The Annals of Statistics
2018, Vol. 46, No. 6A, 2806–2843
https://doi.org/10.1214/17-AOS1640

LOCAL ROBUST ESTIMATION OF THE PICKANDS DEPENDENCE FUNCTION[1]

BY MIKAEL ESCOBAR-BACH*, YURI GOEGEBEUR* AND ARMELLE GUILLOU†

University of Southern Denmark and Université de Strasbourg et CNRS†*

We consider the robust estimation of the Pickands dependence function in the random covariate framework. Our estimator is based on local estimation with the minimum density power divergence criterion. We provide the main asymptotic properties, in particular the convergence of the stochastic process, correctly normalized, towards a tight centered Gaussian process. The finite sample performance of our estimator is evaluated with a simulation study involving both uncontaminated and contaminated samples. The method is illustrated on a dataset of air pollution measurements.

1. Introduction. Modelling dependence among extremes is of primary importance in practical applications where extreme phenomena occur. To this aim, the copula function can be used as a margin-free description of the dependence structure. Indeed, according to the well-known result of Sklar (1959), the distribution function of a pair $(Y^{(1)}, Y^{(2)})$ can be represented in terms of the two margins F_1 and F_2 of $Y^{(1)}$ and $Y^{(2)}$, respectively, and a copula function C as follows:

$$\mathbb{P}\big(Y^{(1)} \le y_1, Y^{(2)} \le y_2\big) = C\big(F_1(y_1), F_2(y_2)\big).$$

This function C characterizes the dependence between $Y^{(1)}$ and $Y^{(2)}$ and is called an extreme value copula if and only if it admits a representation of the form

$$C(y_1, y_2) = \exp\left(\log(y_1 y_2) A\left(\frac{\log(y_2)}{\log(y_1 y_2)}\right)\right),$$

where A: $[0, 1] \to [1/2, 1]$ is the Pickands dependence function, which is convex and satisfies $\max\{t, 1-t\} \le A(t) \le 1$; see Pickands (1981). Statistical inference on the bivariate function C is therefore equivalent to the statistical inference on the one-dimensional function A. Estimating this function A has been extensively studied in the literature. We can mention, among others, Capéraà, Fougères and Genest

Received June 2016; revised May 2017.

[1]Supported by a research grant (VKR023480) from VILLUM FONDEN and an international project for scientific cooperation (PICS-6416). Computation/simulation for the work described in this paper was supported by the DeIC National HPC Centre, SDU.

MSC2010 subject classifications. Primary 62G32, 62G05, 62G20; secondary 60F05, 60G70.

Key words and phrases. Conditional Pickands dependence function, robustness, stochastic convergence.

(1997), Fils-Villetard, Guillou and Segers (2008) or Bücher, Dette and Volgushev (2011).

In this paper, we extend the above framework to the case where the pair $(Y^{(1)}, Y^{(2)})$ is recorded along with a random covariate $X \in \mathbb{R}^p$. In that context, the copula function together with the marginal distribution functions depend on the covariate X. In the sequel, we denote by C_x, $F_1(\cdot|x)$ and $F_2(\cdot|x)$ the conditional copula function and the continuous conditional distribution functions of $Y^{(1)}$ and $Y^{(2)}$ given $X = x$. Our model can thus be written as

$$\mathbb{P}\big(F_1\big(Y^{(1)}|x\big) \leq y_1, F_2\big(Y^{(2)}|x\big) \leq y_2 | X = x\big) = C_x(y_1, y_2), \tag{1.1}$$

where C_x admits a representation of the form

$$C_x(y_1, y_2) = \exp\left(\log(y_1 y_2) A\left(\frac{\log(y_2)}{\log(y_1 y_2)}\Big| x\right)\right),$$

with $A(\cdot|\cdot) : [0, 1] \times \mathbb{R}^p \to [1/2, 1]$ is the conditional Pickands dependence function which is again a convex function satisfying $\max\{t, 1 - t\} \leq A(t|x) \leq 1$ for all $x \in \mathbb{R}^p$. From a practical point of view, the considered family of extreme value distributions has sufficiently large potential for data analysis. First, the family of extreme value distributions is very rich, and includes among others the logistic, the asymmetric logistic, the negative logistic, the Hüsler–Reiss, the t extreme value and Dirichlet model. Second, multivariate extreme value distributions arise naturally as the limiting distributions of properly normalised component-wise maxima, making them a useful approximation to the true, but typically unknown, distribution of these component-wise maxima in practice. We refer to Kotz and Nadarajah (2000) and Gudendorf and Segers (2010) for further motivation and discussion of this class of distributions and additional examples. As a possible application, we consider modelling extremal dependence between air pollutants, like ground-level ozone and particulate matter, conditional on location and time; see Section 5 for more details.

Moreover, in addition to the covariate context, we consider the case of contamination and we propose a robust estimator of the conditional Pickands dependence function $A(\cdot|x)$. To reach this goal, we use the density power divergence method introduced by Basu et al. (1998). In particular, the density power divergence between two density functions g and h is defined as follows:

$$\Delta_\alpha(g, h) := \begin{cases} \displaystyle\int_{\mathbb{R}} \left[h^{1+\alpha}(y) - \left(1 + \frac{1}{\alpha}\right) h^\alpha(y) g(y) + \frac{1}{\alpha} g^{1+\alpha}(y)\right] dy, & \alpha > 0, \\ \displaystyle\int_{\mathbb{R}} \log \frac{g(y)}{h(y)} g(y)\, dy, & \alpha = 0. \end{cases}$$

Here, the density function h is assumed to depend on a parameter vector θ, and if $Z_1, \ldots, Z_n$ is a sample of independent and identically distributed random variables according to the density function g, then the minimum density power divergence estimator (MDPDE) of θ is the point $\widehat{\theta}$ minimizing the empirical version (up to a constant independent of θ)

$$\widehat{\Delta}_\alpha(\theta) := \begin{cases} \int_{\mathbb{R}} h^{1+\alpha}(y)\,dy - \left(1+\frac{1}{\alpha}\right)\frac{1}{n}\sum_{i=1}^{n} h^\alpha(Z_i), & \alpha > 0, \\ -\frac{1}{n}\sum_{i=1}^{n} \log h(Z_i), & \alpha = 0. \end{cases}$$

We can observe that for $\alpha = 0$ one recovers the log-likelihood function, up to the sign. A large value of α allows us to increase the robustness of the estimator, whereas a smaller value implies more efficiency. This parameter α can thus be selected in order to ensure a trade-off between these two antagonist concepts.

The nonparametric estimation of extremal dependence in presence of random covariates is still in its infancy, despite the huge potential of such methods for practical data analysis. Gardes and Girard (2015) introduce an estimator for the tail copula based on a random sample from a distribution in the max-domain of attraction of an extreme value distribution, and provide a finite dimensional convergence result for their estimator, when properly normalised. Portier and Segers (2017) considered model (1.1) but under the simplifying assumption that the dependence between $Y^{(1)}$ and $Y^{(2)}$ does not depend on the value taken by the covariate, that is, $C_x = C$ [see also Gijbels, Omelka and Veraverbeke (2015)]. In the present paper, we introduce a nonparametric and robust estimator for $A(\cdot|x)$ which is obtained by an adjustment of the above introduced density power divergence estimation criterion to the situation of local estimation, and we study the asymptotic properties of the obtained estimator in terms of stochastic process convergence. To the best of our knowledge, nonparametric and robust estimation of the conditional Pickands dependence function has not been considered in the literature.

The remainder of the paper is organized as follows. In Section 2, we simplify the situation to the case where the two marginal distributions are known, we propose a robust estimator for $A(\cdot|x)$ and prove its convergence in terms of a stochastic process. Then, in Section 3, we extend this result to the case of unknown margins. The efficiency and robustness of the estimator are examined with a simulation study, described in Section 4. Finally, in Section 5 we illustrate the practical applicability of the method for modelling extremal dependence between air pollution measurements. Additional simulation results are available in the online Supplementary Material [Escobar-Bach, Goegebeur and Guillou (2018)]. All the proofs are postponed to the Appendix.

2. Case of known margins. We denote by f the density function of the covariate X and by x_0 a reference position such that $x_0 \in \mathrm{Int}(S_X)$, the interior of the support S_X of f. In this section, we restrict our interest to the case where the marginals $F_1(\cdot|x)$ and $F_2(\cdot|x)$ are known, and we denote by $A_0(\cdot|x)$ the true conditional Pickands dependence function associated to the pair $(Y^{(1)}, Y^{(2)})$.

2.1. *Construction of the estimator.* For convenience, we reformulate the model (1.1) into standard exponential margins. After applying the transformations $\widetilde{Y}^{(j)} = -\log F_j(Y^{(j)}|x)$, $j = 1, 2$, we obtain the following bivariate survival function:

$$G(y_1, y_2|x) := \mathbb{P}(\widetilde{Y}^{(1)} > y_1, \widetilde{Y}^{(2)} > y_2|X = x) = \exp\left(-(y_1 + y_2)A_0\left(\frac{y_2}{y_1 + y_2}\Big|x\right)\right)$$

for all $y_1, y_2 > 0$. Let $t \in [0, 1]$. Considering the univariate random variable

$$Z_t := \min\left(\frac{\widetilde{Y}^{(1)}}{1-t}, \frac{\widetilde{Y}^{(2)}}{t}\right),$$

it is clear that

$$\mathbb{P}(Z_t > z|X = x) = e^{-zA_0(t|x)} \qquad \forall z > 0 \text{ and } x \in \mathbb{R}^p.$$

Consequently, the conditional distribution of Z_t given $X = x$ is an exponential distribution with parameter $A_0(t|x)$.

Let $(Z_{t,i}, X_i), i = 1, \dots, n$, be independent copies of the random pair (Z_t, X). In the present paper, we will develop a nonparametric robust estimator for $A_0(t|x_0)$ by fitting this exponential distribution function locally to the variables $Z_{t,i}$, $i = 1, \dots, n$, by means of the MDPD criterion, adjusted to locally weighted estimation, that is, we minimize for $\alpha > 0$

$$\begin{aligned}\widehat{\Delta}_{\alpha,x_0,t}(a) &:= \frac{1}{n}\sum_{i=1}^{n} K_h(x_0 - X_i)\left\{\int_0^\infty (ae^{-az})^{1+\alpha}\,dz - \left(1 + \frac{1}{\alpha}\right)(ae^{-aZ_{t,i}})^\alpha\right\} \\ &= \frac{a^\alpha}{n}\sum_{i=1}^{n} K_h(x_0 - X_i)\left\{\frac{1}{1+\alpha} - \left(1 + \frac{1}{\alpha}\right)e^{-\alpha a Z_{t,i}}\right\}.\end{aligned} \tag{2.1}$$

Here, $K_h(\cdot) := K(\cdot/h)/h^p$ where K is a joint density on $\mathbb{R}^p$ and $h = h_n$ is a positive non-random sequence satisfying $h_n \to 0$ as $n \to \infty$. The MDPDE $\widehat{A}_{\alpha,n}(t|x_0)$ for $A_0(t|x_0)$ satisfies the estimating equation

$$\widehat{\Delta}^{(1)}_{\alpha,x_0,t}(\widehat{A}_{\alpha,n}(t|x_0)) = 0, \tag{2.2}$$

where $\widehat{\Delta}^{(j)}_{\alpha,x_0,t}(\cdot)$ denotes the derivative of order j of $\widehat{\Delta}_{\alpha,x_0,t}(\cdot)$. The minimization of $\widehat{\Delta}_{\alpha,x_0,t}$ is here performed without constraints, which means that $\widehat{A}_{\alpha,n}(\cdot|x_0)$ does not automatically satisfy the conditions of the Pickands dependence function. In fact, this is the case for several of the estimators proposed in the literature; see, for example, Pickands (1981), Deheuvels (1991) or Capéraà, Fougères and Genest (1997). To overcome this, one could follow the idea of Fils-Villetard, Guillou and Segers (2008), and project the obtained estimator onto the space of Pickands dependence functions.

Our aim in this paper is to show the weak convergence of the stochastic process

$$\big\{\sqrt{nh^p}\big(\widehat{A}_{\alpha,n}(t|x_0) - A_0(t|x_0)\big), t \in [0,1]\big\}, \tag{2.3}$$

in the space of all continuous functions on [0, 1], denoted as $\mathcal{C}([0,1])$, when $n \to \infty$.

Our starting point is the estimating equation (2.2). By applying a Taylor series expansion around the true value $A_0(t|x_0)$, we get

$$\begin{aligned} 0 &= \widehat{\Delta}^{(1)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big) + \big(\widehat{A}_{\alpha,n}(t|x_0) - A_0(t|x_0)\big)\widehat{\Delta}^{(2)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big) \\ &\quad + \frac{1}{2}\big(\widehat{A}_{\alpha,n}(t|x_0) - A_0(t|x_0)\big)^2\widehat{\Delta}^{(3)}_{\alpha,x_0,t}\big(\widetilde{A}(t|x_0)\big), \end{aligned}$$

where $\widetilde{A}(t|x_0)$ is a random value between $A_0(t|x_0)$ and $\widehat{A}_{\alpha,n}(t|x_0)$. A straightforward rearrangement of the above display gives

$$\begin{aligned} &\sqrt{nh^p}\big(\widehat{A}_{\alpha,n}(t|x_0) - A_0(t|x_0)\big) \\ &\quad = \frac{-\sqrt{nh^p}\widehat{\Delta}^{(1)}_{\alpha,x_0,t}(A_0(t|x_0))}{\widehat{\Delta}^{(2)}_{\alpha,x_0,t}(A_0(t|x_0)) + \frac{1}{2}\widehat{\Delta}^{(3)}_{\alpha,x_0,t}(\widetilde{A}(t|x_0))(\widehat{A}_{\alpha,n}(t|x_0) - A_0(t|x_0))}. \end{aligned} \tag{2.4}$$

Consequently, in order to obtain the convergence of the stochastic process (2.3), we need to study the properties of the derivatives $\widehat{\Delta}^{(j)}_{\alpha,x_0,t}$, $j = 1,2,3$. According to Appendix A.5, these can be expressed as a linear combination of a key statistic T_n, defined as

$$T_n(K,a,t,\lambda,\beta,\gamma|x_0) := \frac{a^\gamma}{n}\sum_{i=1}^n K_h(x_0 - X_i)Z_{t,i}^\beta e^{-\lambda a Z_{t,i}} \tag{2.5}$$

for $a \in [1/2, 1]$, $t \in [0,1]$, $\lambda, \beta \geq 0$ and $\gamma \in \mathbb{R}$.

2.2. *Asymptotic properties of T_n.* Due to the regression context, we need some Hölder-type conditions on the density function f and on the conditional Pickands dependence function A_0. Let $\|\cdot\|$ be some norm on $\mathbb{R}^p$, and denote by $B_x(r)$ the closed ball with respect to $\|\cdot\|$ centered at x and radius $r > 0$.

ASSUMPTION ($\mathcal{D}$). There exist $M_f > 0$ and $\eta_f > 0$ such that $|f(x) - f(z)| \leq M_f\|x - z\|^{\eta_f}$, for all $(x,z) \in S_X \times S_X$.

ASSUMPTION ($\mathcal{A}_0$). There exist $M_{A_0} > 0$ and $\eta_{A_0} > 0$ such that $|A_0(t|x) - A_0(t|z)| \leq M_{A_0}\|x - z\|^{\eta_{A_0}}$, for all $(x,z) \in B_{x_0}(r) \times B_{x_0}(r)$, $r > 0$ and $t \in [0,1]$.

Also a usual condition is assumed on the kernel K.

ASSUMPTION ($\mathcal{K}_1$). K is a bounded density function on $\mathbb{R}^p$ with support S_K included in the unit ball of $\mathbb{R}^p$ with respect to the norm $\|\cdot\|$.

As a preliminary result, in Lemma 2.1 we prove the convergence in probability of the key statistic T_n.

LEMMA 2.1. *Assume that for all $t \in [0,1]$, $x \to A_0(t|x)$ and the density function f are both continuous at $x_0 \in \mathrm{Int}(S_X)$ nonempty. Under Assumption $(\mathcal{K}_1)$, if $h \to 0$ and $nh^p \to \infty$, then for $a \in [1/2, 1]$, $\lambda, \beta \geq 0$, $\gamma \in \mathbb{R}$ and x_0 such that $f(x_0) > 0$, we have*

$$T_n(K, a, t, \lambda, \beta, \gamma|x_0) \xrightarrow{\mathbb{P}} a^\gamma \Gamma(\beta+1) \frac{A_0(t|x_0)}{(\lambda a + A_0(t|x_0))^{\beta+1}} f(x_0)$$

as $n \to \infty$, where Γ is the gamma function defined as $\Gamma(r) := \int_0^\infty t^{r-1} e^{-t}\, dt$, $\forall r > 0$.

Now, our interest is in the rate of convergence in Lemma 2.1 when a is replaced by $A_0(t|x_0)$. More precisely, we want to show the weak convergence of the stochastic process

$$\left\{\sqrt{nh^p}\left(T_n\big(K, A_0(t|x_0), t, \lambda, \beta, \gamma|x_0\big) - \Gamma(\beta+1)\frac{[A_0(t|x_0)]^{\gamma-\beta}}{(\lambda+1)^{\beta+1}} f(x_0)\right),\right.$$
$$\left. t \in [0,1]\right\}.$$

To establish such a result, we use empirical processes arguments based on the theory of Vapnik–Červonenkis classes (*VC*-classes) of functions as formulated in van der Vaart and Wellner (1996). This allows us to show the following theorem.

