Transactions of A. Razmadze
Mathematical Institute
Vol. **175** (2021), issue 1, 13–32

EXTREME QUANTILE REGRESSION IN A PROPORTIONAL TAIL FRAMEWORK

BENJAMIN BOBBIA, CLÉMENT DOMBRY AND DAVIT VARRON

A paper devoted to the 75th birthday of Estate Khmaladze

Abstract. The model of heteroscedastic extremes initially introduced by Einmahl et al. (JRSSB, 2016) describes the evolution of a nonstationary sequence whose extremes evolve over time. We revisit this model and adapt it into a general extreme quantile regression framework. We provide estimates for the extreme value index and the integrated skedasis function and prove their joint asymptotic normality. Our results are quite similar to those developed for heteroscedastic extremes, but with a different proof approach emphasizing coupling arguments. We also propose a pointwise estimator of the skedasis function and a Weissman estimator of conditional extreme quantiles and prove the asymptotic normality of both estimators.

1. Introduction and Main Results

1.1. **Framework.** One of the main goals of the extreme value theory is to propose estimators of extreme quantiles: given an i.i.d. sample $Y_1, \dots, Y_n$ with distribution F, one wants to estimate the quantile of order $1-\alpha_n$ defined as $q(\alpha_n) := F^{\leftarrow}(1-\alpha_n)$, with $\alpha_n \to 0$ as $n \to \infty$ and

$$F^{\leftarrow}(u) := \inf\{x \in \mathbb{R} : F(x) \geq u\},\ u \in (0,1)$$

denotes the quantile function. The extreme regime corresponds to the case for $\alpha_n < 1/n$ in which case extrapolation beyond the sample maximum is needed. Considering an application in hydrology, these mathematical problems correspond to the following situation: given a record over $n = 50$ years of the level of a river, can we estimate the 100-year return level ? The answer to this question is provided by the univariate extreme value theory and we refer to the monographs by Coles [6], Beirlant et al. [2] or de Haan and Ferreira [8] for a general background.

In many situations, auxiliary information is available and represented by a covariate X taking values in $\mathbb{R}^d$ and, given $x \in \mathbb{R}^d$, one wants to estimate $q(\alpha_n|x)$, the conditional $(1-\alpha_n)$-quantile of Y with respect to some given values of the covariate $X = x$. This is an extreme quantile regression problem. Recent advances in extreme quantile regression include the works by Chernozhukov [5], El Methni et al. [13] or Daouia et al. [7].

In this paper we develop the proportional tail framework for extreme quantile regression. It is an adaptation of the heteroscedastic extremes developed by Einmahl et al. [12], where the authors propose a model for the extremes of independent, but nonstationary observations whose distribution evolves over time, a model which can be viewed as a regression framework with time as covariate and deterministic design with uniformly distributed observation times $1/n, 2/n, \dots, 1$. In our setting, the covariate X takes values in $\mathbb{R}^d$ and is random with arbitrary distribution. The main assumption, directly adapted from Einmmahl et al. [12], is the so-called proportional tail assumption formulated in Equation (1) and stating that the conditional tail function of Y for the given $X = x$ is asymptotically proportional to the unconditional tail. The proportionality factor is given by the so-called skedasis function $\sigma(x)$ that accounts for the dependency of the extremes of Y with respect to the covariate X. Furthermore, as it is standard in the extreme value theory, the unconditional distribution of Y is assumed to be regularly varying. Together with the proportional tail assumption, this implies that

2020 *Mathematics Subject Classification.* 62G32, 62G05.
Key words and phrases. Extreme value theory; Coupling methods; Empirical processes.
Due to technical reason the paper was not printed in the previous volume dedicated to Estate Khmaladze.

all the conditional distributions are regularly varying with the same extreme value index. Hence the proportional tail framework appears suitable for modeling covariate dependent extremes, where the extreme value index is constant, but the scale parameter depends on the covariate X in a nonparametric way related to the skedasis function $\sigma(x)$. Note that this framework is also considered by Gardes [14] for the purpose of estimation of the extreme value index.

Our main results are presented in the following subsections. Section 1.2 considers the estimation of the extreme value index and integrated skedasis function in the proportional tail model, and our results of asymptotic normality are similar to those in Einmahl et al. [9], but with a different proof emphasizing coupling arguments. Section 1.3 considers both the pointwise estimation of the skedasis function and the conditional extreme quantile estimation with Weissman estimators and states their asymptotic normality. Section 2 develops some coupling arguments used in the proofs of the main theorems, proofs gathered in Section 3. Finally, an appendix states a technical lemma and its proof.

1.2. **The proportional tail model.** Let (X, Y) be a generic random couple taking values in $\mathbb{R}^d \times \mathbb{R}$. Define the conditional cumulative distribution function of Y given $X = x$ by

$$F_x(y) := \mathbb{P}(Y \leq y | X = x), \quad y \in \mathbb{R}, \;\; x \in \mathbb{R}^d.$$

The main assumption of *the proportional tail model* is

$$\lim_{y\to\infty} \frac{1 - F_x(y)}{1 - F^0(y)} = \sigma(x) \quad \text{uniformly in } x \in \mathbb{R}^d, \tag{1}$$

where F^0 is some baseline distribution function and σ is the so-called skedasis function following the terminology introduced in [12]. By integration, the unconditional distribution F of Y satisfies

$$\lim_{y\to\infty} \frac{1 - F(y)}{1 - F^0(y)} = \int_{\mathbb{R}^d} \sigma(x) \mathbf{P}_X(dx).$$

We can hence suppose without loss of generality that $F = F^0$ and that $\int \sigma d\mathbf{P}_X = 1$.
We also make the assumption that F is of $1/\gamma$-regular variation,

$$1 - F(y) = y^{-1/\gamma} \ell(y), \quad y \in \mathbb{R},$$

with ℓ, slowly varying at infinity. Together with the proportional tail condition (1) with $F = F^0$, this implies that F_x is also of $1/\gamma$-regular variation for each $x \in \mathbb{R}^d$. This is a strong consequence of the model assumptions. In this model, the extremes are driven by two parameters: the common extreme value index $\gamma > 0$ and the skedasis function $\sigma(\cdot)$. Following [12], we consider the usual ratio estimator (see, e.g., [16, p. 198]) for γ and propose a nonparametric estimator of the integrated (or cumulative) skedasis function

$$C(x) := \int_{\{u \leq x\}} \sigma(u) \mathbf{P}_X(du), \quad x \in \mathbb{R}^d,$$

where $u \leq x$ stands for the componentwise comparison of vectors. Note that - putting aside the case, where X is discrete - the function C is easier to estimate than σ, in the same way that a cumulative distribution function is easier to estimate than a density function. Estimation of C is useful to derive tests, while estimation of σ will be considered later on for the purpose of extreme quantile estimation.

Let $(X_i, Y_i)_{1 \leq i \leq n}$ be i.i.d. copies of (X, Y). The estimators are built with observations (X_i, Y_i) for which Y_i exceeds a high threshold $\mathbf{y}_n$. Note that in this article, $(\mathbf{y}_n)_{n\in\mathbb{N}}$ may be deterministic or data driven. For the purpose of asymptotics, $\mathbf{y}_n$ depends on the sample size $n \geq 1$ in a way such that

$$\mathbf{y}_n \to \infty \quad \text{and} \quad N_n \to \infty \quad \text{in probability},$$

with $N_n := \sum_{i=1}^n \mathbb{1}_{\{Y_i > \mathbf{y}_n\}}$, the (possibly random) number of exceedances. The extreme value index $\gamma > 0$ is estimated by the ratio estimator

$$\hat{\gamma}_n := \frac{1}{N_n} \sum_{i=1}^n \log\left(\frac{Y_i}{\mathbf{y}_n}\right) \mathbb{1}_{\{Y_i > \mathbf{y}_n\}}.$$

The integrated skedasis function C can be estimated by the following empirical pseudo distribution function

$$\widehat{C}_n(x) := \frac{1}{N_n} \sum_{i=1}^{n} \mathbb{1}_{\{Y_i > \mathbf{y}_n,\ X_i \le x\}}, \quad x \in \mathbb{R}^d.$$

When Y is continuous and $\mathbf{y}_n := Y_{n-k_n:n}$ is the (k_n+1)-th highest order statistic, then $N_n = k$ and $\hat{\gamma}_n$ coincides with the usual Hill estimator.

Our first result addresses the joint asymptotic normality of $\hat{\gamma}_n$ and $\widehat{C}_n$, namely,

$$v_n \begin{pmatrix} \widehat{C}_n(\cdot) - C(\cdot) \\ \hat{\gamma}_n - \gamma \end{pmatrix} \xrightarrow{\mathscr{L}} \mathbb{W}, \tag{2}$$

where $\mathbb{W}$ is a Gaussian Borel probability measure on $L^\infty(\mathbb{R}^d) \times \mathbb{R}$, and $v_n \to \infty$ is a deterministic rate. To prove the asymptotic normality, the threshold $\mathbf{y}_n$ must scale suitably with respect to the rates of convergence in the proportional tail and domain of attraction conditions. More precisely, we assume the existence of a positive function A converging to zero and such that as $y \to \infty$,

$$\sup_{x \in \mathbb{R}^d} \left| \frac{\bar{F}_x(y)}{\sigma(x)\bar{F}(y)} - 1 \right| = \mathrm{O}\left(A\left(\frac{1}{\bar{F}(y)} \right) \right), \tag{3}$$

and

$$\sup_{z > \frac{1}{2}} \left| \frac{\bar{F}(zy)}{z^{-1/\gamma}\bar{F}(y)} - 1 \right| = \mathrm{O}\left(A\left(\frac{1}{\bar{F}(y)} \right) \right), \tag{4}$$

with $\bar{F}(y) := 1 - F(y)$ and $\bar{F}_x(y) := 1 - F_x(y)$. Our main result can then be stated as follows. When reading the present article, the reader probably notices that the domain $\{z > 1/2\}$ in (4) can be replaced by any domain $\{z > c\}$ for some $c \in]0,1[$.

