Annals of the Institute of Statistical Mathematics (2023) 75:201–231
https://doi.org/10.1007/s10463-022-00839-1

# Robust estimation of the conditional stable tail dependence function

Yuri Goegebeur[1] · Armelle Guillou[2] · Jing Qin[1]

Received: 18 November 2021 / Revised: 28 March 2022 / Accepted: 24 May 2022 /
Published online: 28 June 2022


**Abstract**
We propose a robust estimator of the stable tail dependence function in the case where random covariates are recorded. Under suitable assumptions, we derive the finite-dimensional weak convergence of the estimator properly normalized. The performance of our estimator in terms of efficiency and robustness is illustrated through a simulation study. Our methodology is applied on a real dataset of sale prices of residential properties.

**Keywords** Empirical processes · Local estimation · Multivariate extreme value statistics · Robustness · Stable tail dependence function

## 1 Introduction

A topic of central interest in multivariate extreme values is to measure the strength of dependence in the extremes. This can be done by using some coefficients of tail dependence or some functions, among them the Pickands dependence function or the stable tail dependence function. In the present paper, we focus on this latter function introduced by Huang (1992), and we estimate it when the random variables of main interest are recorded along with random covariates, related to the target variables. That means that we are in the regression context where our objective is to estimate the stable tail dependence function between the response variables conditional on the covariates. This leads to the concept of conditional stable tail dependence function. Additionally, since in practice some outliers may occur in real datasets with a disturbing effect on the estimates of dependencies, we propose an estimator which is robust

✉ Armelle Guillou
armelle.guillou@math.unistra.fr

1 Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark

2 Institut Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg et CNRS, 7 rue René Descartes, 67084 Strasbourg cedex, France

against observations that are atypical for the extreme dependence structure of the models under consideration. In other words, our contribution in this paper is to introduce a robust estimator of the conditional stable tail dependence function. This topic has been only partially considered in the recent literature, e.g., by Escobar-Bach et al. (2017, 2018b). See also Gardes and Girard (2015), de Carvalho (2016), Castro and de Carvalho (2017), Castro et al. (2018), Mhalla et al. (2019), or Escobar-Bach et al. (2020) and Goegebeur et al. (2020).

Concretely, throughout the paper, we denote by $(Y^{(1)}, Y^{(2)})$ a bivariate random vector recorded along with a random covariate $X \in \mathbb{R}^d$. Let $\|.\|$ denote some norm on $\mathbb{R}^d$ and $\mathcal{B}_x(r)$ the closed ball with respect to $\|.\|$ centered at $x$ and radius $r > 0$. For $j = 1, 2$, we denote by $F_j(.|x)$, the continuous conditional distribution function of $Y^{(j)}$ given $X = x$, by $f_X$ the density function of the covariate $X$ and by $x_0$ a reference position such that $x_0 \in Int(S_X)$, the interior of the support $S_X \subset \mathbb{R}^d$, which is assumed to be non-empty. Our aim in this paper is to estimate the conditional stable tail dependence function defined as

$$\lim_{r\to\infty} r\mathbb{P}\left(1 - F_1\left(Y^{(1)}|X\right) \le r^{-1}y_1 \text{ or } 1 - F_2\left(Y^{(2)}|X\right) \le r^{-1}y_2 \middle| X = x\right) = L(y_1, y_2|x)$$

in a robust way, where we assume that the above limit exists for all $x \in S_X$. By assuming continuous marginal conditional distributions for $Y^{(1)}$ and $Y^{(2)}$, this condition is essentially a condition on the tail behavior of the copula function underlying the joint conditional distribution of $Y^{(1)}$ and $Y^{(2)}$ given $X = x$. As such, the stable tail dependence function contains information about the dependence in extremes.

To reach our goal, we assume that, for all $x \in S_X$, our bivariate random vector $(Y^{(1)}, Y^{(2)})$ satisfies the model

$$\begin{aligned} &\mathbb{P}\left(1 - F_1\left(Y^{(1)}|X\right) \le y_1, 1 - F_2\left(Y^{(2)}|X\right) \le y_2 \middle| X = x\right) \\ &\quad = y_1^{d_1(x)} y_2^{d_2(x)} g(y_1, y_2|x)\left(1 + \delta(y_1, y_2|x)\right), \end{aligned} \tag{1}$$

for any $(y_1, y_2) \in [0, 1]^2 \setminus \{(0, 0)\}$, where $d_1(x), d_2(x)$ are positive and continuous functions such that $d_1(x) + d_2(x) = 1$, $g(y_1, y_2|x)$ is continuous in $(y_1, y_2, x)$ and homogeneous of order 0 in $(y_1, y_2)$, and $\delta(y_1, y_2|x)$ is a function of constant sign in the neighborhood of zero, with $|\delta(., .|x)|$ a bivariate regularly varying function, that is, there exists a function $\xi(., .|x)$ such that

$$\lim_{r\downarrow 0} \frac{|\delta|(ry_1, ry_2|x)}{|\delta|(r, r|x)} = \xi(y_1, y_2|x),$$

for all $(y_1, y_2) \in [0, \infty)^2 \setminus \{(0, 0)\}$, where the convergence is uniform in $(y_1, y_2) \in (0, T]^2$ and $x \in \mathcal{B}_{x_0}(\zeta)$, for any $T > 0$ and $\zeta > 0$. Also, $\xi(y_1, y_2|x)$ is assumed to be continuous in $(y_1, y_2, x)$ and homogeneous of order $\beta(x) > 0$ in $(y_1, y_2)$. Model (1) was introduced in a simpler context without covariates in Dutang et al. (2014) and Escobar-Bach et al. (2017), see also Beirlant et al. (2011), and it has its roots in Ledford and Tawn (1997). Essentially (1) is a further assumption on the tail copula that underlies the joint distribution of $(Y^{(1)}, Y^{(2)})$, conditional on $X = x$. The approach followed in the present paper to the estimation of $L(y_1, y_2|x)$ will be nonparametric and based on local estimation of extreme value models in a neighborhood of the point of interest in the covariate space.

In real data analysis, outliers appear occasionally, and in such contexts robust methods are crucial to avoid poor performance of the usual estimators, like the maximum likelihood estimator. A huge literature exists on outlier detection and robust estimation methods, following the seminal contributions of Huber (1981) and Hampel et al. (1986). In the extreme value context, Dell'Aquila and Embrechts (2006) discussed some methodological aspects related to robust estimation. In particular, they showed how robust methods can improve the quality of data analysis by providing information on the atypical observations, and on the deviation from the structure of the underlying model, while guaranteeing good statistical properties of the resulting estimators computed on the complete dataset. Our aim in this paper is to estimate the conditional stable tail dependence function in a robust way, to prevent possible isolated outliers from completely disturbing the estimate. In the multivariate context, observations can be outlying with respect to the dependency structure, in the sense that they do not follow the pattern set by the majority of the data, and hence they disturb the estimation of the dependence structure. To achieve the robustness, the density power divergence criterion initially proposed by Basu et al. (1998) will be used. It is defined between two density functions $f$ and $h$ as follows

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

Here, $f$ is assumed to be the true (typically unknown) density of the data, whereas $h$ is a parametric model, depending on a vector of parameters which is estimated by minimizing the empirical version of $\Delta_\alpha(f,h)$. This estimator is called the minimum density power divergence (MDPD) estimator. Unlike existing methods such as minimum Hellinger distance estimation, Basu et al.'s (1998) approach avoids the use of nonparametric density estimation and the associated problem of bandwidth selection. This MDPD method only depends on a tuning parameter $\alpha$ which can be viewed as a trade-off between robustness and asymptotic efficiency of the estimators. When $\alpha = 0$, the density power divergence is the Kullback–Leibler divergence (Kullback and Leibler 1951) and the method reduces to the maximum likelihood estimation. When $\alpha = 1$ it corresponds to the mean squared error or $L^2$-divergence. As such the minimum density power divergence represents a whole family of divergences, indexed by the parameter $\alpha \geq 0$.

Thus, we introduce a nonparametric and robust estimator for the conditional stable tail dependence function when the data come from a conditional distribution whose dependence structure converges to that of a conditional extreme value distribution. Compared to related recent literature on estimation of extremal dependence, the differences are as follows. Escobar-Bach et al. (2017) consider robust estimation of the stable tail dependence function though in a context without covariates. Escobar-Bach et al. (2018a) derive a robust estimator for the Pickands dependence function in a context with covariates, where they assume that the underlying conditional distribution has a conditional extreme value copula. In Escobar-Bach et al. (2018b), an estimator for the stable tail dependence function is introduced in a regression context, where it is assumed that the underlying conditional distribution has a dependence function that converges to that of a conditional extreme value distribution, though their estimator is not robust with respect to outlying observations. Goegebeur et al. (2020) discuss a robust estimator for the coefficient of tail dependence in the context of random covariates. In some sense, the present paper can be viewed as a follow-up of the latter paper, although the problem considered in Goegebeur et al. (2020) is simpler than the one considered in the present paper since now the aim is to estimate a dependence function rather than a single parameter. Also, in Goegebeur et al. (2020) a deterministic, i.e., non-random, intermediate threshold is used, while in the present paper we consider the more realistic situation where the intermediate threshold is taken as an intermediate conditional quantile, which complicates the asymptotic analysis considerably.

Our paper is organized as follows. In Sect. 2, we assume that both conditional marginal distribution functions are known, and we propose a robust estimator of the conditional stable tail dependence function for which we establish the finite-dimensional weak convergence. Then, in Sect. 3, we consider the more realistic situation where the conditional marginal distribution functions are unknown. We estimate again in a robust way the conditional stable tail dependence function and we derive similar results as in the previous section, under some additional assumptions. The finite sample performance of our estimator in terms of efficiency and robustness is illustrated in Sect. 4 on a simulation experiment. Finally, in Sect. 5, we apply our methodology to a real dataset of sale prices of residential properties. Some concluding remarks are proposed in Sect. 6. The proofs of some of the main results are postponed to Sect. 7, whereas the others and those of some auxiliary results are given in the online Supplementary Material.

## 2 Estimation of $L(y_1, y_2|x_0)$ in case of known margins

For convenience, assume that the conditional marginal distributions $F_1(.|x)$ and $F_2(.|x)$ are unit Pareto and let $Z_t := \min\{Y^{(1)}, \frac{t}{1-t}Y^{(2)}\}$ for $t := \frac{y_1}{y_1+y_2}$, $0 < t < 1$.

Then, according to model (1), the conditional survival function of $Z_t$ given $X = x$, denoted by $\overline{F}_{Z_t}(.|x)$, is a conditional Pareto-type model of the following form:

$$\overline{F}_{Z_t}(y|x) = G_t(x)y^{-1}\left(1 + \delta_t(y|x)\right), \tag{2}$$

where

$$\begin{aligned} G_t(x) &:= \left(\frac{t}{1-t}\right)^{d_2(x)} g\left(1, \frac{t}{1-t}\Big|x\right), \\ \delta_t(y|x) &:= \delta\left(\frac{1}{y}, \frac{t}{1-t}\frac{1}{y}\Big|x\right). \end{aligned}$$

Note that $\delta_t(.|x)$ is regularly varying at infinity with index $-\beta(x)$, i.e., $\delta_t(uy|x)/\ \delta_t(u|x) \to y^{-\beta(x)}$ as $u \to \infty$ for all $y > 0$. Additionally, we assume in the sequel the following classical condition.

**Assumption** ($\mathcal{D}_t$) For all $x \in S_X$, the conditional survival function of $Z_t$ given by (2) is such that $|\delta_t(.|x)|$ is normalized regularly varying with index $-\beta(x)$, i.e.,

$$\delta_t(y|x) = C_t(x) \exp\left( \int_1^y \frac{\varepsilon_t(u|x)}{u} du \right),$$

with $C_t(x) \in \mathbb{R}$ and $\varepsilon_t(y|x) \to -\beta(x)$ as $y \to \infty$. Moreover, we assume $y \mapsto \varepsilon_t(y|x)$ to be a continuous function.

Under Assumption ($\mathcal{D}_t$), we have that $\delta_t(y|x)$ is differentiable and hence $F_{Z_t}(.|x)$ has a density function. This condition is a restriction of the Karamata representation of regularly varying functions (see, e.g., Corollary 2.1 in Resnick 2007).