THEOREM 2.1. *Let $\gamma \in \mathbb{R}$ and $(\lambda, \beta) \in (0, \infty) \times \mathbb{R}_+$ or $(\lambda, \beta) = (0, 0)$. Under the assumptions of Lemma 2.1 and if $(\mathcal{D})$ and $(\mathcal{A}_0)$ hold with $\sqrt{nh^p}h^{\min(\eta_f, \eta_{A_0})} \to 0$, then the process*

$$\left\{\sqrt{nh^p}\left(T_n\big(K, A_0(t|x_0), t, \lambda, \beta, \gamma|x_0\big) - \Gamma(\beta+1)\frac{[A_0(t|x_0)]^{\gamma-\beta}}{(\lambda+1)^{\beta+1}} f(x_0)\right),\right.$$
$$\left. t \in [0,1]\right\}$$

weakly converges in $\mathcal{C}([0,1])$ towards a tight centered Gaussian process $\{B_t, t \in [0,1]\}$ with covariance structure given by

$$\mathrm{Cov}(B_t, B_s) = \big[A_0(t|x_0)A_0(s|x_0)\big]^\gamma \|K\|_2^2 f(x_0)$$
$$\times \left\{\int_{\mathbb{R}_+^2} g(u,v) G_{t,s}(u,v|x_0)\, du\, dv + \frac{1-\lambda}{1+\lambda}\delta_0(\beta)\right\}$$

for all $(s,t) \in [0,1]^2$, *where* δ_0 *is the Dirac measure on* 0, *and*

$$g(u,v) := u^{\beta-1}(\beta - \lambda A_0(t|x_0)u)e^{-\lambda A_0(t|x_0)u} \times v^{\beta-1}(\beta - \lambda A_0(s|x_0)v)e^{-\lambda A_0(s|x_0)v},$$

$$G_{t,s}(u,v|x_0) := G\big(\max\big((1-t)u,(1-s)v\big), \max(tu,sv)|x_0\big).$$

We now derive the limiting distribution of a vector of statistics of the form (2.5), when properly normalized. Let $\mathbb{T}_n$ be a $(m \times 1)$ vector defined as

$$\mathbb{T}_n := \big(T_n\big(K, A_0(t_1|x_0), t_1, \lambda_1, \beta_1, \gamma_1|x_0\big), \dots, T_n\big(K, A_0(t_m|x_0), t_m, \lambda_m, \beta_m, \gamma_m|x_0\big)\big)^T$$

for some positive integer m and let Σ be a $(m \times m)$ covariance matrix with elements $(\sigma_{j,k})_{1\le j,k\le m}$ defined as

$$\begin{aligned}\sigma_{j,k} &:= \big[A_0(t_j|x_0)\big]^{\gamma_j}\big[A_0(t_k|x_0)\big]^{\gamma_k}\|K\|_2^2 f(x_0) \\ &\quad\times \bigg\{\int_{\mathbb{R}_+^2} g_{j,k}(u,v)G_{t_j,t_k}(u,v|x_0)\,du\,dv \\ &\quad+ \delta_0(\beta_j)\frac{\Gamma(\beta_k+1)}{[\lambda_k+1]^{\beta_k+1}[A_0(t_k|x_0)]^{\beta_k}} \\ &\quad+ \delta_0(\beta_k)\frac{\Gamma(\beta_j+1)}{[\lambda_j+1]^{\beta_j+1}[A_0(t_j|x_0)]^{\beta_j}} - \delta_0(\beta_j)\delta_0(\beta_k)\bigg\},\end{aligned} \tag{2.6}$$

where

$$g_{j,k}(u,v) := u^{\beta_j-1}\big[\beta_j - \lambda_j A_0(t_j|x_0)u\big]e^{-\lambda_j A_0(t_j|x_0)u} \times v^{\beta_k-1}\big[\beta_k - \lambda_k A_0(t_k|x_0)v\big]e^{-\lambda_k A_0(t_k|x_0)v}.$$

The aim of next theorem is to provide the finite dimensional convergence result which will, together with the tightness, allow us to establish the joint convergence of several processes related to the statistic T_n.

THEOREM 2.2. *Under the assumptions of Lemma* 2.1, *we have*

$$\sqrt{nh^p}\big(\mathbb{T}_n - \mathbb{E}[\mathbb{T}_n]\big) \rightsquigarrow \mathcal{N}_m(0,\Sigma),$$

where $\mathcal{N}_m$ *denotes a* m*-dimensional normal distribution.*

We have now all the needed ingredients for proving the asymptotic properties of the MDPDE $\widehat{A}_{\alpha,n}(t|x_0)$.

2.3. *Asymptotic properties of* $\widehat{A}_{\alpha,n}(t|x_0)$. The first result states the existence and uniform consistency of a sequence of solutions to the estimating equation (2.2).

THEOREM 2.3. *Let* $\alpha > 0$. *Under the assumptions of Theorem* 2.1, *with probability tending to* 1, *there exists a sequence* $(\widehat{A}_{\alpha,n}(t|x_0))_{n\in\mathbb{N}}$ *of solutions for the estimating equation* (2.2) *such that*

$$\sup_{t\in[0,1]} \left|\widehat{A}_{\alpha,n}(t|x_0) - A_0(t|x_0)\right| = o_{\mathbb{P}}(1).$$

Now, we come back to our final goal which is the weak convergence of the stochastic process (2.3).

THEOREM 2.4. *Let* $(\widehat{A}_{\alpha,n}(t|x_0))_{n\in\mathbb{N}}$ *be the consistent sequence defined in Theorem* 2.3. *Under the assumptions of Theorem* 2.1, *the process*

$$\left\{\sqrt{nh^p}\left(\widehat{A}_{\alpha,n}(t|x_0) - A_0(t|x_0)\right), t \in [0,1]\right\}$$

weakly converges in $\mathcal{C}([0,1])$ *towards a tight centered Gaussian process* $\{N_t, t \in [0,1]\}$ *with covariance structure given by*

$$\operatorname{Cov}(N_t, N_s) = \frac{\|K\|_2^2 A_0(t|x_0) A_0(s|x_0)}{f(x_0)} \frac{(1+\alpha)^2}{(1+\alpha^2)^2} v_\alpha^T \Sigma(t,s) v_\alpha,$$

where

$$v_\alpha := \begin{pmatrix} \dfrac{\alpha}{1+\alpha} \\ -(1+\alpha) \\ 1+\alpha \end{pmatrix} \quad \text{and} \quad \Sigma(t,s) := \begin{pmatrix} (1+\alpha)^2 & 1+\alpha & 1 \\ 1+\alpha & \Sigma_{2,2}(t,s) & \Sigma_{2,3}(t,s) \\ 1 & \Sigma_{2,3}(s,t) & \Sigma_{3,3}(t,s) \end{pmatrix}$$

with

$$\begin{aligned} \Sigma_{2,2}(t,s) &:= (1-\alpha)(1+\alpha) + \alpha^2(1+\alpha)^2 A_0(t|x_0)A_0(s|x_0) \\ &\quad \times \int_{\mathbb{R}_+^2} e^{-\alpha[A_0(t|x_0)u + A_0(s|x_0)v]} G_{t,s}(u,v|x_0)\,du\,dv, \\ \Sigma_{2,3}(t,s) &:= 1 - \alpha(1+\alpha)^2 A_0(t|x_0)A_0(s|x_0) \\ &\quad \times \int_{\mathbb{R}_+^2} (1-\alpha A_0(s|x_0)v) e^{-\alpha[A_0(t|x_0)u + A_0(s|x_0)v]} G_{t,s}(u,v|x_0)\,du\,dv, \\ \Sigma_{3,3}(t,s) &:= (1+\alpha)^2 A_0(t|x_0)A_0(s|x_0) \\ &\quad \times \int_{\mathbb{R}_+^2} (1-\alpha A_0(t|x_0)u)(1-\alpha A_0(s|x_0)v) \\ &\quad \times e^{-\alpha[A_0(t|x_0)u + A_0(s|x_0)v]} G_{t,s}(u,v|x_0)\,du\,dv. \end{aligned}$$

In particular, for all $t \in [0, 1]$, *we have*

$$\sqrt{nh^p}\big(\widehat{A}_{\alpha,n}(t|x_0) - A_0(t|x_0)\big) \rightsquigarrow \mathcal{N}_1\left(0, \frac{\|K\|_2^2[A_0(t|x_0)]^2}{f(x_0)} \frac{(1+\alpha)^2(1+4\alpha+9\alpha^2+14\alpha^3+13\alpha^4+8\alpha^5+4\alpha^6)}{(1+2\alpha)^3(1+\alpha^2)^2}\right)$$

as $n \to \infty$.

The asymptotic standard deviation is shown as a function of α in Figure S1 of the Supplementary Material. As is clear from this plot, the asymptotic standard deviation is increasing in α. Note that our results could also be obtained under different assumptions by using the local empirical process results of Stute (1986) and Einmahl and Mason (1997), combined with the functional delta method.

3. Case of unknown margins. In this section, we consider the general framework where both $F_1(\cdot|x)$ and $F_2(\cdot|x)$ are unknown conditional distribution functions. We want to mimic what has been done in the previous section and transform to standard exponential margins. To this aim, we consider the triplets

$$\big(-\log\big(F_{n,1}\big(Y_i^{(1)}|X_i\big)\big), -\log\big(F_{n,2}\big(Y_i^{(2)}|X_i\big)\big), X_i\big), \qquad i = 1, \dots, n$$

for suitable estimators $F_{n,j}$ of F_j, $j = 1, 2$, and we compute the univariate random variables

$$\check{Z}_{n,t,i} := \min\left(\frac{-\log(F_{n,1}(Y_i^{(1)}|X_i))}{1-t}, \frac{-\log(F_{n,2}(Y_i^{(2)}|X_i))}{t}\right), \qquad i = 1, \dots, n.$$

Then, similarly as in Section 2, the statistic

$$\check{T}_n(K, a, t, \lambda, \beta, \gamma|x_0) := \frac{a^\gamma}{n} \sum_{i=1}^n K_h(x_0 - X_i)\check{Z}_{n,t,i}^\beta e^{-\lambda a \check{Z}_{n,t,i}} \tag{3.1}$$

is the cornerstone for the MDPDE, denoted $\check{A}_{\alpha,n}(t|x_0)$, which satisfies the estimating equation

$$\check{\Delta}_{\alpha,x_0,t}^{(1)}\big(\check{A}_{\alpha,n}(t|x_0)\big) = 0, \tag{3.2}$$

where $\check{\Delta}_{\alpha,x_0,t}^{(1)}(\cdot)$ is the first derivative of $\check{\Delta}_{\alpha,x_0,t}(\cdot)$ and

$$\check{\Delta}_{\alpha,x_0,t}(a) := \frac{a^\alpha}{n} \sum_{i=1}^n K_h(x_0 - X_i)\left\{\frac{1}{1+\alpha} - \left(1 + \frac{1}{\alpha}\right)e^{-\alpha a \check{Z}_{n,t,i}}\right\}.$$

The final goal is still the same, that is the weak convergence of the stochastic process

$$\big\{\sqrt{nh^p}\big(\check{A}_{\alpha,n}(t|x_0) - A_0(t|x_0)\big), t \in [0, 1]\big\}. \tag{3.3}$$

Again this result relies essentially on the asymptotic properties of the statistic $\check{T}_n$, and so the idea will be to decompose

$$\sqrt{nh^p}\big(\check{T}_n - \mathbb{E}[\check{T}_n]\big)(K, a, t, \lambda, \beta, \gamma|x_0),$$

into the two terms

$$\begin{aligned} &\big\{\sqrt{nh^p}\big(T_n - \mathbb{E}[T_n]\big)(K, a, t, \lambda, \beta, \gamma|x_0)\big\} \\ &\quad + \big\{\sqrt{nh^p}\big([\check{T}_n - T_n] - \mathbb{E}[\check{T}_n - T_n]\big)(K, a, t, \lambda, \beta, \gamma|x_0)\big\}. \end{aligned} \tag{3.4}$$

The first term can be dealt with using the results of Section 2.2, whereas we have to show that the second term is uniformly negligible.

To achieve this objective, let us introduce the following empirical kernel estimator of the unknown conditional distribution functions:

$$F_{n,j}(y|x) := \frac{\sum_{i=1}^n K_c(x - X_i)\mathbb{1}_{\{Y_i^{(j)} \le y\}}}{\sum_{i=1}^n K_c(x - X_i)}, \qquad j = 1, 2,$$

where $c := c_n$ is a positive non-random sequence satisfying $c_n \to 0$ as $n \to \infty$. Here, we kept the same kernel K as in the previous section, but of course any other kernel function can be used.

Before stating our main results, we need to impose again some assumptions, in particular a Hölder-type condition on each marginal conditional distribution function F_j similar to the one imposed on the density function of the covariate.

ASSUMPTION ($\mathcal{F}$). There exist $M_{F_j} > 0$ and $\eta_{F_j} > 0$ such that $|F_j(y|x) - F_j(y|z)| \le M_{F_j} \|x - z\|^{\eta_{F_j}}$, for all $y \in \mathbb{R}$ and all $(x, z) \in S_X \times S_X$, and $j = 1, 2$.

Concerning the kernel K a stronger assumption than ($\mathcal{K}_1$) is needed.

ASSUMPTION ($\mathcal{K}_2$). K satisfies Assumption ($\mathcal{K}_1$), there exists $\delta, m > 0$ such that $B_0(\delta) \subset S_K$ and $K(u) \ge m$ for all $u \in B_0(\delta)$, and K belongs to the linear span (the set of finite linear combinations) of functions $k \ge 0$ satisfying the following property: the subgraph of k, $\{(s, u) : k(s) \ge u\}$, can be represented as a finite number of Boolean operations among sets of the form $\{(s, u) : q(s, u) \ge \varphi(u)\}$, where q is a polynomial on $\mathbb{R}^p \times \mathbb{R}$ and φ is an arbitrary real function.

The latter assumption has already been used in Giné and Guillou (2002) or Giné, Koltchinskii and Zinn (2004). In particular, it allows us to measure the discrepancy between the conditional distribution function F_j and its empirical kernel version $F_{n,j}$.

LEMMA 3.1. *Assume that there exists $b > 0$ such that $f(x) \ge b, \forall x \in S_X \subset \mathbb{R}^p$, f is bounded, and ($\mathcal{K}_2$) and ($\mathcal{F}$) hold. Consider a sequence c tending to 0 as $n \to \infty$ such that for some $q > 1$*

$$\frac{|\log c|^q}{nc^p} \longrightarrow 0.$$

Also assume that there exists an $\varepsilon > 0$ *such that for* n *sufficiently large*

$$\inf_{x \in S_X} \lambda(\{u \in B_0(1) : x - cu \in S_X\}) > \varepsilon, \tag{3.5}$$

where λ *denotes the Lebesgue measure. Then, for any* $0 < \eta < \min(\eta_{F_1}, \eta_{F_2})$, *we have*

$$\sup_{(y,x) \in \mathbb{R} \times S_X} |F_{n,j}(y|x) - F_j(y|x)| = o_{\mathbb{P}}\left(\max\left(\sqrt{\frac{|\log c|^q}{nc^p}}, c^\eta\right)\right) \qquad \textit{for } j = 1, 2.$$

Note that the assumption $f(x) \geq b, \forall x \in S_X$, for some $b > 0$, is similar to the one already used in Gijbels, Omelka and Veraverbeke (2015) and Portier and Segers (2017).

We are now able to study the second term in (3.4).

THEOREM 3.1. *Assume that there exists* $b > 0$ *such that* $f(x) \geq b, \forall x \in S_X \subset \mathbb{R}^p$, f *is bounded, and* $(\mathcal{K}_2)$, $(\mathcal{D})$ *and* $(\mathcal{F})$ *hold together with condition* (3.5). *Consider two sequences* h *and* c *tending to* 0, *such that for* $nh^p \to \infty$ *and for some* $q > 1$ *and any* $0 < \eta < \min(\eta_{F_1}, \eta_{F_2})$

$$\sqrt{nh^p} r_n := \sqrt{nh^p} \max\left(\sqrt{\frac{|\log c|^q}{nc^p}}, c^\eta\right) \longrightarrow 0$$

as $n \to \infty$. *Then, for all* $\gamma \in \mathbb{R}$ *and* $(\lambda, \beta) \in (0, \infty) \times \mathbb{R}_+$ *or* $(\lambda, \beta) = (0, 0)$, *we have*

$$\sup_{t \in [0,1], a \in [1/2,1]} \sqrt{nh^p} |\check{T}_n - T_n - \mathbb{E}[\check{T}_n - T_n]|(K, a, t, \lambda, \beta, \gamma | x_0) = o_{\mathbb{P}}(1).$$

Finally, the decomposition (3.4) combined with Theorem 3.1 and the results from Section 2.2, yields the desired theoretical result of this paper.