Theorem 1.1. *Assume that assumptions* (3) *and* (4) *hold and* $\mathbf{y}_n/y_n \to 1$ *in probability for some deterministic sequence* y_n *such that* $p_n := \bar{F}(y_n)$ *satisfies*

$$p_n \to 0, \quad np_n \to \infty \quad \text{and} \quad \sqrt{np_n}^{1+\varepsilon} A(1/p_n) \to 0 \text{ for some } \varepsilon > 0.$$

Then the asymptotic normality (2) *holds with*

$$v_n := \sqrt{np_n} \quad \text{and} \quad \mathbb{W} \stackrel{\mathscr{L}}{=} \begin{pmatrix} B \\ N \end{pmatrix},$$

with B *a* C*-Brownian bridge on* $\mathbb{R}^d$ *and* N *a centered Gaussian random variable with variance* γ^2 *and independent of* B*.*

Under the C-Brownian bridge we here mean a centered Gaussian process on $\mathbb{R}^d$ with the covariance function

$$\operatorname{cov}(B(x), B(x')) := \int_{\mathbb{R}^d} \mathbb{1}_{]-\infty,x]} \mathbb{1}_{]-\infty,x']} dC - C(x)C(x').$$

Remark. Theorem 1.1 extends Theorem 2.1 of Einmhal *et al.* [12] in two directions: first, it states that their estimators and theoretical results have natural counterparts in the framework of proportional tails. We also could go past their univariate dependency $i/n \to \sigma(i/n)$ to a multivariate dependency $x \to \sigma(x),\ x \in \mathbb{R}^d$. Second, it shows that general data-driven thresholds can be used. Those extensions come at the price of a slightly more stringent condition upon the bias control. Indeed, their condition $\sqrt{k_n} A(n/k_n) \to 0$ corresponds to our condition $\sqrt{np_n}^{1+\varepsilon} A(1/p_n) \to 0$ with $\varepsilon = 0$. We believe that this loss is small in regard to the gain on the practical side: the threshold $\mathbf{y}_n$ in $(\hat{\gamma}_n, \hat{C}_n)$ may be data-driven. Take, for example, $\mathbf{y}_n := Y_{n-k_n:n}$, which is equivalent in probability to $y_n := F^{\leftarrow}(1 - k_n/n)$ is $k_n \to \infty$. As a consequence, Theorem 1.1 holds for this choice of $\mathbf{y}_n$ if

$$k_n \to \infty, \ \frac{k_n}{n} \to 0, \text{ and } \sqrt{k_n}^{1+\varepsilon} A\left(\frac{n}{k_n} \right) \to 0.$$

An example where (3) and (4) hold: The reader might wonder if a model imposing (3) and (4) is not too restrictive for modeling. First, note that condition (4) has been well studied as the second order condition holding uniformly over intervals (see, e.g., [8, p. 383, Section B.3], [1, 11]). A generic example of the regression model, where (3) and (4) hold, is given as follows: take a c.d.f H fulfilling the second order heavy tail condition (4) on any domain $\{z > c\}$. Then assume that the laws of $Y \mid X = x$ obey a location scale model in the sense that

$$F_x(y) = H\left(\frac{y - \mu(x)}{\Delta(x)}\right),$$

for some functions $\mu(\cdot)$ and $\Delta(\cdot)$ that are uniformly bounded on $\mathbb{R}^d$. Then, since $1 - \Delta(x)\mu(x)/y \to 1$ uniformly in x as $y \to \infty$, condition (4) entails

$$\sup_{x\in\mathbb{R}^d}\left|\frac{\overline{F}_x(y)}{\Delta(x)^{1/\gamma}\overline{H}(y)} - 1\right| = O(A(1/\overline{H}(y))), \quad \text{as} \quad y \to \infty.$$

Integrating in x gives $\overline{H}(y) = \theta\overline{F}(y)$ as $y \to \infty$ for some $\theta > 0$, which yields (3) with the choice of $\sigma(\cdot) := \theta\Delta(\cdot)^{1/\gamma}$.

1.3. **Extreme quantile regression.** In this subsection, we restrict ourselves to the case where $\mathbf{y}_n$ is deterministic i.e. $\mathbf{y}_n = y_n$ according to the notations of Theorem 1.1. We now address the estimation of extreme conditional quantiles in the proportional tail model, namely

$$q(\alpha_n|x) := F_x^{\leftarrow}(1 - \alpha_n),$$

for some $x \in \mathbb{R}^d$ that will be fixed once for all in this section, and for a sequence $\alpha_n = O(1/n)$. To that aim, we shall borrow the heuristics behind the Weissman estimator [19], for which we here write a short reminder. It is known that $F \in D(G_\gamma)$ is equivalent to

$$\lim_{t\to\infty}\frac{U(tz)}{U(t)} = z^\gamma, \quad \text{for each} \quad z > 0,$$

with $U(t) = F^{\leftarrow}(1 - 1/t)$, $t > 1$. Recall that $p_n = \bar{F}(y_n)$. Since U is of γ-regular variation, the unconditional quantile $q(\alpha_n) := F^{\leftarrow}(1 - \alpha_n)$ is approximated by

$$q(\alpha_n) = U(1/p_n)\frac{U(1/\alpha_n)}{U(1/p_n)} \approx y_n\left(\frac{p_n}{\alpha_n}\right)^\gamma,$$

leading to the Weissman-type quantile estimator

$$\hat{q}(\alpha_n) := y_n\left(\frac{\hat{p}_n}{\alpha_n}\right)^{\hat{\gamma}_n},$$

where $\hat{p}_n := \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{Y_i > y_n\}}$ is the empirical counterpart of p_n.

Now going back to quantile regression in the proportional tail model, it is readily verified that assumption (1) implies

$$q(\alpha_n \mid x) \sim q\left(\frac{\alpha_n}{\sigma(x)}\right) \quad \text{as } n \to \infty.$$

This immediately leads to the plug-in estimator

$$\hat{q}(\alpha_n|x) := \hat{q}\left(\frac{\alpha_n}{\hat{\sigma}_n(x)}\right) = y_n\left(\frac{\hat{p}_n\hat{\sigma}_n(x)}{\alpha_n}\right)^{\hat{\gamma}_n},$$

where $\hat{\sigma}_n(x)$ denotes a consistent estimator of $\sigma(x)$.

In the following, we propose a kernel estimator of $\sigma(x)$ and prove its asymptotic normality before deriving the asymptotic normality of the extreme conditional quantile estimator $\hat{q}(\alpha_n|x)$. The proportional tail assumption (1) implies

$$\sigma(x) = \lim_{n\to\infty}\frac{\overline{F}_x(y_n)}{\overline{F}(y_n)}.$$

We propose the simplest kernel estimator with bandwidth $h_n > 0$,

$$\frac{\sum_{i=1}^n \mathbb{1}_{\{|x-X_i|<h_n\}}\mathbb{1}_{\{Y_i>y_n\}}}{\sum_{i=1}^n \mathbb{1}_{\{|x-X_i|<h_n\}}}$$

as an estimator of $\overline{F}_x(y_n)$, while the denominator is estimated by $\hat{p}_n$. Combining the two estimators yields

$$\hat{\sigma}_n(x) := n\frac{\sum_{i=1}^n \mathbb{1}_{\{|x-X_i|<h_n\}}\mathbb{1}_{\{Y_i>y_n\}}}{\sum_{i=1}^n \mathbb{1}_{\{|x-X_i|<h_n\}}\sum_{i=1}^n \mathbb{1}_{\{Y_i>y_n\}}}.$$

Our next result states the asymptotic normality of $\hat{\sigma}_n(x)$. The more general case of a random threshold is left for future works.

Theorem 1.2. *Take the notations of Theorem* 1.1*, and let $h_n \to 0$ be deterministic and positive. Assume that*

$$np_nh_n^d \to \infty, \quad \sqrt{np_nh_n^d}A\left(1/p_n\right) \to 0.$$

Assume that the law of X is continuous on a neighborhood of x. Also assume that σ is continuous and positive on a neighborhood of $x \in \mathbb{R}^d$, and that some version f of the density of X also shares those properties. Then, under assumption (3)*, we have*

$$\sqrt{np_nh_n^d}\Big(\hat{\sigma}_n(x) - \sigma(x)\Big) \xrightarrow{\mathscr{L}} \mathcal{N}\left(0, \frac{\sigma(x)}{f(x)}\right).$$

The asymptotic normality of the extreme quantile estimate $\hat{q}(\alpha_n \mid x)$ is deduced from the asymptotic normality of $\hat{\gamma}_n$ and $\hat{\sigma}_n(x)$ stated respectively in Theorems 1.1 and 1.2. This is stated in our next theorem, which has to be seen as the counterpart of [8, p.138, Theorem 4.3.8] for conditional extreme quantiles. See also [16, p. 170, Theorem 9.8] for a similar result when $\log(p_n/\alpha_n) \to d \in \mathbb{R}$.

Theorem 1.3. *Under assumptions of Theorems* 1.1 *and* 1.2*, if $\sqrt{h_n^d}\log(p_n/\alpha_n) \to \infty$, we have*

$$\frac{\sqrt{np_n}}{\log(p_n/\alpha_n)}\log\left(\frac{\hat{q}(\alpha_n|x)}{q(\alpha_n|x)}\right) \xrightarrow{\mathscr{L}} \mathcal{N}\left(0, \gamma^2\right).$$

The condition $\sqrt{h_n^d}\log(p_n/\alpha_n)$ requires the bandwidth to be of larger order than $1/\log(p_n/\alpha_n)$, so the error in the estimation of $\sigma(x)$ is negligible. As a consequence of Theorem 1.3, the consistency

$$\frac{\hat{q}(\alpha_n|x)}{q(\alpha_n|x)} \xrightarrow{\mathbb{P}} 1.$$

That condition seems to state a limit for the extrapolation: α_n cannot be too small or one might lose consistency.

2. A Coupling Approach

We will first prove Theorem 1.1 when $\mathbf{y}_n$ is deterministic (i.e., $\mathbf{y}_n \equiv y_n$). In this case, N_n is binomial (n, p_n). Moreover, $N_n/np_n \to 1$ in probability, since $np_n \to \infty$.
A simple calculus shows that for each A Borel and $t \geq 1$, (1) entails

$$\mathbb{P}\left(\frac{Y}{\mathbf{y}} \geq t, X \in A \middle| Y \geq \mathbf{y}\right) \longrightarrow \int_t^\infty \int_A \mathbf{y}^{-1/\gamma}\sigma(x)\mathrm{d}\mathbf{y}\mathbf{P}_X(\mathrm{d}x), \text{ as } \mathbf{y} \to \infty, \tag{5}$$

defining a "limit model" for $(X, Y/\mathbf{y})$, the law

$$Q := \sigma(x)\mathbf{P}_X \otimes Pareto(1/\gamma)$$

with independent marginals. Fix $n \geq 1$. Using the heuristic of (5), we shall build an explicit coupling between $(X, Y/y_n)$ and the limit model Q. Define the conditional tail quantile function as $U_x(t) := F_x^{\leftarrow}(1 - 1/t)$ and recall that the total variation distance between two Borel probability measures on $\mathbb{R}^d$ is defined as

$$||P_1 - P_2|| := \sup_{B \text{ Borel}} |P_1(B) - P_2(B)|.$$

This distance is closely related to the notion of optimal coupling detailed in [15]. The following fundamental result is due to Dobrushin [10].

Lemma 2.1 (Dobrushin, 1970). *For two probability measures P_1 and P_2 defined on the same measurable space, there exist two random variables (V_1, V_2) on a probability set $(\Omega, \mathcal{A}, \mathbb{P})$ such that*

$$V_1 \sim P_1, \quad V_2 \sim P_2 \quad \text{and} \quad ||P_1 - P_2|| = \mathbb{P}(V_1 \neq V_2).$$

This lemma will be a crucial tool of our coupling construction, which is described as follows.

Coupling construction: Fix $n \geq 1$. Let $(E_{i,n})_{1\leq i\leq n}$ be i.i.d. Bernoulli random variables with $\mathbb{P}(E_{i,n} = 1) = \bar{F}(y_n)$ and $(Z_i)_{1\leq i\leq n}$ i.i.d. with distribution Pareto(1) and independent of $(E_{i,n})_{1\leq i\leq n}$. For each $1 \leq i \leq n$, construct $(\tilde{X}_{i,n}, \tilde{Y}_{i,n}, X^*_{i,n}, Y^*_{i,n})$ as follows.

- ▶ If $E_{i,n} = 1$, then
  - ▷ Take $\tilde{X}_{i,n} \sim P_{X|Y>y_n}$, $X^*_{i,n} \sim \sigma(x)\mathbf{P}_X(dx)$ on the same probability space, satisfying $\mathbb{P}(\tilde{X}_{i,n} \neq X^*_{i,n}) = \|\mathbf{P}_{X|Y>y_n} - \sigma(x)\mathbf{P}_X(dx)\|$. Their existence is guaranteed by Lemma 2.1.
  - ▷ Set $\tilde{Y}_{i,n} := U_{\tilde{X}_{i,n}}(\frac{Z_i}{\bar{F}_{\tilde{X}_{i,n}}(y_n)})$, $Y^*_{i,n} := y_n Z_i^{\gamma}$.
- ▶ If $E_{i,n} = 0$, then
  - ▷ Randomly generate $(\tilde{X}_{i,n}, \tilde{Y}_{i,n}) \sim \mathbf{P}_{(X,Y)|Y\leq y_n}$.
  - ▷ Randomly generate $(X^*_{i,n}, Y^*_{i,n}/y_n) \sim \sigma(x)\mathbf{P}_X(\mathrm{d}x) \otimes Pareto(1/\gamma)$.

