Journal of Multivariate Analysis 178 (2020) 104607

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Robust nonparametric estimation of the conditional tail dependence coefficient

Yuri Goegebeur [a], Armelle Guillou [b,*], Nguyen Khanh Le Ho [a], Jing Qin [a]

[a] Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark
[b] Institut Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg et CNRS, 7 rue René Descartes, 67084 Strasbourg cedex, France

ARTICLE INFO

Article history:
Received 21 July 2019
Received in revised form 2 March 2020
Accepted 2 March 2020
Available online 9 March 2020

AMS 2010 subject classifications:
primary 62G32
secondary 62G35

Keywords:
Coefficient of tail dependence
Empirical process
Local estimation
Robustness

ABSTRACT

We consider robust and nonparametric estimation of the coefficient of tail dependence in presence of random covariates. The estimator is obtained by fitting the extended Pareto distribution locally to properly transformed bivariate observations using the minimum density power divergence criterion. We establish convergence in probability and asymptotic normality of the proposed estimator under some regularity conditions. The finite sample performance is evaluated with a small simulation experiment, and the practical applicability of the method is illustrated on a real dataset of air pollution measurements.

1. Introduction

Many problems involving extreme events are inherently multivariate, and hence they should be handled with appropriate multivariate extreme value methods. Of particular interest is the estimation of the extremal dependence between two or more variables. A full characterization of the extremal dependence between variables can be obtained from functions like the spectral distribution function or the Pickands dependence function. We refer to [5], and [10], and the references therein, for more details about this approach. Alternatively, similar to classical statistics one can try and summarize the extremal dependency in a number of well-chosen coefficients that give a representative picture of the full dependency structure, like, e.g., the coefficient of tail dependence, see [28]. Modelling tail dependence is a critical issue in many scientific disciplines. For instance, in finance and actuarial science an important problem is to estimate very large quantiles of the distribution of the sums of possibly dependent risks, see, e.g., [1]. In environmental science, studying dependence in extreme levels of pollutants like ozone, particulate matter, carbon monoxide and temperature is important as combined high levels of these variables may pose a major threat to human health, see, e.g., [17]. In this paper, we will consider robust and nonparametric estimation of the coefficient of tail dependence when there are random covariates.

Let $(Y^{(1)}, Y^{(2)})$ be a bivariate random vector recorded along with a random covariate $X \in \mathbb{R}^p$. The covariate X has density function f_X with support $S_X \subset \mathbb{R}^p$, having non-empty interior. The continuous conditional marginal distribution

* Corresponding author.
E-mail address: armelle.guillou@math.unistra.fr (A. Guillou).

https://doi.org/10.1016/j.jmva.2020.104607

functions of $Y^{(j)}$ given $X = x$ are denoted by $F_j(\cdot|x), j = 1, 2$, and the joint conditional distribution function of the pair satisfies that for all $x \in S_X$ and $y \in [0, 1]$

$$\Pr\left\{1 - F_1(Y^{(1)}|x) < y, 1 - F_2(Y^{(2)}|x) < y \middle| X = x\right\} = C(x)y^{\frac{1}{\eta(x)}}\left\{1 + \frac{1}{\eta(x)}\delta(y|x)\right\},$$

where $\eta(x) \in (0, 1]$ is the conditional tail dependence coefficient, and $|\delta(\cdot|x)|$ is a regularly varying function in the neighborhood of zero with index $\tau(x) > 0$. This model is the cornerstone of the paper. It is very general and has been already used in [15,16], among others. It can be viewed as a Hall-type model (see [24]) adjusted to the covariate framework, and as such, all the parameters and functions involved in it, depend on the covariate position x. In this paper we focus on the estimation of $\eta(x)$, and introduce a nonparametric estimator, which is obtained from local fits of the above model in a neighborhood of x, a point of interest in the covariate space. In the unconditional case, several practically well performing estimators for the tail dependence coefficient have been introduced, see, e.g., [4,7,14,15,30,31].

Our aim in this paper is to estimate the conditional tail dependence coefficient 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 dependency structure. Note that such outliers are not necessary marginal outliers. To achieve robustness, we will use the idea of the density power divergence introduced by [2]. In particular, the density power divergence between density functions h and g is given by

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

Here h is assumed to be the true (typically unknown) density of the data, whereas g is a parametric model, depending on a parameter vector $\boldsymbol{\theta}$ which is determined by minimizing the empirical version of (1). The resulting estimator will be called, in the sequel, minimum density power divergence (MDPD) estimator. In the present paper we will adjust this criterion to the local estimation context with focus on estimating conditional extreme dependence. At a general level, one can say that the usefulness of robust methods in extreme value statistics has been clearly pointed out by [11]. Robust estimation methods have been successfully applied to tail index estimation (see for instance [12,19,26,27,33,34]) and conditional tail index estimation (for instance [13,22]). [15] used this MDPD criterion to obtain a robust estimator for η, but in a context without covariates. To the best of our knowledge, robust nonparametric estimation of the conditional coefficient of tail dependence has not been considered so far in the extreme value literature.

The remainder of the paper is organized as follows. In Section 2, we simplify the problem to the case where the conditional marginal distributions are known and we prove the existence, convergence in probability and asymptotic normality of the MDPD estimator of the conditional tail dependence coefficient. Then in Section 3, the realistic situation where the margins are unknown is considered and similar results are established. The efficiency and robustness of our MDPD estimator are illustrated in a small simulation study in Section 4 and on a real dataset on air pollution in Section 5. Finally, all the proofs are postponed to the Appendix.

2. Case of known margins

In this section, we assume that the conditional marginal distribution functions $F_1(\cdot|x)$ and $F_2(\cdot|x)$ are known.

Define $Z := \min\left\{\frac{1}{1-F_1(Y^{(1)}|X)}, \frac{1}{1-F_2(Y^{(2)}|X)}\right\}$. Direct computations yield for all $x \in S_X$

$$\overline{F}_Z(z|x) := \Pr(Z > z|X = x) = C(x)z^{-\frac{1}{\eta(x)}}\left\{1 + \frac{1}{\eta(x)}\delta_Z(z|x)\right\}, \tag{2}$$

where

$$\delta_Z(z|x) := \delta\left(\frac{1}{z}\middle|x\right).$$

Here $|\delta_Z(\cdot|x)|$ is a regularly varying function at infinity with index $-\tau(x)$, which is additionally assumed to be normalized, i.e., such that

$$\delta_Z(z|x) = A(x)\exp\left\{\int_1^z \frac{\varepsilon(u|x)}{u}du\right\}, \tag{3}$$

with $A(x) \in \mathbb{R}$ and $\varepsilon(z|x) \to -\tau(x)$ as $z \to \infty$.