THEOREM 3.2. *Let* $\alpha > 0$. *Under the assumptions of Theorem* 3.1 *and* $(\mathcal{A}_0)$, *with probability tending to* 1, *there exists a sequence* $(\check{A}_{\alpha,n}(t|x_0))_{n \in \mathbb{N}}$ *of solutions for the estimating equation* (3.2) *such that*

$$\sup_{t \in [0,1]} |\check{A}_{\alpha,n}(t|x_0) - A_0(t|x_0)| = o_{\mathbb{P}}(1).$$

Moreover, for this consistent sequence, if $\sqrt{nh^p} h^{\min(\eta_f, \eta_{A_0})} \to 0$, *the process*

$$\{\sqrt{nh^p}(\check{A}_{\alpha,n}(t|x_0) - A_0(t|x_0)), t \in [0, 1]\}$$

weakly converges in $\mathcal{C}([0, 1])$ *towards the tight centered Gaussian process* $\{N_t, t \in [0, 1]\}$ *defined in Theorem* 2.4.

4. Simulation study. Our aim in this section is to illustrate the efficiency of the proposed robust estimator for the conditional Pickands dependence function with a simulation study. We assume that the conditional distribution function of $(Y^{(1)}, Y^{(2)})$ given $X = x$ is a mixture model of the form

$$F_\varepsilon(y_1, y_2|x) = (1-\varepsilon)F_\ell(y_1, y_2|x) + \varepsilon F_c(y_1, y_2|x),$$

where $\varepsilon \in [0, 1]$ represents the fraction of contamination in the dataset. The main distribution F_ℓ is the logistic distribution given by

$$F_\ell(y_1, y_2|x) := \exp\{-(y_1^{-1/x} + y_2^{-1/x})^x\} \qquad \text{for } y_1, y_2 \geq 0$$

and

$$A_0(t|x) = (t^{1/x} + (1-t)^{1/x})^x,$$

where the covariate X is a uniformly distributed random variable on $[0, 1]$. For this model, complete dependence is obtained in the limit as $x \downarrow 0$, whereas independence can be reached for $x = 1$. Note also that the conditional marginal distributions of $Y^{(j)}$ given X, $j = 1, 2$, under this logistic model are unit Fréchet distributions. Moreover, we can check that this model satisfies conditions $(\mathcal{D})$, $(\mathcal{A}_0)$ and $(\mathcal{F})$. Two completely different types of distributions F_c will be considered throughout the paper and additional examples will be given in the online Supplementary Material.

• *First type of contamination*: Given $X = x$, the distribution function F_c is

$$F_c(y_1, y_2|x) = \frac{1}{2}\{e^{-y_1^{-1}} + e^{-y_2^{-1}}\}\mathbb{1}_{\{y_1 \geq 0, y_2 \geq 0\}}.$$

The mixture based on this distribution F_c is illustrated in Figure S2 of the Supplementary Material. For this mixture, the contaminated points are on the axes.

• *Second type of contamination*: The distribution function F_c has completely dependent unit exponential margins. Figure S3 in the Supplementary Material shows an example of a simulated dataset from this model. This time, the contaminated points are on the diagonal.

To compute the estimator $\check{A}_{\alpha,n}$, two sequences h and c have to be chosen. Both are determined by a cross validation criterion. Because of the very high computational burden of the cross validations for sample sizes $n \geq 1000$, a random selection of size $n_r := n \wedge 1000$ from the original observations is obtained, denoted $\{(Y_{i,r}^{(1)}, Y_{i,r}^{(2)}, X_{i,r})\}_{i=1,\ldots,n_r}$, and the cross validations are implemented on these random subsamples. Concerning c, we can use the following cross validation criterion, already used in an extreme value context by Daouia et al. (2011):

$$c_j := \arg\min_{\widetilde{c}_j \in \mathcal{C}} \sum_{i=1}^{n_r}\sum_{k=1}^{n_r}[\mathbb{1}_{\{Y_{i,r}^{(j)} \leq Y_{k,r}^{(j)}\}} - \widetilde{F}_{n_r,-i,j}(Y_{k,r}^{(j)}|X_{i,r})]^2, \qquad j = 1, 2,$$

where $\mathcal{C}$ is a grid of values of $\widetilde{c}_j$ and $\widetilde{F}_{n_r,-i,j}(y|x) := \frac{\sum_{k=1,k\neq i}^{n_r} K_{\widetilde{c}_j}(x-X_{k,r})\mathbb{1}_{\{Y_{k,r}^{(j)} \leq y\}}}{\sum_{k=1,k\neq i}^{n_r} K_{\widetilde{c}_j}(x-X_{k,r})}$.

Also the bandwidth parameter h is selected using a cross validation criterion. In particular,

$$h := \arg\min_{\widetilde{h}\in\mathcal{H}} \frac{1}{n_r M} \sum_{i=1}^{n_r} \sum_{j=1}^{M} \check{A}_{\alpha,n,(-i)}(t_j|X_{i,r})^\alpha \times \left(\frac{1}{1+\alpha} - \left(1+\frac{1}{\alpha}\right) e^{-\alpha \check{A}_{\alpha,n,(-i)}(t_j|X_{i,r})\check{Z}_{n,t_j,i,r}}\right) \qquad \text{for } \alpha > 0,$$

$$h := \arg\min_{\widetilde{h}\in\mathcal{H}} \frac{1}{n_r M} \times \sum_{i=1}^{n_r} \sum_{j=1}^{M} -\log\left(\check{A}_{0,n,(-i)}(t_j|X_{i,r}) e^{-\check{A}_{0,n,(-i)}(t_j|X_{i,r})\check{Z}_{n,t_j,i,r}}\right) \qquad \text{for } \alpha = 0,$$

where $\check{A}_{\alpha,n,(-i)}(t|x)$ denotes the estimator of $A_0(t|x)$ obtained on all but observation i, $\check{Z}_{n,t_j,i,r}$ is as $\check{Z}_{n,t_j,i}$ but now calculated for $(Y_{i,r}^{(1)}, Y_{i,r}^{(2)}, X_{i,r})$, and

$$\check{A}_{0,n}(t|x) := \frac{\sum_{i=1}^n K_{\widetilde{h}}(x - X_i)}{\sum_{i=1}^n K_{\widetilde{h}}(x - X_i)\check{Z}_{t,i}}.$$

This criterion can be seen as a generalisation of a commonly used cross validation from the context of local likelihood estimation [see, e.g., Abegaz, Gijbels and Veraverbeke (2012)] to the context of local MDPD estimation.

After extensive simulation studies, we have chosen the grids $\mathcal{C} = \{0.06, 0.12, \ldots, 0.3\}$ and $\mathcal{H} = \{0.02, 0.03, \ldots, 0.06\}$. These choices provide a reasonable trade off between stability of the estimates and accuracy of approximation by asymptotic results.

Concerning the kernel, each time we use the bi-quadratic function

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

As an indicator of efficiency, we compute over a grid the L^2-error in the estimation of the Pickands dependence function $A(\cdot|x)$ as a function of x, that is,

$$\text{MISE}(\varepsilon, \alpha|x) := \frac{1}{NM} \sum_{i=1}^{N} \sum_{m=1}^{M} \left[\check{A}_{\alpha,\varepsilon,n}^{(i)}(t_m|x) - A_0(t_m|x)\right]^2.$$

Here, $\check{A}_{\alpha,\varepsilon,n}^{(i)}(t_m|x)$ is our estimator of $A_0(t_m|x)$ obtained with the ith sample when the contamination is ε. We set $t_m = m/50$, $m = 1, \ldots, 49$. Our simulations are based on datasets of sizes $n = 1000$ and $n = 5000$, and the procedure is repeated $N = 200$ times.

Figure 1 represents the $\text{MISE}(\varepsilon, \alpha|x)$ as a function of $\varepsilon \in \{0, 0.025, 0.05, \ldots, 0.2\}$ for the two types of contamination (rows 1 and 2, and 3 and 4, respectively).

From the left to the right, three positions have been considered: $x = 0.1, 0.3$ and 0.5 for the first type of contamination and $x = 0.5, 0.7$ and 0.9 for the second type. Also, four different values of α have been reported: 0 (black), 0.1 (blue), 0.5 (green) and 1 (red), and two sample sizes: $n = 1000$ and $n = 5000$. Based on these simulations, we can draw the following conclusions:

- As expected, the MISE curves show less variability for $n = 5000$ compared to $n = 1000$.
- If the percentage of contamination ε is very small (in the range 0–0.025), the MISE indicators are typically very similar, whatever the value of α. This result is a nice feature of our method, because if there is almost no contamination then in principle one does not need a robust procedure, but as is clear from these figures, the MDPDE performs similar to the maximum likelihood method (corresponding to $\alpha = 0$), which is efficient (but not robust).
- If we increase the percentage of contamination ε, then it is crucial to increase the value of α to 0.5 or 1 in order to have good results. Indeed, for increasing ε a small value of α implies a drastic increase of the MISE.
- The MISE values are almost constant for $\alpha = 0.5$ and 1, whatever the percentage of contamination. This illustrates again the robustness of our method, since it means that the methodology can handle a quite large percentage of contamination without deterioration of the results.
- For the first type of contamination, the gain in MISE by taking $\alpha = 0.5$ or 1 over $\alpha = 0$ or 0.1 is more important for small x than for large x. In this case, the contamination is on the axes (independently), and this is less disturbing for x close to 1, which corresponds also to independence, than for x close to 0, which corresponds to complete dependence. For the second type of contamination, one can observe the opposite effect. The gain of taking $\alpha = 0.5$ or 1 over $\alpha = 0$ or 0.1 is more important for x close to 1 than for x close to zero. Indeed, the perfectly dependent contamination is less disturbing for small x than for large x.
- Figure 1 gives us also some indications about the breakdown point of our estimator, which is a common concept in the robust framework. Indeed the breakdown point can be interpreted as the smallest ε where the MISE indicator starts to increase. For α small, in the range 0–0.1, the breakdown point is very small, say ε around 0.025, while for $\alpha = 0.5$ and 1, one can go to $\varepsilon = 0.15$ or a larger value, depending on the type of contamination, which illustrates again the nice robustness property of our method.

Next to the above mentioned MISE indicators, we also used our simulated data to compute the empirical coverage probabilities of 90% confidence intervals based on the limiting distribution given in Theorem 2.4. These are given in Tables 1 and 2 for the first and second type of contamination, respectively. From these tables, we can see that:

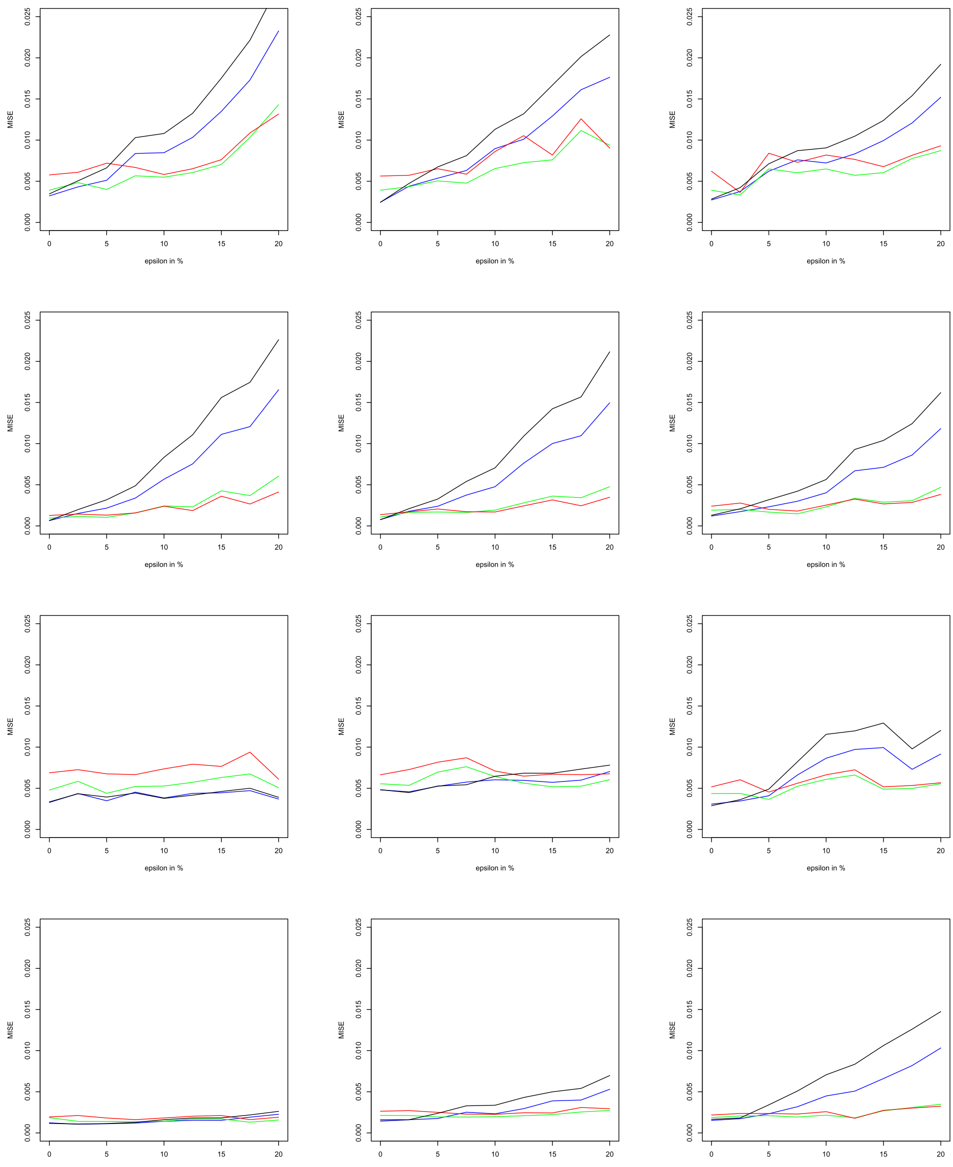

FIG. 1. $\mathrm{MISE}(\varepsilon, \alpha|x)$ *as a function of* $\varepsilon \in \{0, 0.025, 0.05, \ldots, 0.2\}$ *with* $\alpha = 0$ (*black*), $\alpha = 0.1$ (*blue*), $\alpha = 0.5$ (*green*) *and* $\alpha = 1$ (*red*). *First type of contamination*: *rows* 1 ($n = 1000$) *and* 2 ($n = 5000$), $x = 0.1, 0.3, 0.5$ *from the left to the right*. *Second type of contamination*: *rows* 3 ($n = 1000$) *and* 4 ($n = 5000$), $x = 0.5, 0.7, 0.9$ *from the left to the right*.

TABLE 1
First type of contamination—coverage probabilities of 90% confidence intervals

		$t=0.3$				$t=0.5$				$t=0.7$			
	α	**0**	**0.1**	**0.5**	**1**	**0**	**0.1**	**0.5**	**1**	**0**	**0.1**	**0.5**	**1**
						$x=0.1$							
$n=1000$	$\varepsilon=0.0$	0.95	0.95	0.96	0.98	0.95	0.96	0.96	0.97	0.95	0.96	0.96	0.98
	$\varepsilon=0.1$	0.69	0.77	0.92	0.93	0.50	0.64	0.88	0.91	0.67	0.78	0.92	0.93
	$\varepsilon=0.2$	0.26	0.41	0.83	0.91	0.11	0.17	0.64	0.79	0.24	0.42	0.81	0.90
$n=5000$	$\varepsilon=0.0$	0.98	0.97	0.96	0.97	0.96	0.97	0.97	0.98	0.95	0.96	0.97	0.97
	$\varepsilon=0.1$	0.14	0.28	0.84	0.90	0.06	0.12	0.69	0.82	0.13	0.29	0.84	0.90
	$\varepsilon=0.2$	0.02	0.05	0.55	0.83	0.00	0.01	0.16	0.51	0.01	0.03	0.56	0.80
						$x=0.3$							
$n=1000$	$\varepsilon=0.0$	0.97	0.98	0.93	0.93	0.96	0.96	0.95	0.93	0.94	0.95	0.95	0.94
	$\varepsilon=0.1$	0.70	0.80	0.93	0.96	0.71	0.80	0.93	0.94	0.70	0.82	0.95	0.95
	$\varepsilon=0.2$	0.29	0.45	0.86	0.94	0.20	0.40	0.78	0.85	0.27	0.47	0.84	0.92
$n=5000$	$\varepsilon=0.0$	0.96	0.95	0.94	0.94	0.94	0.94	0.93	0.94	0.96	0.95	0.94	0.94
	$\varepsilon=0.1$	0.14	0.33	0.87	0.93	0.15	0.35	0.81	0.86	0.17	0.41	0.85	0.92
	$\varepsilon=0.2$	0.01	0.04	0.57	0.83	0.01	0.02	0.41	0.63	0.01	0.03	0.55	0.83
						$x=0.5$							
$n=1000$	$\varepsilon=0.0$	0.97	0.99	0.98	0.97	0.97	0.97	0.97	0.95	0.96	0.97	0.97	0.98
	$\varepsilon=0.1$	0.77	0.85	0.91	0.95	0.76	0.84	0.91	0.94	0.79	0.83	0.95	0.94
	$\varepsilon=0.2$	0.51	0.69	0.93	0.95	0.53	0.66	0.89	0.92	0.47	0.65	0.92	0.96
$n=5000$	$\varepsilon=0.0$	0.93	0.95	0.95	0.96	0.90	0.92	0.91	0.94	0.94	0.96	0.94	0.94
	$\varepsilon=0.1$	0.30	0.53	0.91	0.95	0.39	0.56	0.90	0.91	0.31	0.54	0.93	0.94
	$\varepsilon=0.2$	0.04	0.10	0.70	0.87	0.06	0.11	0.66	0.80	0.05	0.08	0.69	0.86

- When $\varepsilon = 0$, then the empirical coverage probabilities are generally larger than 0.90, meaning that the confidence interval based on the limiting distribution is conservative;
- For increasing ε, the coverage probabilities generally decrease when α is small (0 or 0.1), while for larger α, especially for $\alpha = 1$, the coverage probabilities do not seem to be as much affected by the contamination.

