right| \leq 2\omega \left( U_{1,m_x}^{(x)}, U_{m_x,m_x}^{(x)}, x, h \right), \quad (36)$$



where, conditionally to  $\{M_n(x, h) = m_x\}$ ,  $U_1^{(x)}, \dots, U_{m_x}^{(x)}$  are independent standard uniform variables. Furthermore, from [1, Theorem 2.1] there exist independent standard exponential variables  $\tilde{E}_1^{(x)}, \dots, \tilde{E}_{k_x-1}^{(x)}$  such that uniformly in  $i \in \{1, \dots, k_x - 1\}$ ,

$$i \log \left( \frac{q(U_{i, m_x}^{(x)} | x)}{q(U_{i+1, m_x}^{(x)} | x)} \right) = \left( \gamma(x) + \Delta \left( \frac{m_x}{k_x} \middle| x \right) \left( \frac{i}{k_x} \right)^{-\rho(x)} \right) \tilde{E}_i^{(x)} + \beta_i^{(x)} + o_{\mathbb{P}}(\Delta(m_x/k_x | x)), \quad (37)$$

where  $\beta_1^{(x)}, \dots, \beta_{k_x-1}^{(x)}$  are random variables satisfying

$$\left| \sum_{i=j}^{k_x-1} \frac{\beta_i^{(x)}}{i} \right| = o_{\mathbb{P}} \left( \Delta \left( \frac{m_x}{k_x} \middle| x \right) \max \left( \log \left( \frac{k_x}{j} \right), 1 \right) \right). \quad (38)$$

Note also that from [1, Lemma 2.3], equation (38) holds uniformly in  $j \in \{1, \dots, k_x - 1\}$ . Hence, (36) and (37) entail that

$$\begin{aligned} \left| \hat{\gamma}_a(x, k_x, h) - \gamma(x) - \frac{\Delta(m_x/k_x | x)}{1 - \rho(x)} \mathcal{AB}(a, x) - S_n(x) \right| &\leq 2\omega \left( U_{1, m_x}^{(x)}, U_{m_x, m_x}^{(x)}, x, h \right) + |R_{n,1}(x)| \\ &\quad + \left| \Delta \left( \frac{m_x}{k_x} \middle| x \right) \right| |R_{n,2}(x)| + o_{\mathbb{P}}(\Delta(m_x/k_x | x)), \end{aligned}$$

where

$$\begin{aligned} R_{n,1}(x) &= \frac{1}{k_x - k_{x,a} + 1} \sum_{l=k_{x,a}}^{k_x} \frac{1}{l-1} \sum_{i=1}^{l-1} \beta_i^{(x)} \\ \text{and } R_{n,2}(x) &= \frac{1}{k_x - k_{x,a} + 1} \sum_{l=k_{x,a}}^{k_x} \frac{1}{l-1} \sum_{i=1}^{l-1} \left( \frac{i}{k_x} \right)^{-\rho(x)} - \frac{\mathcal{AB}(a, x)}{1 - \rho(x)}. \end{aligned}$$

It thus remains to prove that

$$k_x^{1/2} \omega \left( U_{1, m_x}^{(x)}, U_{m_x, m_x}^{(x)}, x, h \right) \xrightarrow{\mathbb{P}} 0, \quad (39a)$$

$$k_x^{1/2} R_{n,1}(x) \xrightarrow{\mathbb{P}} 0, \quad (39b)$$

$$R_{n,2}(x) \rightarrow 0, \quad (39c)$$

with (39c) being sufficient to show that  $k_x^{1/2} \Delta(m_x/k_x | x) R_{n,2}(x) \rightarrow 0$ , since  $k_x^{1/2} \Delta(m_x/k_x | x)$  converges to a finite constant.