Note that the conditional distribution of Z, given $X = x$, satisfies Condition $(\mathcal{R})$ in [13] with second order parameter $\rho(x) := -\tau(x)\eta(x)$. Thus, one can approximate the conditional distribution of Z/u, given $Z > u$, where u denotes a high threshold value, by the extended Pareto distribution given by

$$G(z; \eta, \delta, \rho) = \begin{cases} 1 - \left\{z\left(1 + \delta - \delta z^{\frac{\rho}{\eta}}\right)\right\}^{-\frac{1}{\eta}}, & z > 1, \\ 0, & z \leq 1, \end{cases}$$

and density function

$$g(z;\eta,\delta,\rho)=\begin{cases}\frac{z^{-\frac{1}{\eta}-1}}{\eta}\left\{1+\delta\left(1-z^{\frac{\rho}{\eta}}\right)\right\}^{-\frac{1}{\eta}-1}\left[1+\delta\left\{1-\left(1+\frac{\rho}{\eta}\right)z^{\frac{\rho}{\eta}}\right\}\right], & z>1,\\ 0, & z\leq 1,\end{cases}$$

where $\eta > 0$, $\rho < 0$, and $\delta > \max(-1, \eta/\rho)$.

Indeed, as shown in [6], we have

$$\sup_{z\geq 1}\left|\frac{\overline{F}_Z(uz|x)}{\overline{F}_Z(u|x)}-\overline{G}\{z;\eta(x),\delta_Z(u|x),\rho(x)\}\right| = o(\delta_Z(u|x)) \text{ if } u\to\infty.$$

Clearly, based on this result, one can obtain an estimator for $\eta(x)$ by fitting the extended Pareto distribution to the relative excesses over a high threshold.

Let $(X_1, Z_1), \ldots, (X_n, Z_n)$ be independent copies of the random vector (X, Z). We develop a nonparametric, robust and asymptotically unbiased estimator for $\eta(x)$ by fitting g locally to the relative excesses Z_i/u_n, $i \in \{1, \ldots, n\}$, by means of the MDPD criterion, adjusted to locally weighted estimation, i.e., we minimize

$$\widehat{\Delta}_\alpha(\eta,\delta_Z;\rho) := \frac{1}{n}\sum_{i=1}^{n}K_{h_n}(x-X_i)\left\{\int_1^\infty g^{1+\alpha}(z;\eta,\delta_Z,\rho)dz-\left(1+\frac{1}{\alpha}\right)g^\alpha\left(\frac{Z_i}{u_n};\eta,\delta_Z,\rho\right)\right\}\mathbb{1}_{(Z_i>u_n)},$$

in case $\alpha > 0$ and

$$\widehat{\Delta}_0(\eta,\delta_Z;\rho) := -\frac{1}{n}\sum_{i=1}^{n}K_{h_n}(x-X_i)\ln g\left(\frac{Z_i}{u_n};\eta,\delta_Z,\rho\right)\mathbb{1}_{(Z_i>u_n)},$$

in case $\alpha = 0$, where $K_{h_n}(x) := K(x/h_n)/h_n^p$, K is a joint density function on $\mathbb{R}^p$, h_n is a non-random sequence of bandwidths with $h_n \to 0$ if $n \to \infty$, $\mathbb{1}_{(A)}$ is the indicator function on the event A and u_n is a local non-random threshold sequence satisfying $u_n \to \infty$ if $n \to \infty$. Note that in case $\alpha = 0$, the local empirical density power divergence criterion corresponds with a locally weighted log-likelihood function. The parameter α controls the trade-off between efficiency and robustness of the MDPD criterion: the estimator becomes more efficient but less robust as α gets closer to zero, whereas for increasing α the robustness increases and the efficiency decreases. Note that we only estimate $\eta(x)$ and $\delta_Z(u_n|x)$ with the MDPD criterion, while the second order parameter $\rho(x)$ will be fixed at some value. Fixing second order parameters like $\rho(x)$ here at some value is a common practice in extreme value statistics, and was also proposed in [3,18], and [23]. Alternatively, one can replace $\rho(x)$ by an external consistent estimator. However, the estimation of $\rho(x)$ in a robust way is still an open problem, and moreover, using an external consistent estimator rather than a canonical value, does not, in general, improve the performance of the final MDPD estimator in practice. For all these reasons, we only use a canonical value for the parameter $\rho(x)$ in the sequel.

The MDPD estimators of $\{\eta(x), \delta_Z(u_n|x)\}$ satisfy the estimating equations

$$\begin{aligned}0 &= \frac{1}{n}\sum_{i=1}^{n}K_{h_n}(x-X_i)\mathbb{1}_{(Z_i>u_n)}\int_1^\infty g^\alpha(z;\eta,\delta_Z,\rho)\frac{\partial g(z;\eta,\delta_Z,\rho)}{\partial\eta}dz\\ &\quad-\frac{1}{n}\sum_{i=1}^{n}K_{h_n}(x-X_i)g^{\alpha-1}\left(\frac{Z_i}{u_n};\eta,\delta_Z,\rho\right)\frac{\partial g\left(\frac{Z_i}{u_n};\eta,\delta_Z,\rho\right)}{\partial\eta}\mathbb{1}_{(Z_i>u_n)}, \quad (4)\\ 0 &= \frac{1}{n}\sum_{i=1}^{n}K_{h_n}(x-X_i)\mathbb{1}_{(Z_i>u_n)}\int_1^\infty g^\alpha(z;\eta,\delta_Z,\rho)\frac{\partial g(z;\eta,\delta_Z,\rho)}{\partial\delta_Z}dz\\ &\quad-\frac{1}{n}\sum_{i=1}^{n}K_{h_n}(x-X_i)g^{\alpha-1}\left(\frac{Z_i}{u_n};\eta,\delta_Z,\rho\right)\frac{\partial g\left(\frac{Z_i}{u_n};\eta,\delta_Z,\rho\right)}{\partial\delta_Z}\mathbb{1}_{(Z_i>u_n)}. \quad (5)\end{aligned}$$