We now turn to the estimation of $L(y_1, y_2|x_0)$. The above implies that

$$\begin{aligned}
L(y_1, y_2|x_0) = &\lim_{\Delta \downarrow 0} \frac{1}{\Delta} \Big\{ \mathbb{P}\Big( 1 - F_1(Y^{(1)}|x_0) \le \Delta\, y_1 \Big| X = x_0 \Big) \\
&+ \mathbb{P}\Big( 1 - F_2(Y^{(2)}|x_0) \le \Delta\, y_2 \Big| X = x_0 \Big) \\
&- \mathbb{P}\Big( 1 - F_1(Y^{(1)}|x_0) \le \Delta\, y_1,\ 1 - F_2(Y^{(2)}|x_0) \le \Delta\, y_2 \Big| X = x_0 \Big) \Big\} \\
= &y_1 + y_2 - \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \mathbb{P}\left( Z_{1-t} \ge \frac{1}{\Delta\, y_1} \middle| X = x_0 \right) \\
= &y_1 + y_2 - y_1 G_{1-t}(x_0).
\end{aligned}$$

Thus, estimating the conditional stable tail dependence function requires the estimation of $G_{1-t}(x_0)$. To reach this goal, note that from (2), we deduce that

$$G_{1-t}(x_0) = \frac{k}{n} \frac{U_{Z_{1-t}}(n/k|x_0)}{1 + \delta_{1-t}(U_{Z_{1-t}}(n/k|x_0)|x_0)}, \tag{3}$$

where $U_{Z_{1-t}}(.|x_0)$ is the conditional tail quantile function defined as $U_{Z_{1-t}}(.|x_0) := \inf\{y : F_{Z_{1-t}}(y|x_0) \ge 1 - 1/.\}$, and $k$ is an intermediate sequence such that $k \to \infty$ and $k/n \to 0$. If $\widehat{U}_{Z_{1-t}}(.|x_0)$ and $\widehat{\delta}_{1-t}(.|x_0)$ are estimators for $U_{Z_{1-t}}(.|x_0)$ and $\delta_{1-t}(.|x_0)$, respectively, then by the plug-in method we derive the following estimator for $G_{1-t}(x_0)$:

$$\widehat{G}_{1-t,k}(x_0) := \frac{k}{n} \frac{\widehat{U}_{Z_{1-t}}(n/k|x_0)}{1 + \widehat{\delta}_{1-t}(\widehat{U}_{Z_{1-t}}(n/k|x_0)|x_0)}, \tag{4}$$

which yields a simple estimator for the conditional stable tail dependence function:

$$\widehat{L}_k(y_1, y_2|x_0) := y_1 + y_2 - y_1 \widehat{G}_{1-t,k}(x_0). \tag{5}$$

Recall that we want to propose a robust estimator. To this aim, we will adjust the MDPD criterion to the local estimation context. Remark that $\overline{F}_{Z_{1-t}}$ belongs to the class of distribution functions of Beirlant et al. (2009). Thus, the distribution of the relative excesses $Z_{1-t}/u_n$ given $Z_{1-t} > u_n$ can, for $u_n$ large, be approximated by an extended Pareto distribution (EPD) function given by

$$H(y;\delta_{1-t}(u_n|x_0), \beta(x_0)) := \begin{cases} 1 - y^{-1}\left[1 + \delta_{1-t}(u_n|x_0)\left(1 - y^{-\beta(x_0)}\right)\right]^{-1}, & y > 1, \\ 0, & y \leq 1, \end{cases}$$

where $\delta_{1-t}(u_n|x_0) > \max\{-1, -1/\beta(x_0)\}$. Moreover, using Proposition 2.3 in Beirlant et al. (2009) the approximation error is uniformly $o(\delta_{1-t}(u_n|x_0))$ for $u_n \to \infty$. Using this property, one can estimate $\delta_{1-t}(u_n|x_0)$ with the MDPD approach as follows.

Starting from $(Y_i^{(1)}, Y_i^{(2)}, X_i)$, $i = 1, \dots, n$, independent copies of $(Y^{(1)}, Y^{(2)}, X)$, we obtain $(Z_{1-t,i}, X_i)$, $i = 1, \dots, n$, independent copies of $(Z_{1-t}, X)$, and fit the density function $h$ associated with $H$ and defined for $y > 1$ as

$$\begin{aligned} h\left(y;\delta_{1-t}(u_n|x_0), \beta(x_0)\right) &:= y^{-2}\left[1 + \delta_{1-t}(u_n|x_0)\left(1 - y^{-\beta(x_0)}\right)\right]^{-2} \\ &\quad \times \left[1 + \delta_{1-t}(u_n|x_0)\left(1 - (1 - \beta(x_0))y^{-\beta(x_0)}\right)\right], \end{aligned}$$

locally to the relative excesses $Z_{1-t,i}/\widehat{U}_{Z_{1-t}}(n/k|x_0), i = 1, \dots, n$, given that $Z_{1-t,i} > \widehat{U}_{Z_{1-t}}(n/k|x_0)$. Here, $\widehat{U}_{Z_{1-t}}(n/k|x_0)$ is the natural estimator for $U_{Z_{1-t}}(n/k|x_0)$ defined as $\widehat{U}_{Z_{1-t}}(n/k|x_0) := \inf\{y : \widehat{F}_{Z_{1-t}}(y|x_0) \geq 1 - k/n\}$ where, for $\widehat{F}_{Z_{1-t}}(y|x_0)$, we use the kernel-type estimator

$$\widehat{F}_{Z_{1-t}}(y|x_0) := \frac{\frac{1}{n}\sum_{i=1}^n K_{h_n}(x_0 - X_i)\mathbb{1}_{\{Z_{1-t,i} \leq y\}}}{\frac{1}{n}\sum_{i=1}^n K_{h_n}(x_0 - X_i)}, \tag{6}$$

with $K_{h_n}(.) := K(./h_n)/h_n^d$, $K$ a joint density on $\mathbb{R}^d$ and $h_n$ a positive non-random sequence satisfying $h_n \to 0$ as $n \to \infty$.

This leads to the minimum density power divergence estimator, $\widehat{\delta}_{n,1-t} := \widehat{\delta}_{1-t}(\widehat{U}_{Z_{1-t}}(n/k|x_0)|x_0)$, for $\delta_{1-t}(\widehat{U}_{Z_{1-t}}(n/k|x_0)|x_0)$, and defined as the point minimizing the empirical density power divergence, that is, for $\alpha > 0$

$$\begin{aligned} &\widehat{\Delta}_{\alpha,1-t}(\delta_{1-t}|x_0) \\ &:= \frac{1}{k}\sum_{i=1}^n K_{h_n}(x_0 - X_i)\left\{\int_1^\infty h^{1+\alpha}(y;\delta_{1-t}, \beta)dy - \left(1 + \frac{1}{\alpha}\right)\right. \\ &\quad \left. \times\, h^\alpha\left(\frac{Z_{1-t,i}}{\widehat{U}_{Z_{1-t}}(n/k|x_0)};\delta_{1-t}, \beta\right)\right\}\mathbb{1}_{\{Z_{1-t,i} > \widehat{U}_{Z_{1-t}}(n/k|x_0)\}}. \end{aligned}$$

In our proposed procedure, we only estimate $\delta_{1-t}(\widehat{U}_{Z_{1-t}}(n/k|x_0)|x_0)$ with the density power divergence criterion, while the second order rate parameter will be fixed at a value, denoted $\beta$, which can be either the true value or a mis-specified one. Fixing this second order rate parameter $\beta$ at some value is quite common when fitting second order models like (2) to data, see, e.g., Feuerverger and Hall (1999), Gomes and Martins (2004), Dutang et al. (2014), Escobar-Bach et al. (2017).

To study the asymptotic behavior of our estimator defined in (5), we need to assume some classical conditions due to the regression context, which are nowadays well-known in the conditional extreme value framework.

First, the density $f_X$ and the functions appearing in $\overline{F}_{Z_{1-t}}(y|x)$ need to satisfy the following Hölder conditions.

**Assumption** $(\mathcal{H}_{1-t})$ There exist positive constants $M_{f_X}$, $M_{G_{1-t}}$, $M_{C_{1-t}}$, $M_{\varepsilon_{1-t}}$, $\eta_{f_X}$, $\eta_{G_{1-t}}$, $\eta_{C_{1-t}}$ and $\eta_{\varepsilon_{1-t}}$, such that for all $x, z \in S_X$:

$$\begin{aligned}|f_X(x) - f_X(z)| &\leq M_{f_X} \|x - z\|^{\eta_{f_X}},\\ |G_{1-t}(x) - G_{1-t}(z)| &\leq M_{G_{1-t}} \|x - z\|^{\eta_{G_{1-t}}},\\ |C_{1-t}(x) - C_{1-t}(z)| &\leq M_{C_{1-t}} \|x - z\|^{\eta_{C_{1-t}}},\\ \sup_{y\geq 1} |\varepsilon_{1-t}(y|x) - \varepsilon_{1-t}(y|z)| &\leq M_{\varepsilon_{1-t}} \|x - z\|^{\eta_{\varepsilon_{1-t}}}.\end{aligned}$$

Then, we have also to impose a condition on the kernel function $K$, which is a standard condition in local estimation.

**Assumption** $(\mathcal{K}_1)$ $K$ is a bounded density function on $\mathbb{R}^d$, with support $S_K$ included in the unit ball in $\mathbb{R}^d$.

We have now all the ingredients to state the main result of this section, namely the joint weak convergence of the estimators $\widehat{L}_k(y_{1,j}, y_{2,j}|x_0)$, $j = 1, \dots, J$, after proper normalization. In the sequel, weak convergence is denoted by $\rightsquigarrow$.

**Theorem 1** *Assume* $(\mathcal{D}_{1-t_j})$ *and* $(\mathcal{H}_{1-t_j})$ *for* $j = 1, \dots, J$, $(\mathcal{D}_{0.5})$, $(\mathcal{H}_{0.5})$, $(\mathcal{K}_1)$, $x_0 \in Int(S_X)$ *with* $f_X(x_0) > 0$, *and* $y \mapsto F_{Z_{1-t_j}}(y|x_0)$, $j = 1, \dots, J$, *are strictly increasing. Consider sequences* $k \to \infty$ *and* $h_n \to 0$ *as* $n \to \infty$ *such that* $k/n \to 0$, $kh_n^d \to \infty$, $h_n^{\eta_{\varepsilon_{1-t_1}} \wedge \cdots \wedge \eta_{\varepsilon_{1-t_J}} \wedge \eta_{\varepsilon_{0.5}}} \log \frac{n}{k} \to 0$, $\sqrt{kh_n^d} h_n^{\eta_{f_X} \wedge \eta_{G_{1-t_1}} \wedge \cdots \wedge \eta_{G_{1-t_J}}} \to 0$, $\sqrt{kh_n^d} |\delta_{1-t_j}(U_{Z_{1-t_j}}(\frac{n}{k}|x_0)|x_0)| \to 0$, $j = 1, \dots, J$. *Then, for* $n \to \infty$, *we have,*

$$\sqrt{kh_n^d} \begin{pmatrix} \widehat{L}_k(y_{1,1}, y_{2,1}|x_0) - L(y_{1,1}, y_{2,1}|x_0) \\ \vdots \\ \widehat{L}_k(y_{1,J}, y_{2,J}|x_0) - L(y_{1,J}, y_{2,J}|x_0) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \mathbb{L}_1 \\ \vdots \\ \mathbb{L}_J \end{pmatrix},$$

*where, for* $j = 1, \dots, J$,

$$\begin{aligned}\mathbb{L}_j &:= -y_{1,j} G_{1-t_j}(x_0) \frac{W_{1-t_j}(1)}{f_X(x_0)} + y_{1,j} G_{1-t_j}(x_0) c \left\{ 2\alpha \int_0^1 \left[ \frac{W_{1-t_j}(z)}{z} - W_{1-t_j}(1) \right] z^{2\alpha} \, dz \right.\\ &\quad \left. -(1+\beta)(2\alpha+\beta) \int_0^1 \left[ \frac{W_{1-t_j}(z)}{z} - W_{1-t_j}(1) \right] z^{2\alpha+\beta} \, dz \right\},\\ c &:= \frac{(1+2\alpha)(1+2\alpha+\beta)(1+2\alpha+2\beta)}{\beta^2(1+\beta+4\alpha^2+2\alpha\beta)} \frac{1}{f_X(x_0)},\end{aligned}$$

*and* $W_{1-t_j}(y)$, $j = 1, \dots, J$, *are zero centered Gaussian processes with*

$$
\begin{aligned}
&\mathbb{E}(W_{1-t_j}(y)W_{1-t_{j'}}(\bar{y}))\\
&= \|K\|_2^2 f_X(x_0)\left(\max\left(\frac{G_{1-t_j}(x_0)}{y}, \frac{G_{1-t_{j'}(x_0)}}{\bar{y}}\right)\right)^{-d_1(x_0)}\\
&\quad\times\left(\max\left(\frac{t_j}{1-t_j}\frac{G_{1-t_j}(x_0)}{y}, \frac{t_{j'}}{1-t_{j'}}\frac{G_{1-t_{j'}}(x_0)}{\bar{y}}\right)\right)^{-d_2(x_0)}\\
&\quad\times g\left(\frac{1}{\max\left(\frac{G_{1-t_j}(x_0)}{y}, \frac{G_{1-t_{j'}}(x_0)}{\bar{y}}\right)}, \frac{1}{\max\left(\frac{t_j}{1-t_j}\frac{G_{1-t_j}(x_0)}{y}, \frac{t_{j'}}{1-t_{j'}}\frac{G_{1-t_{j'}}(x_0)}{\bar{y}}\right)}\middle| x_0\right).
\end{aligned}
$$

In practice, the conditional marginal distribution functions $F_1(.|x_0)$ and $F_2(.|x_0)$ are unknown. The aim of the next section is to extend our results to this new framework. The proof of the new theorem will be given in the online Supplementary Material.

## 3 Estimation of $L(y_1, y_2|x_0)$ in case of unknown margins

We consider the general framework where $F_1(.|x)$ and $F_2(.|x)$ are unknown conditional margins. We want to mimic what has been done in the previous section by transforming the margins into approximate unit Pareto distributions. To this aim, we define

$$
\check{Z}_{1-t} := \min\left\{\frac{1}{1-\widehat{F}_{n,1}(Y^{(1)}|X)}, \frac{1-t}{t}\frac{1}{1-\widehat{F}_{n,2}(Y^{(2)}|X)}\right\},
$$

where the estimators $\widehat{F}_{n,j}$, $j = 1, 2$, are defined as

$$
\widehat{F}_{n,j}(y|x_0) := \frac{\frac{1}{n}\sum_{i=1}^n K_{c_n}(x_0 - X_i)\mathbb{1}_{\{Y_i^{(j)}\le y\}}}{\frac{1}{n}\sum_{i=1}^n K_{c_n}(x_0 - X_i)}, \tag{7}
$$

with $c_n$ a positive non-random sequence satisfying $c_n \to 0$ as $n \to \infty$. Note that this estimator has the same form as (6) but with a bandwidth $c_n$ which needs to be different from $h_n$ and the kernel used here is, for simplicity, the same as the one used in the MDPD method.

A similar estimator as the one defined in (5) can be proposed for the robust estimation of the conditional stable tail dependence function in case of unknown margins:

$$
\check{L}_k(y_1, y_2|x_0) := y_1 + y_2 - y_1\check{G}_{1-t,k}(x_0),
$$

where

$$\check{G}_{1-t,k}(x_0) := \frac{k}{n} \frac{\widehat{U}_{\check{Z}_{1-t}}(n/k|x_0)}{1 + \check{\delta}_{1-t}(\widehat{U}_{\check{Z}_{1-t}}(n/k|x_0)|x_0)},$$

with $\check{\delta}_{1-t}(\widehat{U}_{\check{Z}_{1-t}}(n/k|x_0)|x_0)$ the MDPD estimator based on $\check{Z}_{1-t}$.