To prove (39a), let us introduce the event  $\tilde{\mathcal{A}}_n := \{U_{1, m_x}^{(x)} > m_x^{-1-\delta}\} \cap \{U_{m_x, m_x}^{(x)} < 1 - m_x^{-1-\delta}\}$ , where  $\delta > 0$  is taken so that **(A2)** holds. Clearly, since  $m_x \rightarrow \infty$ ,

$$\mathbb{P}(\tilde{\mathcal{A}}_n^C) \leq \mathbb{P}(U_{1, m_x}^{(x)} \leq m_x^{-1-\delta}) + \mathbb{P}(U_{m_x, m_x}^{(x)} \geq 1 - m_x^{-1-\delta}) = 2(1 - (1 - m_x^{-1-\delta})^{m_x}) \rightarrow 0$$

as  $n$  goes to infinity. Furthermore, on the event  $\tilde{\mathcal{A}}_n$ ,

$$k_x^{1/2} \omega \left( U_{1, m_x}^{(x)}, U_{m_x, m_x}^{(x)}, x, h \right) \leq k_x^{1/2} \omega(m_x^{-1-\delta}, 1 - m_x^{-1-\delta}, x, h) \rightarrow 0$$

which, since  $\mathbb{P}(\tilde{\mathcal{A}}_n) \rightarrow 1$ , concludes the proof of (39a).

Let us now show (39b). Changing the order of summation,

$$R_{n,1}(x) = \frac{1}{k_x - k_{x,a} + 1} \sum_{i=1}^{k_x-1} u_i(x) \frac{\beta_i^{(x)}}{i}, \quad \text{where } u_i(x) := \sum_{l=(k_{x,a}-1) \vee i}^{k_x-1} \frac{i}{l}.$$

Remarking that  $u_0(x) = 0$ , some straightforward computations lead to

$$|R_{n,1}(x)| \leq \frac{1}{k_x - k_{x,a} + 1} \sum_{j=1}^{k_x-1} |u_j(x) - u_{j-1}(x)| \left| \sum_{i=j}^{k_x-1} \frac{\beta_i^{(x)}}{i} \right|.$$

Let us first consider the case  $a = 0$  (i.e.  $k_x = k_{x,a}$ ). In this situation,  $u_j(x) - u_{j-1}(x) = 1/(k_x - 1)$  and thus by using (38),

$$|R_{n,1}(x)| \leq \frac{1}{k_x - 1} \sum_{j=1}^{k_x-1} \left| \sum_{i=j}^{k_x-1} \frac{\beta_i^{(x)}}{i} \right| = o_{\mathbb{P}} \left( \Delta \left( \frac{m_x}{k_x} \middle| x \right) \right) \times \frac{1}{k_x - 1} \sum_{j=1}^{k_x-1} \max \left( \log \left( \frac{k_x}{j} \right), 1 \right).$$

Since  $k_x \rightarrow \infty$ , using (29) with  $\varphi(t) = \max(\log(1/t), 1)$ , we get

$$\frac{1}{k_x - 1} \sum_{j=1}^{k_x-1} \max \left( \log \left( \frac{k_x}{j} \right), 1 \right) \rightarrow \int_0^1 \max(\log(1/t), 1) dt < \infty. \quad (40)$$

Using this convergence together with the fact that  $k_x^{1/2} \Delta(m_x/k_x | x)$  converges to a finite constant implies that  $k_x^{1/2} R_{n,1}(x) \xrightarrow{\mathbb{P}} 0$  in the case  $a = 0$ . Consider next the case  $a \in (0, 1)$ . It is easy to check that

$$|u_j(x) - u_{j-1}(x)| = \left| \sum_{l=(k_{x,a}-1) \vee j}^{k_x-1} \frac{1}{l} - \mathbb{I}\{j \geq k_{x,a}\} \right| \leq \sum_{l=k_{x,a}-1}^{k_x-1} \frac{1}{l} + 1.$$

Using together (29) (see the proof of Theorem 2) with  $\varphi(t) = 1/t$ , the fact that  $k_x \rightarrow \infty$  and the convergence  $k_{x,a}/k_x \rightarrow 1 - a$ , one has

$$\sum_{l=k_{x,a}-1}^{k_x-1} \frac{1}{l} + 1 \rightarrow 1 - \log(1 - a).$$

Hence, by using (38), (40) and the fact that  $(k_x - 1)/(k_x - k_{x,a} + 1) \rightarrow 1/a$ , we have for  $a \in (0, 1)$  and for  $n$  large enough,

$$|R_{n,1}(x)| \leq \frac{2(1 - \log(1 - a))}{k_x - k_{x,a} + 1} \sum_{j=1}^{k_x-1} \left| \sum_{i=j}^{k_x-1} \frac{\beta_i^{(x)}}{i} \right| = o_{\mathbb{P}} \left( \Delta \left( \frac{m_x}{k_x} \middle| x \right) \right).$$

In conclusion,  $k_x^{1/2} R_{n,1}(x) \xrightarrow{\mathbb{P}} 0$  for all  $a \in [0, 1)$ .