The following proposition states the properties of our coupling construction, which will play an essential role in our proof of Theorem 1.1.

Proposition 2.2. *For each $n \geq 1$, the coupling $(\tilde{X}_{i,n}, \tilde{Y}_{i,n}, X^*_{i,n}, Y^*_{i,n})_{1\leq i\leq n}$ has the following properties:*

1. *$(\tilde{X}_{i,n}, \tilde{Y}_{i,n})_{1\leq i\leq n}$ has the same law as $(X_i, Y_i)_{1\leq i\leq n}$.*
2. *$(X^*_{i,n}, Y^*_{i,n}/y_n) \rightsquigarrow Q$.*
3. *$(X^*_{i,n}, Y^*_{i,n})_{1\leq i\leq n}$ and $(E_{i,n})_{1\leq i\leq n}$ are independent. Moreover, $(Y^*_{i,n})_{1\leq i\leq n}$ are i.i.d. and independent of $(\tilde{X}_{i,n}, X^*_{i,n})$.*
4. *There exists $M > 0$ such that*

$$\max_{\substack{1\leq i\leq n,\\ E_{i,n}=1}} \left| \frac{Y^*_{i,n}}{\tilde{Y}_{i,n}} - 1 \right| \leq MA\left(1/p_n\right) \tag{6}$$

and

$$\mathbb{P}\left(\tilde{X}_{1,n} \neq X^*_{1,n} | E_{i,n} = 1\right) \leq MA\left(1/p_n\right), \tag{7}$$

where A is given by assumptions (3) and (4).

Proof. To prove Point 1, it is sufficient to see that

$$\mathscr{L}((\tilde{X}_{1,n}, \tilde{Y}_{1,n})|E_{i,n} = 1) = \mathscr{L}((X,Y)|Y > y_n).$$

Since $U_x(z/(1 - F_x(y_n))) \leq y$ if and only if $1 - (1 - F_x(y_n))/z \leq F_x(y)$, then for $y \geq y_n$, we have

$$\begin{aligned}
&\int_1^\infty \mathbb{1}_{\{U_x(z/(1-F_x(y_n)))\leq y\}} \frac{\mathrm{d}z}{z^2} \\
&= \int_1^\infty \mathbb{1}_{\{1-(1-F_x(y_n))/z\leq F_x(y)\}} \frac{\mathrm{d}z}{z^2} \\
&= \int_{F_x(y_n)}^1 \mathbb{1}_{\{t\leq F_x(y)\}} \frac{\mathrm{d}t}{1 - F_x(y_n)}
\end{aligned}$$

$$= \int_{F_x(y_n)}^{F_x(y)} \frac{\mathrm{d}t}{1 - F_x(y_n)} = \frac{F_x(y) - F_x(y_n)}{1 - F_x(y_n)},$$

with the second equality given by the change of variable $t = 1 - (1 - F_x(y_n))/z$. We can deduce from this computation that for a Borel set B and $y \geq y_n$,

$$\begin{aligned}
&\mathbb{P}\Big(\tilde{X}_{1,n} \in B, U_{\tilde{X}_{1,n}}\Big(\frac{Z}{1 - F_{\tilde{X}_{1,n}}(y_n)}\Big) \leq y \Big| E_{1,n} = 1\Big) \\
=& \int_{x \in B} \int_1^{\infty} \mathbb{1}_{\{U_x(z/(1-F_x(y_n))) \leq y\}} \frac{\mathrm{d}z}{z^2} \mathrm{d}P_{X|Y>y_n}(x) \\
=& \int_{x \in B} \frac{F_x(y) - F_x(y_n)}{1 - F_x(y_n)} \mathrm{d}P_{X|Y>y_n}(x) \\
=& \int_{x \in B} \mathbb{P}\big(Y \leq y | Y > y_n, X = x\big) \mathrm{d}P_{X|Y>y_n}(x) \\
=& \mathbb{P}\big(X \in B, Y \leq y | Y > y_n\big).
\end{aligned}$$

□

This proves Point 1. Points 2 and 3 are immediate.

Point 4 will be proved with the two following lemmas.

Lemma 2.3. *Under conditions* (3) *and* (4)*, we have*

$$\sup_{z \geq 1/2} \sup_{x \in \mathbb{R}^p} \left| \frac{1}{z^\gamma y} U_x\left(\frac{z}{\bar{F}_x(y)}\right) - 1 \right| = \mathrm{O}\left(A\left(\frac{1}{\bar{F}(y)}\right)\right), \text{ as } y \to \infty.$$

Proof. According to assumptions (3) and (4), there exists a constant M such that

$$\begin{aligned}
\left| \frac{\bar{F}_x(y)}{\sigma(x)\bar{F}(y)} - 1 \right| &\leq MA\left(\frac{1}{\bar{F}(y)}\right), \text{ uniformly in } x \in \mathbb{R}^d, \text{ and} \\
\left| \frac{\bar{F}(zy)}{z^{-1/\gamma}\bar{F}(y)} - 1 \right| &\leq MA\left(\frac{1}{\bar{F}(y)}\right), \text{ uniformly in } z \geq 1/2.
\end{aligned} \tag{8}$$

From the definition of U_x, we have

$$\begin{aligned}
U_x\big(\tfrac{Z}{\bar{F}_x(y)}\big) &= F_x^{\leftarrow}\Big(1 - \tfrac{\bar{F}_x(y)}{z}\Big) \\
&= \inf\Big\{w \in \mathbb{R} : F_x(w) \geq 1 - \tfrac{\bar{F}_x(y)}{z}\Big\} \\
&= \inf\Big\{w \in \mathbb{R} : z\tfrac{\bar{F}_x(w)}{\bar{F}_x(y)} \leq 1\Big\}.
\end{aligned}$$

Hence for any $w^- < w^+$, one has

$$z\frac{\bar{F}_x(w^+)}{\bar{F}_x(y)} < 1 < z\frac{\bar{F}_x(w^-)}{\bar{F}_x(y)} \Rightarrow U_x\left(\frac{z}{\bar{F}_x(y)}\right) \in \left[w^-, w^+\right]. \tag{9}$$

Now write $\epsilon(y) := MA(1/\bar{F}(y))$ and choose $w^{\pm} := z^\gamma y\,(1 \pm 4\gamma\epsilon(y))$, so one can write

$$\begin{aligned} z\frac{\bar{F}_x(w^-)}{\bar{F}_x(y)} =& z\frac{\sigma(x)\bar{F}(\omega^-)(1-\epsilon(y))}{\sigma(x)\bar{F}(y)(1+\epsilon(y))} \\ \geq& z\frac{1-\epsilon(y)}{1+\epsilon(y)}\frac{1}{\bar{F}(y)}\bar{F}\left(z^\gamma y(1-4\gamma\epsilon(y))\right) \\ \geq& z\frac{1-\epsilon(y)}{1+\epsilon(y)}\frac{1}{\bar{F}(y)}\bar{F}(y)(1-\epsilon(y))\left(z^\gamma(1-4\gamma\epsilon(y))\right)^{-1/\gamma}, \text{ by (8)} \\ \geq& \frac{(1-\epsilon(y))^2}{1+\epsilon(y)}\left(1-4\gamma\epsilon(y)\right)^{-1/\gamma}. \end{aligned}$$

A similar computation gives

$$z\frac{\bar{F}_x(w^+)}{\bar{F}_x(y)} \leq \frac{(1+\epsilon(y))^2}{1-\epsilon(y)}\left(1+4\gamma\epsilon(y)\right)^{-1/\gamma}.$$

As a consequence, the condition before "⇒" in (9) holds if

$$4\gamma \geq \frac{1}{\epsilon(y)}\max\left\{1-\left(\frac{(1-\epsilon(y))^2}{1+\epsilon(y)}\right)^\gamma ; \left(\frac{(1+\epsilon(y))^2}{1-\epsilon(y)}\right)^\gamma - 1\right\}.$$

But a Taylor expansion of the right hand side shows that it is $3\gamma + o(1)$ as $y \to \infty$. This concludes the proof of Lemma 2.3. □

Applying Lemma 2.3 with $z := Z_i$ and $y := y_n$, we have

$$\max_{i:E_{i,n}=1}\left|\frac{Y^*_{i,n}}{\tilde{Y}_{i,n}} - 1\right| = \mathrm{O}\left(A\left(1/p_n\right)\right).$$

Now, by the construction of $(\tilde{X}_{1,n}, X^*_{1,n})$, when $E_{1,n} = 1$, we see that (7) is a consequence of the following

Lemma 2.4. *Under conditions* (3) *and* (4), *we have*

$$||P_{X|Y>y} - \sigma(x)\mathbf{P}_X(dx)|| = O\left(A\left(\frac{1}{\bar{F}(y)}\right)\right), \text{ as } y \to \infty.$$

Proof. For $B \in \mathbb{R}^d$, we have

$$\begin{aligned} &|P(X \in B|Y > y) - \int_B \sigma(x)\mathbf{P}_X(dx)| \\ =&\left|\frac{\int_B \bar{F}_x(y)P_X(dx)}{\bar{F}(y)} - \int_B \sigma(x)\mathbf{P}_X(dx)\right| \\ \leq& \int_B \left|\frac{\bar{F}_x(y)}{\bar{F}(y)} - \sigma(x)\right|\mathbf{P}_X(dx) \\ =& O\left(A\left(\frac{1}{\bar{F}(y)}\right)\right), \text{ by (3)}. \end{aligned}$$

□

This proves (7) and hence concludes the proof of Proposition 2.2.

3. Proofs

3.1. **Proof of Theorem 1.1. Change of notation:** Since for each n, the law of $(\tilde{X}_{i,n}, \tilde{Y}_{i,n})_{i=1,\dots,n}$ is $\mathbf{P}^{\otimes n}_{X,Y}$, we shall confound them with $(X_i, Y_i)_{i=1,\dots,n}$ to unburden notations.