To establish the joint weak convergence of the estimators $\check{L}_k(y_{1,j}, y_{2,j}|x_0)$, $j = 1, \dots, J$, after proper normalization, we need to impose again some assumptions, in particular a Hölder-type condition on each marginal conditional distribution function $F_j$, $j = 1, 2$.

**Assumption** $(\mathcal{F}_m)$. There exist $M_{F_j} > 0$ and $\eta_{F_j} > 0$ such that $|F_j(y|x) - F_j(y|z)| \leq M_{F_j} \|x - z\|^{\eta_{F_j}}$, for all $y \in \mathbb{R}$, 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 $\mathcal{B}_0(\delta) \subset S_K$ and $K(u) \geq m$ for all $u \in \mathcal{B}_0(\delta)$, and $K$ belongs to the linear span (the set of finite linear combinations) of functions $k \geq 0$ satisfying the following property: the subgraph of $k$, $\{(s, u) : k(s) \geq u\}$, can be represented as a finite number of Boolean operations among sets of the form $\{(s, u) : q(s, u) \geq \varphi(u)\}$, where $q$ is a polynomial on $\mathbb{R}^d \times \mathbb{R}$ and $\varphi$ is an arbitrary real function.

The latter assumption has already been used in Giné and Guillou (2002) or Giné et al. (2004). As stated in these contributions, it is satisfied by $K(x) = \phi\{a(x)\}$, $a$ being a polynomial and $\phi$ a bounded real function of bounded variation (see, e.g., Nolan and Pollard, 1987). This is also the case, e.g., if the graph of $K$ is a pyramid (truncated or not), or if $K = \mathbb{1}_{[-1,1]^d}$, etc.

The main result of the paper is given in the below theorem.

**Theorem 2** *Assume that there exists $b > 0$ such that $f_X(x) \geq b, \forall x \in S_X \subset \mathbb{R}^d$, $f_X$ is bounded, $(\mathcal{D}_{1-t_j})$, $(\mathcal{H}_{1-t_j})$ for $j = 1, \dots, J$, $(\mathcal{D}_{0.5})$, $(\mathcal{H}_{0.5})$, $(\mathcal{K}_2)$, $(\mathcal{F}_m)$ hold, and that $y \mapsto F_{Z_{1-t_j}}(y|x_0)$, $j = 1, \dots, J$, are strictly increasing at $x_0 \in Int(S_X)$ non-empty.*

*Consider sequences $k \to \infty$, $h_n \to 0$ and $c_n \to 0$ as $n \to \infty$, such that $k/n \to 0$, $kh_n^d \to \infty$, $h_n^{\eta_{\varepsilon_{1-t_1}} \wedge \cdots \wedge \eta_{\varepsilon_{1-t_J}} \wedge \eta_{\varepsilon_{0.5}}} \log \frac{n}{k} \to 0$, $\sqrt{kh_n^d} h_n^{\eta_{f_X} \wedge \eta_{G_{1-t_1}} \wedge \cdots \wedge \eta_{G_{1-t_J}}} \to 0$, $\sqrt{kh_n^d} |\delta_{1-t_j}(U_{Z_{1-t_j}}(\frac{n}{k}|x_0)|x_0)| \to 0$, $j = 1, \dots, J$. Assume also that there exists an $\varepsilon > 0$ such that for $n$ sufficiently large*

$$\inf_{x \in S_X} \lambda\big(\{u \in \mathcal{B}_0(1) : x - c_n u \in S_X\}\big) > \varepsilon, \tag{8}$$

*where $\lambda$ denotes the Lebesgue measure, and for some $q > 1$ and $0 < \eta < \min(\eta_{F_1}, \eta_{F_2})$*

$$n \sqrt{\frac{h_n^d}{k}} \max\left(\sqrt{\frac{|\log c_n|^q}{n c_n^d}}, c_n^\eta\right) \longrightarrow 0, \text{ as } n \to \infty. \tag{9}$$

*Then, we have*

$$\sqrt{kh_n^d}\begin{pmatrix} \check{L}_k(y_{1,1}, y_{2,1}|x_0) - L(y_{1,1}, y_{2,1}|x_0) \\ \vdots \\ \check{L}_k(y_{1,J}, y_{2,J}|x_0) - L(y_{1,J}, y_{2,J}|x_0) \end{pmatrix} \rightsquigarrow \begin{pmatrix} \mathbb{L}_1 \\ \vdots \\ \mathbb{L}_J \end{pmatrix},$$

*where* $\mathbb{L}_j$*,* $j = 1, \dots, J$*, are defined as in Theorem* 1.

Note that the conditions (8) and (9) are needed to measure the discrepancy between the conditional distribution function $F_j(y|x)$ and its empirical kernel version $\widehat{F}_{n,j}(y|x), j = 1, 2$, uniformly in $(x, y)$, see, e.g., Lemma 3.1 in Escobar-Bach et al. (2018a).

## 4 Simulation study

The aim of this section is to illustrate the performance of our robust estimator $\check{L}_k(y_1, y_2|x_0)$ with a simulation study. The two following models will be considered.

**Model 1.** The logistic copula model

$$C(u_1, u_2|x) = e^{-[(-\ln u_1)^x + (-\ln u_2)^x]^{1/x}}, u_1, u_2 \in [0, 1], x \geq 2.$$

We take $X \sim U[2, 10]$, and combine this copula model with Fréchet distributions for $Y^{(1)}$ and $Y^{(2)}$:

$$F_j(y) = e^{-y^{-1/\gamma_j}}, y > 0, j = 1, 2.$$

We set $\gamma_1 = 0.25$ and $\gamma_2 = 0.5$. This model corresponds to $L(y_1, y_2|x) = (y_1^x + y_2^x)^{1/x}$.

**Model 2.** The conditional distribution of $(Y^{(1)}, Y^{(2)})$ given $X = x$ is that of

$$(|Z_1|^{\gamma_1(x)}, |Z_2|^{\gamma_2(x)}),$$

where $(Z_1, Z_2)$ follow a bivariate standard Cauchy distribution with density function

$$f(z_1, z_2) = \frac{1}{2\pi}(1 + z_1^2 + z_2^2)^{-3/2}, \quad (z_1, z_2) \in \mathbb{R}^2.$$

We take $X \sim U[0, 1]$, and set

$$\gamma_1(x) = 0.25 + 0.125 \sin(2\pi x),$$
$$\gamma_2(x) = 0.1 + 0.1x.$$

This model corresponds to $L(y_1, y_2|x) = \sqrt{y_1^2 + y_2^2}$.

Contamination will be introduced by adding observations that follow a different dependency structure, namely contamination according to the following mixture model

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

where $\varepsilon$ denotes the fraction of contamination, $F$ is the distribution function of Model 1 or Model 2 described above, and $F_c$ is the contamination distribution function. Given $X = x$, the distribution function $F_c$ used with Model 1 is

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

which corresponds to a contamination on the axes, whereas for Model 2, where the dependence is not very strong, we propose to use a diagonal contamination to highlight the effect of contamination. Concretely, that means a distribution function $F_c$ of the following form

$$F_c(y_1, y_2|x) = e^{-\{\min(y_1, y_2)\}^{-1}}, y_1, y_2 > 0,$$

corresponding to the distribution function of completely dependent unit Fréchet random variables. We want to estimate the extreme dependence structure of $F(y_1, y_2|x)$ in presence of contamination coming from $F_c(y_1, y_2|x)$.

Note that the logistic and Cauchy models have already been considered in Escobar-Bach et al. (2017) with a similar model (1) as ours, but in a framework without covariates. This model, naturally extended to the regression context in the present paper, is also satisfied for these two conditional models, with $\beta(x_0) = 1$ for Model 1 and $\beta(x_0) = 2$ for Model 2. Additionally, we can also check that the Hölder-type conditions $(\mathcal{H}_{1-t})$ are satisfied. Concerning the conditional marginal distribution functions in the two models, they are standard heavy-tailed distributions (see, e.g., Beirlant et al., 2009), and satisfy our Assumption $(\mathcal{F}_m)$.

To compute our estimates $\check{L}_k(y_1, y_2|x_0)$, first, we have to transform the margins into approximate unit Pareto distributions using the kernel-type empirical distribution functions given in (7). To this aim, we need to choose a kernel $K$ and to select the bandwidths $c_n$ for each of the margins. Since the kernel has almost no impact on the results, we use in (7) and also in our MDPD procedure, the same biquadratic function

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

which satisfies our Assumption $(\mathcal{K}_2)$. Concerning the bandwidth $c_n$, a cross-validation criterion, already used in an extreme value context by Daouia et al. (2011), is performed, where

$$c_{n,j} := \arg\min_{\widetilde{c}_j \in \mathcal{C}} \sum_{i=1}^{n}\sum_{k=1}^{n}\left[\mathbb{1}_{\left\{Y_i^{(j)} \leq Y_k^{(j)}\right\}} - \widetilde{F}_{n,-i,j}(Y_k^{(j)}|X_i)\right]^2, \ j = 1, 2,$$

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

The bandwidth $h_n$ is taken as $h_n = \min(c_{n,1}, c_{n,2})\left(\frac{k}{n}\right)^{1/d} |\log[\min(c_{n,1}, c_{n,2})]|^{-\xi}$, where $\xi d > q$, in order to satisfy the condition:

$$n\sqrt{\frac{h_n^d}{k}}\sqrt{\frac{|\log c_n|^q}{nc_n^d}} \to 0,$$

coming from (9) in our main theorem.

In the minimization of the empirical density power divergence, we fix $\beta$ at the value 1, i.e., the true value for Model 1 and a mis-specified value for Model 2. The parameter $k$ which determines the threshold $\widehat{U}_{\check{Z}_{1-t}}(n/k|x_0)$ is selected by an automated procedure based on minimizing the standard deviation of the estimates $\check{L}_k(y_1, y_2|x_0)$ computed in a moving window over the range for $k$, see, e.g., Goegebeur et al. (2019).

In all the settings, $\mathcal{C} = R_X \times \{0.05, 0.075, \dots, 0.3\}$, where $R_X$ is the range of the covariate $X$, and $\xi = 1.1$. Figures 1, 2, 3 illustrate the boxplots of the estimates $\check{L}_k(y_1, y_2|x_0)$ based on 500 samples of size $n = 1\,000$ for $(y_1, y_2) \in \{(l/10, 1 - l/10), l = 1, \dots, 9\}$ and for three values of the covariate: $x_0 = 3$, 5 and 9, in case of Model 1. The columns of the figures represent the two fractions of contamination: $\varepsilon = 0\%$ (left) and $\varepsilon = 10\%$ (right), and the rows the three values of $\alpha$, namely, from the top to the bottom, $\alpha = 0.1, 0.5$ and 1. Figures 4, 5 and 6 are constructed similarly for Model 2 and the three covariate values: $x_0 = 0.2$, 0.5 and 0.8. Each time, the true function $L(y_1, y_2|x_0)$ is computed at the same positions $\{(l/10, 1 - l/10), l = 1, \dots, 9\}$ and connected with a blue line.

Based on these simulations, we can draw the following conclusions:

- Overall, our robust estimator performs quite well, but of course, the results depend on the model, the covariate position and the fraction of contamination. In Model 1, $L(y_1, y_2|x_0)$ depends on the covariate, but the marginal distributions do not. On the contrary, for Model 2, $L(y_1, y_2|x_0)$ does not depend on the value of $x_0$ but the marginal distributions do;
- For all models, when $\varepsilon = 0$, the best results are obtained when $\alpha = 0.1$. This result was expected since this value is close to 0, the value which leads to the maximum likelihood estimator, which is efficient (but not robust). On the contrary, in case of contamination, increasing $\alpha$ is crucial to get more robustness, the central box remaining closer to the true value for large values of $\alpha$ compared to $\alpha = 0.1$;
- For Model 1, the contamination on the axes pulls slightly the estimates up, whereas, on the contrary, for Model 2, the diagonal contamination pulls the estimates a bit down, as expected. The estimation results are good for all covariate positions but exhibit more variability at $(y_1, y_2)$ close to (1/2, 1/2).
- We have also considered data with 20% contamination but for such a high percentage of contamination the estimation procedure did not perform well anymore.