Finally, let us prove (39c). Using (29) with  $\varphi(t) = t^{-\rho(x)}$ , it is clear that uniformly in  $l \in \{k_{x,a}, \dots, k_x\}$ ,

$$\frac{1}{l-1} \sum_{i=1}^{l-1} \left( \frac{i}{l-1} \right)^{-\rho(x)} = \frac{1}{1 - \rho(x)} (1 + o(1)),$$

leading to

$$(1 - \rho(x)) R_{n,2}(x) = \frac{1}{k_x - k_{x,a} + 1} \sum_{l=k_{x,a}}^{k_x} \left( \frac{l-1}{k_x} \right)^{-\rho(x)} (1 + o(1)) - \mathcal{AB}(a, x).$$

In the case  $a = 0$ ,  $k_x = k_{x,a}$  and thus

$$(1 - \rho(x)) R_{n,2}(x) = 1 + o(1) - \mathcal{AB}(0, x) = o(1), \quad (41)$$

since  $\mathcal{AB}(0, x) = 1$ . Now, if  $a \in (0, 1)$ , using again (29) with  $\varphi(t) = t^{-\rho(x)}$  together with the fact that  $k_x/(k_x - k_{x,a} + 1) \rightarrow 1/a$  leads to

$$(1 - \rho(x))R_{n,2}(x) = \frac{1 - (1 - a)^{1-\rho(x)}}{a(1 - \rho(x))}(1 + o(1)) - \mathcal{AB}(a, x) = o(1). \quad (42)$$

In conclusion, (41) and (42) imply that  $R_{n,2}(x) \rightarrow 0$  for all  $a \in [0, 1)$  and the proof is complete. ■

Situation	Smoothed Hill estimator $\hat{\gamma}_a$	Estimator $\tilde{\gamma}_D$ of Daouia <i>et al.</i>	Estimator $\tilde{\gamma}_G$ of Goegebeur <i>et al.</i>	Estimator $\tilde{\gamma}_{G,BC}$ of Goegebeur <i>et al.</i>
$\gamma = \gamma_1$				
$\rho = -0.8$	0.0115	0.0116	0.0241	0.00490
$\rho = -1$	0.00753	0.00668	0.0157	0.00410
$\rho = -1.2$	0.00512	0.00416	0.00972	0.00359
$\gamma = \gamma_2$				
$\rho = -0.8$	0.0164	0.0189	0.0321	0.00715
$\rho = -1$	0.0102	0.0102	0.0198	0.00714
$\rho = -1.2$	0.00724	0.00679	0.0148	0.00606

Table 1: MSEs associated to the estimators in all cases.