The following statistic is crucial for studying the asymptotic behavior of the estimators. Set $\ln_+ x := \ln\max(x, 1)$, $x > 0$, and

$$T_n(K,s,t|x) := \frac{1}{n}\sum_{i=1}^{n}K_{h_n}(x-X_i)\left(\frac{Z_i}{u_n}\right)^s\left(\ln_+\frac{Z_i}{u_n}\right)^t\mathbb{1}_{(Z_i>u_n)},$$

where $s \leq 0$ and $t \geq 0$. The motivation for considering this type of statistic is that the estimating equations (4) and (5) only depend on statistics of this form. Note that $\ln_+ x$ is introduced to ensure that $(\ln_+ Z_i/u_n)^t$ is always well defined (t is nonnegative, not necessary integer).

Due to the regression context, we need the following classical Hölder-type conditions. Here $\|\cdot\|$ denotes some norm on $\mathbb{R}^p$.

Assumption $(\mathcal{H})$. There exist positive constants M_{f_X}, M_C, M_A, M_η, M_ε, δ_{f_X}, δ_C, δ_A, δ_η, and δ_ε, such that for all $(x,z) \in S_X \times S_X$: $|f_X(x) - f_X(z)| \le M_{f_X}\|x-z\|^{\delta_{f_X}}$, $|C(x) - C(z)| \le M_C\|x-z\|^{\delta_C}$, $|A(x)-A(z)| \le M_A\|x-z\|^{\delta_A}$, $|\eta(x)-\eta(z)| \le M_\eta\|x-z\|^{\delta_\eta}$ and $\sup_{y\ge 1}|\varepsilon(y|x) - \varepsilon(y|z)| \le M_\varepsilon\|x-z\|^{\delta_\varepsilon}$.

Also, the following assumption, standard in the context of local estimation, is required on the kernel function.

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

In order to establish the asymptotic normality of the consistent sequence of solutions $\{\widehat{\eta}_n(x), \widehat{\delta}_{Z,n}(x)\}$ of the estimating equations (4) and (5), we introduce

$$\mathbb{S}_n^{(j)}(s) := \sqrt{nh_n^p\overline{F}_Z(u_n|x)f_X(x)}\left[\frac{T_n(K,s,j|x)}{\overline{F}_Z(u_n|x)f_X(x)} - \frac{j!\eta_0^j(x)}{\{1-s\eta_0(x)\}^{j+1}}\right], \quad j \in \{0,1,2,3\},$$

where we denote by $\eta_0(x)$, resp. $\rho_0(x)$, the true conditional tail dependence coefficient, resp. second order parameter. Hereafter, '$\rightsquigarrow$' denotes the convergence in distribution.

Theorem 2.1. *Let $(X_1,Z_1),\ldots,(X_n,Z_n)$ be a sample of independent copies of the random vector (X,Z) where the distribution of Z, given $X=x$, satisfies (2) and (3), X follows a distribution with density function f_X, and assume $(\mathcal{H})$ and $(\mathcal{K}_1)$ hold. For all $x \in Int(S_X)$ with $f_X(x) > 0$, we assume that $u_n \to \infty$ and $h_n \to 0$ in such a way that $h_n^{\delta_\varepsilon}\ln u_n \to 0$, $nh_n^p\overline{F}_Z(u_n|x) \to \infty$, $\sqrt{nh_n^p\overline{F}_Z(u_n|x)}\delta_Z(u_n|x) \to \lambda \in \mathbb{R}$, $\sqrt{nh_n^p\overline{F}_Z(u_n|x)}h_n^{\delta_{f_X}\wedge\delta_C} \to 0$, $\sqrt{nh_n^p\overline{F}_Z(u_n|x)}h_n^{\delta_\eta}\ln u_n \to 0$. Then in $\mathcal{C}^4([S,0])$, $S<0$,*

$$(\mathbb{S}_n^{(0)}, \mathbb{S}_n^{(1)}, \mathbb{S}_n^{(2)}, \mathbb{S}_n^{(3)}) \rightsquigarrow (\mathbb{S}^{(0)}, \mathbb{S}^{(1)}, \mathbb{S}^{(2)}, \mathbb{S}^{(3)}), \qquad \text{for } n \to \infty,$$

a Gaussian process, with, for $s \in [S,0]$, mean functions

$$E\{\mathbb{S}^{(j)}(s)\} = -\lambda\sqrt{f_X(x)}j!\eta_0^{j-1}(x)\left[\frac{1}{\{1-s\eta_0(x)\}^{j+1}} - \frac{1-\rho_0(x)}{\{1-\rho_0(x)-s\eta_0(x)\}^{j+1}}\right], \quad j \in \{0,1,2,3\},$$

and covariance functions given by

$$Cov\{\mathbb{S}^{(j)}(s_1), \mathbb{S}^{(k)}(s_2)\} = \frac{(j+k)!\eta_0^{j+k}(x)\|K\|_2^2}{\{1-(s_1+s_2)\eta_0(x)\}^{1+j+k}}, \quad (j,k) \in \{0,1,2,3\}^2.$$

Note that [13] obtained a similar result, though under their high level assumption called $(\mathcal{M})$, which is avoided in the present paper.

Based on this theorem, one can now establish the existence, convergence in probability and asymptotic normality of the MDPD estimators of $\{\eta_0(x), \delta_Z(u_n|x)\}$, when suitably normalized. This theorem is similar to Theorems 2 and 3 in [13] with our new conditions given in our Theorem 2.1, and thus the proof is omitted.