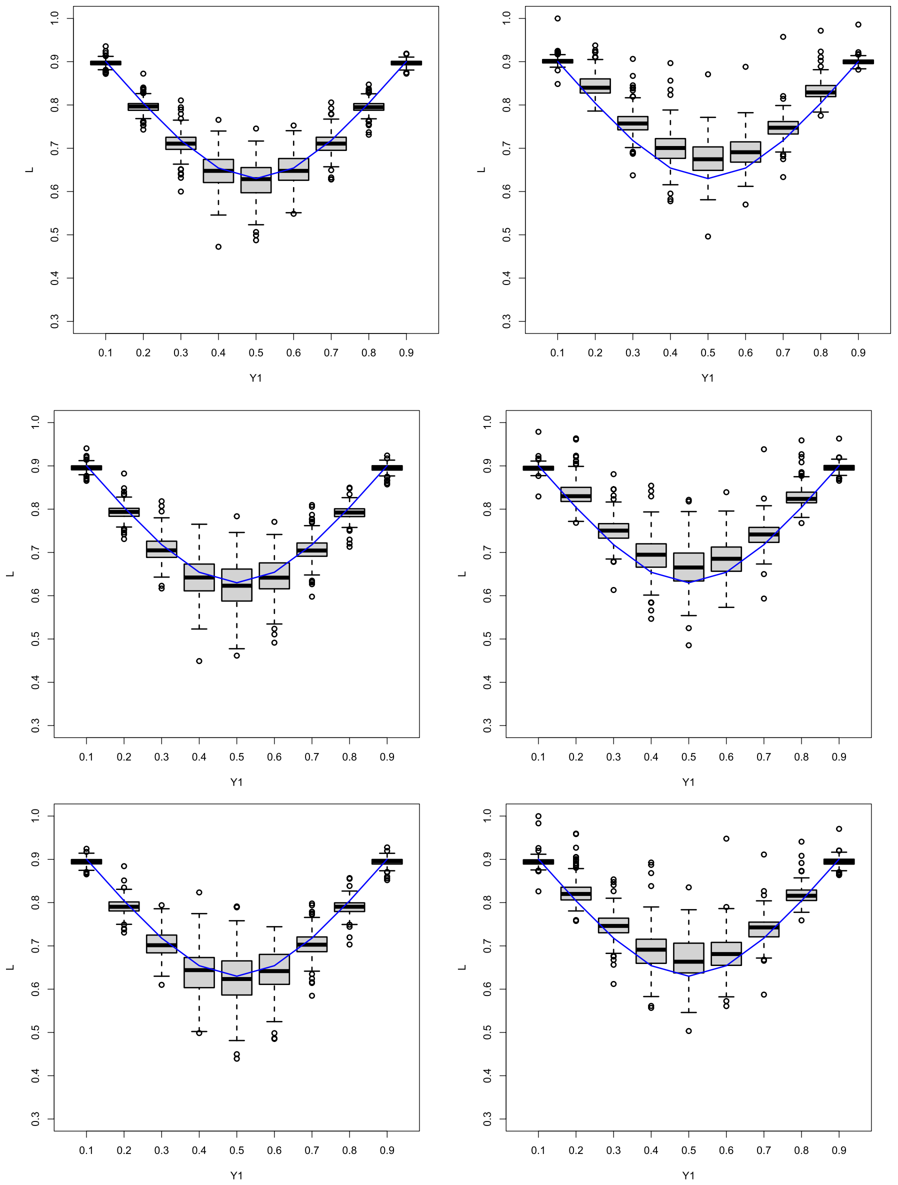



**Fig. 1** Logistic model, $x_0 = 3$: no contamination (left) and 10% axis contamination (right), and $\alpha = 0.1$ (first row), $\alpha = 0.5$ (second row) and $\alpha = 1$ (third row)

## 5 Real data analysis

In this section, we illustrate the robust estimator for $L(y_1, y_2|x_0)$ on the Ames housing dataset (De Cock 2011), which is publicly available at https://www.kaggle.com/c/house-prices-advanced-regression-techniques.

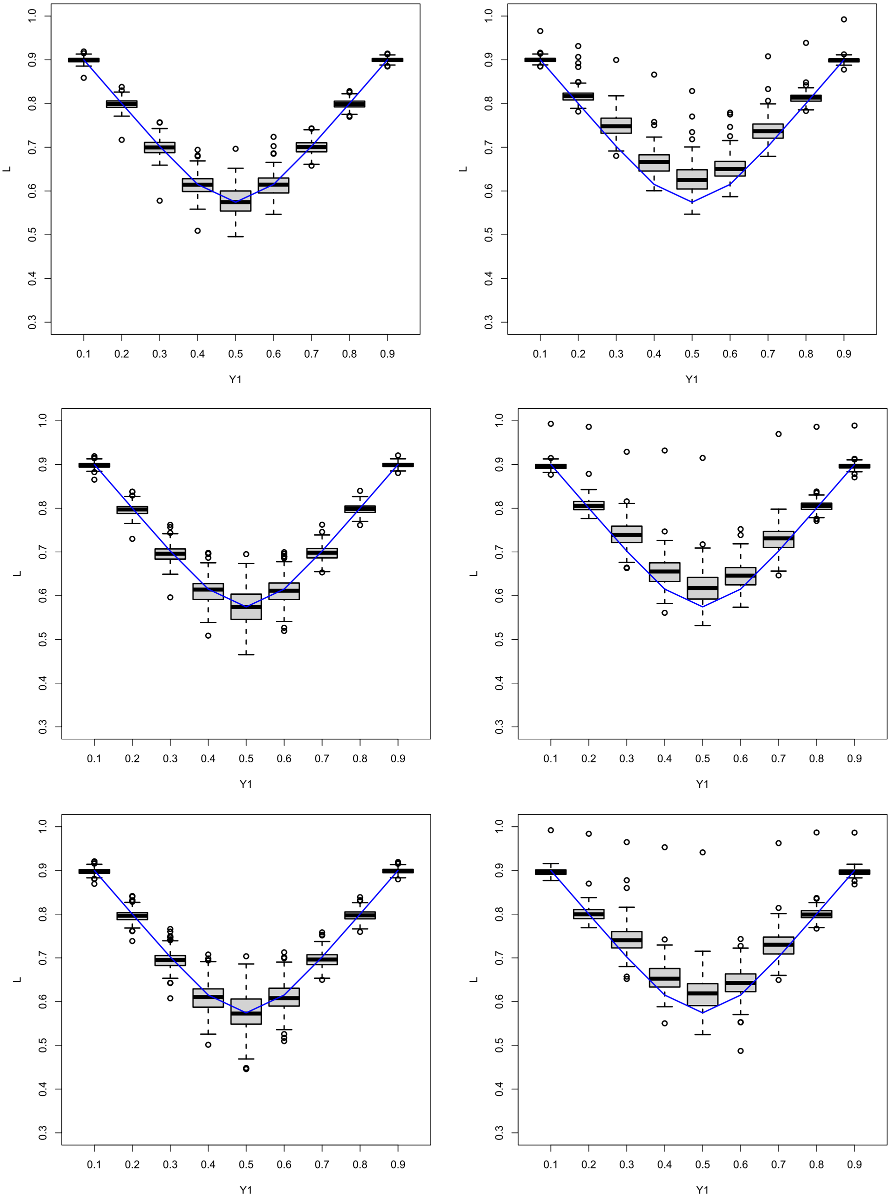



**Fig. 2** Logistic model, $x_0 = 5$: no contamination (left) and 10% axis contamination (right), and $\alpha = 0.1$ (first row), $\alpha = 0.5$ (second row) and $\alpha = 1$ (third row)

This dataset contains information on the sale of individual residential property in Ames, Iowa, from 2006 to 2010. The dataset has $n = 2930$ observations on a large number of variables (23 nominal, 23 ordinal, 14 discrete and 20 continuous) involved in assessing home values. We estimate the conditional stable tail

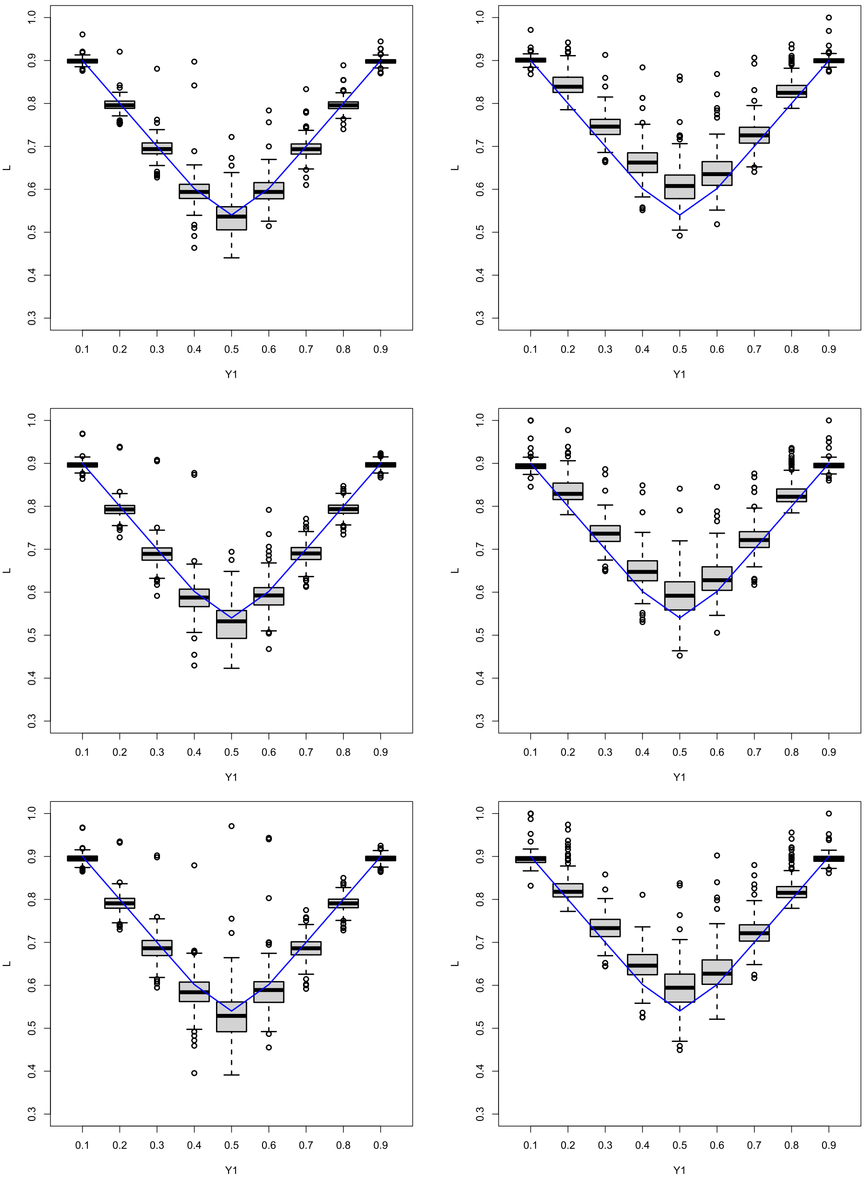



**Fig. 3** Logistic model, $x_0 = 9$: no contamination (left) and 10% axis contamination (right), and $\alpha = 0.1$ (first row), $\alpha = 0.5$ (second row) and $\alpha = 1$ (third row)

dependence function of the variables sale price ($Y^{(1)}$) and above grade living area in square feet ($Y^{(2)}$) conditional on the original construction year of the property ($X$). When estimating a residential property's market value, living area is an important

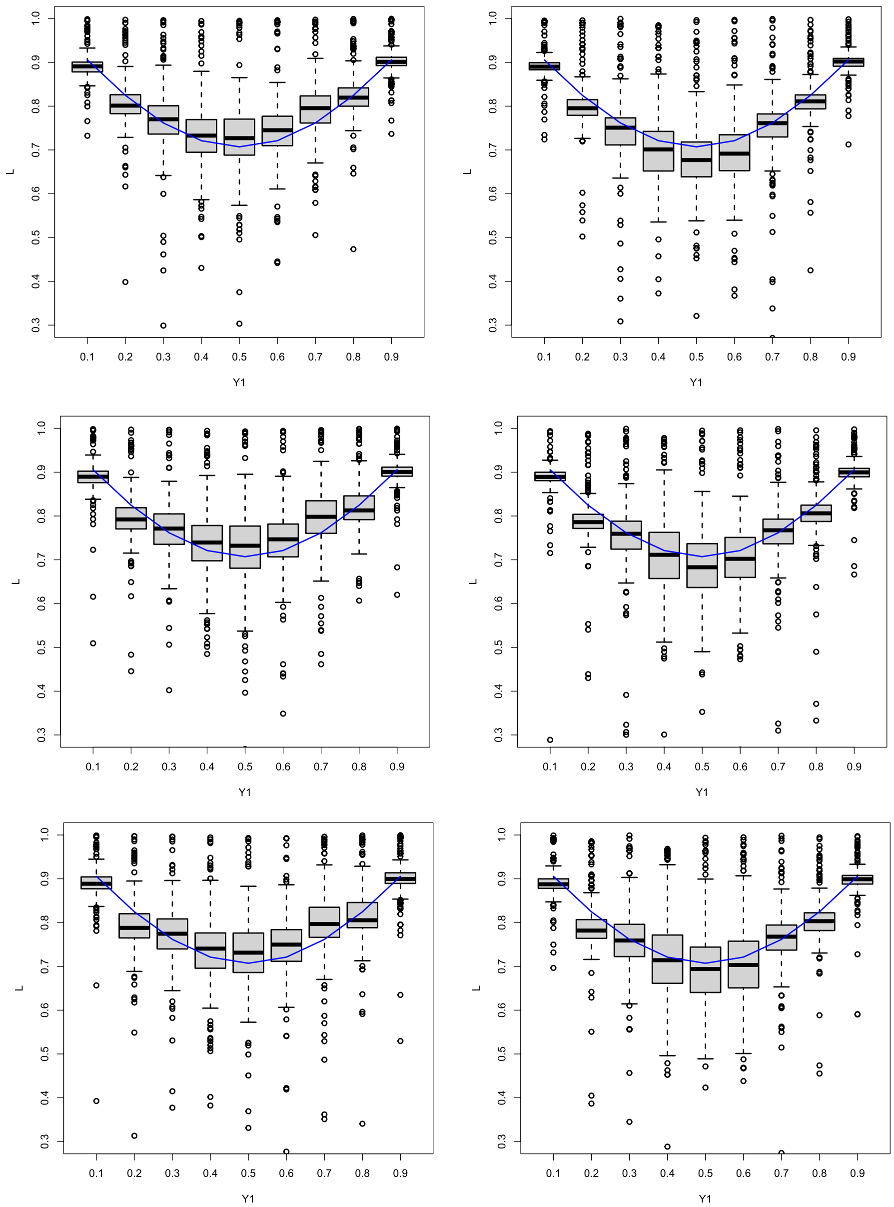



**Fig. 4** Cauchy model, $x_0 = 0.2$: no contamination (left) and 10% diagonal contamination (right), and $\alpha = 0.1$ (first row), $\alpha = 0.5$ (second row) and $\alpha = 1$ (third row)

element to consider, since a bigger property will positively impact its valuation. Indeed, many buyers look at the sales price divided by the square footage of a property, which is a usual indicator of the value of a property. This thus motivates the study of the measure of dependence between the sale price of a residential property