3.1.1. *Proof when* $\mathbf{y}_n = y_n$ *is deterministic.* Fix $0 < \varepsilon < \frac{1}{2}$ and $0 < \beta < \varepsilon/(2\gamma)$. We consider the empirical process defined for every $x \in \mathbb{R}^d$ and $y \geq 1/2$ as

$$\mathbb{G}_n(x, y) := \sqrt{np_n}(\mathbb{F}_n(x, y) - \mathbb{F}(x, y)),$$

with

$$\mathbb{F}_n(x, y) := \frac{1}{N_n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}\mathbb{1}_{\{Y_i/y_n > y\}}E_{i,n},$$

and

$$\mathbb{F}(x, y) := C(x)V_\gamma(y) = Q\left(]-\infty, x]\times]y, +\infty[\right),$$

where $V_\gamma(y) := y^{-1/\gamma}$ for $y \geq 1$ and $V_\gamma(y) := 1$, otherwise.
Note that neither $\mathbb{F}$, nor any realisation of $\mathbb{F}_n$ is a cumulative distribution function in the strict sense, since they are decreasing in y. Their roles should, however, be seen as the same as for c.d.f. Now denote by $L^{\infty,\beta}(\mathbb{R}^d \times [1/2, \infty[)$ the (closed) subspace of $L^\infty(\mathbb{R}^d \times [1/2, \infty[)$ of all f satisfying

$$\|f\|_{\infty,\beta} := \sup_{x\in\mathbb{R}^d, y\geq 1/2} |y^\beta f(x, y)| < \infty,$$
$$f(\infty, y) := \lim_{\min\{x_1,\dots,x_d\}\to\infty} f(x, y) \text{ exists for each } y \geq 1,$$
$$\{y \mapsto f(\infty, y)\} \text{ is Càdlàg (see e.g., [4], p. 121).}$$

Simple arguments show that $\mathbb{G}_n$ takes values in $L^{\infty,\beta}(\mathbb{R}^d \times [1/2, \infty[)$.
First note that $\widehat{C}_n - C$ and $\hat{\gamma}_n - \gamma$ are images of $\mathbb{G}_n$ by the following map φ.

$$\varphi : L^{\infty,\beta}(\mathbb{R}^d \times [1/2, \infty[) \to L^\infty(\mathbb{R}^d) \times \mathbb{R}$$
$$f \mapsto \left(\{x \mapsto f(x, 1)\}, \int_1^\infty y^{-1} f(\infty, y)\mathrm{d}y\right),$$

and remark that φ is continuous, since $\beta > 0$. By the continuous mapping theorem, we hence see that Theorem 1.1 will be a consequence of

$$\mathbb{G}_n \overset{\mathscr{L}}{\to} \mathbf{W} \text{ in } L^{\infty,\beta}(\mathbb{R}^d \times [1/2, \infty[), \tag{10}$$

where $\mathbf{W}$ is the centered Gaussian process with a covariance function

$$\operatorname{cov}\big(\mathbf{W}(x_1, y_1), \mathbf{W}(x_2, y_2)\big) = C(x_1 \wedge x_2)V_\gamma(y_1) \wedge V_\gamma(y_2) - C(x_1)C(x_2)V_\gamma(y_1)V_\gamma(y_2),$$

and where $x_1 \wedge x_2$ is understood componentwise.

The proof is divided into two steps. In *step 1*, we prove (10) for the counterpart of $\mathbb{G}_n$, that is, we build on the Q sample $(X^*_{i,n}, Y^*_{i,n})_{1\leq i\leq n}$. Our proof relies on a standard argument from empirical processes. In *step 2*, we use the coupling properties of Proposition (2.2) to deduce (10) for the original sample $(X_i, Y_i)_{1\leq i\leq n}$.
Step 1: Define

$$\mathbb{F}^*_n(x, y) := \frac{1}{N_n}\sum_{i=1}^{n} \mathbb{1}_{\{X^*_i \leq x\}}\mathbb{1}_{\{Y^*_{i,n}/y_n > y\}}E_{i,n} \; x \in \mathbb{R}^d,\; y \geq 1/2.$$

The following proposition is a Donsker theorem in weighted topology for $\mathbb{G}^*_n := \sqrt{np_n}(\mathbb{F}^*_n - \mathbb{F})$.

Proposition 3.1. *If* (3) *and* (4) *hold, then*

$$\mathbb{G}^*_n \overset{\mathscr{L}}{\to} \mathbf{W}, \text{ in } L^{\infty,\beta}(\mathbb{R}^d \times [1/4, \infty[).$$

Proof. Write $\delta_x(A) = 1$ if $x \in A$ and 0, otherwise.
Since $(X^*_{i,n}, Y^*_{i,n})_{1\leq i\leq n}$ is independent of $(E_{i,n})_{1\leq i\leq n}$, Lemma 4.1 entails the following equality in laws

$$\sum_{i=1}^{n} \delta_{\left(X^*_{i,n}, \frac{Y^*_{i,n}}{y_n}\right)} E_{i,n} \overset{\mathscr{L}}{=} \sum_{i=1}^{\nu(n)} \delta_{\left(X^*_{i,n}, \frac{Y^*_{i,n}}{y_n}\right)},$$

where $\nu(n) \sim \mathcal{B}(n, p_n)$ is independent of $(X^*_{i,n}, Y^*_{i,n})_{1\le i\le n}$.
Since $(X^*_{i,n}, Y^*_{i,n}/y_n) \rightsquigarrow Q$ and since $\nu(n) \xrightarrow{\mathbb{P}} \infty$, $\nu(n)/np_n \xrightarrow{\mathbb{P}} 1$ and $\nu(n)$ independent of $(X^*_{i,n}, Y^*_{i,n})_{1\le i\le n}$, we see that $\mathbb{G}_n \xrightarrow{\mathscr{L}} \mathbf{W}$ will be a consequence of

$$\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{k} \mathbb{1}_{\{U_i\le .,V_i>.\}} - \mathbb{F}(.,.)\right) \xrightarrow{\mathscr{L}}_{k\to\infty} \mathbf{W} \text{ in } L^{\infty,\beta}\big(\mathbb{R}^d \times [1/4,\infty[\big),$$

where the (U_i, V_i) are i.i.d. with distribution Q. Now consider the following class of functions on $\mathbb{R}^d \times [1/4,\infty[$:

$$\mathcal{F}_\beta := \big\{f_{x,y} : (u,v) \mapsto y^\beta \mathbb{1}_{(-\infty,x]}(u)\mathbb{1}_{]y,\infty[}(v), x \in \mathbb{R}^d, y \ge 1/4\big\}.$$

Using the isometry

$$L^{\infty,\beta}(\mathbb{R}^d \times [1/4,\infty[\,) \to L^\infty(\mathcal{F}_\beta)$$
$$g \mapsto \{\Psi : f_{x,y} \mapsto g(x,y)\},$$

it is enough to prove that the abstract empirical process indexed by $\mathcal{F}_\beta$ converges weakly to the Q-Brownian bridge indexed by $\mathcal{F}_\beta$. In other words, we need to verify that $\mathcal{F}_\beta$ is Q-Donsker. This property can be deduced from two remarks:

(1) $\mathcal{F}_\beta$ is a VC-subgraph class of functions (see, e.g., Van der Vaart and Wellner [18], p.141). To see this, note that

$$\mathcal{F}_\beta \subset \big\{f_{x,s,z} : (u,v) \mapsto z\mathbb{1}_{(-\infty,x]}(u)\mathbb{1}_{]y,\infty[}(v), x \in \mathbb{R}^d, s \in [1/4,\infty[, z \in \mathbb{R}\big\}$$

which is a VC-subgraph class: the subgraph of each of its members is a hypercube of $\mathbb{R}^{d+2}$.

(2) $\mathcal{F}_\beta$ has a square integrable envelope F. This is proved by noting that for fixed $(u,v) \in \mathbb{R}^d \times [1/4,\infty[$.

$$F^2(u,v) = \sup_{x\in\mathbb{R}^d,\ y\ge 1/4} y^{2\beta}\mathbb{1}_{[0,x]}(u)\mathbb{1}_{]y,\infty[}(v) = v^{2\beta}$$

as a consequence F^2 is Q-integrable, since $\beta < (2\gamma)^{-1}$.

This concludes the proof of Proposition 3.1. □

Step 2: We show here that the two empirical processes $\mathbb{G}_n$ and $\mathbb{G}^*_n$ must have the same weak limit by proving the next proposition.

Proposition 3.2. *Under Assumptions* (3) *and* (4)*, we have*

$$\sup_{x\in\mathbb{R}^d,\ y\ge 1/2} y^\beta \sqrt{np_n}|\mathbb{F}^*_n(x,y) - \mathbb{F}_n(x,y)| = o_{\mathbb{P}}(1).$$

Proof. Adding and subtracting

$$\mathbb{F}^\sharp_n(x,y) := \frac{1}{N_n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i\le x\}}\mathbb{1}_{\{Y^*_{i,n}/y_n>y\}}E_{i,n}$$

in $|\mathbb{F}_n(x,y) - \mathbb{F}^*_n(x,y)|$, the triangle inequality entails, almost surely,

$$\begin{aligned}
&|\mathbb{F}_n(x,y) - \mathbb{F}^*_n(x,y)|\\
=&|\mathbb{F}_n(x,y) - \mathbb{F}^\sharp_n(x,y) + \mathbb{F}^\sharp_n(x,y) - \mathbb{F}^*_n(x,y)|\\
\le&\frac{1}{N_n}\sum_{i=1}^{n}|\mathbb{1}_{\{X_i\le x\}} - \mathbb{1}_{\{X^*_{i,n}\le x\}}|\mathbb{1}_{\left\{\frac{Y^*_{i,n}}{y_n}>y\right\}}E_{i,n}\\
&+ \frac{1}{N_n}\sum_{i=1}^{n}|\mathbb{1}_{\left\{\frac{Y_i}{y_n}>y\right\}} - \mathbb{1}_{\left\{\frac{Y^*_{i,n}}{y_n}>y\right\}}|\mathbb{1}_{\{X_i\le x\}}E_{i,n}\\
\le&\frac{1}{N_n}\sum_{i=1}^{n}\mathbb{1}_{\{X_i\ne X^*_{i,n}\}}\mathbb{1}_{\left\{\frac{Y^*_{i,n}}{y_n}>y\right\}}E_{i,n} + \frac{1}{N_n}\sum_{i=1}^{n}|\mathbb{1}_{\left\{\frac{Y_i}{y_n}>y\right\}} - \mathbb{1}_{\left\{\frac{Y^*_{i,n}}{y_n}>y\right\}}|E_{i,n}.
\end{aligned}$$

Let us first focus on the first term. Notice that

$$\begin{aligned}
&\sup_{x\in\mathbb{R}^d,\ y\geq 1/2} \frac{y^\beta\sqrt{np_n}}{N_n}\sum_{i=1}^n \mathbb{1}_{\{X_i\neq X^*_{i,n}\}}\mathbb{1}_{\left\{\frac{Y^*_{i,n}}{y_n}>y\right\}}E_{i,n}\\
=\ &\sup_{y\geq 1/2} \frac{y^\beta\sqrt{np_n}}{N_n}\sum_{i=1}^n \mathbb{1}_{\{X_i\neq X^*_{i,n}\}}\mathbb{1}_{\left\{\frac{Y^*_{i,n}}{y_n}>y\right\}}E_{i,n}\\
\leq\ &\sup_{y\geq 1/2} \frac{y^\beta\sqrt{np_n}}{N_n}\left(\max_{i=1,\dots,n}\mathbb{1}_{\left\{\frac{Y^*_{i,n}}{y_n}>y\right\}}E_{i,n}\right)\sum_{i=1}^n \mathbb{1}_{\{X_i\neq X^*_{i,n}\}}E_{i,n}.
\end{aligned}$$

Now notice that

$$\begin{aligned}
\sup_{y\geq 1/2}\max_{i=1,\dots,n} y^\beta \mathbb{1}_{\left\{\frac{Y^*_{i,n}}{y_n}>y\right\}}E_{i,n} &= \max_{i=1,\dots,n}\sup_{y\geq 1/2} y^\beta\mathbb{1}_{[1,Y^*_{i,n}/y_n]}(y)E_{i,n}\\
&= \max_{i=1,\dots,n}\left(\frac{Y^*_{i,n}}{y_n}\right)^\beta E_{i,n}.
\end{aligned}$$

By the independence between $E_{i,n}$ and $Y^*_{i,n}/y_n$, Lemma 4.1 in the Appendix gives

$$\max_{i=1,\dots,n}\left(\frac{Y^*_{i,n}}{y_n}\right)^\beta E_{i,n} \overset{\mathscr{L}}{=} \max_{i=1,\dots,\nu(n)}\left(\frac{Y^*_{i,n}}{y_n}\right)^\beta$$

where $Y^*_{i,n}/y_n$ in the right-hand side have a Pareto$(1/\gamma)$ distribution, whence

$$\max_{i=1,\dots,n}\left(\frac{Y^*_{i,n}}{y_n}\right)^\beta E_{i,n} = O_{\mathbb{P}}(\nu(n)^{\beta\gamma}) = O_{\mathbb{P}}((np_n)^{\beta\gamma}). \tag{11}$$