Since the main objective of this paper is to estimate the conditional Pickands dependence function $A(\cdot|x)$, we also provide in the online Supplementary Material the boxplots of our estimator $\check{A}_{\alpha,n}(\cdot|x)$ based on 200 replications for the two examples of contamination introduced above and two additional examples. These figures emphasize again the robustness properties of our estimator.

5. Application to air pollution data. In this section, we illustrate the practical applicability of our method on a dataset of air pollution measurements. Extreme temperature and high levels of pollutants like ground-level ozone and particu-

TABLE 2
Second type of contamination—coverage probabilities of 90% confidence intervals

		$t = 0.3$				$t = 0.5$				$t = 0.7$			
	α	**0**	**0.1**	**0.5**	**1**	**0**	**0.1**	**0.5**	**1**	**0**	**0.1**	**0.5**	**1**
						$x = 0.5$							
$n = 1000$	$\varepsilon = 0.0$	0.96	0.96	0.97	0.96	0.94	0.93	0.96	0.96	0.96	0.97	0.96	0.96
	$\varepsilon = 0.1$	0.97	0.96	0.96	0.95	0.91	0.93	0.94	0.93	0.96	0.97	0.98	0.96
	$\varepsilon = 0.2$	0.99	0.99	0.96	0.96	0.80	0.88	0.92	0.94	0.99	0.99	0.96	0.95
$n = 5000$	$\varepsilon = 0.0$	0.93	0.94	0.95	0.96	0.93	0.93	0.95	0.97	0.95	0.96	0.95	0.96
	$\varepsilon = 0.1$	0.93	0.98	0.96	0.96	0.48	0.69	0.94	0.96	0.95	0.99	0.99	0.98
	$\varepsilon = 0.2$	0.92	0.98	0.96	0.95	0.20	0.33	0.88	0.96	0.87	0.97	0.96	0.94
						$x = 0.7$							
$n = 1000$	$\varepsilon = 0.0$	0.97	0.97	1.00	1.00	0.94	0.94	0.96	0.99	0.94	0.96	1.00	1.00
	$\varepsilon = 0.1$	0.98	0.98	1.00	1.00	0.76	0.84	0.96	0.98	0.96	0.95	1.00	1.00
	$\varepsilon = 0.2$	0.98	0.98	1.00	1.00	0.55	0.70	0.94	0.99	0.98	0.98	1.00	1.00
$n = 5000$	$\varepsilon = 0.0$	0.92	0.93	0.94	0.94	0.92	0.94	0.95	0.95	0.94	0.95	0.95	0.96
	$\varepsilon = 0.1$	0.81	0.89	0.95	0.96	0.24	0.54	0.94	0.95	0.79	0.88	0.95	0.95
	$\varepsilon = 0.2$	0.57	0.67	0.90	0.94	0.06	0.11	0.77	0.90	0.60	0.69	0.91	0.93
						$x = 0.9$							
$n = 1000$	$\varepsilon = 0.0$	1.00	1.00	1.00	1.00	0.99	0.99	1.00	1.00	0.98	0.98	0.99	0.99
	$\varepsilon = 0.1$	0.91	0.95	0.99	0.99	0.64	0.81	0.97	0.99	0.91	0.93	0.99	0.99
	$\varepsilon = 0.2$	0.89	0.92	0.98	1.00	0.40	0.63	0.96	0.98	0.89	0.93	0.99	1.00
$n = 5000$	$\varepsilon = 0.0$	0.98	0.98	0.99	0.98	0.98	0.97	0.98	0.98	0.97	0.97	0.98	0.98
	$\varepsilon = 0.1$	0.61	0.77	0.97	0.98	0.15	0.39	0.95	0.97	0.60	0.82	0.98	0.99
	$\varepsilon = 0.2$	0.26	0.44	0.92	0.95	0.01	0.09	0.74	0.91	0.24	0.42	0.91	0.96

late matter pose a major threat to human health. We consider the data collected by the United States Environmental Protection Agency (EPA), publicly available at https://aqsdr1.epa.gov/aqsweb/aqstmp/airdata/download_files.html. The dataset contains daily measurements on, among others, maximum temperature, and ground-level ozone, carbon monoxide and particulate matter concentrations, for the time period 1999 to 2013. These data are collected at stations spread over the U.S. We focus the analysis on the ground-level ozone and particulate matter concentrations. In order to estimate the extremal dependence between these, we calculate the component-wise monthly maximum of daily maximum concentrations, and estimate the Pickands dependence function conditional on the covariates time and location, where the latter is expressed by latitude and longitude. The estimation method was implemented with the same cross validation criteria as in the simulation section, including the same choices for $\mathcal{C}$ and $\mathcal{H}$, after standardising the covariates to the interval [0, 1]. As kernel function K^*, we use the following generalisation of the bi-quadratic kernel K:

$$K^*(x_1, x_2, x_3) := \prod_{i=1}^{3} K(x_i),$$

where x_1, x_2, x_3, refer to the covariates time, latitude and longitude, respectively, in standardised form. Note that K^* has as support the unit ball with respect to the max-norm on $\mathbb{R}^3$. We report here only the results for the city of Houston. Similar results can though be obtained for other cities or regions in the U.S. In the left panel of Figure 2, we show the time plot of the estimates for the conditional extremal coefficient over the observation period. The conditional extremal coefficient is defined as $\eta(x) = 2A_0(0.5|x)$, and is often used as a summary measure of extremal dependence. Its range is $[1, 2]$, where 1 corresponds with perfect dependence and 2 with independence. The time plot shows a seasonal pattern in the extremal dependence, and moreover, the extremal dependence seems to decrease with time. We also observe that the estimates for $\alpha = 0$ and 0.1 are similar, but different from those obtained with $\alpha = 0.5$ and 1 (which are also similar), indicating that the dataset contains contamination with respect to the dependence structure. In order to get a better idea about the extremal dependence, we show in the right panel of Figure 2 the estimate of $A_0(t|x)$ for a particular month (April 2002). This plot shows again estimates which are similar for $\alpha = 0$ and 0.1, but different from those obtained with $\alpha = 0.5$ and 1 (which are similar), confirming our earlier observation that there are observations which are contaminating with respect to the dependence structure.

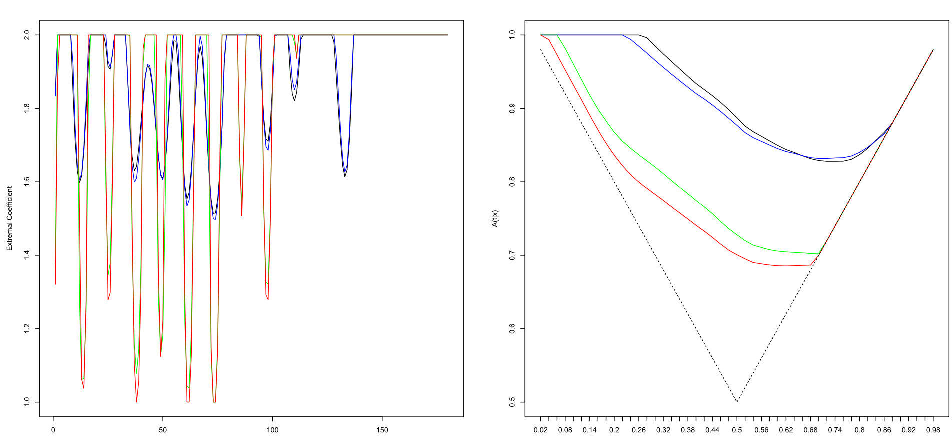

FIG. 2. *Air pollution data*: *time plot of the estimate for the conditional extremal coefficient* (*left*) *and estimate for the conditional Pickands dependence function in April* 2002 (*right*), $\alpha = 0$ (*black*), $\alpha = 0.1$ (*blue*), $\alpha = 0.5$ (*green*) *and* $\alpha = 1$ (*red*).

APPENDIX: PROOFS OF THE RESULTS

A.1. Proof of Lemma 2.1. Using the fact that for any $x \in \mathbb{R}^p$ the conditional distribution function of Z_t given $X = x$ is an exponential distribution with parameter $A_0(t|x)$ and since $\lambda a + A_0(t|x) > 0$, we have

$$\mathbb{E}\big[Z_t^{\beta} e^{-\lambda a Z_t} | X = x\big] = \Gamma(\beta+1) \frac{A_0(t|x)}{(\lambda a + A_0(t|x))^{\beta+1}}. \tag{A.1}$$

Then

$$\begin{aligned}
&\mathbb{E}\big[K_h(x_0 - X) Z_t^{\beta} e^{-\lambda a Z_t}\big] \\
&\quad = \mathbb{E}\bigg[K_h(x_0 - X)\Gamma(\beta+1)\frac{A_0(t|X)}{(\lambda a + A_0(t|X))^{\beta+1}}\bigg] \\
&\quad = \Gamma(\beta+1)\int_{\mathbb{R}^p} K_h(x_0 - y)\frac{A_0(t|y)}{(\lambda a + A_0(t|y))^{\beta+1}} f(y)\,dy \\
&\quad = \Gamma(\beta+1)\int_{S_K} K(z)\frac{A_0(t|x_0 - zh)}{(\lambda a + A_0(t|x_0 - zh))^{\beta+1}} f(x_0 - hz)\,dz \\
&\quad = \Gamma(\beta+1)\frac{A_0(t|x_0)}{(\lambda a + A_0(t|x_0))^{\beta+1}} f(x_0)\big(1 + o(1)\big).
\end{aligned} \tag{A.2}$$

The last transition in the above display follows since $z \in S_K$ and for n large enough, using the continuity of $A_0(t|\cdot)$ and f at $x_0 \in \mathrm{Int}(S_X)$ nonempty, we have boundedness in a neighborhood of x_0, allowing us to use Lebesgue's dominated convergence theorem. Consequently,

$$\mathbb{E}\big[T_n(K, a, t, \lambda, \beta, \gamma|x_0)\big] = a^{\gamma}\Gamma(\beta+1)\frac{A_0(t|x_0)}{(\lambda a + A_0(t|x_0))^{\beta+1}} f(x_0)\big(1 + o(1)\big).$$

Also, similar arguments yield

$$\begin{aligned}
\mathrm{Var}\big(T_n(K, a, t, \lambda, \beta, \gamma|x_0)\big) &= \frac{1}{nh^p}\frac{\|K\|_2^2 A_0(t|x_0) a^{2\gamma}\Gamma(2\beta+1) f(x_0)}{(2\lambda a + A_0(t|x_0))^{2\beta+1}}\big(1 + o(1)\big) \\
&= o(1),
\end{aligned}$$

from which the convergence in probability simply follows.

A.2. Asymptotic covariance matrix of the finite dimensional vector $\mathbb{T}_n$. Our aim in this section is to compute the explicit expression of the elements of the covariance matrix $\Sigma = (\sigma_{j,k})_{1 \le j,k \le m}$ given in (2.6). In this section, we work under the assumptions of Lemma 2.1. According to (A.2), we have

$$\mathbb{E}\big[K_h(x_0 - X) Z_{t_j}^{\beta_j} e^{-\lambda_j A_0(t_j|x_0) Z_{t_j}}\big] = \frac{f(x_0)\Gamma(\beta_j+1)}{[\lambda_j + 1]^{\beta_j+1}[A_0(t_j|x_0)]^{\beta_j}}\big(1 + o(1)\big)$$

for $1 \le j \le m$. In order to compute the cross expectation, we need to derive the conditional distribution function of the pair (Z_{t_j}, Z_{t_k}) given $X = x$. Let $u, v > 0$

$$\begin{aligned}
&\mathbb{P}(Z_{t_j} > u, Z_{t_k} > v | X = x) \\
&\quad = \mathbb{P}\big(Y^{(1)} > \max\big((1-t_j)u, (1-t_k)v\big), Y^{(2)} > \max(t_j u, t_k v) | X = x\big) \\
&\quad = G\big(\max\big((1-t_j)u, (1-t_k)v\big), \max(t_j u, t_k v) | x\big).
\end{aligned}$$

Hence, for $j, k \in \{1, \dots, m\}^2$, we have

$$\begin{aligned}
&\mathbb{E}\big[K_h(x_0 - X) Z_{t_j}^{\beta_j} e^{-\lambda_j A_0(t_j|x_0) Z_{t_j}} K_h(x_0 - X) Z_{t_k}^{\beta_k} e^{-\lambda_k A_0(t_k|x_0) Z_{t_k}}\big] \\
&\qquad = \mathbb{E}\big[K_h^2(x_0 - X) \mathbb{E}\big[Z_{t_j}^{\beta_j} e^{-\lambda_j A_0(t_j|x_0) Z_{t_j}} Z_{t_k}^{\beta_k} e^{-\lambda_k A_0(t_k|x_0) Z_{t_k}} | X\big]\big].
\end{aligned} \tag{A.3}$$

We focus now on the conditional expectation. Using (A.1) and the fact that

$$z^{\beta} e^{-a\lambda z} - \delta_0(\beta) = \int_{\mathbb{R}_+} \mathbb{1}_{\{z>u\}} u^{\beta-1}(\beta - a\lambda u) e^{-a\lambda u}\, du, \tag{A.4}$$

we have

$$\begin{aligned}
&\mathbb{E}\big[Z_{t_j}^{\beta_j} e^{-\lambda_j A_0(t_j|x_0) Z_{t_j}} Z_{t_k}^{\beta_k} e^{-\lambda_k A_0(t_k|x_0) Z_{t_k}} | X\big] \\
&\quad = \mathbb{E}\big[\big(Z_{t_j}^{\beta_j} e^{-\lambda_j A_0(t_j|x_0) Z_{t_j}} - \delta_0(\beta_j)\big)\big(Z_{t_k}^{\beta_k} e^{-\lambda_k A_0(t_k|x_0) Z_{t_k}} - \delta_0(\beta_k)\big) | X\big] \\
&\qquad - \delta_0(\beta_j)\delta_0(\beta_k) \\
&\qquad + \delta_0(\beta_j)\mathbb{E}\big[Z_{t_k}^{\beta_k} e^{-\lambda_k A_0(t_k|x_0) Z_{t_k}} | X\big] + \delta_0(\beta_k)\mathbb{E}\big[Z_{t_j}^{\beta_j} e^{-\lambda_j A_0(t_j|x_0) Z_{t_j}} | X\big] \\
&\quad = \int_{\mathbb{R}_+^2} g_{j,k}(u, v) G_{t_j,t_k}(u, v | X)\, du\, dv - \delta_0(\beta_j)\delta_0(\beta_k) \\
&\qquad + \delta_0(\beta_j) \frac{\Gamma(\beta_k + 1)}{[\lambda_k + 1]^{\beta_k+1}[A_0(t_k|X)]^{\beta_k}} + \delta_0(\beta_k) \frac{\Gamma(\beta_j + 1)}{[\lambda_j + 1]^{\beta_j+1}[A_0(t_j|X)]^{\beta_j}}.
\end{aligned} \tag{A.5}$$

Combining the continuity at x_0 and boundedness of the functions f, $A_0(t|\cdot)$ and $G(u, v|\cdot)$, the expression of $\sigma_{j,k}$ in (2.6) follows.