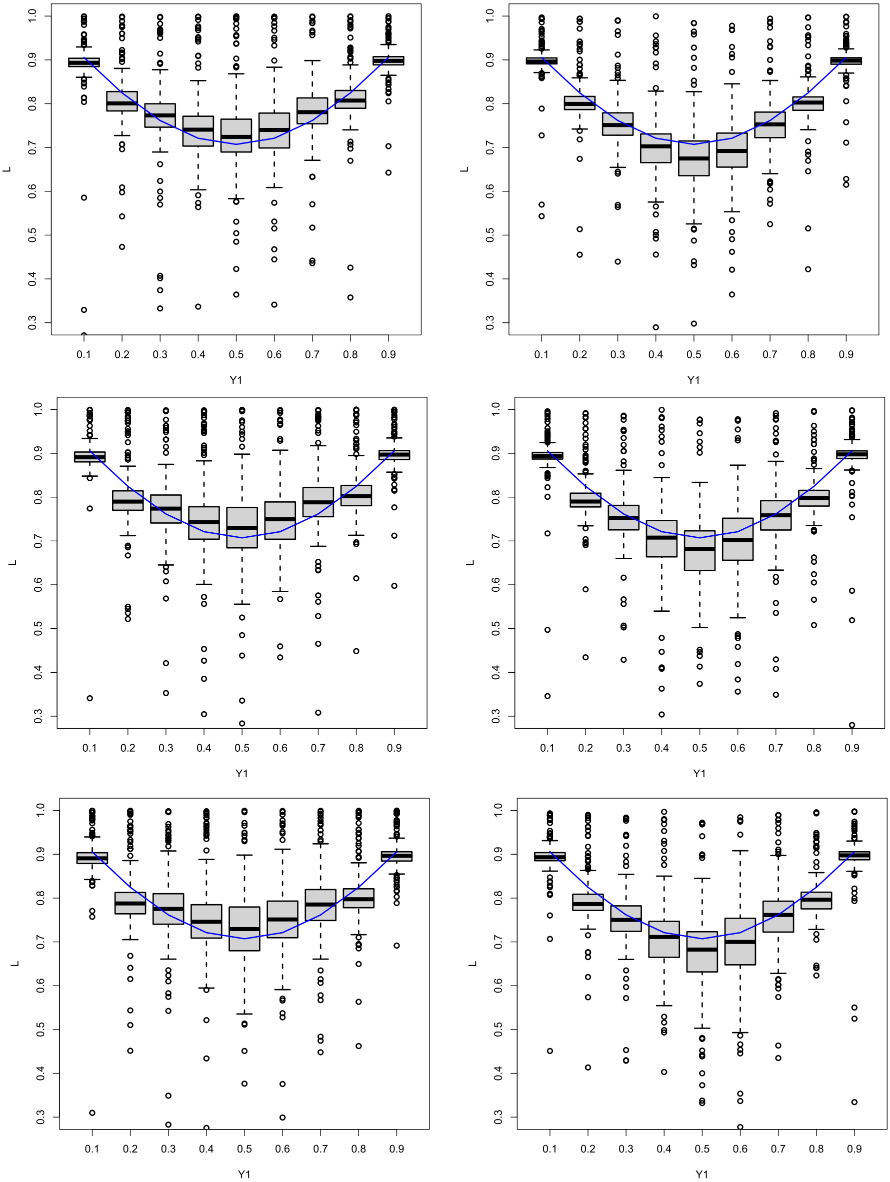



**Fig. 5** Cauchy model, $x_0 = 0.5$: no contamination (left) and 10% diagonal contamination (right), and $\alpha = 0.1$ (first row), $\alpha = 0.5$ (second row) and $\alpha = 1$ (third row)

and the above grade living area in square feet. In Fig. 7, we show the scatterplot of sale price versus above grade living area. The scatterplot shows overall a positive association between the two variables, though there are also some observations that are atypical for the dependence structure and hence may disturb the estimation of the extremal dependence. One knows that properties that are newer often appraise

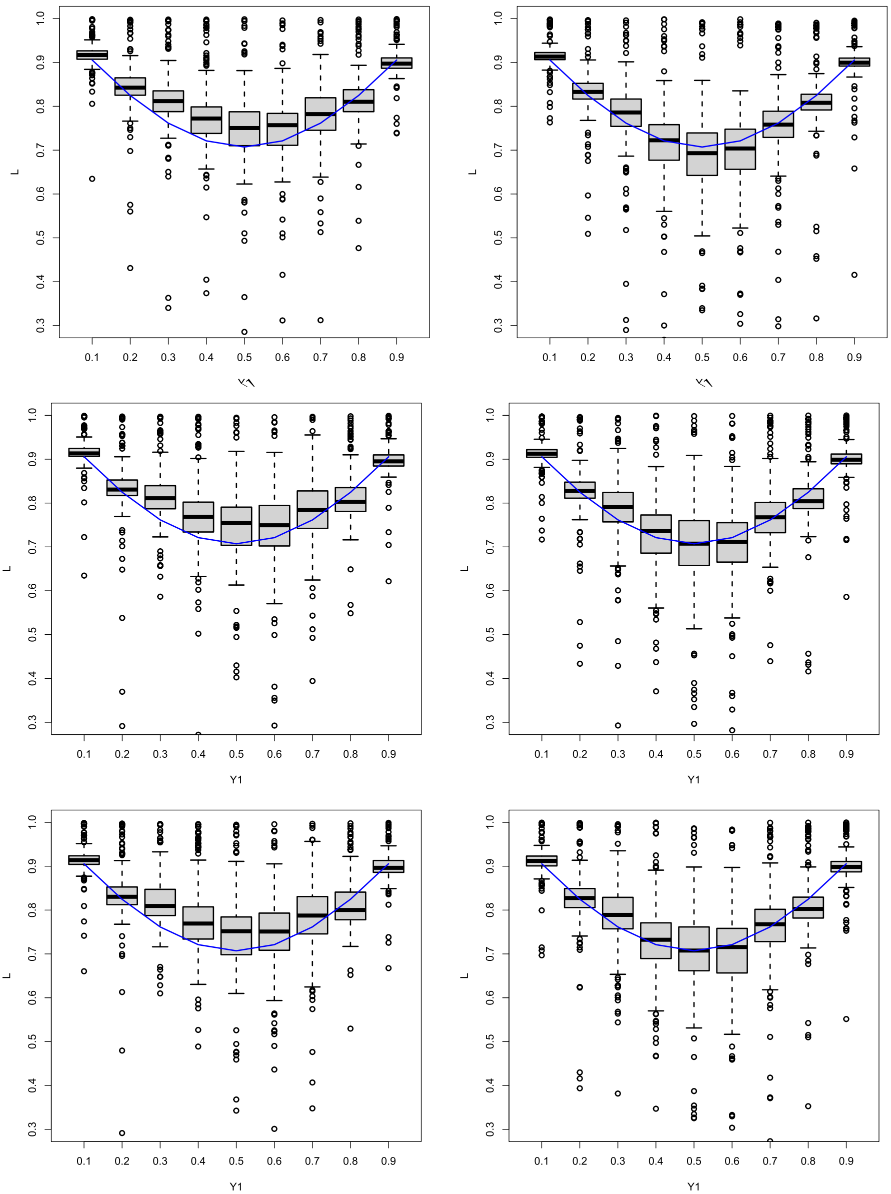
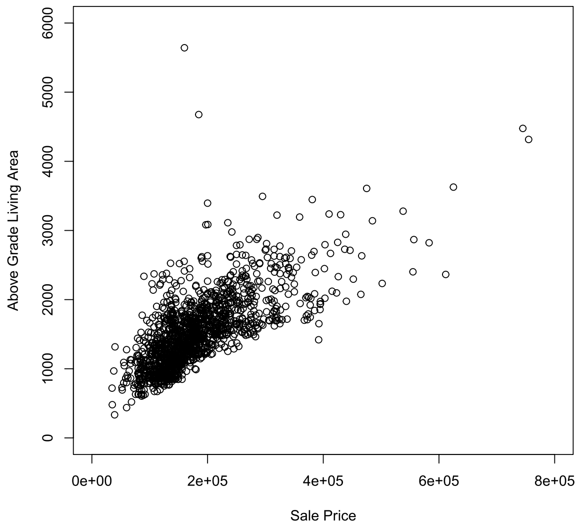



**Fig. 6** Cauchy model, $x_0 = 0.8$: no contamination (left) and 10% diagonal contamination (right), and $\alpha = 0.1$ (first row), $\alpha = 0.5$ (second row) and $\alpha = 1$ (third row)

at a higher value. Indeed, the fact that some parts of the property, like the plumbing, the electrical installations, and the roof are newer can generate savings for a buyer. For example, if a roof has a 20-year warranty, then that is money an owner will save over the next two decades, compared to an older home that may need a roof replaced in just a few years. This is illustrated in Fig. 9 where the sale prices tend

**Fig. 7** Ames housing dataset: scatterplot of sale price versus above grade living area




to be larger in case of an original construction year in 2004 (right panel) compared to 1946 (left panel) for a given living area. Thus, the original construction year is an important covariate which should be taken into account when estimating the measure of dependence between $Y^{(1)}$ and $Y^{(2)}$.

We estimate $L(y_1, y_2|x_0)$ in a robust way with the proposed local minimum density power divergence method. The estimation is implemented with the biquadratic kernel function and the same cross-validation criterion for $c_1$ and $c_2$ as described in the simulation section. Also the bandwidth $h_n$ is here determined by cross-validation. In Fig. 8, we show the estimates $\check{L}_k(y_1, y_2|x_0)$ with $(y_1, y_2) \in (l/20, 1 - l/20), l = 1, \ldots, 19$, for $\alpha = 0.1$ (blue), 0.5 (black) and 1 (green), for the years of original construction 1946 (left) and 2004 (right). For the

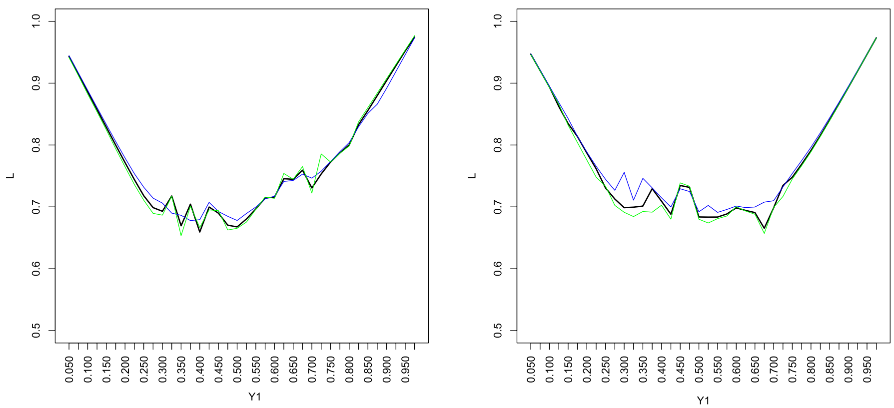



**Fig. 8** Ames housing dataset: estimates of $L(y_1, y_2|x_0)$ with $x_0 = 1946$ (left) and $x_0 = 2004$ (right) for $\alpha = 0.1$ (blue), 0.5 (black) and 1 (green)

construction year 1946, we see that the non-robust estimates, corresponding to $\alpha = 0.1$, are in line with the robust estimates obtained with $\alpha = 0.5$ and $\alpha = 1$, which indicates that the data used for the estimation did not contain disturbing observations. For the construction year 2004, the non-robust estimates with $\alpha = 0.1$ are somehow different from those obtained with $\alpha = 0.5$ and $\alpha = 1$, where the latter two are similar, which indicates potential outliers in the data used for the estimation. This is confirmed by the scatterplots of the data used in the local estimation, given in Fig. 9. Indeed, for construction year 2004, there are two observations with an above grade living area greater than 4000 square feet (which is very high) for a corresponding sale price lower than 200 000 dollars (which is not credible). These two observations are clearly outliers, and since they are far away from the main cloud, that are atypical for the dependence structure, while the scatterplot of the data used for estimation at $x_0 = 1946$ does not indicate outliers. For $x_0 = 2004$, these two outlying observations were removed and the estimates for $L(y_1, y_2|x_0)$ were calculated again. The result of this is shown in Fig. 10, where we see that the estimates obtained with the three values of $\alpha$ are now closer together, as expected.

## 6 Concluding remarks

In this paper, we introduced a robust nonparametric estimator for the stable tail dependence function when next to the variables of main interest, $Y^{(1)}$ and $Y^{(2)}$, there is also a random covariate, $X$. The work proposed here provides a series of interesting open questions which will lead to further investigations, among them:

- Outlier detection. To reach the goal of robustness, we adapted the idea of MDPD estimation to our context. As illustrated in the original Basu et al. (1998) paper, in the density power divergence criterion, the estimating equations consist generally of likelihood score functions with a relative-to-the model down-weighting for outlying observations. Thus, if an observation is unusual relative to the pro-

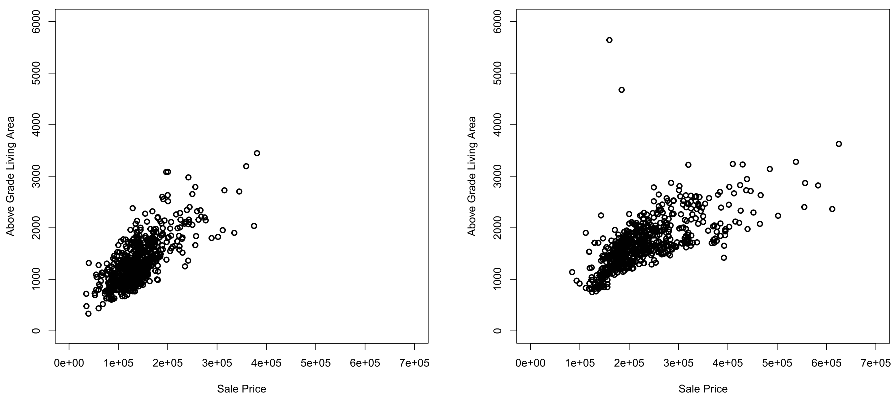



**Fig. 9** Ames housing dataset: scatterplot of data used for local estimation at $x_0 = 1946$ (left) and $x_0 = 2004$ (right)