Theorem 2.2. *Let $(X_1,Z_1),\ldots,(X_n,Z_n)$ be a sample of independent copies of the random vector (X,Z) where the distribution of Z, given $X=x$, satisfies (2) and (3), X follows a distribution with density function f_X, and assume $(\mathcal{H})$ and $(\mathcal{K}_1)$ hold.*

For all $x \in Int(S_X)$ with $f_X(x)>0$, let $u_n \to \infty$ and $h_n \to 0$ in such a way that $nh_n^p\overline{F}_Z(u_n|x) \to \infty$ and $h_n^{\delta_\eta\wedge\delta_\varepsilon}\ln u_n \to 0$. Then with probability tending to 1, there exist sequences of solutions $\{\widehat{\eta}_n(x), \widehat{\delta}_{Z,n}(x)\}$ of the estimating equations (4) and (5), with ρ fixed at $\rho_0(x)$ such that $\{\widehat{\eta}_n(x), \widehat{\delta}_{Z,n}(x)\} \xrightarrow{\Pr} \{\eta_0(x), 0\}$.

If additionally,

$$\sqrt{nh_n^p\overline{F}_Z(u_n|x)}\,\delta_Z(u_n|x) \longrightarrow \lambda \in \mathbb{R}, \tag{6}$$

$$\sqrt{nh_n^p\overline{F}_Z(u_n|x)}\,h_n^{\delta_{f_X}\wedge\delta_C} \longrightarrow 0, \tag{7}$$

$$\sqrt{nh_n^p\overline{F}_Z(u_n|x)}\,h_n^{\delta_\eta}\ln u_n \longrightarrow 0, \tag{8}$$

then

$$\sqrt{nh_n^p\overline{F}_Z(u_n|x)f_X(x)}\begin{bmatrix}\widehat{\eta}_n(x) - \eta_0(x) \\ \widehat{\delta}_{Z,n}(x) - \delta_Z(u_n|x)\end{bmatrix} \rightsquigarrow \mathcal{N}_2\left(\boldsymbol{0}, \boldsymbol{C}^{-1}\{\rho_0(x)\}\boldsymbol{B}\{\rho_0(x)\}\boldsymbol{\Sigma}\{\rho_0(x)\}\boldsymbol{B}^\top\{\rho_0(x)\}\boldsymbol{C}^{-1}\{\rho_0(x)\}\right),$$

for $n \to \infty$, where the matrix $\boldsymbol{B}\{\rho_0(x)\}$ is defined by

$$\boldsymbol{B}\{\rho_0(x)\} := \eta_0^{-\alpha-2}(x)\begin{bmatrix} -\frac{\alpha\eta_0(x)\{1+\eta_0(x)\}}{[1+\alpha\{1+\eta_0(x)\}]^2} & \eta_0(x) & 0 & -1 \\ -\frac{\alpha\eta_0(x)\rho_0(x)\{1+\eta_0(x)\}}{[1+\alpha\{1+\eta_0(x)\}][1-\rho_0(x)+\alpha\{1+\eta_0(x)\}]} & \eta_0(x) & -\eta_0(x)\{1-\rho_0(x)\} & 0\end{bmatrix},$$

the elements of the symmetric (4×4) *matrix* $\boldsymbol{\Sigma}\{\rho_0(x)\}$ *are given by*

$$\begin{aligned}
\sigma_{11}(\rho_0(x)) &:= \|K\|_2^2,\ \sigma_{21}(\rho_0(x)) := \frac{\|K\|_2^2}{1+\alpha(1+\eta_0(x))},\ \sigma_{22}(\rho_0(x)) := \frac{\|K\|_2^2}{1+2\alpha(1+\eta_0(x))},\\
\sigma_{31}(\rho_0(x)) &:= \frac{\|K\|_2^2}{1-\rho_0(x)+\alpha(1+\eta_0(x))},\ \sigma_{32}(\rho_0(x)) := \frac{\|K\|_2^2}{1-\rho_0(x)+2\alpha(1+\eta_0(x))},\\
\sigma_{33}(\rho_0(x)) &:= \frac{\|K\|_2^2}{1-2\rho_0(x)+2\alpha(1+\eta_0(x))},\ \sigma_{41}(\rho_0(x)) := \frac{\eta_0(x)\|K\|_2^2}{[1+\alpha(1+\eta_0(x))]^2},\\
\sigma_{42}(\rho_0(x)) &:= \frac{\eta_0(x)\|K\|_2^2}{[1+2\alpha(1+\eta_0(x))]^2},\ \sigma_{43}(\rho_0(x)) := \frac{\eta_0(x)\|K\|_2^2}{[1-\rho_0(x)+2\alpha(1+\eta_0(x))]^2},\\
\sigma_{44}(\rho_0(x)) &:= \frac{2\eta_0^2(x)\|K\|_2^2}{[1+2\alpha(1+\eta_0(x))]^3},
\end{aligned}$$

and those of the symmetric (2×2) *matrix* $\mathbb{C}(\rho_0(x))$ *by*

$$\begin{aligned}
C_{11}\{\rho_0(x)\} &:= \eta_0^{-\alpha-2}(x)\,\frac{1+\alpha^2\{1+\eta_0(x)\}^2}{[1+\alpha\{1+\eta_0(x)\}]^3},\\
C_{21}\{\rho_0(x)\} &:= \eta_0^{-\alpha-2}(x)\,\frac{\rho_0(x)\{1-\rho_0(x)\}[1+\alpha\{1+\eta_0(x)\}+\alpha^2\{1+\eta_0(x)\}^2]+\alpha^3\rho_0(x)\{1+\eta_0(x)\}^3}{[1+\alpha\{1+\eta_0(x)\}]^2[1-\rho_0(x)+\alpha\{1+\eta_0(x)\}]^2},\\
C_{22}\{\rho_0(x)\} &:= \eta_0^{-\alpha-2}(x)\,\frac{\{1-\rho_0(x)\}\rho_0^2(x)+\alpha\rho_0^2(x)\{1+\eta_0(x)\}[\alpha\{1+\eta_0(x)\}-\rho_0(x)]}{[1+\alpha\{1+\eta_0(x)\}][1-\rho_0(x)+\alpha\{1+\eta_0(x)\}][1-2\rho_0(x)+\alpha\{1+\eta_0(x)\}]}.
\end{aligned}$$

Note that the expected value of the limiting random vector in Theorem 2.2 is zero, whatever the value of λ. The estimator is therefore said to be asymptotically unbiased.

The following proposition deals with the behavior of the MDPD estimators $\{\widehat{\eta}_n(x), \widehat{\delta}_{Z,n}(x)\}$ when the parameter $\rho(x)$ is fixed at some value $\widetilde{\rho} < 0$, possibly mis-specified.