A.3. Proof of Theorem 2.1. First, remark that to show Theorem 2.1, it is sufficient to look at the weak convergence of the process

$$\begin{aligned}
&\big\{\sqrt{nh^p}\big(T_n\big(K, A_0(t|x_0), t, \lambda, \beta, \gamma | x_0\big) - \mathbb{E}\big[T_n\big(K, A_0(t|x_0), t, \lambda, \beta, \gamma | x_0\big)\big]\big), \\
&\qquad t \in [0, 1]\big\},
\end{aligned} \tag{A.6}$$

since

$$\begin{aligned}
\lim_{n\to\infty} \sup_{t\in[0,1]} \sqrt{nh^p} \bigg| &\mathbb{E}\big[T_n\big(K, A_0(t|x_0), t, \lambda, \beta, \gamma | x_0\big)\big] \\
&- \Gamma(\beta + 1) \frac{[A_0(t|x_0)]^{\gamma-\beta}}{(\lambda + 1)^{\beta+1}} f(x_0) \bigg| = 0.
\end{aligned}$$

Indeed, according to (A.2), we have

$$\begin{aligned}
&\left|\mathbb{E}\big[T_n\big(K, A_0(t|x_0), t, \lambda, \beta, \gamma|x_0\big)\big] - \Gamma(\beta+1)\frac{[A_0(t|x_0)]^{\gamma-\beta}}{(\lambda+1)^{\beta+1}} f(x_0)\right| \\
&\quad\le \Gamma(\beta+1) A_0^{\gamma}(t|x_0) \\
&\qquad \times \int_{S_K} K(y)\left|\frac{A_0(t|x_0-yh)}{(\lambda A_0(t|x_0)+A_0(t|x_0-yh))^{\beta+1}} f(x_0-hy)\right. \\
&\qquad \left. - \frac{A_0^{-\beta}(t|x_0)}{(\lambda+1)^{\beta+1}} f(x_0)\right| dy.
\end{aligned}$$

Now, using Assumptions ($\mathcal{D}$) and ($\mathcal{A}_0$), we deduce that

$$\begin{aligned}
&\left|\frac{A_0(t|x_0-yh)}{(\lambda A_0(t|x_0)+A_0(t|x_0-yh))^{\beta+1}} f(x_0-hy) - \frac{A_0^{-\beta}(t|x_0)}{(\lambda+1)^{\beta+1}} f(x_0)\right| \\
&\quad\le \frac{A_0(t|x_0-yh)}{(\lambda A_0(t|x_0)+A_0(t|x_0-yh))^{\beta+1}}\big|f(x_0-yh)-f(x_0)\big| \\
&\qquad + \left|\frac{A_0(t|x_0-yh)}{(\lambda A_0(t|x_0)+A_0(t|x_0-yh))^{\beta+1}} - \frac{A_0^{-\beta}(t|x_0)}{(\lambda+1)^{\beta+1}}\right| f(x_0) \\
&\quad= O\big(h^{\min(\eta_f,\eta_{A_0})}\big)
\end{aligned}$$

for n large enough such that $h \le 1$, with a bound which is uniform in t.

Then, to show the weak convergence of the stochastic process (A.6), we will use Theorem 19.28 in van der Vaart (1998). To apply this result, we need to introduce some notation. Define the covering number $N(\mathcal{F}, L_2(Q), \tau)$ as the minimal number of $L_2(Q)$-balls of radius τ needed to cover the class of functions $\mathcal{F}$ and the uniform entropy integral as

$$J(\delta, \mathcal{F}, L_2) := \int_0^{\delta} \sqrt{\log \sup_{\mathcal{Q}} N\big(\mathcal{F}, L_2(Q), \tau\|F\|_{Q,2}\big)}\, d\tau,$$

where $\mathcal{Q}$ is the set of all probability measures Q for which $0 < \|F\|_{Q,2}^2 := \int F^2\, dQ < \infty$ and F is an envelope function of the class $\mathcal{F}$.

Let P denote the law of the vector $(Y^{(1)}, Y^{(2)}, X)$ and define the expectation under P, the empirical version and empirical process as follows:

$$Pf := \int f\, dP,$$

$$\mathbb{P}_n f := \frac{1}{n}\sum_{i=1}^{n} f\big(Y_i^{(1)}, Y_i^{(2)}, X_i\big), \qquad \mathbb{G}_n f := \sqrt{n}(\mathbb{P}_n - P) f$$

for any real-valued measurable function f.

For any $\gamma \in \mathbb{R}$ and $(\lambda, \beta) \in (0, \infty) \times \mathbb{R}_+$ or $(\lambda, \beta) = (0, 0)$, we introduce our sequence of classes $\mathcal{F}_n$ as

$$\begin{aligned}\mathcal{F}_n &:= \{(y_1, y_2, z) \to f_{n,t}(y_1, y_2, z), t \in [0, 1]\} \\ &:= \{(y_1, y_2, z) \\ &\qquad \to \sqrt{h^p} K_h(x_0 - z)[A_0(t|x_0)]^{\gamma-\beta}[A_0(t|x_0) Z_t(y_1, y_2)]^{\beta} \\ &\qquad \times e^{-\lambda A_0(t|x_0) Z_t(y_1,y_2)}, t \in [0, 1]\},\end{aligned}$$

where $Z_t(y_1, y_2) := \min(\frac{y_1}{1-t}, \frac{y_2}{t})$. Remark that $Z_t = Z_t(\widetilde{Y}^{(1)}, \widetilde{Y}^{(2)})$. Denote now by F_n an envelope function of the class $\mathcal{F}_n$ and for any $y \in S_X$, define the bivariate function $\rho_y : [0, 1]^2 \to \mathbb{R}_+$ as

$$\rho_y(t, s) := \mathbb{E}[(A_0^{\gamma}(t|x_0) Z_t^{\beta} e^{-\lambda A_0(t|x_0) Z_t} - A_0^{\gamma}(s|x_0) Z_s^{\beta} e^{-\lambda A_0(s|x_0) Z_s})^2 | X = y].$$

Naturally, ρ_y defines a semi-metric on $[0, 1]^2$ and since it is bi-continuous, it makes $[0, 1]$ totally bounded.

Now, according to Theorem 19.28 in van der Vaart (1998), the weak convergence of the stochastic process (A.6) follows from the four following conditions:

$$\sup_{\rho_{x_0}(t,s) \le \delta_n} P(f_{n,t} - f_{n,s})^2 \longrightarrow 0 \qquad \text{for every } \delta_n \searrow 0, \tag{A.7}$$

$$P F_n^2 = O(1), \tag{A.8}$$

$$P F_n^2 \{F_n > \varepsilon \sqrt{n}\} \longrightarrow 0 \qquad \text{for every } \varepsilon > 0, \tag{A.9}$$

$$J(\delta_n, \mathcal{F}_n, L_2) \longrightarrow 0 \qquad \text{for every } \delta_n \searrow 0. \tag{A.10}$$

We start to prove (A.7). By definition, we have

$$\begin{aligned}P(f_{n,t} - f_{n,s})^2 &= \int_{\mathbb{R}^p} h^{-p} K^2\left(\frac{x_0 - u}{h}\right) \rho_u(t, s) f(u)\, du \\ &= \int_{S_K} K^2(u) \rho_{x_0 - hu}(t, s) f(x_0 - hu)\, du \\ &= \|K\|_2^2 f(x_0) \rho_{x_0}(t, s) \\ &\quad + \int_{S_K} K^2(u) f(x_0 - hu)[\rho_{x_0 - hu}(t, s) - \rho_{x_0}(t, s)]\, du \\ &\quad + \rho_{x_0}(t, s) \int_{S_K} K^2(u)[f(x_0 - hu) - f(x_0)]\, du.\end{aligned}$$

By the Assumptions ($\mathcal{D}$), ($\mathcal{K}_1$) and since ρ_{x_0} is bounded, it remains to show that

$$\sup_{\rho_{x_0}(t,s) \le \delta_n} |\rho_{x_0 - hu}(t, s) - \rho_{x_0}(t, s)| \to 0. \tag{A.11}$$

Recall that

$$\begin{aligned}\rho_y(t,s) = {} & \big[A_0(t|x_0)\big]^{2\gamma}\mathbb{E}\big[Z_t^{2\beta}e^{-2\lambda A_0(t|x_0)Z_t}|X=y\big] \\ & + \big[A_0(s|x_0)\big]^{2\gamma}\mathbb{E}\big[Z_s^{2\beta}e^{-2\lambda A_0(s|x_0)Z_s}|X=y\big] \\ & - 2\big[A_0(t|x_0)A_0(s|x_0)\big]^{\gamma}\mathbb{E}\big[Z_t^{\beta}e^{-\lambda A_0(t|x_0)Z_t}Z_s^{\beta}e^{-\lambda A_0(s|x_0)Z_s}|X=y\big].\end{aligned}$$

Such expectations have been computed in (A.1) and (A.5). Using the mean value theorem combined with the boundedness of $A_0(\cdot|\cdot)$ and Assumption $(\mathcal{A}_0)$, we can easily infer that for all $(y,y') \in B_{x_0}(r) \times B_{x_0}(r)$, we have

$$\sup_{(t,s)\in[0,1]^2} \big|\rho_y(t,s) - \rho_{y'}(t,s)\big| \le C\|y-y'\|^{\eta_{A_0}}$$

for some positive constant C. This implies (A.11), and thus (A.7) is established.

Now, we move to the proof of (A.8) and (A.9). Since the function $x \to x^{\beta}e^{-\lambda x}$ is bounded over $\mathbb{R}_+$ by $(\beta/\lambda)^{\beta}e^{-\beta}$ and $A_0(t|x_0) \in [1/2, 1]$, $\mathcal{F}_n$ admits the natural envelope function

$$(y_1,y_2,z) \to F_n(y_1,y_2,z) := \sqrt{h^p}K_h(x_0-z)M, \tag{A.12}$$

where $M := (\frac{\beta}{\lambda})^{\beta}e^{-\beta}\max(1, 2^{\beta-\gamma})$. Consequently,

$$\begin{aligned}PF_n^2 &= M^2\int_{\mathbb{R}^p} h^{-p}K^2\left(\frac{x_0-u}{h}\right)f(u)\,du \\ &= M^2\int_{S_K} K^2(u)f(x_0-hu)\,du = M^2\|K\|_2^2 f(x_0)\big(1+o(1)\big),\end{aligned}$$

$$PF_n^2\{F_n > \varepsilon\sqrt{n}\} = M^2\int_{\{K(u)>M^{-1}\varepsilon\sqrt{nh^p}\}} K^2(u)f(x_0-hu)\,du = 0$$

for all $\varepsilon > 0$ and n sufficiently large, since $nh^p \to \infty$, K satisfies Assumption $(\mathcal{K}_1)$ and f is continuous.

Finally, it remains to prove (A.10). First, we introduce the class of functions $\mathcal{W} := \{(y_1,y_2) \to A_0(t|x_0)Z_t(y_1,y_2), t \in [0,1]\}$ and its subgraph σ_t in $\mathbb{R}_+^2 \times \mathbb{R}$ as

$$\begin{aligned}\sigma_t &:= \big\{(u,v,w) : A_0(t|x_0)Z_t(u,v) > w\big\} \\ &= \left\{(u,v,w) : \frac{A_0(t|x_0)}{1-t}u > w\right\} \cap \left\{(u,v,w) : \frac{A_0(t|x_0)}{t}v > w\right\}.\end{aligned}$$

We can show that $\{\sigma_t : t \in [0,1]\}$ is a *VC*-class of sets. Indeed, if we look more generally, at the collection of sets $\mathcal{C} := \{\{(x,y) : \delta x > y\}, \delta > 0\}$ in $\mathbb{R}_+ \times \mathbb{R}$ and if we define two points (x_1,y_1) and (x_2,y_2) such that, without loss of generality, $\frac{y_1}{x_1} \le \frac{y_2}{x_2}$. Then, for any $\delta > 0$, $\delta x_2 \ge y_2$ implies that $\delta x_1 \ge y_1$. Thus, $\mathcal{C}$ cannot shatter the set $\{(x_1,y_1),(x_2,y_2)\}$ and by consequence it is a *VC*-class of sets. Now, the collection of one set $\mathbb{R}_+$ is naturally a *VC*-class of sets. According to Lemma 2.6.17(vii) in van der Vaart and Wellner (1996), $\mathcal{C} \times \mathbb{R}_+$ is a *VC*-class of sets as well. Invoking Lemma 2.6.17(ii), $\{\sigma_t : t \in [0,1]\}$ belongs to a *VC*-class and as such is *VC*. Define now for all $z \in \mathbb{R}_+$

$$\phi_{\lambda,\beta}(z) := z^{\beta} e^{-\lambda z}.$$

We can easily check that $\phi_{\lambda,\beta}$ is of bounded variation. This implies that $\phi_{\lambda,\beta}$ can be decomposed as the sum of two monotone functions, say $\phi_{\lambda,\beta}^{(1)}$ and $\phi_{\lambda,\beta}^{(2)}$. Thus, according to Lemma 2.6.18(viii) in van der Vaart and Wellner (1996), $\phi_{\lambda,\beta}^{(1)} \circ \mathcal{W}$ and $\phi_{\lambda,\beta}^{(2)} \circ \mathcal{W}$ are *VC*. Now, according to Theorem 2.6.7 in van der Vaart and Wellner (1996), there exists a universal constant C such that for any $j = 1, 2$ and $0 < \tau < 1$

$$\sup_{Q} N\big(\phi_{\lambda,\beta}^{(j)} \circ \mathcal{W}, L_2(Q), \tau \|\mathcal{W}_j\|_{Q,2}\big) \le C V_j (16e)^{V_j} \left(\frac{1}{\tau}\right)^{2(V_j-1)},$$

where V_j is the *VC*-index of $\phi_{\lambda,\beta}^{(j)} \circ \mathcal{W}$ and $\mathcal{W}_j$ its envelope function. Now, consider the sequence of class of functions

$$\mathcal{F}_{n,j} := \big\{z \to \sqrt{h^p} K_h(x_0 - z)\big\} \otimes \phi_{\lambda,\beta}^{(j)} \circ \mathcal{W}$$

for $j = 1, 2$, where $\otimes$ denotes the direct product between the two classes involved. Since we only update the previous sets with one single function and only one ball is needed to cover the class $\{z \to \sqrt{h^p} K_h(x_0 - z)\}$ whatever the measure Q, we have

$$\sup_{Q} N\big(\mathcal{F}_{n,j}, L_2(Q), \tau \|\kappa F_n\|_{Q,2}\big) \le C V_j (16e)^{V_j} \left(\frac{1}{\tau}\right)^{2(V_j-1)},$$

where κ is a suitable constant. Moreover since $\sup_{t\in[0,1]}[A_0(t|x_0)]^{\gamma-\beta} = \max(1, 2^{\beta-\gamma})$, for any $0 < \tau < 1$, the minimal number of balls of radius $\tau \max(1, 2^{\beta-\gamma})$ needed to cover the interval $[0, \max(1, 2^{\beta-\gamma})]$ is $\lceil 1/2\tau \rceil$. Hence

$$\sup_{Q} N\big(\{[A_0(t|x_0)]^{\gamma-\beta}, t \in [0,1]\}, L_2(Q), \tau \max(1, 2^{\beta-\gamma})\big) = \left\lceil \frac{1}{2\tau} \right\rceil \le \frac{3}{2}\left(\frac{1}{\tau}\right)^2.$$

Consequently, we have

$$\begin{aligned} &\sup_{Q} N\big(\{[A_0(t|x_0)]^{\gamma-\beta}, t \in [0,1]\} \otimes \mathcal{F}_{n,j}, L_2(Q), \tau \max(1, 2^{\beta-\gamma}) \|\kappa F_n\|_{Q,2}\big) \\ &\quad \le \frac{3C}{2} V_j (16e)^{V_j} \left(\frac{1}{\tau}\right)^{2V_j}. \end{aligned}$$

Finally, since our class of interest $\mathcal{F}_n$ is included in the class of functions

$$\widetilde{\mathcal{F}}_n := \big\{[A_0(t|x_0)]^{\gamma-\beta}, t \in [0,1]\big\} \otimes \mathcal{F}_{n,1} + \big\{[A_0(t|x_0)]^{\gamma-\beta}, t \in [0,1]\big\} \otimes \mathcal{F}_{n,2}$$

with envelope function $2\max(1, 2^{\beta-\gamma})\kappa F_n$, using Lemma 16 in Nolan and Pollard (1987), we have

$$\begin{aligned}
&\sup_Q N(\mathcal{F}_n, L_2(Q), 2\tau \max(1, 2^{\beta-\gamma})\|\kappa F_n\|_{Q,2}) \\
&\quad \le \sup_Q N(\widetilde{\mathcal{F}}_n, L_2(Q), 2\tau \max(1, 2^{\beta-\gamma})\|\kappa F_n\|_{Q,2}) \\
&\quad \le \frac{9C^2}{4} V_1 V_2 (16e)^{V_1+V_2}\left(\frac{4}{\tau}\right)^{2(V_1+V_2)} \\
&\quad =: L\left(\frac{1}{\tau}\right)^V.
\end{aligned}$$

Thus, (A.10) is established since for any sequence $\delta_n \searrow 0$ and n large enough, we have

$$J(\delta_n, \mathcal{F}_n, L_2) \le \int_0^{\delta_n} \sqrt{\log([2\kappa \max(1, 2^{\beta-\gamma})]^V L) - V\log(\tau)}\, d\tau = o(1).$$

This achieves the proof of Theorem 2.1 since the covariance structure follows from (2.6).