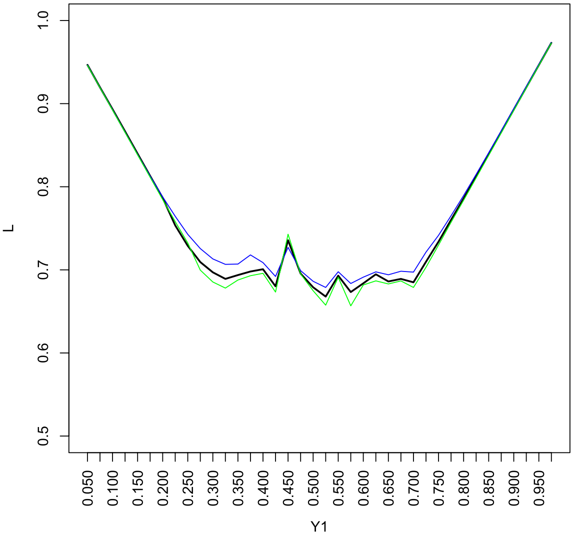



**Fig. 10** Ames housing dataset: estimates of $L(y_1, y_2|x_0)$ with $x_0 = 2004$ for $\alpha = 0.1$ (blue), 0.5 (black) and 1 (green) after removal of the two outlying observations

posed model then its contribution to the estimating equations gets less weight and as such its influence on the estimation results becomes dampened. However, although the methodology avoids that possible isolated outliers can completely disturb the estimation, it does not allow to identify which observations are the outliers in the dataset. In other words, the takeaway message of our paper is to compute our estimator for several values of $\alpha$, among them some small $\alpha$ like $\alpha = 0.1$. If there is almost no difference in the estimates, that means that there are no outliers in the dataset and in that case maximum likelihood method should be used since it is efficient. On the contrary, if there are differences, that means that we have to take care because of the presence of outliers. If we want to know which observation is an outlier, a heuristic approach would be to draw the scatterplot of the data and to try to visualize the observations which seem to be far away from the main cloud. Then, we could remove them one after one, and compute again our estimator for different values of $\alpha$. If the estimates become this time close to each other, that means that the observations removed were indeed outliers. This is the strategy used in our real dataset for the year 2004. As an alternative to this heuristic approach, we could investigate in future research outlier detection on the basis of the empirical influence function of a robust estimator for tail dependence (like the one introduced in the present paper). This was pursued in Hubert et al. (2013) in the context of identifying influential observations for the Hill estimator in univariate extreme value statistics.

- Change-points. In some applications, we are faced with events that can cause structural changes in the underlying model. The statistical analysis for detecting such changes is referred to as change-point analysis. It has been recently considered in the multivariate extreme value framework, see, e.g., de Carvalho et al. (2020) or Drees (2022). Since the traditional methods for identifying change-points can struggle with the presence of outliers, robust methods should be developed in that context, based on, e.g., the MDPD method used in this paper. A starting point for this new topic of research might be the recent paper of Song (2021) to be adapted to the context of extreme values.

- Parametric models. In the present paper, we have applied the MDPD estimation method locally in order to obtain a nonparametric estimator for tail dependence. The MDPD method can also be used for fitting completely parametric extreme value models to data. In the nonparametric approach, one lets 'the data speak for themselves' and as such we get a preview of the extreme dependence structure. This nonparametric estimate could also be useful to evaluate the fit of parametric models.
- Other types of divergences. In the paper, we used the density power divergence method of Basu et al. (1998) to obtain a robust estimate. The basic idea of the density power divergence is to introduce a density power weight in the estimation procedure. This idea is also at the basis of other robust methods like those based on the $\gamma$ divergence (Fujisawa and Eguchi 2008) and the $\beta$ divergence (Minami and Eguchi 2002). The development of estimation procedures for extreme value problems based on the latter types of divergences, and a comparison of their performance with that of estimators based on the density power divergence is a topic of future research.

## Appendix

The minimization of the empirical density power divergence $\widehat{\Delta}_{\alpha,1-t}(\delta_{1-t}|x_0)$ is based on its derivative. Direct computations show that all the terms appearing in this derivative have the following form

$$S_{n,1-t}(s|x_0) := \frac{1}{k}\sum_{i=1}^{n} K_{h_n}(x_0 - X_i)\left(\frac{Z_{1-t,i}}{\widehat{U}_{Z_{1-t}}(n/k|x_0)}\right)^s \mathbb{1}_{\{Z_{1-t,i} > \widehat{U}_{Z_{1-t}}(n/k|x_0)\}}$$

for $s < 0$.

Assuming $F_{Z_{1-t}}(y|x_0)$ is strictly increasing in $y$, we can rewrite this main statistic as follows:

$$
\begin{aligned}
&S_{n,1-t}(s|x_0)\\
&= \frac{1}{k}\sum_{i=1}^{n} K_{h_n}(x_0 - X_i)\left\{1 + \int_{\widehat{U}_{Z_{1-t}}(n/k|x_0)}^{Z_{1-t,i}} \frac{s\,u^{s-1}}{\widehat{U}^s_{Z_{1-t}}(n/k|x_0)}du\right\}\mathbb{1}_{\{Z_{1-t,i}>\widehat{U}_{Z_{1-t}}(n/k|x_0)\}}\\
&= \frac{1}{k}\sum_{i=1}^{n} K_{h_n}(x_0 - X_i)\mathbb{1}_{\{Z_{1-t,i}>\widehat{U}_{Z_{1-t}}(n/k|x_0)\}}\\
&\quad + \frac{1}{k}\sum_{i=1}^{n} K_{h_n}(x_0 - X_i)\left\{\int_{\widehat{U}_{Z_{1-t}}(n/k|x_0)}^{Z_{1-t,i}} \frac{s\,u^{s-1}}{\widehat{U}^s_{Z_{1-t}}(n/k|x_0)}du\right\}\mathbb{1}_{\{Z_{1-t,i}>\widehat{U}_{Z_{1-t}}(n/k|x_0)\}}\\
&= \frac{1}{k}\sum_{i=1}^{n} K_{h_n}(x_0 - X_i)\mathbb{1}_{\{Z_{1-t,i}>\widehat{U}_{Z_{1-t}}(n/k|x_0)\}}\\
&\quad + \int_{\widehat{U}_{Z_{1-t}}(n/k|x_0)}^{\infty} \frac{1}{k}\sum_{i=1}^{n} K_{h_n}(x_0 - X_i)\frac{s\,u^{s-1}}{\widehat{U}^s_{Z_{1-t}}(n/k|x_0)}\mathbb{1}_{\{u<Z_{1-t,i}\}}du\\
&= \frac{1}{k}\sum_{i=1}^{n} K_{h_n}(x_0 - X_i)\mathbb{1}_{\{Z_{1-t,i}>\widehat{U}_{Z_{1-t}}(n/k|x_0)\}}\\
&\quad + \int_{\widehat{U}_{Z_{1-t}}(n/k|x_0)}^{\infty} \frac{1}{k}\sum_{i=1}^{n} K_{h_n}(x_0 - X_i)\frac{s\,u^{s-1}}{\widehat{U}^s_{Z_{1-t}}(n/k|x_0)}\mathbb{1}_{\{\overline{F}_{Z_{1-t}}(Z_{1-t,i}|x_0)<\frac{k}{n}\frac{n}{k}\overline{F}_{Z_{1-t}}(u|x_0)\}}du\\
&= \frac{1}{k}\sum_{i=1}^{n} K_{h_n}(x_0 - X_i)\mathbb{1}_{\{Z_{1-t,i}>\widehat{U}_{Z_{1-t}}(n/k|x_0)\}}\\
&\quad + \int_0^1 \frac{1}{k}\sum_{i=1}^{n} K_{h_n}(x_0 - X_i)sz^{-1-s}\mathbb{1}_{\{\overline{F}_{Z_{1-t}}(Z_{1-t,i}|x_0)<\frac{k}{n}\frac{n}{k}\overline{F}_{Z_{1-t}}(z^{-1}\widehat{U}_{Z_{1-t}}(n/k|x_0)|x_0)\}}dz\\
&= T_{n,1-t}\big(s_{n,1-t}(1|x_0)|x_0\big) + \int_0^1 T_{n,1-t}\big(s_{n,1-t}(z|x_0)|x_0\big)\,s\,z^{-1-s}\,dz,
\end{aligned}
\tag{10}
$$

where

$$
\begin{aligned}
T_{n,1-t}(y|x_0) &:= \frac{1}{k}\sum_{i=1}^{n} K_{h_n}(x_0 - X_i)\mathbb{1}_{\{\overline{F}_{Z_{1-t}}(Z_{1-t,i}|x_0)<\frac{k}{n}y\}}, y \in (0, T],\\
s_{n,1-t}(z|x_0) &:= \frac{n}{k}\overline{F}_{Z_{1-t}}\left(z^{-1}\widehat{U}_{Z_{1-t}}(n/k|x_0)\middle|x_0\right).
\end{aligned}
$$

Thus, we start this appendix with some auxiliary results allowing us to study the statistic $T_{n,1-t}(y|x_0)$ and subsequently in Section 7.2 we establish the weak convergence of $S_{n,1-t}(s|x_0)$. Finally, in Sect. 7.3, Theorem 1 will be established. The proof of Theorem 2 from Sect. 3 is deferred to the online Supplementary Material.

### Auxiliary results in case of known margins

First, we establish the joint weak convergence of processes $W_{n,1-t_j} := \{\sqrt{kh_n^d}[T_{n,1-t_j}(y|x_0) - yf_X(x_0)]; y \in (0,T]\}$, $j = 1, \ldots, J$.

**Lemma 1** *Assume* $(\mathcal{D}_{1-t_j})$ *and* $(\mathcal{H}_{1-t_j})$ *for* $j = 1, \ldots, J$, $(\mathcal{D}_{0.5})$, $(\mathcal{H}_{0.5})$, $(\mathcal{K}_1)$, $x_0 \in Int(S_X)$ *with* $f_X(x_0) > 0$, *and* $y \mapsto F_{Z_{1-t_j}}(y|x_0)$, $j = 1, \ldots, J$, *are strictly increasing. Consider sequences* $k \to \infty$ *and* $h_n \to 0$ *as* $n \to \infty$ *such that* $k/n \to 0$, $kh_n^d \to \infty$, $h_n^{\eta_{\varepsilon_{1-t_1}} \wedge \cdots \wedge \eta_{\varepsilon_{1-t_J}} \wedge \eta_{\varepsilon_{0.5}}} \log \frac{n}{k} \to 0$, $\sqrt{kh_n^d} h_n^{\eta_{f_X} \wedge \eta_{G_{1-t_1}} \wedge \cdots \wedge \eta_{G_{1-t_J}}} \to 0$, *and for* $j = 1, \ldots, J$, $\sqrt{kh_n^d} |\delta_{1-t_j}(U_{Z_{1-t_j}}(\frac{n}{k}|x_0)|x_0)| h_n^{\eta_{C_{1-t_j}}} \to 0$ *and* $\sqrt{kh_n^d} |\delta_{1-t_j}(U_{Z_{1-t_j}}(\frac{n}{k}|x_0)|x_0)| h_n^{\eta_{\varepsilon_{1-t_j}}} \log \frac{n}{k} \to 0$. *Then, for* $n \to \infty$, *we have*

$$(W_{n,1-t_1}, \ldots, W_{n,1-t_J}) \rightsquigarrow (W_{1-t_1}, \ldots, W_{1-t_J}),$$

*in* $\ell^J((0,T])$, *for any* $T > 0$.

**Lemma 2** *Under the assumptions of Lemma* 1, *for any sequence* $u_n^{(j)}$ *satisfying*

$$\sqrt{kh_n^d}\left(\frac{\overline{F}_{Z_{1-t_j}}(U_{Z_{1-t_j}}(n/k|x_0)|x_0)}{\overline{F}_{Z_{1-t_j}}(u_n^{(j)}|x_0)} - 1\right) \to c_j \in \mathbb{R},$$

*as* $n \to \infty$, $j = 1, \ldots, J$, *we have*

$$\begin{pmatrix} \sqrt{nh_n^d \overline{F}_{Z_{1-t_1}}(u_n^{(1)}|x_0)}\left(\frac{\hat{\overline{F}}_{Z_{1-t_1}}(u_n^{(1)}|x_0)}{\overline{F}_{Z_{1-t_1}}(u_n^{(1)}|x_0)} - 1\right) \\ \vdots \\ \sqrt{nh_n^d \overline{F}_{Z_{1-t_J}}(u_n^{(J)}|x_0)}\left(\frac{\hat{\overline{F}}_{Z_{1-t_J}}(u_n^{(J)}|x_0)}{\overline{F}_{Z_{1-t_J}}(u_n^{(J)}|x_0)} - 1\right) \end{pmatrix} \rightsquigarrow \frac{1}{f_X(x_0)} \begin{pmatrix} W_{1-t_1}(1) \\ \vdots \\ W_{1-t_J}(1) \end{pmatrix}.$$

**Lemma 3** *Assume* $(\mathcal{D}_{1-t_j})$ *and* $(\mathcal{H}_{1-t_j})$ *for* $j = 1, \ldots, J$, $(\mathcal{D}_{0.5})$, $(\mathcal{H}_{0.5})$, $(\mathcal{K}_1)$, $x_0 \in Int(S_X)$ *with* $f_X(x_0) > 0$, *and* $y \mapsto F_{Z_{1-t_j}}(y|x_0)$, $j = 1, \ldots, J$, *are strictly increasing. Consider sequences* $k \to \infty$ *and* $h_n \to 0$ *as* $n \to \infty$ *such that* $k/n \to 0$, $kh_n^d \to \infty$, $h_n^{\eta_{\varepsilon_{1-t_1}} \wedge \cdots \wedge \eta_{\varepsilon_{1-t_J}} \wedge \eta_{\varepsilon_{0.5}}} \log \frac{n}{k} \to 0$, $\sqrt{kh_n^d} h_n^{\eta_{f_X} \wedge \eta_{G_{1-t_1}} \wedge \cdots \wedge \eta_{G_{1-t_J}}} \to 0$, $\sqrt{kh_n^d} |\delta_{1-t_j}(U_{Z_{1-t_j}}(\frac{n}{k}|x_0)|x_0)| \to 0$ , $j = 1, \ldots, J$. *Then, we have*

$$\sqrt{kh_n^d}\begin{pmatrix} \frac{\hat{U}_{Z_{1-t_1}}(n/k|x_0)}{U_{Z_{1-t_1}}(n/k|x_0)} - 1 \\ \vdots \\ \frac{\hat{U}_{Z_{1-t_J}}(n/k|x_0)}{U_{Z_{1-t_J}}(n/k|x_0)} - 1 \end{pmatrix} \rightsquigarrow \frac{1}{f_X(x_0)} \begin{pmatrix} W_{1-t_1}(1) \\ \vdots \\ W_{1-t_J}(1) \end{pmatrix}.$$

**Joint weak convergence of $S_{n,1-t_j}(s_j|x_0), j = 1, \ldots, M$**

We have now all the ingredients to state the joint weak convergence of $S_{n,1-t_j}(s_j|x_0)$, $j = 1, \ldots, M$. Note that we allow for the possibility that $t_j = t_{j'}$ for $j \neq j'$, but of course the statistics $S_{n,1-t_j}(s_j|x_0)$, $j = 1, \ldots, M$, must be different. This is due to the fact that, for a given value of $t$, the study of the MDPD estimator $\widehat{\delta}_{n,1-t}$ requires the joint convergence in distribution of several statistics $S_{n,1-t}(s|x_0)$, with different values of $s$.