Proposition 2.1. *Under the assumptions of Theorem 2.2, but now with ρ fixed at $\widetilde{\rho}$ in the estimating equations (4) and (5), with probability tending to 1, there exist sequences of solutions $\{\widehat{\eta}_n(x), \widehat{\delta}_{Z,n}(x)\}$ of the estimating equations such that $\{\widehat{\eta}_n(x), \widehat{\delta}_{Z,n}(x)\} \xrightarrow{\Pr} \{\eta_0(x), 0\}$.*

If additionally (6), (7) and (8) hold, then

$$\sqrt{nh_n^p\overline{F}_Z(u_n|x)f_X(x)}\begin{bmatrix}\widehat{\eta}_n(x)-\eta_0(x)\\ \widehat{\delta}_{Z,n}(x)\end{bmatrix} \rightsquigarrow \mathcal{N}_2\left(-\lambda\sqrt{f_X(x)}\boldsymbol{C}^{-1}(\widetilde{\rho})\boldsymbol{B}(\widetilde{\rho})\boldsymbol{d},\ \boldsymbol{C}^{-1}(\widetilde{\rho})\boldsymbol{B}(\widetilde{\rho})\boldsymbol{\Sigma}(\widetilde{\rho})\boldsymbol{B}^\top(\widetilde{\rho})\boldsymbol{C}^{-1}(\widetilde{\rho})\right),$$

for $n \to \infty$, where the elements of the vector $\boldsymbol{d}$ are the following

$$\begin{aligned}
d_1 &:= 0,\ d_2 := -\frac{\alpha\rho_0(x)\{1+\eta_0(x)\}}{\eta_0(x)[1+\alpha\{1+\eta_0(x)\}][1-\rho_0(x)+\alpha\{1+\eta_0(x)\}]},\\
d_3 &:= -\frac{[\alpha\{1+\eta_0(x)\}-\widetilde{\rho}]\rho_0(x)}{\eta_0(x)[1-\widetilde{\rho}+\alpha\{1+\eta_0(x)\}][1-\rho_0(x)-\widetilde{\rho}+\alpha\{1+\eta_0(x)\}]},\\
d_4 &:= \frac{\rho_0(x)\{1-\rho_0(x)\}-\alpha^2\rho_0(x)\{1+\eta_0(x)\}^2}{[1+\alpha\{1+\eta_0(x)\}]^2[1-\rho_0(x)+\alpha\{1+\eta_0(x)\}]^2}.
\end{aligned}$$

Again the proof of Proposition 2.1 is similar to the one of Proposition 1 in [13] and thus is omitted. Note that in case $\widetilde{\rho}$ is mis-specified, then the mean of the limiting normal distribution is not necessarily zero, and hence one possible loses the asymptotic unbiasedness. However, as will be clear from the simulations, even though $\widetilde{\rho}$ is mis-specified, the proposed MDPD estimator performs well with respect to bias. Also note that the asymptotic variance expression in Proposition 2.1 is the same as that in Theorem 2.2, though with $\rho_0(x)$ replaced by $\widetilde{\rho}$.

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. To this aim, we define

$$\breve{Z} := \min\left\{\frac{1}{1-F_{n,1}(Y^{(1)}|X)}, \frac{1}{1-F_{n,2}(Y^{(2)}|X)}\right\},$$

for suitable estimators $F_{n,j}$ of F_j, $j=1,2$. Then similarly as in the previous section, the statistic

$$\breve{T}_n(K,s,t|x) := \frac{1}{n}\sum_{i=1}^{n} K_{h_n}(x-X_i)\left(\frac{\breve{Z}_i}{u_n}\right)^s \left(\ln_+ \frac{\breve{Z}_i}{u_n}\right)^t \mathbb{1}_{(\breve{Z}_i>u_n)},$$

is the cornerstone for the MDPD estimator, denoted $\breve{\eta}_n(x)$. In particular, the main result relies essentially on the asymptotic properties of this statistic, and so the idea will be to decompose

$$\sqrt{nh_n^p \overline{F}_Z(u_n|x)f_X(x)}\left[\frac{\breve{T}_n(K,s,j|x)}{\overline{F}_Z(u_n|x)f_X(x)} - \frac{j!\eta_0^j(x)}{\{1-s\eta_0(x)\}^{j+1}}\right]$$

into the two terms

$$\sqrt{nh_n^p \overline{F}_Z(u_n|x)f_X(x)}\left[\frac{T_n(K,s,j|x)}{\overline{F}_Z(u_n|x)f_X(x)} - \frac{j!\eta_0^j(x)}{\{1-s\eta_0(x)\}^{j+1}}\right]$$
$$+\sqrt{nh_n^p \overline{F}_Z(u_n|x)f_X(x)}\left\{\frac{\breve{T}_n(K,s,j|x)}{\overline{F}_Z(u_n|x)f_X(x)} - \frac{T_n(K,s,j|x)}{\overline{F}_Z(u_n|x)f_X(x)}\right\}. \quad (9)$$

The first term can be dealt with using the results from the previous section, whereas we have to show that the second term is negligible for all $s<0$ with $j\in\{0,1,2,3\}$ or $(s,j)=(0,0)$.

In the sequel, we will use the empirical kernel estimator of the unknown distribution functions

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

where the bandwidth $c:=c_n$ 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 divergence objective function, 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 conditional marginal distribution function F_j similar to those imposed in Section 2.

Assumption $(\mathcal{F})$. There exist $M_{F_j}>0$ and $\delta_{F_j}>0$ such that $|F_j(y|x)-F_j(y|z)|\le M_{F_j}\|x-z\|^{\delta_{F_j}}$, 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. Denote by $B_z(r)$ the closed ball with center z and radius r with respect to $\|.\|$.

Assumption $(\mathcal{K}_2)$. K satisfies Assumption $(\mathcal{K}_1)$, there exist $\delta,m>0$ 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 [20] or [21]. As stated in these contributions, it is satisfied by $K(x)=\phi\{a(x)\}$, a being a polynomial and ϕ a bounded real function of bounded variation (see, e.g., [29]). This is also the case, e.g., if the graph of K is a pyramid (truncated or not), or if $K=\mathbb{1}_{[-1,1]^p}$, etc. In particular, this Assumption $(\mathcal{K}_2)$ allows us to measure the discrepancy between the conditional distribution function F_j and its empirical kernel version $F_{n,j}$, as stated in the following lemma established by [17].