A.4. Proof of Theorem 2.2. To prove this theorem, we will make use of the Cramér–Wold device [see, e.g., Severini (2005), page 337], according to which it is sufficient to show that

$$\Lambda_n := \xi^T \sqrt{nh^p}(\mathbb{T}_n - \mathbb{E}[\mathbb{T}_n]) \rightsquigarrow \mathcal{N}_1(0, \xi^T \Sigma \xi)$$

for all $\xi \in \mathbb{R}^m$. A straightforward rearrangement of the terms leads to

$$\begin{aligned}
\Lambda_n &= \frac{1}{n}\sum_{i=1}^n \sqrt{nh^p}\Bigg\{\sum_{j=1}^m \xi_j [A_0(t_j|x_0)]^{\gamma_j} K_h(x_0 - X_i) Z_{t_j,i}^{\beta_j} e^{-\lambda_j A_0(t_j|x_0) Z_{t_j,i}} \\
&\qquad - \mathbb{E}\Bigg[\sum_{j=1}^m \xi_j [A_0(t_j|x_0)]^{\gamma_j} K_h(x_0 - X_i) Z_{t_j,i}^{\beta_j} e^{-\lambda_j A_0(t_j|x_0) Z_{t_j,i}}\Bigg]\Bigg\} \\
&=: \frac{1}{n}\sum_{i=1}^n W_i.
\end{aligned}$$

Since $W_1, \dots, W_n$ are independent and identically distributed random variables, $\operatorname{Var}(\Lambda_n) = \frac{\operatorname{Var}(W_1)}{n}$ with

$$\operatorname{Var}(W_1) = nh^p \sum_{j=1}^m \sum_{k=1}^m \xi_j \xi_k \mathbb{C}_{j,k},$$

where

$$\mathbb{C}_{j,k} := \mathbb{E}\big[\big(A_0(t_j|x_0)\big)^{\gamma_j}\big(A_0(t_k|x_0)\big)^{\gamma_k} K_h^2(x_0 - X) Z_{t_j}^{\beta_j} \\ \times e^{-\lambda_j A_0(t_j|x_0) Z_{t_j}} Z_{t_k}^{\beta_k} e^{-\lambda_k A_0(t_k|x_0) Z_{t_k}}\big] \\ - \mathbb{E}\big[\big(A_0(t_j|x_0)\big)^{\gamma_j} K_h(x_0 - X) Z_{t_j}^{\beta_j} e^{-\lambda_j A_0(t_j|x_0) Z_{t_j}}\big] \\ \times \mathbb{E}\big[\big(A_0(t_k|x_0)\big)^{\gamma_k} K_h(x_0 - X) Z_{t_k}^{\beta_k} e^{-\lambda_k A_0(t_k|x_0) Z_{t_k}}\big].$$

According to the computations in Appendix A.2, $\operatorname{Var}(\Lambda_n) = \xi^T \Sigma \xi(1 + o(1))$. Hence, to ensure the convergence in distribution of Λ_n to a normal random variable, we have to verify the Lyapounov condition for triangular arrays of random variables [Billingsley (1995), page 362]. In the present context, this simplifies to verifying $\frac{1}{n^2}\mathbb{E}(|W_1|^3) \to 0$. We have

$$\mathbb{E}\big(|W_1|^3\big) \le n^{3/2} h^{3p/2} \Biggl\{ \mathbb{E}\Biggl[\Biggl(\sum_{j=1}^m |\xi_j| \big(A_0(t_j|x_0)\big)^{\gamma_j} K_h(x_0 - X) Z_{t_j}^{\beta_j} e^{-\lambda_j A_0(t_j|x_0) Z_{t_j}}\Biggr)^3\Biggr] \\ + 3\mathbb{E}\Biggl[\Biggl(\sum_{j=1}^m |\xi_j| \big(A_0(t_j|x_0)\big)^{\gamma_j} K_h(x_0 - X) Z_{t_j}^{\beta_j} e^{-\lambda_j A_0(t_j|x_0) Z_{t_j}}\Biggr)^2\Biggr] \\ \times \mathbb{E}\Biggl[\sum_{j=1}^m |\xi_j| \big(A_0(t_j|x_0)\big)^{\gamma_j} K_h(x_0 - X) Z_{t_j}^{\beta_j} e^{-\lambda_j A_0(t_j|x_0) Z_{t_j}}\Biggr] \\ + 4\Biggl[\mathbb{E}\Biggl(\sum_{j=1}^m |\xi_j| \big(A_0(t_j|x_0)\big)^{\gamma_j} K_h(x_0 - X) Z_{t_j}^{\beta_j} e^{-\lambda_j A_0(t_j|x_0) Z_{t_j}}\Biggr)\Biggr]^3\Biggr\}.$$

A similar treatment as for (A.3) yields for all positive integer q

$$\mathbb{E}\Biggl[\Biggl(\sum_{j=1}^m |\xi_j| \big(A_0(t_j|x_0)\big)^{\gamma_j} K_h(x_0 - X) Z_{t_j}^{\beta_j} e^{-\lambda_j A_0(t_j|x_0) Z_{t_j}}\Biggr)^q\Biggr] \\ = \mathbb{E}\Biggl(\mathbb{E}\Biggl[\Biggl(\sum_{j=1}^m |\xi_j| \big(A_0(t_j|x_0)\big)^{\gamma_j} K_h(x_0 - X) Z_{t_j}^{\beta_j} e^{-\lambda_j A_0(t_j|x_0) Z_{t_j}}\Biggr)^q \Bigg| X\Biggr]\Biggr) \\ =: \mathbb{E}\big[K_h^q(x_0 - X) Q(X)\big],$$

where the explicit expression of $Q(X)$ can be obtained similarly as for (A.5). Hence

$$\mathbb{E}\Biggl[\Biggl(\sum_{j=1}^m |\xi_j| \big(A_0(t_j|x_0)\big)^{\gamma_j} K_h(x_0 - X) Z_{t_j}^{\beta_j} e^{-\lambda_j A_0(t_j|x_0) Z_{t_j}}\Biggr)^q\Biggr] \\ = \frac{1}{h^{qp}} \int_{\mathbb{R}^p} K^q\biggl(\frac{x_0 - u}{h}\biggr) Q(u) f(u)\, du$$

$$= (h^p)^{1-q} \int_{S_K} K^q(z) Q(x_0 - zh) f(x_0 - zh)\, dz$$
$$= \mathcal{O}\big((h^p)^{1-q}\big)$$

by continuity and boundedness of the functions. Consequently,

$$\frac{1}{n^2}\mathbb{E}\big(|W_1|^3\big) = \mathcal{O}\big((\sqrt{nh^p})^{-1}\big) = o(1).$$

A.5. The derivatives of $\widehat{\Delta}_{\alpha,x_0,t}$ and their asymptotic properties. Straightforward computations for $a \in [1/2, 1]$, $\alpha > 0$, give

$$\widehat{\Delta}^{(1)}_{\alpha,x_0,t}(a) = \alpha a^{-1}\widehat{\Delta}_{\alpha,x_0,t}(a) + a^{\alpha}(1+\alpha)\frac{1}{n}\sum_{i=1}^{n} K_h(x_0 - X_i) Z_{t,i} e^{-\alpha a Z_{t,i}},$$

$$\widehat{\Delta}^{(2)}_{\alpha,x_0,t}(a) = \alpha a^{-1}\widehat{\Delta}^{(1)}_{\alpha,x_0,t}(a) - \alpha a^{-2}\widehat{\Delta}_{\alpha,x_0,t}(a) + \alpha(\alpha+1)a^{\alpha-1}\frac{1}{n}\sum_{i=1}^{n} K_h(x_0 - X_i)(1 - aZ_{t,i}) Z_{t,i} e^{-\alpha a Z_{t,i}},$$

$$\begin{aligned}\widehat{\Delta}^{(3)}_{\alpha,x_0,t}(a) &= \alpha\big(2a^{-3}\widehat{\Delta}_{\alpha,x_0,t}(a) + a^{-1}\widehat{\Delta}^{(2)}_{\alpha,x_0,t}(a) - 2a^{-2}\widehat{\Delta}^{(1)}_{\alpha,x_0,t}(a)\big) \\ &\quad + (\alpha-1)\alpha(\alpha+1)\frac{a^{\alpha-2}}{n}\sum_{i=1}^{n} K_h(x_0 - X_i)(1 - aZ_{t,i}) Z_{t,i} e^{-\alpha a Z_{t,i}} \\ &\quad - \alpha(\alpha+1)\frac{a^{\alpha-1}}{n}\sum_{i=1}^{n} K_h(x_0 - X_i)\big(\alpha(1 - aZ_{t,i}) + 1\big) Z^2_{t,i} e^{-\alpha a Z_{t,i}}.\end{aligned}$$

The convergence in probability of the two first derivatives of $\widehat{\Delta}_{\alpha,x_0,t}$ is therefore a direct application of Lemma 2.1, which yields as $n \to \infty$

$$\widehat{\Delta}^{(1)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big) \xrightarrow{\mathbb{P}} \ell^{(1)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big) := 0,$$

$$\widehat{\Delta}^{(2)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big) \xrightarrow{\mathbb{P}} \ell^{(2)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big) := \frac{1+\alpha^2}{(1+\alpha)^2}\big[A_0(t|x_0)\big]^{\alpha-2} f(x_0).$$

Now the rate of convergence of $\widehat{\Delta}^{(j)}_{\alpha,x_0,t}(A_0(t|x_0))$, $j \in \{1, 2\}$, to its limit is also useful to study (2.4), and thus to reach our final goal. The aim of the next corollary is to provide such a rate.

COROLLARY A.1. *Under the assumptions of Theorem 2.1, then for any $j \in \{1, 2\}$, the process*

$$\big\{\sqrt{nh^p}\big(\widehat{\Delta}^{(j)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big) - \ell^{(j)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big)\big), t \in [0, 1]\big\}$$

weakly converges in $\mathcal{C}([0,1])$ *towards a tight centered Gaussian process. In particular, we have*

$$\sup_{t\in[0,1]} \big|\widehat{\Delta}^{(j)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big) - \ell^{(j)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big)\big| = o_{\mathbb{P}}(1).$$

PROOF OF COROLLARY A.1. As usual, it is sufficient to show the finite dimensional convergence and the tightness of the process. Using Theorem 2.2, we directly solve the finite dimensional convergence issue. Next, Theorem 2.1 combined with (A.6) implies tightness for any process $t \to \sqrt{nh^p}(T_n - \mathbb{E}[T_n])(K, A_0(t|x_0), t, \lambda, \beta, \gamma|x_0)$ and similarly as in Lemma 1 in Bai and Taqqu (2013), we have tightness for any multivariate process with similar coordinates. Corollary A.1 then follows. □

A.6. Proof of Theorem 2.3. To prove the theorem, we will adjust the arguments used to prove existence and consistency of solutions of the likelihood estimating equation; see, for example, Theorem 3.7 and Theorem 5.1 in Chapter 6 of Lehmann and Casella (1998), to the MDPD framework. Let $\zeta, b > 0$, $C(\cdot|\cdot) : [0,1] \times S_X \to [1/2-\zeta, 1+\zeta]$ and $\forall t \in [0,1]$, $r(t) := |A_0(t|x_0) - C(t|x_0)|$. Define in addition the b-level of r as

$$T_b := \{t \in [0,1], r(t) > b\}.$$

We first show that for any $b > 0$

$$\mathbb{P}\big(\forall t \in T_b, \widehat{\Delta}_{\alpha,x_0,t}\big(A_0(t|x_0)\big) < \widehat{\Delta}_{\alpha,x_0,t}\big(C(t|x_0)\big)\big) \to 1 \tag{A.13}$$

as $n \to \infty$, for any function $C(\cdot|x_0)$ different from but close enough to $A_0(\cdot|x_0)$. By applying a Taylor series expansion, we have

$$\begin{aligned}\widehat{\Delta}_{\alpha,x_0,t}\big(C(t|x_0)\big) - \widehat{\Delta}_{\alpha,x_0,t}\big(A_0(t|x_0)\big) &= \big(C(t|x_0) - A_0(t|x_0)\big)\widehat{\Delta}^{(1)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big)\\ &\quad + \frac{1}{2}\big(C(t|x_0) - A_0(t|x_0)\big)^2\widehat{\Delta}^{(2)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big)\\ &\quad + \frac{1}{6}\big(C(t|x_0) - A_0(t|x_0)\big)^3\widehat{\Delta}^{(3)}_{\alpha,x_0,t}\big(\widetilde{C}(t|x_0)\big),\end{aligned}$$

where $\widetilde{C}(t|x_0)$ is an intermediate value between $C(t|x_0)$ and $A_0(t|x_0)$. According to Appendix A.5, as $n \to \infty$,

$$\sup_{t\in[0,1]} \big|\widehat{\Delta}^{(1)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big)\big| = \sup_{t\in[0,1]} \big|\widehat{\Delta}^{(1)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big) - \ell^{(1)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big)\big| \xrightarrow{\mathbb{P}} 0.$$

This convergence implies, that for all $0 < \varepsilon \le b^2$

$$\begin{aligned}&\mathbb{P}\big(\forall t \in T_b, r(t)\big|\widehat{\Delta}^{(1)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big)\big| \le r^3(t)\big)\\ &\quad \ge \mathbb{P}\Big(\forall t \in T_b, \big|\widehat{\Delta}^{(1)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big)\big| \le r^2(t), \sup_{t\in[0,1]}\big|\widehat{\Delta}^{(1)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big)\big| \le \varepsilon\Big)\\ &\quad = \mathbb{P}\Big(\sup_{t\in[0,1]}\big|\widehat{\Delta}^{(1)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big)\big| \le \varepsilon\Big) \longrightarrow 1\end{aligned}$$

as $n \to \infty$. Now, concerning $\widehat{\Delta}^{(2)}_{\alpha,x_0,t}(A_0(t|x_0))$, we have

$$\sup_{t\in[0,1]} \big|\widehat{\Delta}^{(2)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big) - \ell^{(2)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big)\big| \xrightarrow{\mathbb{P}} 0$$

as $n \to \infty$. Consequently, there exists $\delta_1 > 0$ such that

$$\forall t \in [0, 1], \qquad \frac{r^2(t)}{2}\widehat{\Delta}^{(2)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big) > \delta_1 r^2(t)$$

with probability tending to 1.

Finally, since $x \to x^\lambda e^{-x}$ is bounded $\forall \lambda \geq 1$ on $\mathbb{R}^+$ and by Lemma 2.1

$$T_n(K, a, t, 0, 0, 0|x_0) = \frac{1}{n}\sum_{i=1}^{n} K_h(x_0 - X_i) \xrightarrow{\mathbb{P}} f(x_0)$$

as $n \to \infty$, we have for any $\varepsilon > 0$, $n^{-1}\sum_{i=1}^n K_h(x_0 - X_i) \leq f(x_0) + \varepsilon$ with probability tending to 1. This implies that

$$\text{(A.14)} \qquad \sup_{a\in[1/2-\zeta,1+\zeta],t\in[0,1]} \big|\widehat{\Delta}^{(3)}_{\alpha,x_0,t}(a)\big| =: M < \infty$$

with probability tending to 1. We can therefore conclude that

$$\forall t \in [0, 1], \qquad \frac{r^3(t)}{6}\big|\widehat{\Delta}^{(3)}_{\alpha,x_0,t}\big(\widetilde{C}(t|x_0)\big)\big| \leq \frac{M}{6} r^3(t)$$

with probability tending to 1.

Overall, we have shown that

$$\mathbb{P}\biggl(\forall t \in T_b, \widehat{\Delta}_{\alpha,x_0,t}\big(C(t|x_0)\big) - \widehat{\Delta}_{\alpha,x_0,t}\big(A_0(t|x_0)\big) > \delta_1 r^2(t) - \biggl(1 + \frac{M}{6}\biggr) r^3(t)\biggr) \longrightarrow 1$$

as $n \to \infty$, where the right-hand side of the inequality is positive for $r(t) < \delta_1/(1 + M/6)$. Thus, setting

$$\sup_{t\in[0,1]} r(t) < \delta_1/(1 + M/6),$$

(A.13) follows.

To complete the proof, we adjust the line of argumentation of Theorem 3.7 in Chapter 6 of Lehmann and Casella (1998). Take $0 < \delta < \zeta$ and define the event

$$S_n(\delta) := \big\{\forall t \in [0, 1], \widehat{\Delta}_{\alpha,x_0,t}\big(A_0(t|x_0)\big) < \widehat{\Delta}_{\alpha,x_0,t}\big(A_0(t|x_0) \pm \delta\big)\big\}.$$

For $\upsilon \in S_n(\delta)$, since $\widehat{\Delta}_{\alpha,x_0,t}(a)$ is differentiable with respect to a, there exists $\widetilde{A}_{\alpha,n,\delta}(t|x_0) \in (A_0(t|x_0) - \delta, A_0(t|x_0) + \delta)$ where $\widehat{\Delta}_{\alpha,x_0,t}(a)$ achieves a local minimum, so $\widehat{\Delta}^{(1)}_{\alpha,x_0,t}(\widetilde{A}_{\alpha,n,\delta}(t|x_0)) = 0$.

By (A.13), $\mathbb{P}(S_n(\delta)) \to 1$ for any small enough δ, and hence there exists a sequence $\delta_n \downarrow 0$, such that $\mathbb{P}(S_n(\delta_n)) \to 1$, as $n \to \infty$. Now, let $\widehat{A}_{\alpha,n}(t|x_0) := \widetilde{A}_{\alpha,n,\delta_n}(t|x_0)$ if $\upsilon \in S_n(\delta_n)$ and arbitrary otherwise. Since $\upsilon \in S_n(\delta_n)$ implies $\widehat{\Delta}^{(1)}_{\alpha,x_0,t}(\widehat{A}_{\alpha,n}(t|x_0)) = 0$, we have that

$$\mathbb{P}(\widehat{\Delta}^{(1)}_{\alpha,x_0,t}(\widehat{A}_{\alpha,n}(t|x_0)) = 0) \geq \mathbb{P}(S_n(\delta_n)) \to 1$$

as $n \to \infty$, which establishes the existence part. Note that the measurability of the local minimum can be verified in the same way as it is done in the framework of maximum likelihood estimation [see, e.g., Serfling (1980), page 147].