**Theorem 3** *Under the conditions of Theorem 1, we have, for $s_1, \ldots, s_M < 0$,*

$$\sqrt{kh_n^d}\begin{pmatrix} S_{n,1-t_1}(s_1|x_0) - \frac{1}{1-s_1}f_X(x_0) \\ \vdots \\ S_{n,1-t_M}(s_M|x_0) - \frac{1}{1-s_M}f_X(x_0) \end{pmatrix} \rightsquigarrow \begin{pmatrix} s_1 \int_0^1 \left[\frac{W_{1-t_1}(z)}{z} - W_{1-t_1}(1)\right] z^{-s_1} dz \\ \vdots \\ s_M \int_0^1 \left[\frac{W_{1-t_M}(z)}{z} - W_{1-t_M}(1)\right] z^{-s_M} dz \end{pmatrix}.$$

To prove this Theorem 3, we start to establish the weak convergence of an individual statistic $S_{n,1-t}(s|x_0)$, properly normalized. We have the following decomposition

$$\begin{aligned}
&\sqrt{kh_n^d}\Big(S_{n,1-t}(s|x_0) - \frac{1}{1-s}f_X(x_0)\Big) \\
&= \int_0^1 [W_{1-t}(z) - W_{1-t}(1)]\, s\, z^{-1-s} dz \\
&\quad + \left\{ \sqrt{kh_n^d}\big[T_{n,1-t}(s_{n,1-t}(1|x_0)|x_0) - s_{n,1-t}(1|x_0) f_X(x_0)\big] - W_{1-t}\big(s_{n,1-t}(1|x_0)\big) \right\} \\
&\quad + \left\{ W_{1-t}\big(s_{n,1-t}(1|x_0)\big) - W_{1-t}(1) \right\} \\
&\quad + \sqrt{kh_n^d}\big(s_{n,1-t}(1|x_0) - 1\big) f_X(x_0) \\
&\quad + \int_0^1 \left\{ \sqrt{kh_n^d}\big[T_{n,1-t}(s_{n,1-t}(z|x_0)|x_0) - s_{n,1-t}(z|x_0) f_X(x_0)\big] - W_{1-t}\big(s_{n,1-t}(z|x_0)\big) \right\} s\, z^{-1-s}\, dz \\
&\quad + \int_0^1 \big[W_{1-t}\big(s_{n,1-t}(z|x_0)\big) - W_{1-t}(z)\big]\, s\, z^{-1-s}\, dz \qquad (11) \\
&\quad + f_X(x_0) \sqrt{kh_n^d} \int_0^1 \big[s_{n,1-t}(z|x_0) - z\big]\, s\, z^{-1-s}\, dz \\
&=: \int_0^1 [W_{1-t}(z) - W_{1-t}(1)]\, s\, z^{-1-s} dz + \sum_{i=1}^{6} T_{i,k}. \qquad (12)
\end{aligned}$$

We study the terms separately. Clearly, using Lemma 5.2 from Goegebeur et al. (2021) we have that for $n$ large, with arbitrary large probability,

$$|T_{1,k}| \leq \sup_{y \in (0,2]} \left| \sqrt{kh_n^d}\big[T_{n,1-t}(y|x_0) - y f_X(x_0)\big] - W_{1-t}(y) \right|, \qquad (13)$$

$$\text{and } |T_{4,k}| \leq \sup_{y\in(0,2]} \left| \sqrt{kh_n^d}\left[T_{n,1-t}(y|x_0) - yf_X(x_0)\right] - W_{1-t}(y) \right| \left| \int_0^1 s\, z^{-1-s} dz \right|, \tag{14}$$

and hence, by Lemma 1 combined with the Skorohod construction we obtain $T_{1,k} = o_{\mathbb{P}}(1)$ and $T_{4,k} = o_{\mathbb{P}}(1)$.

Using again Lemma 5.2 in Goegebeur et al. (2021) with continuity, we have

$$|T_{2,k}| = o_{\mathbb{P}}(1). \tag{15}$$

Concerning $T_{3,k}$, we can use the following decomposition:

$$\begin{aligned}
T_{3,k} &= \sqrt{kh_n^d}\left[\frac{\overline{F}_{Z_{1-t}}(\widehat{U}_{Z_{1-t}}(n/k|x_0)|x_0)}{\overline{F}_{Z_{1-t}}(U_{Z_{1-t}}(n/k|x_0)|x_0)} - 1\right] f_X(x_0) \\
&= \sqrt{kh_n^d}\left[\left(\frac{\widehat{U}_{Z_{1-t}}(n/k|x_0)}{U_{Z_{1-t}}(n/k|x_0)}\right)^{-1} - 1\right] \frac{1+\delta_{1-t}(\widehat{U}_{Z_{1-t}}(n/k|x_0)|x_0)}{1+\delta_{1-t}(U_{Z_{1-t}}(n/k|x_0)|x_0)} f_X(x_0) \\
&\quad + \sqrt{kh_n^d}\left[\frac{1+\delta_{1-t}(\widehat{U}_{Z_{1-t}}(n/k|x_0)|x_0)}{1+\delta_{1-t}(U_{Z_{1-t}}(n/k|x_0)|x_0)} - 1\right] f_X(x_0) \\
&= \sqrt{kh_n^d}\left[\left(\frac{\widehat{U}_{Z_{1-t}}(n/k|x_0)}{U_{Z_{1-t}}(n/k|x_0)}\right)^{-1} - 1\right] \frac{1+\delta_{1-t}(\widehat{U}_{Z_{1-t}}(n/k|x_0)|x_0)}{1+\delta_{1-t}(U_{Z_{1-t}}(n/k|x_0)|x_0)} f_X(x_0) \\
&\quad + \sqrt{kh_n^d}\, \frac{\delta_{1-t}(U_{Z_{1-t}}(n/k|x_0)|x_0)}{1+\delta_{1-t}(U_{Z_{1-t}}(n/k|x_0)|x_0)} \\
&\quad \times \left\{\left[\frac{\delta_{1-t}(\widehat{U}_{Z_{1-t}}(n/k|x_0)|x_0)}{\delta_{1-t}(U_{Z_{1-t}}(n/k|x_0)|x_0)} - \left(\frac{\widehat{U}_{Z_{1-t}}(n/k|x_0)}{U_{Z_{1-t}}(n/k|x_0)}\right)^{-\beta(x_0)}\right]\right. \\
&\quad \left. + \left[\left(\frac{\widehat{U}_{Z_{1-t}}(n/k|x_0)}{U_{Z_{1-t}}(n/k|x_0)}\right)^{-\beta(x_0)} - 1\right]\right\}. \\
&=: -\sqrt{kh_n^d}\left[\frac{\widehat{U}_{Z_{1-t}}(n/k|x_0)}{U_{Z_{1-t}}(n/k|x_0)} - 1\right] f_X(x_0)(1+o_{\mathbb{P}}(1)) \\
&\quad + \sqrt{kh_n^d}\, \frac{\delta_{1-t}(U_{Z_{1-t}}(n/k|x_0)|x_0)}{1+\delta_{1-t}(U_{Z_{1-t}}(n/k|x_0)|x_0)} T_{3,k}^{(1)}.
\end{aligned}$$

By Proposition B.1.10 in de Haan and Ferreira (2006), for $n$ large, with arbitrary large probability, we have for $\varepsilon, \xi > 0$

$$|T_{3,k}^{(1)}| \leq \varepsilon \left( \frac{\widehat{U}_{Z_{1-t}}(n/k|x_0)}{U_{Z_{1-t}}(n/k|x_0)} \right)^{-\beta(x_0)\pm\xi} + \left( \frac{\widehat{U}_{Z_{1-t}}(n/k|x_0)}{U_{Z_{1-t}}(n/k|x_0)} \right)^{-\beta(x_0)} + 1. \tag{16}$$

In the above, the notation $a^{\pm\bullet}$ means $a^{\bullet}$ if $a \geq 1$ and $a^{-\bullet}$ if $a < 1$. This implies by Lemma 3 and our conditions that

$$T_{3,k} \rightsquigarrow -W_{1-t}(1). \tag{17}$$

Concerning now $T_{5,k}$, we have for any $\delta \in (0,1)$ small

$$\begin{aligned}
|T_{5,k}| \leq & \int_0^{\delta} \left| W_{1-t}\big(s_{n,1-t}(z|x_0)\big) - W_{1-t}(z) \right| |s|\, z^{-1-s}\, dz \\
& + \int_{\delta}^{1} \left| W_{1-t}\big(s_{n,1-t}(z|x_0)\big) - W_{1-t}(z) \right| |s|\, z^{-1-s}\, dz \\
\leq & |s| \left\{ \sup_{z\in(0,\delta]} \left| W_{1-t}\big(s_{n,1-t}(z|x_0)\big) \right| + \sup_{z\in(0,\delta]} |W_{1-t}(z)| \right\} \int_0^{\delta} z^{-1-s}\, dz \\
& + |s| \sup_{z\in(\delta,1]} \left| W_{1-t}\big(s_{n,1-t}(z|x_0)\big) - W_{1-t}(z) \right| \int_{\delta}^{1} z^{-1-s}\, dz \\
= & o_{\mathbb{P}}(1).
\end{aligned} \tag{18}$$

Finally, concerning $T_{6,k}$, we have

$$\begin{aligned}
T_{6,k} = & f_X(x_0) \sqrt{kh_n^d} \int_0^1 \left[ \frac{\overline{F}_{Z_{1-t}}(z^{-1}\widehat{U}_{Z_{1-t}}(n/k|x_0)|x_0)}{\overline{F}_{Z_{1-t}}(U_{Z_{1-t}}(n/k|x_0)|x_0)} - z \right] s\, z^{-1-s}\, dz \\
= & f_X(x_0) \sqrt{kh_n^d} \left\{ \left( \frac{\widehat{U}_{Z_{1-t}}(n/k|x_0)}{U_{Z_{1-t}}(n/k|x_0)} \right)^{-1} - 1 \right\} s \int_0^1 z^{-s} dz \\
& + f_X(x_0) \sqrt{kh_n^d} \left( \frac{\widehat{U}_{Z_{1-t}}(n/k|x_0)}{U_{Z_{1-t}}(n/k|x_0)} \right)^{-1} \\
& \times \int_0^1 \left( \frac{1 + \delta_{1-t}(z^{-1}\widehat{U}_{Z_{1-t}}(n/k|x_0)|x_0)}{1 + \delta_{1-t}(U_{Z_{1-t}}(n/k|x_0)|x_0)} - 1 \right) s\, z^{-s} dz \\
=: & -f_X(x_0) \frac{s}{1-s} \sqrt{kh_n^d} \left( \frac{\widehat{U}_{Z_{1-t}}(n/k|x_0)}{U_{Z_{1-t}}(n/k|x_0)} - 1 \right) (1 + o_{\mathbb{P}}(1)) \\
& + f_X(x_0) \sqrt{kh_n^d} \left( \frac{\widehat{U}_{Z_{1-t}}(n/k|x_0)}{U_{Z_{1-t}}(n/k|x_0)} \right)^{-1} \frac{\delta_{1-t}(U_{Z_{1-t}}(n/k|x_0)|x_0)}{1 + \delta_{1-t}(U_{Z_{1-t}}(n/k|x_0)|x_0)} T_{6,k}^{(1)}
\end{aligned}$$

with

$$
\begin{aligned}
|T_{6,k}^{(1)}| \leq & |s| \int_0^1 \left| \frac{\delta_{1-t}(z^{-1}\widehat{U}_{Z_{1-t}}(n/k|x_0)|x_0)}{\delta_{1-t}(U_{Z_{1-t}}(n/k|x_0)|x_0)} - \left( z^{-1} \frac{\widehat{U}_{Z_{1-t}}(n/k|x_0)}{U_{Z_{1-t}}(n/k|x_0)} \right)^{-\beta(x_0)} \right| z^{-s} dz \\
& + |s| \int_0^1 \left| \left( z^{-1} \frac{\widehat{U}_{Z_{1-t}}(n/k|x_0)}{U_{Z_{1-t}}(n/k|x_0)} \right)^{-\beta(x_0)} - 1 \right| z^{-s} dz \\
= & O_{\mathbb{P}}(1),
\end{aligned}
$$

using arguments similar to those for $T_{3,k}^{(1)}$. Consequently, using again Lemma 3, we deduce that

$$
T_{6,k} \rightsquigarrow -\frac{s}{1-s} W_{1-t}(1). \tag{19}
$$

Combining decomposition (12) with (13)–(19), the proof of the marginal weak convergence of $S_{n,1-t}(s|x_0)$, properly normalized, is achieved.