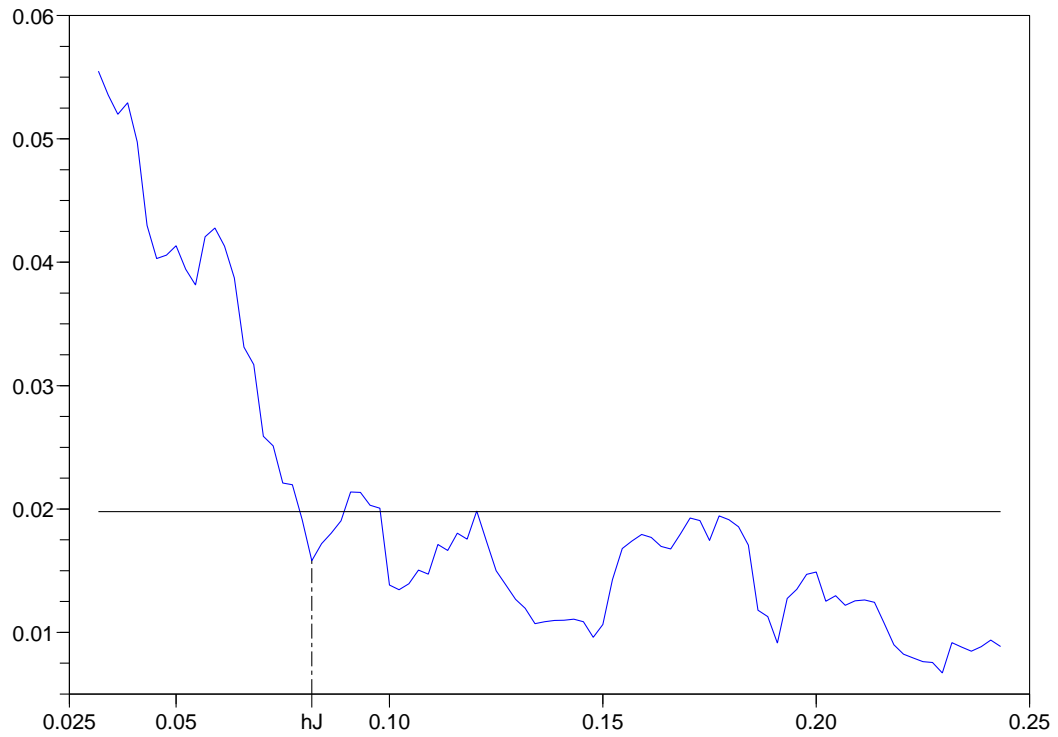
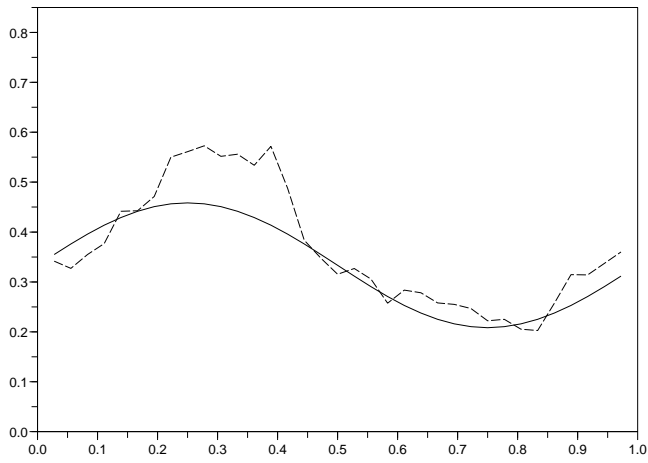
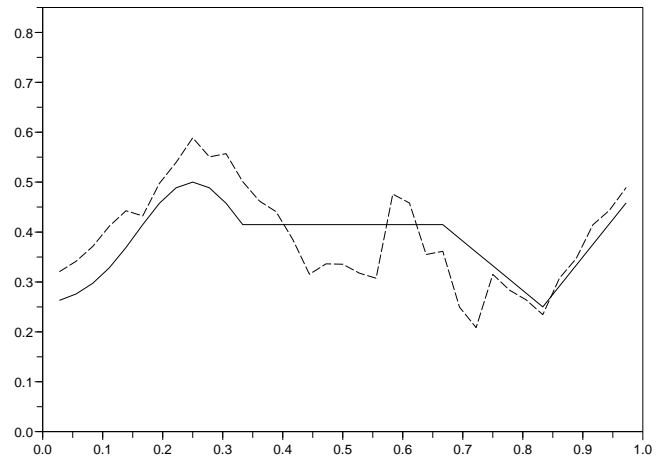


Figure 1: Choice of  $h_J$  on a given sample:  $x$ -axis: bandwidth  $h$ , blue line: moving average  $\bar{\sigma}$ , black line: mean of  $\bar{\sigma}$ .