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 $q>1$*

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

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

$$\inf_{x\in S_X}\lambda\left[\{u\in B_0(1): x-cu\in S_X\}\right]>\varepsilon,$$

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

$$\sup_{(y,x)\in\mathbb{R}\times S_X}\left|F_{n,j}(y|x)-F_j(y|x)\right| = o_P\left(\max\left(\sqrt{\frac{|\ln c|^q}{nc^p}},c^\delta\right)\right), j\in\{1,2\}.$$

We are now able to study the second term in (9).

Theorem 3.1. *Let $(X_1,Z_1),\dots,(X_n,Z_n)$ be a sample of independent copies of the random vector (X,Z) where the distribution of Z, given $X=x$, satisfies (2) and (3), X follows a distribution with a bounded density function f_X, and such that there exists $b>0$ satisfying $f(x)\ge b$, $\forall x\in S_X\subset\mathbb{R}^p$. Assume also Assumptions $(\mathcal{H})$, $(\mathcal{K}_2)$ and $(\mathcal{F})$.*

Consider now a sequence c tending to 0 as $n \to \infty$ such that for some $q > 1$

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

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

$$\inf_{x \in S_X} \lambda\left[\{u \in B_0(1) : x - cu \in S_X\}\right] > \varepsilon,$$

where λ denotes the Lebesgue measure. Let $u_n \to \infty$ and $h_n \to 0$ in such a way that for any $\delta \in (0, \min(\delta_{F_1}, \delta_{F_2}))$

$$nh_n^p r_n := nh_n^p \max\left(\sqrt{\frac{|\ln c|^q}{nc^p}}, c^\delta\right) \longrightarrow 0, \tag{10}$$

$$nh_n^p \overline{F}_Z(u_n|x) \longrightarrow \infty, \tag{11}$$

then for any $s < 0$ with $j \in \{0, 1, 2, 3\}$ or $(s, j) = (0, 0)$, we have

$$\sqrt{\frac{nh_n^p}{\overline{F}_Z(u_n|x) f_X(x)}} \left\{\breve{T}_n(K, s, j|x) - T_n(K, s, j|x)\right\} = o_P(1).$$

Using Theorem 3.1 we can now establish the main theorem of this paper, stating consistency and asymptotic normality of the conditional η estimator, in case of general conditional marginal distribution functions, which are estimated with kernel estimators.

Theorem 3.2. *Under the same assumptions as in Theorem 3.1, let $x \in Int(S_X)$ and suppose that $h_n^{\delta_\eta \wedge \delta_\varepsilon} \ln u_n \to 0$. Then with probability tending to 1, there exist sequences of solutions $\{\breve{\eta}_n(x), \breve{\delta}_{Z,n}(x)\}$ of the estimating equations (4) and (5) such that $\{\breve{\eta}_n(x), \breve{\delta}_{Z,n}(x)\} \xrightarrow{\Pr} \{\eta_0(x), 0\}$.*

If additionally (6), (7) and (8) hold, then

$$\sqrt{nh_n^p \overline{F}_Z(u_n|x) f_X(x)} \begin{bmatrix} \breve{\eta}_n(x) - \eta_0(x) \\ \breve{\delta}_{Z,n}(x) \end{bmatrix} \rightsquigarrow \mathcal{N}_2\left(-\lambda\sqrt{f_X(x)}\boldsymbol{C}^{-1}(\widetilde{\rho})\boldsymbol{B}(\widetilde{\rho})\boldsymbol{d}, \boldsymbol{C}^{-1}(\widetilde{\rho})\boldsymbol{B}(\widetilde{\rho})\boldsymbol{\Sigma}(\widetilde{\rho})\boldsymbol{B}^\top(\widetilde{\rho})\boldsymbol{C}^{-1}(\widetilde{\rho})\right).$$

The result of Theorem 3.2 follows directly from the decomposition (9) and Theorem 3.1, and therefore we omit the proof of it from the paper.

4. A simulation study

Our aim in this section is to illustrate the performance of our robust conditional tail dependence coefficient estimator with a small simulation study in case $p = 1$. The joint conditional distribution function of the pair has the following form:

$$\Pr\left\{1 - F_1(Y^{(1)}|x) < y_1, 1 - F_2(Y^{(2)}|x) < y_2 \middle| X = x\right\} = C(y_1, y_2|x),$$

where $C(., .|x)$ is one of the three copulas:

Case 1: The BB6 copula in Joe [25, p. 152] defined for $\theta(x) \geq 1$ and $\zeta(x) \geq 1$, as follows

$$C(y_1, y_2|x) = 1 - \left[1 - \exp\left\{-\left(\left[-\ln\left\{1 - (1 - y_1)^{\theta(x)}\right\}\right]^{\zeta(x)} + \left[-\ln\left\{1 - (1 - y_2)^{\theta(x)}\right\}\right]^{\zeta(x)}\right)^{\frac{1}{\zeta(x)}}\right\}\right]^{\frac{1}{\theta(x)}}.$$

For this model exact independence is obtained for $\theta(x) = 1$ with $\zeta(x) = 1$, and perfect dependence is achieved if either $\theta(x) \to \infty$ or $\zeta(x) \to \infty$. We can easily see that in case $\theta(x) > 1$, this model satisfies our model assumption (2) with $\eta(x) = 2^{-\frac{1}{\zeta(x)}}$, $C(x) = \{\theta(x)\}^{2^{\frac{1}{\zeta(x)}} - 1}$ and $\tau(x) = 1$. We take $X \sim U(1, 6)$, $\theta(x) = 2$ and $\zeta(x) = x$.

Case 2: The Farlie Gumbel Morgenstern (FGM) copula defined for $\zeta(x) \in (-1, 1]$, as follows

$$C(y_1, y_2|x) = y_1 y_2 \left\{1 + \zeta(x)(1 - y_1)(1 - y_2)\right\}.$$

Exact independence is obtained for $\zeta(x) = 0$, and perfect dependence is not attainable under this model. Clearly, for $\zeta(x) \neq 0$, our model assumption (2) is also satisfied, with $\eta(x) = 1/2$, $C(x) = 1 + \zeta(x)$ and $\tau(x) = 1$. We take $X \sim U(-0.9, 1)$ and $\zeta(x) = x$.