Moreover, writing $A_n := A(1/p_n)$, one has

$$\mathbb{E}\left(\sum_{i=1}^n \mathbb{1}_{\{X_i\neq X^*_{i,n}\}}E_{i,n}\right) = np_nA_n,$$

which entails

$$\frac{1}{np_nA_n}\sum_{i=1}^n \mathbb{1}_{\{X_i\neq X^*_{i,n}\}}E_{i,n} = O_{\mathbb{P}}(1). \tag{12}$$

As a consequence,

$$\begin{aligned}
&\frac{\sqrt{np_n}}{N_n}\max_{i=1,\dots,n}\left(\frac{Y^*_{i,n}}{y_n}\right)^\beta E_{i,n}\left(\sum_{i=1}^n \mathbb{1}_{\{X_i\neq X^*_{i,n}\}}E_{i,n}\right)\\
&= \frac{np_n}{N_n}\max_{i=1,\dots,n}\left(\frac{Y^*_{i,n}}{y_n}\right)^\beta E_{i,n}\left(\frac{1}{np_nA_n}\sum_{i=1}^n \mathbb{1}_{\{X_i\neq X^*_{i,n}\}}E_{i,n}\right)\sqrt{np_n}A_n\\
&= O_{\mathbb{P}}(1)O_{\mathbb{P}}((np_n)^{\beta\gamma})O_{\mathbb{P}}(1)\sqrt{np_n}A_n,\ \text{by (11) and (12)}\\
&= o_{\mathbb{P}}(1),\ \text{by the assumption of Theorem 1.1, and since } \beta\gamma<\frac{\varepsilon}{2}.
\end{aligned}$$

Let us now focus on the convergence

$$\sup_{x\in\mathbb{R}^d,\ y\geq 1/2} y^\beta\sqrt{np_n}\frac{1}{N_n}\sum_{i=1}^n\left|\mathbb{1}_{\left\{\frac{Y_i}{y_n}>y\right\}} - \mathbb{1}_{\left\{\frac{Y^*_{i,n}}{y_n}>y\right\}}\right|E_{i,n}\xrightarrow{\mathbb{P}} 0.$$

We deduce from Proposition 2.2 that, almost surely, writing $\epsilon_n := MA_n$:

$$(1-\epsilon_n)\frac{Y_i}{y_n}E_{i,n} \leq \frac{Y^*_{i,n}}{y_n}E_{i,n} \leq (1+\epsilon_n)\frac{Y_i}{y_n}E_{i,n},$$

which entails, almost surely, for all $y\geq 1$:

$$E_{i,n}\mathbb{1}_{\left\{\frac{Y^*_{i,n}}{y_n}\geq(1+\epsilon_n)y\right\}} \leq E_{i,n}\mathbb{1}_{\left\{\frac{Y_i}{y_n}\geq y\right\}} \leq E_{i,n}\mathbb{1}_{\left\{\frac{Y^*_{i,n}}{y_n}\geq(1-\epsilon_n)y\right\}},$$

implying

$$\left|\mathbb{1}_{\left\{\frac{Y_i}{y_n}>y\right\}}-\mathbb{1}_{\left\{\frac{Y^*_{i,n}}{y_n}>y\right\}}\right|E_{i,n}\leq\left|\mathbb{1}_{\left\{\frac{Y^*_{i,n}}{y_n}>(1-\epsilon_n)y\right\}}-\mathbb{1}_{\left\{\frac{Y^*_{i,n}}{y_n}>(1+\epsilon_n)y\right\}}\right|E_{i,n}.$$

This entails

$$\begin{aligned}&\sup_{x\in\mathbb{R}^d,\ y\geq 1/2} y^\beta\sqrt{np_n}\frac{1}{N_n}\sum_{i=1}^n\left|\mathbb{1}_{\left\{\frac{Y_i}{y_n}>y\right\}}-\mathbb{1}_{\left\{\frac{Y^*_{i,n}}{y_n}>y\right\}}\right|E_{i,n}\\ \leq&\sup_{x\in\mathbb{R}^d,\ y\geq 1/2} y^\beta\sqrt{np_n}\left|\mathbb{F}^*_n(\infty,(1-\epsilon_n)y)-\mathbb{F}^*_n(\infty,(1+\epsilon_n)y)\right|.\end{aligned}$$

Consequently, we have, adding and subtracting expectations:

$$\begin{aligned}&\sup_{x\in\mathbb{R}^d,\ y\geq 1/2} y^\beta\sqrt{np_n}\frac{1}{N_n}\sum_{i=1}^n\left|\mathbb{1}_{\{\frac{Y_i}{y_n}>y\}}-\mathbb{1}_{\{\frac{Y^*_{i,n}}{y_n}>y\}}\right|E_{i,n}\\ \leq&\sup_{x\in\mathbb{R}^d,\ y\geq 1/2} y^\beta\left|\tilde{\mathbb{G}}^*_n((1-\epsilon_n)y)-\tilde{\mathbb{G}}^*_n((1+\epsilon_n)y)\right| \qquad (13)\\ &+\sqrt{np_n}\sup_{y\geq 1/2} y^\beta(V_\gamma((1-\epsilon_n)y)-V_\gamma((1+\epsilon_n)y)), \qquad (14)\end{aligned}$$

where we write $\tilde{\mathbb{G}}^*_n(y):=\mathbb{G}^*_n(\infty,y)$.
We first prove that (14) converges to 0. For $y\geq 1$, we can bound

$$\begin{aligned}&y^\beta(V_\gamma((1-\epsilon_n)y)-V_\gamma((1+\epsilon_n)y))\\ \leq& y^\beta|1-((1+\epsilon_n)y)^{-1/\gamma}|\mathbb{1}_{\{(1-\epsilon_n)y<1\}}\\ &+y^\beta|((1-\epsilon_n)y)^{-1/\gamma}-((1+\epsilon_n)y)^{-1/\gamma}|\mathbb{1}_{\{(1-\epsilon_n)y\geq 1\}}. \qquad (15)\end{aligned}$$

In the first term of the right-hand side, since $(1-\epsilon_n)y<1$, we can write

$$\begin{aligned}&y^\beta|1-((1+A_n)y)^{-1/\gamma}|\mathbb{1}_{\{(1-\epsilon_n)y<1\}}\\ &\leq y^{\beta-1/\gamma}|y^{1/\gamma}-(1+A_n)^{-1/\gamma}|\mathbb{1}_{\{(1-\epsilon_n)y<1\}}\\ &\leq y^{\beta-1/\gamma}|(1-\epsilon_n)^{-1/\gamma}-(1+A_n)^{-1/\gamma}|\mathbb{1}_{\{(1-\epsilon_n)y<1\}}\\ &\leq 4\gamma^{-1}\epsilon_n,\ \text{since } \beta-1/\gamma<0.\end{aligned}$$

The second term of (15) is bounded by similar arguments, from where we have

$$\begin{aligned}\sqrt{np_n}&\sup_{x\in\mathbb{R}^d,\ y\geq 1/2} y^\beta|V_\gamma((1-\epsilon_n)y)-V_\gamma((1+\epsilon_n)y)|\\ &\leq 8\gamma^{-1}M\sqrt{np_n}A_n,\end{aligned}$$

which converges in probability to 0 by assumptions of Theorem 1.1.
We now prove that (13) converges to zero in probability. By Proposition 3.1, the continuous mapping theorem together with the Portmanteau theorem entail

$$\begin{aligned}\forall\varepsilon>0,\forall\rho>0,\ \overline{\lim}\ \mathbb{P}\Bigg(&\sup_{y\geq 1/2,\delta<\rho} y^\beta|\tilde{\mathbb{G}}^*_n((1-\delta)y)-\tilde{\mathbb{G}}^*_n((1+\delta)y)|\geq\varepsilon\Bigg)\\ &\leq\mathbb{P}\Bigg(\sup_{y\geq 1/2,\delta<\rho} y^\beta|\tilde{\mathbf{W}}((1-\delta)y)-\tilde{\mathbf{W}}((1+\delta)y)|\geq\varepsilon\Bigg),\end{aligned}$$

where $\tilde{\mathbf{W}}(y):=\mathbf{W}(\infty,y)$ is the centered Gaussian process with the covariance function

$$\mathrm{cov}(\tilde{\mathbf{W}}(y_1),\tilde{\mathbf{W}}(y_2)):=V_\gamma(y_1)\wedge V_\gamma(y_2)-V_\gamma(y_1)V_\gamma(y_2),\ (y_1,y_2)\in[1/4,\infty[^2.$$

With Proposition 3.1 together with the continuous mapping theorem, we see that the proof of Proposition 3.2 will be concluded if we establish the following lemma.

Lemma 3.3. *We have*

$$\sup_{y\geq 1/2,\delta<\rho} y^{\beta}|\tilde{\mathbf{W}}((1-\delta)y)-\tilde{\mathbf{W}}((1+\delta)y)| \xrightarrow[\rho\to 0]{\mathbb{P}} 0.$$

Proof. Let $\mathbb{B}_0$ be the standard Brownian bridge with $\mathbb{B}_0$ identically zero on $[1,\infty[$). $\tilde{\mathbf{W}}$ has the same law as $\{y\mapsto \mathbb{B}_0(y^{-1/\gamma})\}$ (see [17], p. 99), from where

$$\begin{aligned}\sup_{y\geq 1/2,\delta<\rho} & y^{\beta}|\tilde{\mathbf{W}}((1-\delta)y)-\tilde{\mathbf{W}}((1+\delta)y)| \\ \overset{\mathscr{L}}{=} & \sup_{y\geq 1/2,\delta<\rho} y^{\beta}\left|\mathbb{B}_0(((1-\delta)y)^{-1/\gamma})-\mathbb{B}_0(((1+\delta)y)^{-1/\gamma})\right| \\ \leq & \sup_{0\leq y\leq 2,\delta<\rho} y^{-\beta\gamma}\left|\mathbb{B}_0((1-\delta)^{-1/\gamma}y)-\mathbb{B}_0((1+\delta)^{-1/\gamma}y)\right|, \text{ almost surely.}\end{aligned}$$

Since $\beta\gamma<1/2$, the process $\mathbb{B}_0$ is a.s-$\beta\gamma$-Hölder continuous on $[0,+\infty[$. Consequently, for an a.s finite random variable H one has with probability one:

$$\begin{aligned}\sup_{0\leq y\leq 2,\delta<\rho} & y^{-\beta\gamma}|\mathbb{B}_0((1-\delta)^{-1/\gamma}y)-\mathbb{B}_0((1+\delta)^{-1/\gamma}y)| \\ \leq & \sup_{0\leq y\leq 2} y^{-\beta\gamma}|(1-\rho)^{-1/\gamma}-(1+\rho)^{-1/\gamma}|^{\beta\gamma}y^{\beta\gamma}H \\ = & |2(1-\rho)^{-1/\gamma}-2(1+\rho)^{-1/\gamma}|^{\beta\gamma}H \\ = & (4\tfrac{\rho}{\gamma})^{\beta\gamma}H.\end{aligned}$$

□

The preceding lemma concludes the proof of Proposition 3.2, which, combined with Proposition (3.1), proves (10). This concludes the proof of Theorem 1.1 when $\mathbf{y}_n\equiv y_n$. □