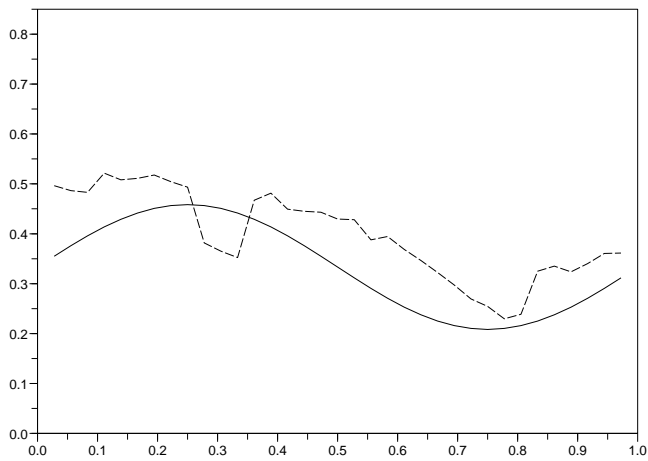


(a) Case  $\gamma = \gamma_1$

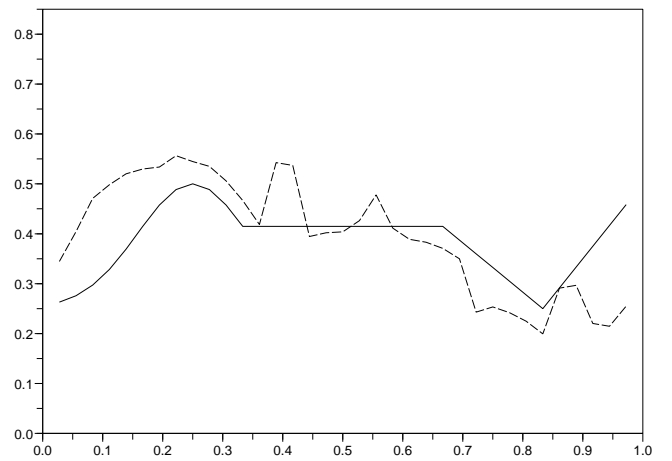


(b) Case  $\gamma = \gamma_2$

Figure 2: Case  $\rho = -1$ : the true function  $\gamma$  (solid line) and its smoothed Hill estimator  $\hat{\gamma}_a$  (dashed line), each corresponding to the 10% quantile of the MSE.

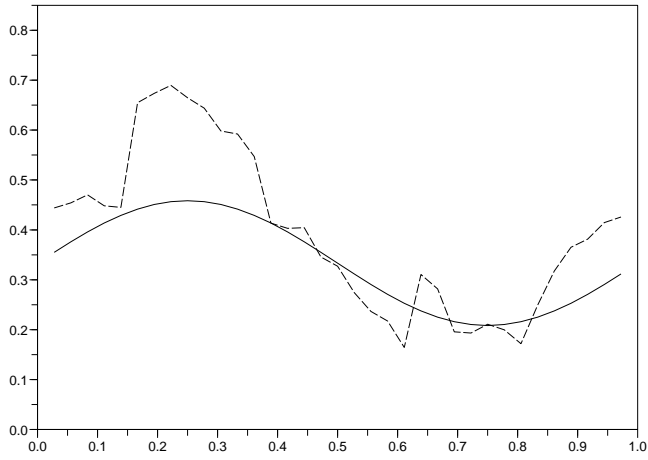


(a) Case  $\gamma = \gamma_1$

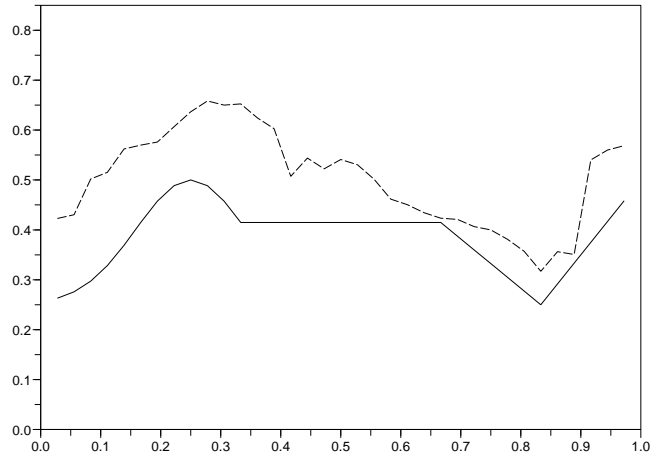


(b) Case  $\gamma = \gamma_2$

Figure 3: Case  $\rho = -1$ : the true function  $\gamma$  (solid line) and its smoothed Hill estimator  $\hat{\gamma}_a$  (dashed line), each corresponding to the median of the MSE.