Concerning now the uniform consistency of the solution sequence, note that for any $\varepsilon > 0$ and n large enough such that $\delta_n \leq \varepsilon$, we have

$$\begin{aligned}\mathbb{P}\Big(\sup_{t\in[0,1]} \big|\widehat{A}_{\alpha,n}(t|x_0) - A_0(t|x_0)\big| \leq \varepsilon\Big) &\geq \mathbb{P}\Big(\sup_{t\in[0,1]} \big|\widehat{A}_{\alpha,n}(t|x_0) - A_0(t|x_0)\big| \leq \delta_n\Big) \\ &\geq \mathbb{P}\big(S_n(\delta_n)\big) \to 1\end{aligned}$$

as $n \to \infty$, whence the uniform consistency of the estimator sequence.

A.7. Proof of Theorem 2.4. The starting point is (2.4). According to Corollary A.1, $\{\sqrt{nh^p}\widehat{\Delta}^{(1)}_{\alpha,x_0,t}(A_0(t|x_0)), t \in [0,1]\}$ weakly converges, as $n \to \infty$, towards a tight centered Gaussian process and

$$\big\{\widehat{\Delta}^{(2)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big), t \in [0,1]\big\} \xrightarrow{\mathbb{P}} \big\{\ell^{(2)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big), t \in [0,1]\big\}.$$

Combining these results with (A.14), we have, as $n \to \infty$,

$$\begin{aligned}&\left\{\left[\widehat{\Delta}^{(2)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big) + \frac{1}{2}\widehat{\Delta}^{(3)}_{\alpha,x_0,t}\big(\widetilde{A}(t|x_0)\big)\big(\widehat{A}_{\alpha,n}(t|x_0) - A_0(t|x_0)\big)\right]^{-1}, t \in [0,1]\right\} \\ &\quad\xrightarrow{\mathbb{P}} \big\{\big[\ell^{(2)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big)\big]^{-1}, t \in [0,1]\big\}.\end{aligned}$$

Concerning the covariance structure, it follows from Theorem 2.2 and the fact that

$$\widehat{\Delta}^{(1)}_{\alpha,x_0,t}\big(A_0(t|x_0)\big) = \upsilon_\alpha^T T_n^{(3)}(t|x_0),$$

where

$$T_n^{(3)}(t|x_0) := \begin{pmatrix} T_n\big(K, A_0(t|x_0), t, 0, 0, \alpha - 1|x_0\big) \\ T_n\big(K, A_0(t|x_0), t, \alpha, 0, \alpha - 1|x_0\big) \\ T_n\big(K, A_0(t|x_0), t, \alpha, 1, \alpha|x_0\big) \end{pmatrix}.$$

A.8. Proof of Lemma 3.1. We use the following decomposition:

$$\begin{aligned}
&F_{n,j}(y|x) - F_j(y|x)\\
&= \frac{1}{\widehat{f}_n(x)}\left\{\frac{1}{n}\sum_{i=1}^n K_c(x - X_i)\mathbb{1}_{\{Y_i^{(j)}\le y\}} - \mathbb{E}\big[K_c(x - X)\mathbb{1}_{\{Y^{(j)}\le y\}}\big]\right\}\\
&\quad - \frac{1}{\widehat{f}_n(x)}\left\{\frac{1}{n}\sum_{i=1}^n K_c(x - X_i)\mathbb{E}[\mathbb{1}_{\{Y_i^{(j)}\le y\}}|X_i] - \mathbb{E}\big[K_c(x - X)\mathbb{1}_{\{Y^{(j)}\le y\}}\big]\right\}\\
&\quad + \frac{1}{\widehat{f}_n(x)}\left\{\frac{1}{n}\sum_{i=1}^n K_c(x - X_i)\big[F_j(y|X_i) - F_j(y|x)\big]\right\}\\
&=: \frac{1}{\widehat{f}_n(x)}\big\{T_1(y|x) - T_2(y|x) + T_3(y|x)\big\},
\end{aligned}$$

where

$$\widehat{f}_n(x) := \frac{1}{n}\sum_{i=1}^n K_c(x - X_i)$$

denotes the kernel density estimator of f.

We start by showing that, for some $q > 1$,

$$\begin{aligned}
&\sup_{(y,x)\in\mathbb{R}\times S_X}\left|n^{-1}\sum_{i=1}^n K_c(x - X_i)\mathbb{1}_{\{Y_i^{(j)}\le y\}} - \mathbb{E}\big[K_c(x - X)\mathbb{1}_{\{Y^{(j)}\le y\}}\big]\right|\\
&\qquad = o_{\mathbb{P}}\left(\sqrt{\frac{|\log c|^q}{nc^p}}\right),
\end{aligned} \tag{A.15}$$

$$\sup_{x\in S_X}\left|n^{-1}\sum_{i=1}^n K_c(x - X_i) - \mathbb{E}\big[K_c(x - X)\big]\right| = o_{\mathbb{P}}\left(\sqrt{\frac{|\log c|^q}{nc^p}}\right), \tag{A.16}$$

$$\begin{aligned}
&\sup_{(y,x)\in\mathbb{R}\times S_X}\left|n^{-1}\sum_{i=1}^n K_c(x - X_i)\mathbb{E}[\mathbb{1}_{\{Y_i^{(j)}\le y\}}|X_i] - \mathbb{E}\big[K_c(x - X)\mathbb{1}_{\{Y^{(j)}\le y\}}\big]\right|\\
&\qquad = o_{\mathbb{P}}\left(\sqrt{\frac{|\log c|^q}{nc^p}}\right).
\end{aligned} \tag{A.17}$$

To this aim, let us introduce the class:

$$\begin{aligned}
\mathcal{G} &:= \left\{(u, v) \to K\left(\frac{x - v}{d}\right)\mathbb{1}_{\{u\le y\}};\, y \in \mathbb{R}, x \in S_X, d > 0\right\}\\
&= \left\{K\left(\frac{x - \cdot}{d}\right); x \in S_X, d > 0\right\} \otimes \{\mathbb{1}_{\{\cdot\le y\}}; y \in \mathbb{R}\}\\
&=: \mathcal{G}_1 \otimes \mathcal{G}_2.
\end{aligned}$$

Under Assumption $(\mathcal{K}_2)$, $\mathcal{G}_1$ is a uniformly bounded *VC*-class of measurable functions [see, e.g., Giné and Guillou (2002)]. Next, since the collection of all cells $\{(-\infty, a], a \in \mathbb{R}\}$ is a *VC*-class of sets, it follows that $\mathcal{G}_2$ is also a uniformly bounded *VC*-class of measurable functions. Now, using the fact that the covering number of the direct product of two *VC*-classes is bounded by the product of the respective covering numbers,

$$\mathcal{G}_n := \left\{(u, v) \to K\left(\frac{x-v}{c}\right)\mathbb{1}_{\{u \le y\}};\, y \in \mathbb{R}, x \in S_X, c = c_n > 0\right\},$$

admits the same bound for the covering number as $\mathcal{G}$, that is,

$$N(\mathcal{G}_n, L_2(Q), \tau\|K\|_\infty) \le C V_{\mathcal{G}}(16e)^{V_{\mathcal{G}}}\left(\frac{1}{\tau}\right)^{2(V_{\mathcal{G}}-1)} =: \left(\frac{A_{\mathcal{G}}}{\tau}\right)^{\nu_{\mathcal{G}}},$$

where C is a universal constant, $\tau \in (0, 1)$ and $V_{\mathcal{G}}$ is the *VC*-index of $\mathcal{G}$ [see Theorem 2.6.7 in van der Vaart and Wellner (1996)]. Now, according to Proposition 2.1 in Giné and Guillou (2001) [see also Theorem 2.1 in Giné and Guillou (2002)] for $\sigma^2 \ge \sup_{g \in \mathcal{G}_n} \operatorname{Var}(g)$, $U \ge \|K\|_\infty$ and $0 < \sigma \le U$, there exists a universal constant B such that

$$\begin{aligned}
&\mathbb{E}\left[\sup_{(y,x)\in\mathbb{R}\times S_X}\left|n^{-1}\sum_{i=1}^n K_c(x-X_i)\mathbb{1}_{\{Y_i^{(j)} \le y\}} - \mathbb{E}\left[K_c(x-X)\mathbb{1}_{\{Y^{(j)} \le y\}}\right]\right|\right] \\
&\quad \le [nc^p]^{-1} B\left[U\nu_{\mathcal{G}}\log\left(\frac{UA_{\mathcal{G}}}{\sigma}\right) + \sqrt{\nu_{\mathcal{G}} n\sigma^2 \log\left(\frac{UA_{\mathcal{G}}}{\sigma}\right)}\right].
\end{aligned}$$

Since

$$\operatorname{Var}\left(K\left(\frac{x-X}{c}\right)\mathbb{1}_{\{Y^{(j)} \le y\}}\right) \le c^p \int_{S_K} K^2(u) f(x-cu)\,du \le c^p \|f\|_\infty \|K\|_2^2,$$

the choices $\sigma^2 = \sigma_n^2 := c^p\|f\|_\infty\|K\|_2^2$ and $U = \|K\|_\infty$ imply that $\sigma_n^2 \le U^2$ for n large enough. This yields (A.15). Similar arguments can be used in order to show (A.16). Also (A.17) can be shown similarly, though with a refinement as used in Portier and Segers (2017), page 23.

As for $\widehat{f}_n(x)$, we use (A.16), and obtain $\widehat{f}_n(x) = \mathbb{E}(\widehat{f}_n(x)) + o_{\mathbb{P}}(\sqrt{|\log c|^q/(nc^p)})$, where the $o_{\mathbb{P}}$ term is uniform in $x \in S_X$. By using the assumptions on K, f and (3.5) we derive, for n sufficiently large, the following uniform lower bound:

$$\begin{aligned}
\mathbb{E}(\widehat{f}_n(x)) &= \int_{\{z \in S_K : x-cz \in S_X\}} K(z) f(x-zc)\,dz \\
&\ge b\int_{\{z \in B_0(\delta) : x-cz \in S_X\}} K(z)\,dz \\
&\ge bm\lambda(\{z \in B_0(\delta) : x-cz \in S_X\}) \\
&\ge bm\delta^p\varepsilon.
\end{aligned}$$

Whence $\{T_1(y|x) - T_2(y|x)\}/\widehat{f}_n(x) = o_{\mathbb{P}}(\sqrt{|\log c|^q/(nc^p)})$, uniformly in $(y, x) \in \mathbb{R} \times S_X$.

Concerning $T_3(y|x)$, we obtain for $x \in S_X$ the following direct bound:

$$\begin{aligned}\frac{|T_3(y|x)|}{\widehat{f}_n(x)} &\leq \frac{1}{\widehat{f}_n(x)}\left\{\frac{1}{n}\sum_{i:\|x-X_i\|\leq c} K_c(x - X_i)\big|F_j(y|X_i) - F_j(y|x)\big|\right\}\\ &\leq M_{F_j} c^{\eta_{F_j}}.\end{aligned}$$

Combining the above results establishes the lemma.

A.9. Proof of Theorem 3.1. Let

$$\mathcal{I}_n := \{g_{\theta,\boldsymbol{\delta},n} : \theta \in \Theta, \boldsymbol{\delta} \in H\},$$

where for $\theta := (t, a) \in \Theta := [0, 1] \times [1/2, 1]$, and $\boldsymbol{\delta} \in H := \{\boldsymbol{\delta} = (\delta_1, \delta_2); \boldsymbol{\delta} : \mathbb{R} \times \mathbb{R} \times S_X \to \mathbb{R}^2\}$,

$$\begin{aligned}g_{\theta,\boldsymbol{\delta},n}(y_1, y_2, u) &:= \sqrt{h^p} K_h(x_0 - u) q_{\theta,\boldsymbol{\delta}}(y_1, y_2, u)\\ &:= \sqrt{h^p} K_h(x_0 - u) a^{\gamma}\big[Z_{\theta,\boldsymbol{\delta}}(y_1, y_2, u)\big]^{\beta} \exp\big(-\lambda a Z_{\theta,\boldsymbol{\delta}}(y_1, y_2, u)\big)\end{aligned}$$

with

$$Z_{\theta,\boldsymbol{\delta}}(y_1, y_2, u) := \min\left(\frac{-\log(|\delta_1(y_1, y_2, u)|)}{1-t}, \frac{-\log(|\delta_2(y_1, y_2, u)|)}{t}\right).$$

For convenience, denote $\boldsymbol{\delta}_n := (F_{n,1}, F_{n,2})$ and $\boldsymbol{\delta}_0 := (F_1, F_2)$. According to Lemma 3.1, $r_n^{-1}|\boldsymbol{\delta}_n - \boldsymbol{\delta}_0|$ converges in probability towards the null function $H_0 := \{0\}$ in H, endowed with the norm $\|\boldsymbol{\delta}\|_H := \|\delta_1\|_\infty + \|\delta_2\|_\infty$ for any $\boldsymbol{\delta} \in H$. In order to apply Theorem 2.3 in van der Vaart and Wellner (2007), we have now to show

ASSERTION 1: $\sup_{\theta\in\Theta} \sqrt{n} P G_n(\theta, b_n) \longrightarrow 0$ for every $b_n \to 0$

and

ASSERTION 2: $\sup_{\theta\in\Theta} |\mathbb{G}_n G_n(\theta, b)| \xrightarrow{\mathbb{P}} 0$ for every $b > 0$,

where $G_n(\theta, b)$ is the minimal envelope function for the class

$$\mathcal{E}_n(\theta, b) := \{g_{\theta,\boldsymbol{\delta}_0+r_n\boldsymbol{\delta},n} - g_{\theta,\boldsymbol{\delta}_0,n} : \boldsymbol{\delta} \in H, \|\boldsymbol{\delta}\|_H \leq b\},$$

that is,

$$\begin{aligned}G_n(\theta, b) &:= \sup_{\|\boldsymbol{\delta}\|_H\leq b} |g_{\theta,\boldsymbol{\delta}_0+r_n\boldsymbol{\delta},n} - g_{\theta,\boldsymbol{\delta}_0,n}|\\ &= \sqrt{h^p} K_h(x_0 - \cdot) \sup_{\|\boldsymbol{\delta}\|_H\leq b} |q_{\theta,\boldsymbol{\delta}_0+r_n\boldsymbol{\delta}} - q_{\theta,\boldsymbol{\delta}_0}|.\end{aligned} \tag{A.18}$$

Now, remark that $\forall(y_1, y_2, u) \in \mathbb{R} \times \mathbb{R} \times S_X$

$$\begin{aligned}&\sup_{\|\boldsymbol{\delta}\|_H\leq b} |q_{\theta,\boldsymbol{\delta}_0+r_n\boldsymbol{\delta}} - q_{\theta,\boldsymbol{\delta}_0}|(y_1, y_2, u)\\ &\quad= \sup_{(\delta_1(y_1,y_2,u),\delta_2(y_1,y_2,u))\in B} |q_{\theta,\boldsymbol{\delta}_0+r_n\boldsymbol{\delta}} - q_{\theta,\boldsymbol{\delta}_0}|(y_1, y_2, u),\end{aligned}$$

where $B := \{(x, y) \in \mathbb{R}^2 : |x| + |y| \leq b\}$. Since B is compact and $\boldsymbol{\delta} \to q_{\theta,\boldsymbol{\delta}}(y_1, y_2, u)$ is continuous, (A.18) reaches its supremum on at least one position $\boldsymbol{\delta}^*_{\theta,b}(y_1, y_2, u) = (\delta^*_{1,\theta,b}(y_1, y_2, u), \delta^*_{2,\theta,b}(y_1, y_2, u))$ in B. Thus, according to Theorem 18.19 in Aliprantis and Border (2006), one can find a measurable function $\boldsymbol{\delta}^*_{\theta,b}$ bounded by b in H such that

$$G_n(\theta, b) = |g_{\theta,\boldsymbol{\delta}_0+r_n\boldsymbol{\delta}^*_{\theta,b},n} - g_{\theta,\boldsymbol{\delta}_0,n}|.$$

PROOF OF ASSERTION 1. For any positive sequence $b_n \to 0$, we have

$$\sqrt{n} P G_n(\theta, b_n) = \sqrt{nh^p} \int_{S_K} K(u) \mathbb{E}\big[|q_{\theta,\boldsymbol{\delta}_0+r_n\boldsymbol{\delta}^*_{\theta,b_n}} - q_{\theta,\boldsymbol{\delta}_0}| \big| X = x_0 - hu\big] f(x_0 - hu)\, du.$$

Note that for any $(\boldsymbol{\delta}, \boldsymbol{\delta}') \in H \times H$, using (A.4)

$$|q_{\theta,\boldsymbol{\delta}} - q_{\theta,\boldsymbol{\delta}'}| \leq a^\gamma \int_0^{+\infty} |\beta - \lambda a s| s^{\beta-1} \times e^{-\lambda a s} \mathbb{1}_{\{s \in [\min(Z_{\theta,\boldsymbol{\delta}}, Z_{\theta,\boldsymbol{\delta}'}), \max(Z_{\theta,\boldsymbol{\delta}}, Z_{\theta,\boldsymbol{\delta}'})]\}}\, ds. \tag{A.19}$$