3.1.2. *Proof of Theorem* 1.1 *in the general case.* We now drop the assumption $\mathbf{y}_n\equiv y_n$ and relax it to $\frac{\mathbf{y}_n}{y_n}\xrightarrow{\mathbb{P}}1$ to achieve the proof of Theorem 1.1 in its full generality. We use the results of §3.1.1. Define

$$\overset{\vee}{\mathbb{F}}_n(x,y) := \frac{1}{\sum_{i=1}^{n}\mathbb{1}_{\{Y_i>y_n\}}}\sum_{i=1}^{n}\mathbb{1}_{\{X_i\leq x\}}\mathbb{1}_{\{Y_i/\mathbf{y}_n>y\}}$$

and

$$\overset{\vee}{\mathbb{G}}_n(x,y) := \sqrt{np_n}\left(\overset{\vee}{\mathbb{F}}_n(x,y)-\mathbb{F}(x,y)\right).$$

Now write $u_n := \frac{\mathbf{y}_n}{y_n}$. From §3.1.1, we know that

$$\left(\overset{\vee}{\mathbb{G}}_n,u_n\right)\xrightarrow{\mathscr{L}}(\mathbf{W},1) \text{ in } \mathbf{D}\times]0,+\infty[, \text{ where } \mathbf{D}:=L^{\infty,\beta}(\mathbb{R}^d\times[1/2,\infty[).$$

Moreover, as pointed out in Lemma 3.3, $\mathbf{W}$ almost surely belongs to

$$\mathbf{D}_0=\left\{\varphi\in L^{\infty,\beta}(\mathbb{R}^d\times[1/2,\infty[), \sup_{x\in\mathbb{R}^d,y,y'>1/2}\frac{|\varphi(x,y)-\varphi(x,y')|}{|y-y'|^{\beta\gamma}}<\infty\right\}.$$

Consider the followings maps $(g_n)_{n\in\mathbb{N}}$ and g from $\mathbf{D}$ to $L^{\infty,\beta}(\mathbb{R}^d\times[1,\infty[)$

$$g_n:(\varphi,u)\mapsto\sqrt{np_n}\left(\frac{\mathbb{F}(.,u.)+\frac{1}{\sqrt{np_n}}\varphi(.,u.)}{\mathbb{F}(\infty,u)+\frac{1}{\sqrt{np_n}}\varphi(\infty,u)}-\mathbb{F}(.,.)\right),$$

and

$$g:(\varphi,u)\mapsto u^{1/\gamma}\Big(\varphi(.,u.)-\varphi(\infty,u)\mathbb{F}(.,.)\Big).$$

Notice that $\mathbb{G}_n=g_n(\overset{\vee}{\mathbb{G}}_n,u_n)$ and $g(\mathbf{W},1)=\mathbf{W}$. The achievement of the proof of Theorem 1.1 hence boils down to making use of the extended continuous mapping theorem (see, e.g., Theorem 1.11.1, p. 67 in [18]) which is applicable to the sequence $(g_n,\mathbb{G}_n)$ provided that we establish the following

Lemma 3.4. *For any sequence φ_n of elements of $\mathbf{D}$ that converges to some $\varphi \in \mathbf{D}_0$, and for any sequence $u_n \to 1$ one has $g_n(\varphi_n, u_n) \to g(\varphi, 1)$ in $L^{\infty,\beta}(\mathbb{R}^d \times [1, \infty[)$. Here, the convergence in $L^{\infty,\beta}(\mathbb{R}^d \times [1, \infty[)$ is understood as with respect to $\|\cdot\|$, the restriction of $\|\cdot\|_{\infty,\beta}$ to $\mathbb{R}^d \times [1, \infty[$.*

Proof. For fixed $(x, y) \in \mathbb{R}^d \times [1/2, \infty[$ and $n \geq 1$, with the writing $t_n := (np_n)^{-1/2}$, we have

$$\begin{aligned}
&|g_n(\varphi_n, u_n)(x, y) - g(\varphi, 1)(x, y)| \\
&= \left| \frac{1}{t_n} \left(\frac{\mathbb{F}(x, u_n y) + t_n \varphi_n(x, u_n y)}{\mathbb{F}(\infty, u_n) + t_n \varphi_n(\infty, u_n)} - \mathbb{F}(x, y) \right) - \Big(\varphi(x, y) - \varphi(\infty, 1)\mathbb{F}(x, y) \Big) \right|.
\end{aligned}$$

Now, elementary algebra using $\mathbb{F}(x, yu_n)/\mathbb{F}(\infty, u_n) = \mathbb{F}(x, y)$ shows that

$$\begin{aligned}
&\frac{\mathbb{F}(x, u_n y) + t_n \varphi_n(x, u_n y)}{\mathbb{F}(\infty, u_n) + t_n \varphi_n(\infty, u_n)} - \mathbb{F}(x, y) \\
&= \mathbb{F}(x, y) \left(\frac{1 + t_n \frac{\varphi_n(x, u_n y)}{\mathbb{F}(x, u_n y)}}{1 + t_n \frac{\varphi_n(\infty, u_n)}{\mathbb{F}(\infty, u_n)}} - 1 \right) \\
&= \mathbb{F}(x, y) \left(\left(1 + t_n \frac{\varphi_n(x, u_n y)}{\mathbb{F}(x, u_n y)} \right) \left(1 - t_n u_n^{1/\gamma} \varphi_n(\infty, u_n)(1 + \epsilon_n) \right) (1 + \theta_n(x, y)) - 1 \right) \\
&= \mathbb{F}(x, y) \left(t_n \left(\frac{\varphi_n(x, u_n y)}{\mathbb{F}(x, u_n y)} - u_n^{1/\gamma} \varphi_n(\infty, u_n) \right) + R_n(x, y) \right),
\end{aligned}$$

with $\epsilon_n \to 0$ a sequence of real numbers, not depending on x and y, and with

$$R_n(x, y) := t_n u_n^{1/\gamma} \varphi_n(\infty, u_n)\epsilon_n + (t_n u_n^{1/\gamma})^2 \varphi_n(\infty, u_n) \frac{\varphi_n(x, u_n y)}{\mathbb{F}(x, y)} (1 + \epsilon_n).$$

This implies that

$$\|g_n(\varphi_n, u_n) - g(\varphi, 1)\| \leq B_{1,n} + B_{2,n} + B_{3,n} + B_{4,n},$$

where the four terms $B_{1,n}, \dots, B_{4,n}$ are detailed below and will be proved to converge to zero as $n \to \infty$.

First term

$$\begin{aligned}
B_{1,n} :=& \|u_n^{1/\gamma} \varphi_n(., u_n.) - \varphi(., .)\| \\
\leq& \|u_n^{1/\gamma} \varphi_n(., u_n.) - \varphi_n(., u_n.) + \|\varphi_n(., u_n.) - \varphi(., .)\| \\
=& |u_n^{1/\gamma} - 1| \|\varphi_n(., u_n.)\| + \|\varphi_n(., u_n.) - \varphi(., .)\| \\
\leq& |u_n^{1/\gamma} - 1| \|\varphi_n(., u_n.)\| + \|\varphi_n(., u_n.) - \varphi(., u_n.)\| \\
&+ \|\varphi(., u_n.) - \varphi(., .)\| \\
\leq& |u_n^{1/\gamma} - 1| \|\varphi_n(., u_n.)\| + u_n^{-\beta} ||\varphi_n(x, y) - \varphi(x, y)||_{\infty,\beta} \\
&+ H_\varphi |u_n - u|^{\beta\gamma},
\end{aligned}$$

where $H_\varphi := \sup\{|y - y'|^{-\beta\gamma} |\varphi(x, y) - \varphi(x, y')|, x \in \mathbb{R}^d, y, y' \geq 1/2\}$ is finite since $\varphi \in \mathbf{D}_0$. The first two terms converge to 0, since $u_n \to 1$ and $\varphi_n \to \varphi$ in $\mathbf{D}$. The third term converges to zero, since H_φ is finite.

Second term

$$\begin{aligned}
B_{2,n} :=& \|(u_n^{1/\gamma} \varphi_n(\infty, u_n) - \varphi(\infty, 1))\mathbb{F}\| \\
&\leq \left(|u_n^{1/\gamma} \varphi_n(\infty, u_n) - \varphi_n(\infty, u_n)| + |\varphi_n(\infty, u_n) - \varphi(\infty, 1)| \right) \|\mathbb{F}\|.
\end{aligned}$$

But $\|\mathbb{F}\|$ is finite since $\beta\gamma < \varepsilon < 1/2$, from where $B_{2,n} \to 0$ by similar arguments as those used for $B_{1,n}$.

Third term

$$B_{3,n} := \| u_n^{1/\gamma} \varphi_n(\infty, u_n) \epsilon_n \mathbb{F} \| \leq |u_n^{1/\gamma} \varphi_n(\infty, u_n)| \times |\epsilon_n| \times \|\mathbb{F}\|.$$

Since $\|\mathbb{F}\|$ is finite, since $|u_n^{1/\gamma}\varphi_n(\infty, u_n)|$ is a converging sequence, and since $|\epsilon_n| \to 0$, we deduce that $B_{3,n} \to 0$.

Fourth term

$$\begin{aligned} B_{4,n} := & \big(1 + |\epsilon_n|\big) \Big\| (t_n u_n^{1/\gamma})^2 \varphi_n(\infty, u_n) \varphi_n(., u_n .) \Big\| \\ \leq & \big(1 + |\epsilon_n|\big) \Big| (t_n u_n^{1/\gamma})^2 \varphi_n(\infty, u_n) \Big| \times \|\varphi_n(., u_n .)\|. \end{aligned}$$

Since $\varphi_n \to \varphi$ in $L^{\infty,\beta}(\mathbb{R}^d \times [1/2, \infty[)$, the same arguments as for $B_{3,n}$ entail the convergence to zero of $B_{4,n}$. □

3.2. **Proof of Theorem 1.2.** Let $x \in \mathbb{R}^d$, which will be kept fixed in all this section. To prove the asymptotic normality of $\hat{\sigma}_n(x)$, we first establish the asymptotic normality of the numerator and the denominator separately. Note that we don't need to study their joint asymptotic normality, because only the numerator will rule the asymptotic normality of $\hat{\sigma}_n(x)$, as its rate of convergence is the slowest.