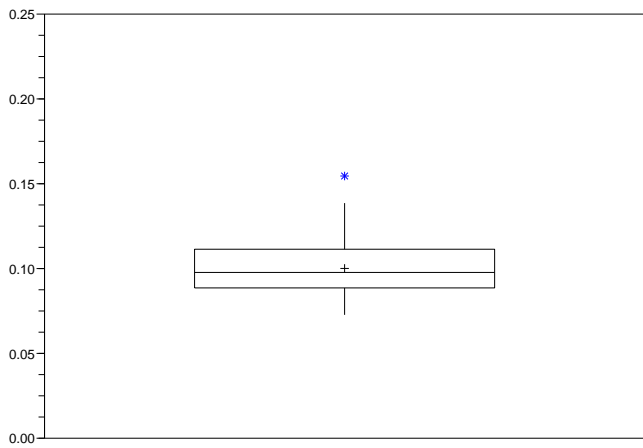


(a) Case  $\gamma = \gamma_1$

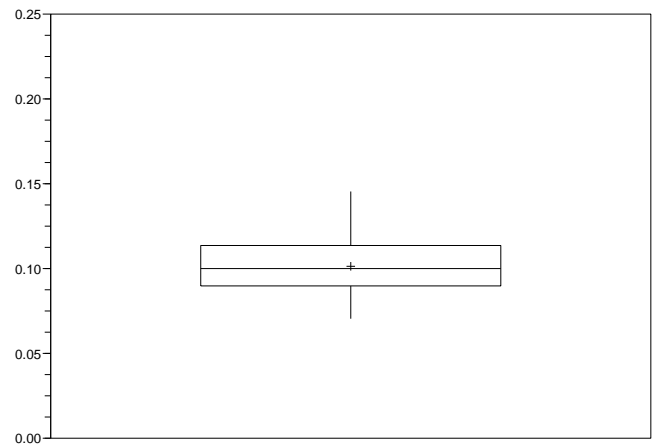


(b) Case  $\gamma = \gamma_2$

Figure 4: Case  $\rho = -1$ : the true function  $\gamma$  (solid line) and its smoothed Hill estimator  $\hat{\gamma}_a$  (dashed line), each corresponding to the 90% quantile of the MSE.

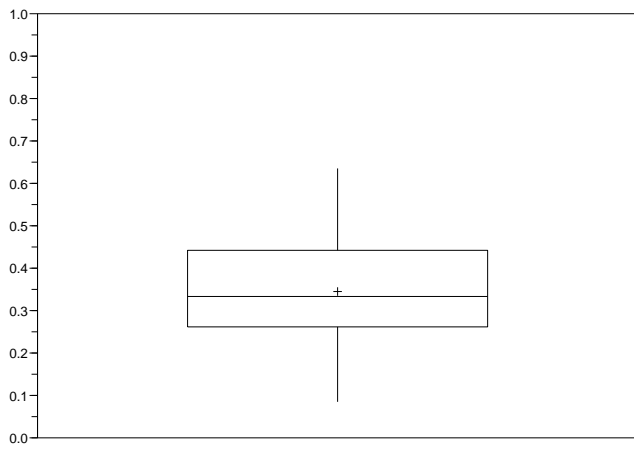


(a) Case  $\gamma = \gamma_1$

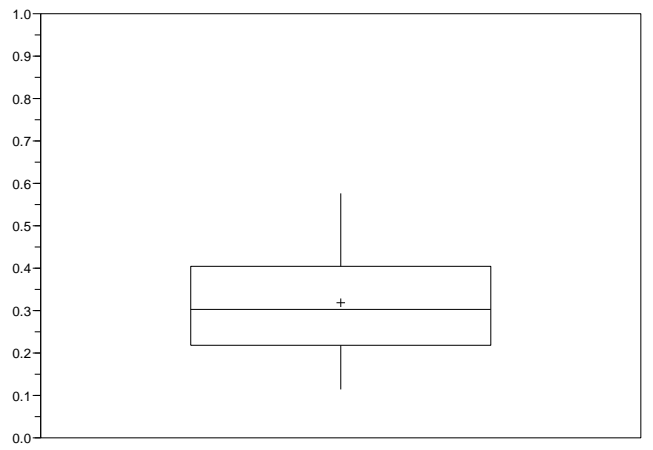


(b) Case  $\gamma = \gamma_2$

Figure 5: Case  $\rho = -1$ : boxplots of the bandwidths for the smoothed Hill estimator  $\hat{\gamma}_a$ .



(a) Case  $\gamma = \gamma_1$



(b) Case  $\gamma = \gamma_2$

Figure 6: Case  $\rho = -1$ : boxplots of the ratios  $k_x^*/M_n(x, h^*)$  for the smoothed Hill estimator  $\hat{\gamma}_a$ .