Consequently,

$$\begin{aligned}
&\mathbb{E}\big[|q_{\theta,\boldsymbol{\delta}_0+r_n\boldsymbol{\delta}^*_{\theta,b_n}} - q_{\theta,\boldsymbol{\delta}_0}| \big| X = x_0 - hu\big] \\
&\quad \leq a^\gamma \int_0^{+\infty} |\beta - \lambda a s| s^{\beta-1} \\
&\qquad \times e^{-\lambda a s} \mathbb{P}\big(s \in \big[\min(Z_{\theta,\boldsymbol{\delta}_0+r_n\boldsymbol{\delta}^*_{\theta,b_n}}, Z_{\theta,\boldsymbol{\delta}_0}), \max(Z_{\theta,\boldsymbol{\delta}_0+r_n\boldsymbol{\delta}^*_{\theta,b_n}}, Z_{\theta,\boldsymbol{\delta}_0})\big] \big| \\
&\qquad\quad X = x_0 - hu\big)\, ds.
\end{aligned}$$

Remark now that

$$\begin{aligned}
&\big\{s \in \big[\min(Z_{\theta,\boldsymbol{\delta}_0+r_n\boldsymbol{\delta}^*_{\theta,b_n}}, Z_{\theta,\boldsymbol{\delta}_0}), \max(Z_{\theta,\boldsymbol{\delta}_0+r_n\boldsymbol{\delta}^*_{\theta,b_n}}, Z_{\theta,\boldsymbol{\delta}_0})\big]\big\} \\
&\quad = \big\{e^{-s} \in \big[\min\big(\max\big(|F_1 + r_n\delta^*_{1,\theta,b_n}|^{\frac{1}{1-t}}, |F_2 + r_n\delta^*_{2,\theta,b_n}|^{\frac{1}{t}}\big), \max\big(F_1^{\frac{1}{1-t}}, F_2^{\frac{1}{t}}\big)\big), \\
&\qquad \max\big(\max\big(|F_1 + r_n\delta^*_{1,\theta,b_n}|^{\frac{1}{1-t}}, |F_2 + r_n\delta^*_{2,\theta,b_n}|^{\frac{1}{t}}\big), \max\big(F_1^{\frac{1}{1-t}}, F_2^{\frac{1}{t}}\big)\big)\big]\big\} \\
&\quad \subset \big\{e^{-s} \in \big[\min\big(|F_1 + r_n\delta^*_{1,\theta,b_n}|^{\frac{1}{1-t}}, F_1^{\frac{1}{1-t}}\big), \max\big(|F_1 + r_n\delta^*_{1,\theta,b_n}|^{\frac{1}{1-t}}, F_1^{\frac{1}{1-t}}\big)\big]\big\} \\
&\qquad \cup \big\{e^{-s} \in \big[\min\big(|F_2 + r_n\delta^*_{2,\theta,b_n}|^{\frac{1}{t}}, F_2^{\frac{1}{t}}\big), \max\big(|F_2 + r_n\delta^*_{2,\theta,b_n}|^{\frac{1}{t}}, F_2^{\frac{1}{t}}\big)\big]\big\} \\
&\quad \subset \big\{e^{-(1-t)s} \in \big[\min\big(|F_1 + r_n\delta^*_{1,\theta,b_n}|, F_1\big), \max\big(|F_1 + r_n\delta^*_{1,\theta,b_n}|, F_1\big)\big]\big\} \\
&\qquad \cup \big\{e^{-ts} \in \big[\min\big(|F_2 + r_n\delta^*_{2,\theta,b_n}|, F_2\big), \max\big(|F_2 + r_n\delta^*_{2,\theta,b_n}|, F_2\big)\big]\big\} \\
&\quad \subset \big\{e^{-(1-t)s} \in [F_1 - r_n b_n, F_1 + r_n b_n]\big\} \cup \big\{e^{-ts} \in [F_2 - r_n b_n, F_2 + r_n b_n]\big\} \\
&\quad =: A_{n,1}(s) \cup A_{n,2}(s).
\end{aligned}$$

Since for any subsets A and B, we have $\mathbb{1}_{\{A\cup B\}} \le \mathbb{1}_{\{A\}} + \mathbb{1}_{\{B\}}$, we can deduce that

$$\begin{aligned}
&\mathbb{P}\big(s \in \big[\min(Z_{\theta,\delta_0+r_n\delta^*_{\theta,b_n}}, Z_{\theta,\delta_0}), \max(Z_{\theta,\delta_0+r_n\delta^*_{\theta,b_n}}, Z_{\theta,\delta_0})\big]|X = x_0 - hu\big)\\
&\quad\le \mathbb{P}\big(A_{n,1}(s)|X = x_0 - hu\big) + \mathbb{P}\big(A_{n,2}(s)|X = x_0 - hu\big)\\
&\quad= \int_0^1 \mathbb{1}_{\{e^{-(1-t)s}\in[v-r_nb_n, v+r_nb_n]\}}\, dv + \int_0^1 \mathbb{1}_{\{e^{-ts}\in[v-r_nb_n, v+r_nb_n]\}}\, dv\\
&\quad\le 2r_nb_n + 2r_nb_n = 4r_nb_n.
\end{aligned} \tag{A.20}$$

This implies that

$$\begin{aligned}
&\sqrt{nh^p}\,\mathbb{E}\big[|q_{\theta,\delta_0+r_n\delta^*_{\theta,b_n}} - q_{\theta,\delta_0}|\big|X = x_0 - hu\big]\\
&\quad\le 4\sqrt{nh^p}\, r_nb_n \sup_{a\in[1/2,1]} \int_0^\infty a^\gamma |\beta - \lambda as| s^{\beta-1} e^{-\lambda as}\, ds.
\end{aligned}$$

This achieves the proof of Assertion 1 since K is bounded, $\sup_{a\in[1/2,1]} \int_0^\infty a^\gamma \times |\beta - \lambda as| s^{\beta-1} e^{-\lambda as}\, ds < +\infty$, $\sqrt{nh^p}\, r_n \to 0$ and $b_n \to 0$.

PROOF OF ASSERTION 2. The idea is to apply Lemma 2.2 in van der Vaart and Wellner (2007). To this aim, first observe that the class $\mathcal{E}_n(\theta, b)$ admits an envelope function E_n of the same form as F_n in (A.12), for some suitable constant $M > 0$. Thus E_n satisfies the conditions (A.8) and (A.9), with F_n replaced by E_n. Consequently, it remains to show the two following convergences

$$\sup_{\theta\in\Theta} PG_n^2(\theta, b) \longrightarrow 0, \tag{A.21}$$

$$J\big(d_n, \{G_n(\theta, b) : \theta \in \Theta\}, L_2\big) \longrightarrow 0 \qquad \text{for all } d_n \searrow 0. \tag{A.22}$$

We start to show (A.21). Since

$$PG_n^2(\theta, b) = \int_{S_K} K^2(u)\mathbb{E}\big(|q_{\theta,\delta_0+r_n\delta^*_{\theta,b}} - q_{\theta,\delta_0}|^2 |X = x_0 - hu\big) f(x_0 - hu)\, du,$$

and (A.19), (A.21) follows from the proof of Assertion 1.

Now, to deal with the uniform entropy integral, we can adjust the lines of proof of Theorem 2.1 by considering the classes of functions defined on $\mathbb{R} \times \mathbb{R} \times S_X$

$$\phi^{(j)}_{\lambda,\beta} \circ \mathcal{W} \circ \Psi, \qquad j = 1, 2,$$

where Ψ is either the function

$$(y_1, y_2, u) \to \big(-\log\big(F_1(y_1|u)\big), -\log\big(F_2(y_2|u)\big)\big)$$

or

$$\begin{aligned}
(y_1, y_2, u) \to \big(&-\log\big(\big|F_1(y_1|u) + r_n\delta^*_{1,\theta,b}(y_1, y_2, u)\big|\big),\\
&-\log\big(\big|F_2(y_2|u) + r_n\delta^*_{2,\theta,b}(y_1, y_2, u)\big|\big)\big),
\end{aligned}$$

which are VC-classes. This allows us to prove that there exist positive constants C and V such that

$$\sup_{Q} N\big(\{G_n(\theta, b) : \theta \in \Theta\}, L_2(Q), \tau \|E_n\|_{Q,2}\big) \le C\left(\frac{1}{\tau}\right)^V,$$

from which (A.22) follows. This achieves the proof of Theorem 3.1.

A.10. Proof of Theorem 3.2. One can check that the proof of Theorems 2.3 and 2.4 are mainly due to the asymptotic properties of $\widehat{\Delta}^{(j)}_{\alpha,x_0,t}$, $j = 1, 2$ and 3. Thus, if we are able to prove that the two key statistics T_n and $\check{T}_n$ are sufficiently close enough, in the sense that

$$\sup_{t\in[0,1],a\in[1/2,1]} \sqrt{nh^p}|\check{T}_n - T_n|(K, a, t, \lambda, \beta, \gamma|x_0) = o_{\mathbb{P}}(1), \tag{A.23}$$

and

$$\sup_{t\in[0,1],a\in[1/2,1]} \sqrt{nh^p}\mathbb{E}\big[|\check{T}_n - T_n|\big](K, a, t, \lambda, \beta, \gamma|x_0) = o(1), \tag{A.24}$$

then we can swap $\widehat{\Delta}^{(j)}_{\alpha,x_0,t}$ by $\check{\Delta}^{(j)}_{\alpha,x_0,t}$, $j = 1, 2$ and 3. According to Theorem 3.1, (A.23) is a direct consequence of (A.24). So it remains to prove (A.24). Note that

$$\begin{aligned}
&\sqrt{nh^p}\mathbb{E}\big[|\check{T}_n - T_n|\big](K, a, t, \lambda, \beta, \gamma|x_0)\\
&\quad = \sqrt{n}\mathbb{E}\Bigg[\Bigg|\frac{1}{n}\sum_{i=1}^{n}\big[\sqrt{h^p}K_h(x_0 - X_i)a^\gamma \check{Z}^\beta_{n,t,i}e^{-\lambda a\check{Z}_{n,t,i}}\\
&\qquad - \sqrt{h^p}K_h(x_0 - X_i)a^\gamma Z^\beta_{t,i}e^{-\lambda a Z_{t,i}}\big]\Bigg|\Bigg]\\
&\quad \le \sqrt{n}\mathbb{E}\big[|g_{\theta,\boldsymbol{\delta}_n,n}(Y^{(1)}, Y^{(2)}, X) - g_{\theta,\boldsymbol{\delta}_0,n}(Y^{(1)}, Y^{(2)}, X)|\big]\\
&\quad \le \sqrt{n}PG_n(\theta, b),
\end{aligned}$$

since $\boldsymbol{\delta}_n \in \boldsymbol{\delta}_0 + r_n\mathcal{B}(0, b)$ where $\mathcal{B}(0, b) := \{\boldsymbol{\delta} : \|\boldsymbol{\delta}\|_H \le b\}$. This implies that

$$\begin{aligned}
&\sup_{t\in[0,1],a\in[1/2,1]} \sqrt{nh^p}\mathbb{E}\big[|\check{T}_n - T_n|\big](K, a, t, \lambda, \beta, \gamma|x_0)\\
&\quad \le \sup_{t\in[0,1],a\in[1/2,1]} \sqrt{n}PG_n(\theta, b) = o(1)
\end{aligned}$$

by Assertion 1 since it is clear from its proof that $b_n \to 0$ can be replaced by any fixed value b in (A.20) without changing the conclusion. This achieves the proof of Theorem 3.2.

Acknowledgements. The authors sincerely thank the Editor, Associate Editor and the referees for their helpful comments and suggestions that led to substantial improvement of the paper.

SUPPLEMENTARY MATERIAL

Supplementary material to "Local robust estimation of the Pickands dependence function" (DOI: 10.1214/17-AOS1640SUPP; .pdf). This document contains additional simulation results.

REFERENCES

ABEGAZ, F., GIJBELS, I. and VERAVERBEKE, N. (2012). Semiparametric estimation of conditional copulas. *J. Multivariate Anal.* **110** 43–73. MR2927509

ALIPRANTIS, C. D. and BORDER, K. C. (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*, 3rd ed. Springer, Berlin. MR2378491

BAI, S. and TAQQU, M. S. (2013). Multivariate limit theorems in the context of long-range dependence. *J. Time Series Anal.* **34** 717–743. MR3127215

BASU, A., HARRIS, I. R., HJORT, N. L. and JONES, M. C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika* **85** 549–559. MR1665873

BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd ed. Wiley, New York. MR1324786

BÜCHER, A., DETTE, H. and VOLGUSHEV, S. (2011). New estimators of the Pickands dependence function and a test for extreme-value dependence. *Ann. Statist.* **39** 1963–2006. MR2893858

CAPÉRAÀ, P., FOUGÈRES, A.-L. and GENEST, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* **84** 567–577. MR1603985

DAOUIA, A., GARDES, L., GIRARD, S. and LEKINA, A. (2011). Kernel estimators of extreme level curves. *TEST* **20** 311–333. MR2834049

DEHEUVELS, P. (1991). On the limiting behavior of the Pickands estimator for bivariate extreme-value distributions. *Statist. Probab. Lett.* **12** 429–439. MR1142097

EINMAHL, U. and MASON, D. M. (1997). Gaussian approximation of local empirical processes indexed by functions. *Probab. Theory Related Fields* **107** 283–311.

ESCOBAR-BACH, M., GOEGEBEUR, Y. and GUILLOU, A. (2018). Supplement to "Local robust estimation of the Pickands dependence function." DOI:10.1214/17-AOS1640SUPP.

FILS-VILLETARD, A., GUILLOU, A. and SEGERS, J. (2008). Projection estimators of Pickands dependence functions. *Canad. J. Statist.* **36** 369–382.

GARDES, L. and GIRARD, S. (2015). Nonparametric estimation of the conditional tail copula. *J. Multivariate Anal.* **137** 1–16.

GIJBELS, I., OMELKA, M. and VERAVERBEKE, N. (2015). Estimation of a copula when a covariate affects only marginal distributions. *Scand. J. Stat.* **42** 1109–1126. MR3426313

GINÉ, E. and GUILLOU, A. (2001). On consistency of kernel density estimators for randomly censored data: Rates holding uniformly over adaptive intervals. *Ann. Inst. Henri Poincaré Probab. Stat.* **37** 503–522.

GINÉ, E. and GUILLOU, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. Henri Poincaré Probab. Stat.* **38** 907–921.

GINÉ, E., KOLTCHINSKII, V. and ZINN, J. (2004). Weighted uniform consistency of kernel density estimators. *Ann. Probab.* **32** 2570–2605. MR2078551

GUDENDORF, G. and SEGERS, J. (2010). Extreme-value copulas. In *Copula Theory and Its Applications* (P. Jaworski, F. Durante, W. Härdle and W. Rychlik, eds.) 127–145. Springer, Berlin.

KOTZ, S. and NADARAJAH, S. (2000). *Extreme Value Distributions—Theory and Applications*. Imperial College Press, London.

LEHMANN, E. L. and CASELLA, G. (1998). *Theory of Point Estimation*, 2nd ed. Springer, New York.

NOLAN, D. and POLLARD, D. (1987). *U*-Processes: Rates of convergence. *Ann. Statist.* **15** 780–799. MR0888439

PICKANDS, J. III (1981). Multivariate extreme value distributions. *Bull. Inst. Internat. Statist.* **49** 859–878, 894–902. MR0820979

PORTIER, F. and SEGERS, J. (2017). On the weak convergence of the empirical conditional copula under a simplifying assumption. Available at arXiv:1511.06544.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

SEVERINI, T. A. (2005). *Elements of Distribution Theory*. *Cambridge Series in Statistical and Probabilistic Mathematics* **17**. Cambridge Univ. Press, Cambridge. MR2168237

SKLAR, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris* **8** 229–231. MR0125600

STUTE, W. (1986). Conditional empirical processes. *Ann. Statist.* **14** 638–647. MR0840519

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York. MR1385671

VAN DER VAART, A. W. and WELLNER, J. A. (2007). Empirical processes indexed by estimated functions. In *Asymptotics: Particles, Processes and Inverse Problems*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **55** 234–252. IMS, Beachwood, OH. MR2459942

M. ESCOBAR-BACH
Y. GOEGEBEUR
DEPARTMENT OF MATHEMATICS AND
COMPUTER SCIENCE
UNIVERSITY OF SOUTHERN DENMARK
CAMPUSVEJ 55
5230 ODENSE M
DENMARK
E-MAIL: escobar@imada.sdu.dk
yuri.goegebeur@imada.sdu.dk

A. GUILLOU
INSTITUT RECHERCHE MATHÉMATIQUE AVANCÉE,
UMR 7501
UNIVERSITÉ DE STRASBOURG ET CNRS
7 RUE RENÉ DESCARTES
67084 STRASBOURG CEDEX
FRANCE
E-MAIL: armelle.guillou@math.unistra.fr