The joint weak convergence of $(\sqrt{kh_n^d}[S_{n,1-t_j}(s_j|x_0) - f_X(x_0)/(1-s_j)], j = 1, \ldots, M)$ follows from Lemmas 1 and 3, respectively. $\square$

### Proof of Theorem 1

Again we first consider the case of a single estimator $\widehat{L}_k(y_1, y_2|x_0)$. From (3), (4) and (5), we deduce that

$$
\begin{aligned}
& \sqrt{kh_n^d}\left(\widehat{L}_k(y_1, y_2|x_0) - L(y_1, y_2|x_0)\right) \\
& = -y_1 \sqrt{kh_n^d}\left(\widehat{G}_{1-t,k}(x_0) - G_{1-t}(x_0)\right) \\
& = -y_1 \sqrt{kh_n^d}\left( \frac{k}{n} \frac{\widehat{U}_{Z_{1-t}}(n/k|x_0)}{1+\widehat{\delta}_{n,1-t}} - G_{1-t}(x_0) \right) \\
& = -y_1 G_{1-t}(x_0) \sqrt{kh_n^d}\left( \frac{\widehat{U}_{Z_{1-t}}(n/k|x_0)}{U_{Z_{1-t}}(n/k|x_0)} \frac{1+\delta_{1-t}(U_{Z_{1-t}}(n/k|x_0)|x_0)}{1+\widehat{\delta}_{n,1-t}} - 1 \right) \\
& = -y_1 G_{1-t}(x_0) \sqrt{kh_n^d}\left( \frac{\widehat{U}_{Z_{1-t}}(n/k|x_0)}{U_{Z_{1-t}}(n/k|x_0)} - 1 \right) \\
& \quad + y_1 G_{1-t}(x_0) \sqrt{kh_n^d}\left( \widehat{\delta}_{n,1-t} - \delta_{1-t}(U_{Z_{1-t}}(n/k|x_0)|x_0) \right) \frac{1}{1+\widehat{\delta}_{n,1-t}} \\
& \quad + y_1 G_{1-t}(x_0) \frac{\widehat{\delta}_{n,1-t} - \delta_{1-t}(U_{Z_{1-t}}(n/k|x_0)|x_0)}{1+\widehat{\delta}_{n,1-t}} \sqrt{kh_n^d}\left( \frac{\widehat{U}_{Z_{1-t}}(n/k|x_0)}{U_{Z_{1-t}}(n/k|x_0)} - 1 \right).
\end{aligned}
$$

Now remark that

$$
\begin{aligned}
&\sqrt{kh_n^d}\left|\delta_{1-t}(\widehat{U}_{Z_{1-t}}(n/k|x_0)|x_0) - \delta_{1-t}(U_{Z_{1-t}}(n/k|x_0)|x_0)\right| \\
&\quad = \sqrt{kh_n^d}\left|\delta_{1-t}(U_{Z_{1-t}}(n/k|x_0)|x_0)\right|\left|\frac{\delta_{1-t}(\widehat{U}_{Z_{1-t}}(n/k|x_0)|x_0)}{\delta_{1-t}(U_{Z_{1-t}}(n/k|x_0)|x_0)} - 1\right| \\
&\quad = o_{\mathbb{P}}(1),
\end{aligned}
$$

by (16). This implies that

$$
\begin{aligned}
&\sqrt{kh_n^d}\left(\widehat{L}_k(y_1, y_2|x_0) - L(y_1, y_2|x_0)\right) \\
&\quad = -y_1 G_{1-t}(x_0)\sqrt{kh_n^d}\left(\frac{\widehat{U}_{Z_{1-t}}(n/k|x_0)}{U_{Z_{1-t}}(n/k|x_0)} - 1\right) \\
&\qquad + y_1 G_{1-t}(x_0)\sqrt{kh_n^d}\left(\widehat{\delta}_{n,1-t} - \delta_{1-t}(\widehat{U}_{Z_{1-t}}(n/k|x_0)|x_0)\right) + o_{\mathbb{P}}(1).
\end{aligned}
$$

Using the fact that

$$
\begin{aligned}
&\sqrt{kh_n^d}\begin{pmatrix} \frac{\widehat{U}_{Z_{1-t}}(n/k|x_0)}{U_{Z_{1-t}}(n/k|x_0)} - 1 \\ \widehat{\delta}_{n,1-t} - \delta_{1-t}(\widehat{U}_{Z_{1-t}}(n/k|x_0)|x_0) \end{pmatrix} \\
&\rightsquigarrow \begin{pmatrix} \frac{W_{1-t}(1)}{f_X(x_0)} \\ c\left(2\alpha\int_0^1\left[\frac{W_{1-t}(z)}{z} - W_{1-t}(1)\right]z^{2\alpha}\,dz - (1+\beta)(2\alpha+\beta)\int_0^1\left[\frac{W_{1-t}(z)}{z} - W_{1-t}(1)\right]z^{2\alpha+\beta}\,dz\right) \end{pmatrix},
\end{aligned}
$$

we can deduce that

$$
\begin{aligned}
&\sqrt{kh_n^d}\left(\widehat{L}_k(y_1, y_2|x_0) - L(y_1, y_2|x_0)\right) \\
&\quad \rightsquigarrow -y_1 G_{1-t}(x_0)\frac{W_{1-t}(1)}{f_X(x_0)} + y_1 G_{1-t}(x_0)c\left\{2\alpha\int_0^1\left[\frac{W_{1-t}(z)}{z} - W_{1-t}(1)\right]z^{2\alpha}\,dz\right. \\
&\qquad \left. -(1+\beta)(2\alpha+\beta)\int_0^1\left[\frac{W_{1-t}(z)}{z} - W_{1-t}(1)\right]z^{2\alpha+\beta}\,dz\right\}.
\end{aligned}
$$

Now, concerning the finite-dimensional convergence, it follows from Lemma 3 combined with the following theorem which states the joint behavior of the MDPD estimator $\widehat{\delta}_{n,1-t_j}, j = 1, \ldots, J$, and whose proof is deferred to the online Supplementary Material:

**Theorem 4** *Under the conditions of Theorem 1, with probability tending to one, there exists sequences of solutions $(\widehat{\delta}_{n,1-t_j})_{n\geq 1}$, $j = 1, \ldots, J$, to the MDPD estimating equations such that*

$$\begin{pmatrix} \widehat{\delta}_{n,1-t_1} - \delta_{1-t_1}(\widehat{U}_{Z_{1-t_1}}(n/k|x_0)|x_0) \\ \vdots \\ \widehat{\delta}_{n,1-t_J} - \delta_{1-t_J}(\widehat{U}_{Z_{1-t_J}}(n/k|x_0)|x_0) \end{pmatrix} \xrightarrow{\mathbb{P}} \mathbf{0}.$$

*Moreover, for the consistent solution sequences one has that*

$$\sqrt{kh_n^d} \begin{pmatrix} \widehat{\delta}_{n,1-t_1} - \delta_{1-t_1}(\widehat{U}_{Z_{1-t_1}}(n/k|x_0)|x_0) \\ \vdots \\ \widehat{\delta}_{n,1-t_J} - \delta_{1-t_J}(\widehat{U}_{Z_{1-t_J}}(n/k|x_0)|x_0) \end{pmatrix}$$
$$\rightsquigarrow c \begin{pmatrix} 2\alpha \int_0^1 \left[\frac{W_{1-t_1}(z)}{z} - W_{1-t_1}(1)\right] z^{2\alpha}\, dz - (1+\beta)(2\alpha+\beta) \int_0^1 \left[\frac{W_{1-t_1}(z)}{z} - W_{1-t_1}(1)\right] z^{2\alpha+\beta}\, dz \\ \vdots \\ 2\alpha \int_0^1 \left[\frac{W_{1-t_J}(z)}{z} - W_{1-t_J}(1)\right] z^{2\alpha}\, dz - (1+\beta)(2\alpha+\beta) \int_0^1 \left[\frac{W_{1-t_J}(z)}{z} - W_{1-t_J}(1)\right] z^{2\alpha+\beta}\, dz \end{pmatrix},$$

*where c is defined in Theorem 1.* □

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s10463-022-00839-1.

**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. The research of Armelle Guillou was supported by the French National Research Agency under the grant ANR-19-CE40-0013-01/ExtremReg project and an International Emerging Action (IEA-00179). Computation/simulation for the work described in this paper was supported by the DeIC National HPC Centre, SDU.

**Declaration**

**Conflict of interest** The authors declare no conflicts of interest.

## References

Basu, A., Harris, I. R., Hjort, N. L., Jones, M. C. (1998). Robust and efficient estimation by minimizing a density power divergence. *Biometrika, 85*, 549–559.

Beirlant, J., Joossens, E., Segers, J. (2009). Second-order refined peaks-over-threshold modelling for heavy-tailed distributions. *Journal of Statistical Planning and Inference, 139*, 2800–2815.

Beirlant, J., Dierckx, G., Guillou, A. (2011). Bias-reduced estimators for bivariate tail modelling. *Insurance: Mathematics and Economics, 49*, 18–26.

Castro, D., de Carvalho, M. (2017). Spectral density regression for bivariate extremes. *Stochastic Environmental Research and Risk Assessment, 31,* 1603–1613.

Castro, D., de Carvalho, M., Wadsworth, J. L. (2018). Time-varying extreme value dependence with application to leading European stock markets. *Annals of Applied Statistics, 12,* 283–309.

Daouia, A., Gardes, L., Girard, S., Lekina, A. (2011). Kernel estimators of extreme level curves. *TEST, 20*, 311–333.

de Carvalho, M. (2016). Statistics of extremes: Challenges and opportunities. In F. Longin (Ed.), *Extreme events in finance: A handbook of extreme value theory and its applications*. Hoboken: Wiley.

de Carvalho, M., Leonelli, M., Rossi, A. (2020). Tracking change-points in multivariate extreme. arXiv:2011.05067.

De Cock, D. (2011). Ames, Iowa: Alternative to the Boston housing data as an end of semester regression project. *Journal of Statistics Education*. https://doi.org/10.1080/10691898.2011.11889627.

de Haan, L., Ferreira, A. (2006). *Extreme value theory: An introduction*. New York: Springer.

Dell'Aquila, R., Embrechts, P. (2006). Extremes and robustness: A contradiction? *Financial Markets and Portfolio Management, 20*, 103–118.
Drees, H. (2022). Statistical inference on a changing extreme value dependence structure. arXiv:2201.06389v2.
Dutang, C., Goegebeur, Y., Guillou, A. (2014). Robust and bias-corrected estimation of the coefficient of tail dependence. *Insurance*: *Mathematics and Economics*, *57*, 46–57.
Escobar-Bach, M., Goegebeur, Y., Guillou, A., You, A. (2017). Bias-corrected and robust estimation of the bivariate stable tail dependence function. *TEST, 26*, 284–307.
Escobar-Bach, M., Goegebeur, Y., Guillou, A. (2018a). Local robust estimation of the Pickands dependence function. *Annals of Statistics, 46*, 2806–2843.
Escobar-Bach, M., Goegebeur, Y., Guillou, A. (2018b). Local estimation of the conditional stable tail dependence function. *Scandinavian Journal of Statistics, 45*, 590–617.
Escobar-Bach, M., Goegebeur, Y., Guillou, A. (2020). Bias correction in conditional multivariate extremes. *Electronic Journal of Statistics, 14*, 1773–1795.
Feuerverger, A., Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution. *Annals of Statistics, 27*, 760–781.
Fujisawa, H., Eguchi, S. (2008). Robust parameter estimation with a small bias against heavy contamination. *Journal of Multivariate Analysis, 99*, 2053–2081.
Gardes, L., Girard, S. (2015). Nonparametric estimation of the conditional tail copula. *Journal of Multivariate Analysis, 137*, 1–16.
Giné, E., Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, 38*, 907–921.
Giné, E., Koltchinskii, V., Zinn, J. (2004). Weighted uniform consistency of kernel density estimators. *Annals of Probability, 32*, 2570–2605.
Goegebeur, Y., Guillou, A., Qin, J. (2019). Bias-corrected estimation for conditional Pareto- type distributions with random right censoring. *Extremes, 22*, 459–498.
Goegebeur, Y., Guillou, A., Ho, N. K. L., Qin, J. (2020). Robust nonparametric estimation of the conditional tail dependence coefficient. *Journal of Multivariate Analysis*. https://doi.org/10.1016/j.jmva.2020.104607.
Goegebeur, Y., Guillou, A., Ho, N. K. L., Qin, J. (2021). A Weissman-type estimator of the conditional marginal expected shortfall. *Econometrics and Statistics*. https://doi.org/10.1016/j.ecosta.2021.09.006
Gomes, M. I., Martins, M. J. (2004). Bias-reduction and explicit semi-parametric estimation of the tail index. *Journal of Statistical Planning and Inference, 124*, 361–378.
Hampel, F., Ronchetti, E., Rousseeuw, P., Stahel, W. (1986). *Robust statistics*: *The approach based on influence functions*. New York: Wiley.
Huang, X. (1992). Statistics of bivariate extremes. PhD Thesis, Erasmus University Rotterdam, Tinbergen Institute Research series No. 22.
Huber, P. (1981). *Robust statistics*. New York: Wiley.
Hubert, M., Dierckx, G., Vanpaemel, D. (2013). Detecting influential data points for the Hill estimator in Pareto-type distributions. *Computational Statistics and Data Analysis, 65*, 13–28.
Kullback, S., Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics, 22*, 79–86.
Ledford, A. W., Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society*: *Series B*, *59*, 475–499.
Mhalla, L., de Carvalho, M., Chavez-Demoulin, V. (2019). Regression type models for extremal dependence. *Scandinavian Journal of Statistics, 46*, 1141–1167.
Minami, M., Eguchi, S. (2002). Robust blind source separation by beta divergence. *Neural Computation, 14*, 1859–1886.
Nolan, D., Pollard, D. (1987). U-processes: Rates of convergence. *Annals of Statistics, 15,* 780–799.
Resnick, S. I. (2007). *Heavy-tail phenomena. Probabilistic and statistical modeling*. New York: Springer.
Song, J. (2021). Sequential change point test in the presence of outliers: The density power divergence based approach. *Electronic Journal of Statistics, 15,* 3504–3550.