Case 3: The BB9 or Crowder copula in Joe [25, p. 154] defined for $\alpha(x) \geq 0$ and $\theta(x) \geq 1$, as follows

$$C(y_1, y_2|x) = \exp\left(-\left[\{\alpha(x) - \ln(y_1)\}^{\theta(x)} + \{\alpha(x) - \ln(y_2)\}^{\theta(x)} - \{\alpha(x)\}^{\theta(x)}\right]^{\frac{1}{\theta(x)}} + \alpha(x)\right).$$

Exact independence is obtained for $\theta(x) = 1$ or $\alpha(x) \to \infty$, and perfect dependence for $\theta(x) \to \infty$. We can check that this model has the form of (2) with $\eta(x) = 2^{-\frac{1}{\theta(x)}}$, $C(x) = \exp\left[\alpha(x)\left\{1 - 2^{\frac{1}{\theta(x)}}\right\}\right]$, but $\tau(x) = 0$. That means that this case does

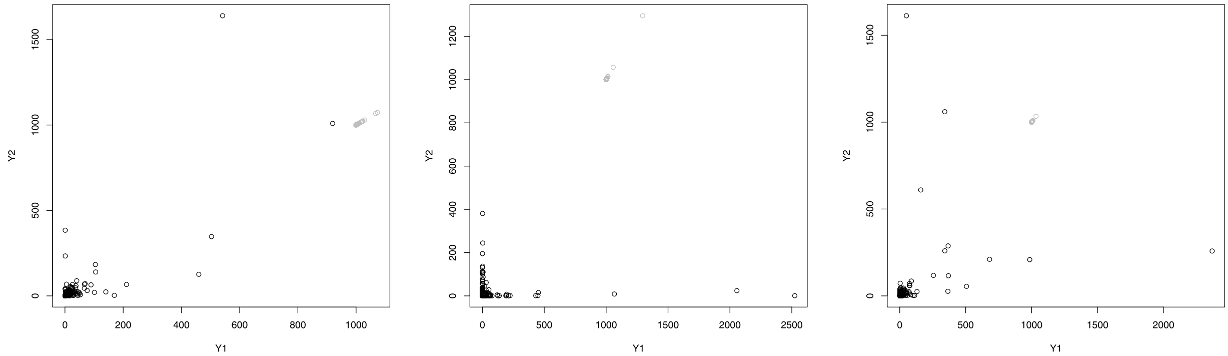


Fig. 1. Scatterplots of simulated datasets from BB6 (left), FGM (middle) and BB9 (right), when there is 5% contamination (grey circles).

not fit our model assumption, but we use it here to show the robustness of our approach in case our main assumption is violated. We set $X \sim U(1,6)$, $\alpha(x) = 1$ and $\theta(x) = x$.

These copula models are combined with unit Fréchet marginal distributions, leading to

$$F(y_1, y_2|x) = \exp(-1/y_1) + \exp(-1/y_2) - 1 + C\{1 - \exp(-1/y_1), 1 - \exp(-1/y_2)|x\}.$$

Contamination will be introduced 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 ε denotes the fraction of contamination, and F_c is the contaminating distribution function. We take here

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

i.e., the distribution function of completely dependent unit Fréchet random variables, translated by a. This choice for F_c means that we put contamination along the diagonal, which corresponds to perfect dependent contamination. We take for a quantile 0.999 of the unit Fréchet distribution, and consider $\varepsilon = 0$, 5% and 10%. In Fig. 1 we show the scatterplots of datasets generated from the BB6, FGM and BB9 models, with $\varepsilon = 0.05$. In these plots, the non-contaminated sample is represented as black circles whereas the contaminated pairs are represented as grey circles. When focusing on the main data, the BB6 and BB9 models have clearly stronger dependence in their extremes than the FGM model, where for the latter the large values occur closer to the axes. For all models this contamination is quite severe, but the situation is worse for the FGM model, which has weaker dependence in its extremes compared to the BB6 and BB9 models. We refer to the discussion of the simulation results below.

Concerning the kernel function K, we take the bi-quadratic function

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

We have also considered other kernels like the Epanichnikov and triweight kernel functions, and the results are insensitive with respect to the kernel choice. To compute our estimator $\breve{\eta}_n(x)$, two sequences h_n and c have to be chosen. Concerning c, we can use the following cross validation criterion introduced by [35], and used in an extreme value context by [8,9] and [17]: for $j \in \{1,2\}$

$$c_j := \arg\min_{c\in\mathcal{C}_g} \sum_{i=1}^n \sum_{k=1}^n \left\{ \mathbb{1}_{\left(Y_i^{(j)} \le Y_k^{(j)}\right)} - \widetilde{F}_{n,-i,j}(Y_k^{(j)}|X_i) \right\}^2, \text{ where } \widetilde{F}_{n,-i,j}(y|x) := \frac{\sum_{k=1,k\ne i}^n K_c(x-X_k)\mathbb{1}_{(Y_k^{(j)}\le y)}}{\sum_{k=1,k\ne i}^n K_c(x-X_k)},$$

and $\mathcal{C}_g$ is a grid of values of c. We take $\mathcal{C}_g = R_X \times \{0.05, 0.10, \dots, 0.30\}$, where R_X is the range of the covariate X. The bandwidth parameter h_n is determined from the condition $nh_n\sqrt{|\ln c|^q/(nc)} \longrightarrow 0$, by taking $h_n = R_X\sqrt{c/(n|\ln c|^\kappa)}$, where $\kappa > q$ and $c := \min(c_1, c_2)$. Next to h_n and c, our estimation procedure also requires the selection of a threshold parameter u_n. As usual in extreme value statistics, this parameter will be set at the $(k+1)$th largest of the $\breve{Z}$ for which the X coordinate is in $B(x, h_n)$.

As mentioned before, we only estimate $\eta(x)$ and $\delta_Z(u_n|x)$ with the MDPD method, while the parameter ρ is fixed at some value. Here we set $\rho = -1$, which is a mis-specification.