Proposition 3.5. *Assume that $(p_n)_{n\geq 1}$ and $(h_n)_{n\geq 1}$ both converge to 0 and satisfy $np_n h_n^d \to 0$. We have*

$$\frac{1}{\sqrt{np_n h_n^d}} \sum_{i=1}^n \frac{\mathbb{1}_{\{|X_i - x| \leq h_n, Y_i > \mathbf{y}_n\}} - \mathbb{P}(|X_i - x| \leq h_n, Y_i > \mathbf{y}_n)}{\sqrt{\sigma(x) f(x)}} \xrightarrow{\mathscr{L}} \mathcal{N}(0,1), \tag{16}$$

$$\frac{1}{\sqrt{nh_n^d}} \sum_{i=1}^n \frac{\mathbb{1}_{\{|X_i - x| \leq h_n\}} - \mathbb{P}(|X_i - x| \leq h_n)}{\sqrt{f(x)}} \xrightarrow{\mathscr{L}} \mathcal{N}(0,1), \tag{17}$$

and

$$\frac{1}{\sqrt{np_n}} \sum_{i=1}^n (\mathbb{1}_{\{Y_i > \mathbf{y}_n\}} - p_n) \xrightarrow{\mathscr{L}} \mathcal{N}(0,1). \tag{18}$$

Proof. Note that (18) is the central limit theorem for binomial(n, p_n) sequences with $p_n \to 0$ and $np_n \to \infty$, while (17) is the well known pointwise asymptotic normality of the Parzen-Rosenblatt density estimator. The proof of (16) is a straghtforward use of the Lindeberg-Levy Theorem (see, e.g [3], Theorem 27.2 p. 359). First, we define

$$Z_{i,n} := \frac{\mathbb{1}_{\{|X_i - x| \leq h_n, Y_i > \mathbf{y}_n\}} - \mathbb{P}(|X_i - x| \leq h_n, Y_i > \mathbf{y}_n)}{\sqrt{np_n h_n^d}\sqrt{\sigma(x) f(x)}}$$

and remark that $\mathbb{E}(Z_{i,n}) = 0$. Moreover, we can write

$$\begin{aligned} \mathbb{E}\left(\mathbb{1}_{\{|X_i - x| \leq h_n, Y_i > \mathbf{y}_n\}}\right) = & \int_{B(x,h)} \mathbb{P}(Y_i > \mathbf{y}_n | X_i = z) P_X(dz) \\ \approx & \int_{B(x,h)} \sigma(z) p_n P_X(dz) && \text{(a)} \\ \approx & \sigma(x) f(x) p_n h_n^d, && \text{(b)} \end{aligned}$$

where (a) is a consequence of the uniformity in assumption (3), while equivalence (b) holds by our assumptions upon the regularity of both f and σ in Theorem 1.2. We conclude that $\sup\{|n\text{Var}(Z_{i,n}) - 1|,\ i = 1, \ldots, n\} \to 0$. Note that we can invoke the Lindeberg-Levy Theorem if for all $\varepsilon > 0$, we have

$$\sum_{i=1}^n \int_{\{Z_{i,n} > \varepsilon\}} Z_{i,n}^2 P_X(dx) \to 0.$$

This convergence holds since the set $\{Z_{i,n} > \varepsilon\}$ can be rewritten as

$$\Big\{|\mathbb{1}_{\{|X_i-x|\le h_n, Y_i>\mathbf{y}_n\}} - \mathbb{P}(|X_i - x| \le h_n, Y_i > \mathbf{y}_n)| \ge \varepsilon\sqrt{\sigma(x)f(x)}\sqrt{np_nh_n^d}\Big\},$$

which is empty when n is large enough, since $np_nh_n^d \to \infty$. This proves (16). □

Now, writing

$$\hat{\sigma}_n(x) = \frac{n}{\sum_{i=1}^n \mathbb{1}_{\{Y_i>\mathbf{y}_n\}}} \times \frac{\sum_{i=1}^n \mathbb{1}_{\{|x-X_i|<h_n\}}\mathbb{1}_{\{Y_i>\mathbf{y}_n\}}}{\sum_{i=1}^n \mathbb{1}_{\{|x-X_i|<h_n\}}},$$

we have

$$\hat{\sigma}_n(x) = \frac{1}{1 + \frac{1}{\sqrt{np_n}}\sum_{i=1}^n Z^\sharp_{i,n}} \times \frac{\frac{\mathbb{P}(|X-x|\le h_n, Y>\mathbf{y}_n)}{p_nh_n^d} + \sqrt{\frac{f(x)\sigma(x)}{np_nh_n^d}}\sum_{i=1}^n Z_{i,n}}{\frac{\mathbb{P}(|X-x|\le h_n)}{h_n^d} + \sqrt{\frac{f(x)}{nh_n^d}}\sum_{i=1}^n \tilde{Z}_{i,n}},$$

where

$$\tilde{Z}_{i,n} := \frac{\mathbb{1}_{\{|X_i-x|\le h_n\}} - \mathbb{P}(|X_i - x| \le h_n)}{\sqrt{f(x)}\sqrt{nh_n^d}},$$

and

$$Z^\sharp_{i,n} := \frac{\mathbb{1}_{\{Y_i>\mathbf{y}_n\}} - p_n}{\sqrt{np_n}}.$$

Now, we write

$$\sigma_{h_n}(x) := \frac{\mathbb{P}(|X - x| \le h_n, Y > \mathbf{y}_n)}{p_nh_n^d f(x)}.$$

Since f is continuous and bounded away from zero on a neighbourhood of x, we have

$$\hat{\sigma}_n(x) = \frac{1}{1 + \frac{1}{\sqrt{np_n}}\sum_{i=1}^n Z^\sharp_{i,n}} \frac{\sigma_{h_n}(x)f(x)(1 + \varepsilon_{n,1}) + \sqrt{\frac{f(x)\sigma(x)}{np_nh_n^d}}\sum_{i=1}^n Z_{i,n}}{f(x)(1 + \varepsilon_{n,2}) + \sqrt{\frac{f(x)}{nh_n^d}}\sum_{i=1}^n \tilde{Z}_{i,n}},$$

with $|\varepsilon_{n,1}| \vee |\varepsilon_{n,2}| \to 0$. Now a Taylor expansion of the denominator gives

$$\hat{\sigma}_n(x) = \frac{1}{1 + \frac{1}{\sqrt{np_n}}\sum_{i=1}^n Z^\sharp_{i,n}} \left(\sigma_{h_n}(x) + \sqrt{\frac{\sigma(x)}{np_nh_n^d f(x)}}\sum_{i=1}^n Z_{i,n}\right) \times \left(1 - \sqrt{\frac{1}{nh_n^d f(x)}}\sum_{i=1}^n \tilde{Z}_{i,n} + o_{\mathbb{P}}\left(\sqrt{\frac{1}{nh_n^d f(x)}}\right)\right).$$

By similar arguments, remarking that $(nh_n^d)^{-1} = o\Big((np_nh_n^d)^{-1}\Big)$, by (16) and (17), we have

$$\hat{\sigma}_n(x) = \frac{1}{1 + \frac{1}{\sqrt{np_n}}\sum_{i=1}^n Z^\sharp_{i,n}} \left(\sigma_{h_n}(x) + \sqrt{\frac{\sigma(x)}{np_nh_n^d f(x)}}\sum_{i=1}^n Z_{i,n} + o_{\mathbb{P}}\left(\frac{1}{\sqrt{np_nh_n^d}}\right)\right).$$

Moreover, with one more Taylor expansion of the denominator, by (18), we have

$$\hat{\sigma}_n(x) = \sigma_{h_n}(x) + \sqrt{\frac{\sigma(x)}{np_nh_n^d f(x)}}\sum_{i=1}^n Z_{i,n} + o_{\mathbb{P}}\left(\frac{1}{\sqrt{np_nh_n^d}}\right),$$

which entails

$$\sqrt{np_nh_n^d}(\hat{\sigma}_n(x) - \sigma_{h_n}(x)) = \sqrt{\frac{\sigma(x)}{f(x)}}\sum_{i=1}^n Z_{i,n} + o_{\mathbb{P}}(1).$$

The asymptotic normality of $\sum_{i=1}^{n} Z_{i,n}$ gives

$$\sqrt{np_n h_n^d}(\hat{\sigma}_n(x) - \sigma_{h_n}(x)) \xrightarrow{\mathscr{L}} \mathcal{N}\left(0, \frac{\sigma(x)}{f(x)}\right).$$

The proof is achieved by noticing that assumption (3) entails

$$\begin{aligned}
\sqrt{np_n h_n^d}|\sigma_{h_n}(x) - \sigma(x)| &= \sqrt{np_n h_n^d}\left|\frac{\mathbb{P}(|X-x| \le h_n, Y > \mathbf{y}_n)}{f(x)h_n^d \mathbb{P}(Y > \mathbf{y}_n)} - \sigma(x)\right| \\
&= \sqrt{np_n h_n^d}\left|\frac{\mathbb{P}(Y > \mathbf{y}_n | X \in B(x, h_n))}{\mathbb{P}(Y > \mathbf{y}_n)} - \sigma(x)\right| \\
&= O\left(\sqrt{np_n h_n^d} A\left(1/p_n\right)\right) \to 0.
\end{aligned}$$

3.3. **Proof of Theorem 1.3.** For the sake of clarity, we first express conditions (3) and (4) in terms of the tail quantile function U: we have, uniformly in x,

$$\left|\frac{U_x(1/\alpha_n)}{U(\sigma(x)/\alpha_n)} - 1\right| = O(A_n) \text{ and } \left|\frac{U(1/\alpha_n)}{xU(x^{-1/\gamma}/\alpha_n)} - 1\right| = O(A_n),$$

where $A_n := A(1/p_n)$. Start the proof by splitting the quantity of interest into four parts,

$$\begin{aligned}
\log\left(\frac{\hat{q}(\alpha_n|x)}{q(\alpha_n|x)}\right) &= \log\left(\frac{\mathbf{y}_n}{q(\alpha_n|x)}\left(\frac{\hat{p}_n\hat{\sigma}_n(x)}{\alpha_n}\right)^{\hat{\gamma}_n}\right) \\
&= \log\left(\frac{\mathbf{y}_n}{q(\alpha_n|x)}\left(\frac{p_n\hat{\sigma}_n(x)}{\alpha_n}\right)^{\hat{\gamma}_n}\left(\frac{\hat{p}_n}{p_n}\right)^{\hat{\gamma}_n}\right) \\
&= \log\left(\frac{\mathbf{y}_n}{q(\alpha_n|x)}\right) + \hat{\gamma}_n \log\left(\frac{p_n}{\alpha_n}\right) + \hat{\gamma}_n \log(\hat{\sigma}_n(x)) + \hat{\gamma}_n \log\left(\frac{\hat{p}_n}{p_n}\right) \\
&= \log\left(\frac{\mathbf{y}_n}{q(\alpha_n|x)}\left(\frac{p_n}{\alpha_n}\right)^{\gamma}\right) + (\hat{\gamma}_n - \gamma)\log\left(\frac{p_n}{\alpha_n}\right) \\
&\quad + \hat{\gamma}_n \log(\hat{\sigma}_n(x)) + \hat{\gamma}_n \log\left(\frac{\hat{p}_n}{p_n}\right).
\end{aligned}$$

Moreover, we can see that

$$\begin{aligned}
\log\left(\frac{\mathbf{y}_n}{q(\alpha_n|x)}\left(\frac{p_n}{\alpha_n}\right)^{\gamma}\right) &= \log\left(\frac{U(1/p_n)}{U_x(1/\alpha_n)}\left(\frac{p_n}{\alpha_n}\right)^{\gamma}\right) \\
&= \log\left(\frac{U(1/p_n)}{U(1/\alpha_n)}\left(\frac{p_n}{\alpha_n}\right)^{\gamma}\right) + \log\left(\frac{U(1/\alpha_n)}{U_x(1/\alpha_n)}\right)
\end{aligned}$$

Further, we write

$$\frac{\sqrt{np_n}}{\log(p_n/\alpha_n)}\log\left(\frac{\hat{q}(\alpha_n|x)}{q(\alpha_n|x)}\right) = Q_{1,n} + Q_{2,n} + Q_{3,n} + Q_{4,n}, \text{ with}$$

$$\begin{aligned}
Q_{1,n} &:= \frac{\sqrt{np_n}}{\log(p_n/\alpha_n)}\log\left(\frac{U(1/p_n)}{U(1/\alpha_n)}\left(\frac{p_n}{\alpha_n}\right)^{\gamma}\right), \\
Q_{2,n} &:= \sqrt{np_n}(\hat{\gamma}_n - \gamma), \\
Q_{3,n} &:= \frac{\sqrt{np_n}}{\log(p_n/\alpha_n)}\left(\hat{\gamma}_n \log(\hat{\sigma}_n(x)) + \log\left(\frac{U(1/\alpha_n)}{U_x(1/\alpha_n)}\right)\right), \\
Q_{4,n} &:= \frac{\sqrt{np_n}}{\log(p_n/\alpha_n)}\hat{\gamma}_n \log\left(\frac{\hat{p}_n}{p_n}\right).
\end{aligned}$$

First, condition (4) entails

$$\begin{aligned} Q_{1,n} \sim & \frac{\sqrt{np_n}}{\log(p_n/\alpha_n)} \left(\frac{U(1/p_n)}{U(1/\alpha_n)} \left(\frac{p_n}{\alpha_n} \right)^\gamma - 1 \right) \\ \sim & \frac{\sqrt{np_n}}{\log(p_n/\alpha_n)} \left(\frac{U((\alpha_n/p_n)^{\alpha\gamma}/\alpha_n)}{U(1/\alpha_n)} \left(\frac{p_n}{\alpha_n} \right)^\gamma - 1 \right) \\ = & \frac{\sqrt{np_n}}{\log(p_n/\alpha_n)} O(A_n). \end{aligned}$$

Since $\alpha_n = o(p_n)$, we see that $\log(p_n/\alpha_n)^{-1} \to 0$ together with $\sqrt{np_n} A_n \to 0$ entails that $Q_{1,n} \to 0$.
Second, we know by Theorem 1.1 that $Q_{2,n} \xrightarrow{\mathscr{L}} \mathcal{N}(0, \gamma^2)$.
Now $Q_{3,n}$ is studied remarking that

$$\log \left(\frac{U(1/\alpha_n)}{U_x(1/\alpha_n)} \right) = \log \left(\frac{U(\sigma(x)/\alpha_n)}{U_x(1/\alpha_n)} \right) + \log \left(\frac{U(1/\alpha_n)}{\sigma(x)^{-\gamma} U(\sigma(x)/\alpha_n)} \right) - \gamma \log(\sigma(x)).$$

Together with (3) and (4), one has

$$\log \left(\frac{U(1/\alpha_n)}{U_x(1/\alpha_n)} \right) = O(A_n) - \gamma \log(\sigma(x)).$$

Consequently,

$$Q_{3,n} = \frac{\sqrt{np_n}}{\log(p_n/\alpha_n)} O(A_n) + \frac{\sqrt{np_n}}{\log(p_n/\alpha_n)} \Big(\hat{\gamma}_n \log(\hat{\sigma}_n(x)) - \gamma \log(\sigma(x)) \Big).$$

Hence, the asymptotic behavior of $Q_{3,n}$ is ruled by that of $\hat{\gamma}_n \log(\hat{\sigma}_n(x)) - \gamma \log(\sigma(x))$, which we split into

$$(\hat{\gamma}_n - \gamma) \log(\hat{\sigma}_n(x)) + \gamma \log(\hat{\sigma}_n(x)) - \gamma \log(\sigma(x)).$$

Now, Theorem 1.1 entails

$$\frac{\log(\hat{\sigma}_n(x))}{\log(p_n/\alpha_n)} \sqrt{np_n} (\hat{\gamma}_n - \gamma) \xrightarrow{\mathbb{P}} 0.$$

Moreover, Theorem 1.2 together with the delta-method show that

$$\begin{aligned} & \frac{\sqrt{np_n}}{\log(p_n/\alpha_n)} (\gamma \log(\hat{\sigma}_n(x)) - \gamma \log(\sigma(x))) \\ & = \frac{\sqrt{np_n h_n^d}}{\sqrt{h_n^d} \log(p_n/\alpha_n)} (\gamma \log(\hat{\sigma}_n(x)) - \gamma \log(\sigma(x))) \xrightarrow{\mathbb{P}} 0. \end{aligned}$$

Finally, using the notation introduced in the proof of Theorem 1.2, we have

$$\begin{aligned} \frac{\sqrt{np_n}}{\log(p_n/\alpha_n)} \log \left(\frac{\hat{p}_n}{p_n} \right) = & \frac{\sqrt{np_n}}{\log(p_n/\alpha_n)} \log \left(1 + \frac{1}{\sqrt{np_n}} \sum_{i=1}^n Z_{i,n}^\sharp \right) \\ \sim & \frac{1}{\log(p_n/\alpha_n)} \sum_{i=1}^n Z_{i,n}^\sharp + o_{\mathbb{P}} \left(\frac{1}{\log(p_n/\alpha_n)} \right) \\ & \xrightarrow{\mathbb{P}} 0, \end{aligned}$$

which proves that $Q_{4,n} \xrightarrow{\mathbb{P}} 0$, since $\hat{\gamma}_n \xrightarrow{\mathbb{P}} \gamma$.

4. Appendix

Lemma 4.1. *For fixed $n \geq 1$, let $(Y_i)_{1 \leq i \leq n}$ be a sequence of i.i.d. random variables taking values in $(\mathfrak{X}, \mathcal{X})$. Let $E = (E_i)_{1 \leq i \leq n}$ be an n-uple of independent Bernoulli random variables independent of Y_i. Write*

$$\nu(k) := \sum_{i=1}^k E_i, \ k \leq n.$$

Then we have

$$\sum_{i=1}^{n} \delta_{Y_i} E_i \overset{\mathscr{L}}{=} \sum_{i=1}^{\nu(n)} \delta_{Y_i}, \tag{19}$$

where the equality in law is understood as on the sigma algebra spanned by all Borel positive functions on $(\mathfrak{X}, \mathcal{X})$*. Moreover, if the* (Y_i) *are almost surely positive, then*

$$\max_{i=1,\ldots,n} Y_i E_i \overset{\mathscr{L}}{=} \max_{i=1,\ldots,\nu(n)} Y_i. \tag{20}$$

Proof. Note that (19) is exactly Khinchin's equality (see [16, p. 307, (14.6)]). We shall now prove (20). $e \in \{0,1\}^n$, and let g be a real measurable and positive function. Since the variables $(Y_i)_{1\le i\le n}$ are i.i.d. and independent of E, for any given permutation σ of $[\![1,n]\!]$,

$$wehave (Y_1, \ldots, Y_n) \overset{\mathscr{L}}{=} (Y_1, \ldots, Y_n)|_{E=e} \overset{\mathscr{L}}{=} (Y_{\sigma(1)}, \ldots, Y_{\sigma(n)})|_{E=e}$$

by exchangeability. Now, define σ by

$$\sigma(k) := \begin{cases} \sum_{j=1}^{i} e_j \text{ if } e_i = 1 \\ n - \sum_{j=1}^{i} (1-e_j) \text{ if } e_i = 0 \end{cases} \qquad 1 \le i \le n.$$

Write $s(e) := \sum_{i=1}^{n} s(e_i)$ for the total number of ones in $(e_1, \ldots, e_n)$. By construction, the indices i for which $e_i = 1$ are mapped injectively to the set of first indices $[\![1, s(e)]\!]$, while those for which $e_i = 0$ are injectively mapped into $[\![s(e)+1, n]\!]$. Since e has fixed and nonrandom coordinates, we have

$$(Y_1 e_1, \ldots, Y_n e_n)|_{E=e} \overset{\mathscr{L}}{=} (Y_{\sigma(1)} e_1, \ldots, Y_{\sigma(n)} e_n)|_{E=e}.$$

Hence

$$\begin{aligned} \max_{i=1,\ldots,n} Y_i e_i \;|_{E=e} &\overset{\mathscr{L}}{=} \max_{i=1,\ldots,n} Y_i e_i \\ &\overset{\mathscr{L}}{=} \max_{i=1,\ldots,n} Y_{\sigma(i)} e_{\sigma(i)} \\ &\overset{\mathscr{L}}{=} \max_{i=1,\ldots,s(e)} Y_{\sigma(i)} && \text{(a)} \\ &\overset{\mathscr{L}}{=} \max_{i=1,\ldots,s(e)} Y_i && \text{(b)} \\ &\overset{\mathscr{L}}{=} \max_{i=1,\ldots,s(e)} Y_i \;\;|_{E=e} \\ &\overset{\mathscr{L}}{=} \max_{i=1,\ldots,s(E)} Y_i \;\;|_{E=e}, \end{aligned}$$

where (a) holds because $e_{\sigma(i)} = 0$ for $i > s(e)$ by construction and the Y_i are a.s. positive, while (b) is obtained by noticing that $F_e(Y_{\sigma(1)}, \ldots, Y_{\sigma(n)}) \overset{\mathscr{L}}{=} F_e(Y_1, \ldots, Y_n)$ with

$$F_e : (y_1, \ldots, y_n) \mapsto \max_{i=1,\ldots,s(e)} y_i.$$

Unconditioning upon E gives (20). □

References

1. M. F. Alves, M. I. Gomes, L. de Haan, C. Neves, A note on second order conditions in extreme value theory: linking general and heavy tail conditions. *REVSTAT* **5** (2007), no. 3, 285–304.
2. J. Beirlant, Y. Goegebeur, J. Teugels, J. Segers, *Statistics of Extremes.* Theory and applications. With contributions from Daniel De Waal and Chris Ferro. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 2004.
3. P. Billingsley, *Probability and Measure.* Third edition. Wiley Series in Probability and Mathematical Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1995.

4. P. Billingsley, *Convergence of Probability Measures*. Second edition. Wiley Series in Probability and Statistics: Probability and Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1999.
5. V. Chernozhukov, Extremal quantile regression. *Ann. Statist.* **33** (2005), no. 2, 806–839.
6. S. Coles, *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. Springer-Verlag London, Ltd., London, 2001.
7. A. Daouia, L. Gardes, S. Girard, On kernel smoothing for extremal quantile regression. *Bernoulli* **19** (2013), no. 5B, 2557–2589.
8. L. de Haan, A. Ferreira, *Extreme Value Theory*. An introduction. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006.
9. A. L. M. Dekkers, J. H. J. Einmahl, L. de Haan, A moment estimator for the index of an extreme-value distribution. *Ann. Statist.* **17** (1989), no. 4, 1833–1855.
10. R. L. Dobrushin, Prescribing a system of random variables by conditional distributions. *Theory of Probability & Its Applications* **15** (1970), no. 3, 458–486.
11. H. Drees, L. de Haan, D. Li, Approximations to the tail empirical distribution function with application to testing extreme value conditions. *J. Statist. Plann. Inference* **136** (2006), no. 10, 3498–3538.
12. J. H. J. Einmahl, L. de Haan, C. Zhou, Statistics of heteroscedastic extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** (2016), no. 1, 31–51.
13. J. El Methni, L. Gardes, S. Girard, A. Guillou, Estimation of extreme quantiles from heavy and light tailed distributions. *J. Statist. Plann. Inference* **142** (2012), no. 10, 2735–2747.
14. L. Gardes, A general estimator for the extreme value index: applications to conditional and heteroscedastic extremes. *Extremes* **18** (2015), no. 3, 479–510.
15. T. Lindvall, *Lectures on the Coupling Method*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1992.
16. S. Y. Novak, *Extreme Value Methods with Applications to Finance*. CRC Press, 2011.
17. G. R. Shorack, J. A. Wellner, *Empirical Processes with Applications to Statistics*. Reprint of the 1986 original [MR0838963]. Classics in Applied Mathematics, 59. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2009.
18. A. W. van der Vaart, J. A. Wellner, *Weak Convergence and Empirical Processes*. With applications to statistics. Springer Series in Statistics. Springer-Verlag, New York, 1996.
19. I. Weissman, Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Assoc.* **73** (1978), no. 364, 812–815.

(Received 21.11.2019)

Université de Bourgogne-Franche-Comté, Laboratoire de Mathématiques de Besançon, UMR 6623
E-mail address: benjamin.bobbia@univ-fcomte.fr
E-mail address: clement.dombry@univ-fcomte.fr
E-mail address: davit.varron@univ-fcomte.fr