For each of the above distributions we simulate $N = 500$ samples of size $n = 1000$. The results of the simulation experiment are reported in Figs. 2 till 7. In Fig. 2 we show the mean of $\breve{\eta}_n(x)$ as a function of k for $\alpha = 0$ (solid line), $\alpha = 0.5$ (dashed line) and $\alpha = 1$ (dotted line) for the BB6 copula. The true value of η is represented by the horizontal reference line. The columns of the figure represent three different values of x, while the rows correspond with the contamination percentages, 0%, 5% and 10% from top to bottom. Fig. 3 displays the empirical mean squared error (MSE) as a function of

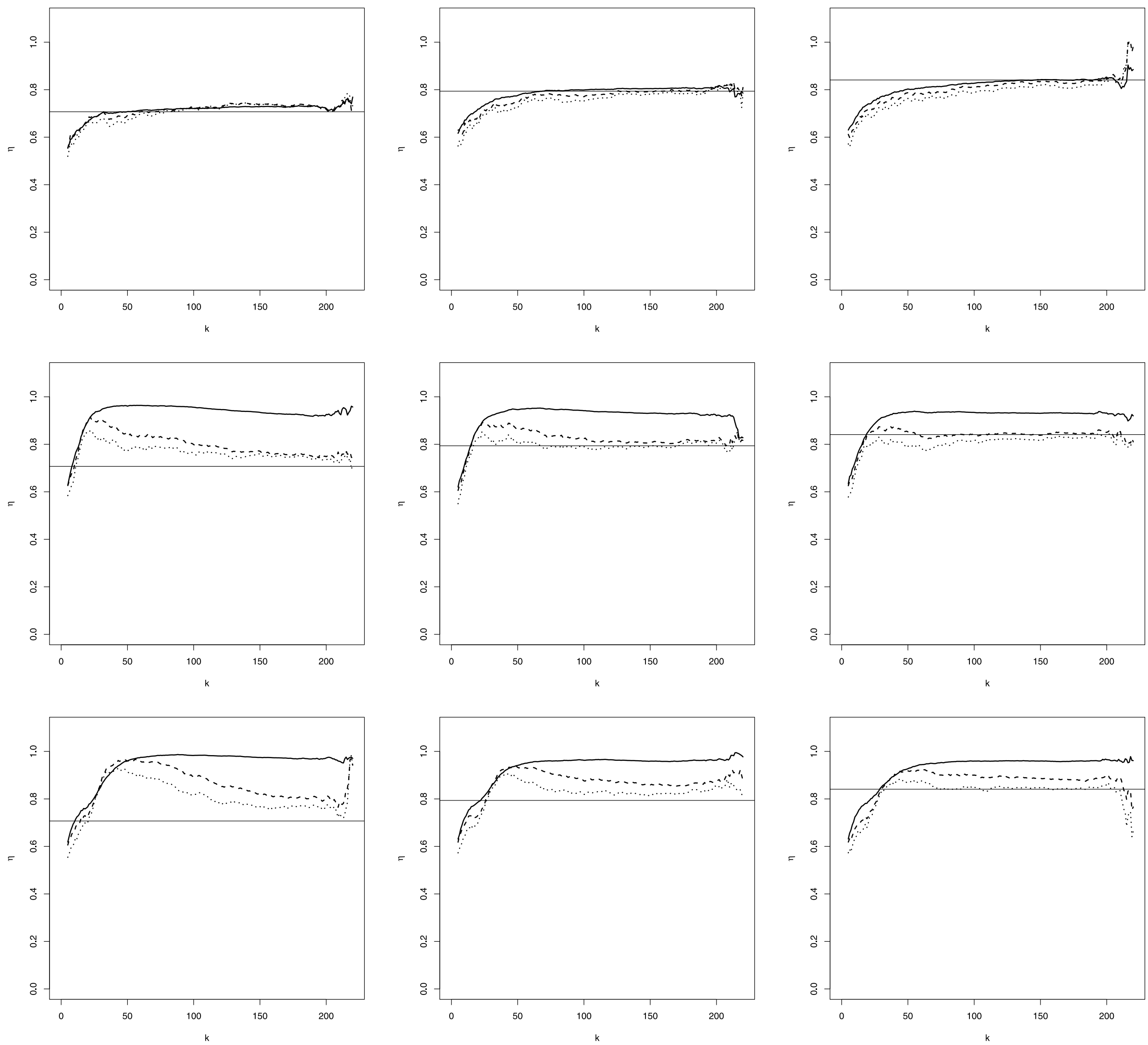


Fig. 2. BB6 simulation with (shifted) diagonal contamination. Mean of $\check{\eta}_n(x)$ with $\alpha = 0$ (solid line), $\alpha = 0.5$ (dashed line) and $\alpha = 1$ (dotted line), as a function of k at $x = 2$ (left), $x = 3$ (middle) and $x = 4$ (right). From top to bottom: 0%, 5% and 10% of contamination.

k, but has otherwise a layout that is similar to Fig. 2. Concerning the selection of h_n and c, we note the following. In a first step we compute the optimal h_n and c for each dataset using the above mentioned cross-validation criterion. This implies that the range of k varies from one dataset to the other, so means and MSE's would be based on a different number of observations when plotted as a function of k. In order to avoid this we take the median of the h_n and c values obtained in the 500 simulations and use this for all estimations. Figs. 4 and 5, and Figs. 6 and 7, show the corresponding results for the Farlie Gumbel Morgenstern and BB9 copula, respectively. From the simulation we can draw the following conclusions:

- In absence of contamination, the estimators show generally a quite stable pattern for a wide range of k, close to the true value of η, despite the mis-specification of the parameter ρ. In terms of MSE we see that, the estimator with $\alpha = 0$, which corresponds to maximum likelihood, performs best, followed by $\alpha = 0.5$ and $\alpha = 1$. This can be explained as follows: in terms of bias the estimators with different values of α perform similarly, while for the variance we have that $\alpha = 0$, corresponding to maximum likelihood, performs best. It is well-known that the efficiency of the MDPDE decreases with increasing α, see, e.g., [2].
- When there is contamination, then the non-robust estimator ($\alpha = 0$) is clearly affected, with a sample mean that can be far from the true value, while the robust estimators generally stay closer to the true value. The estimator

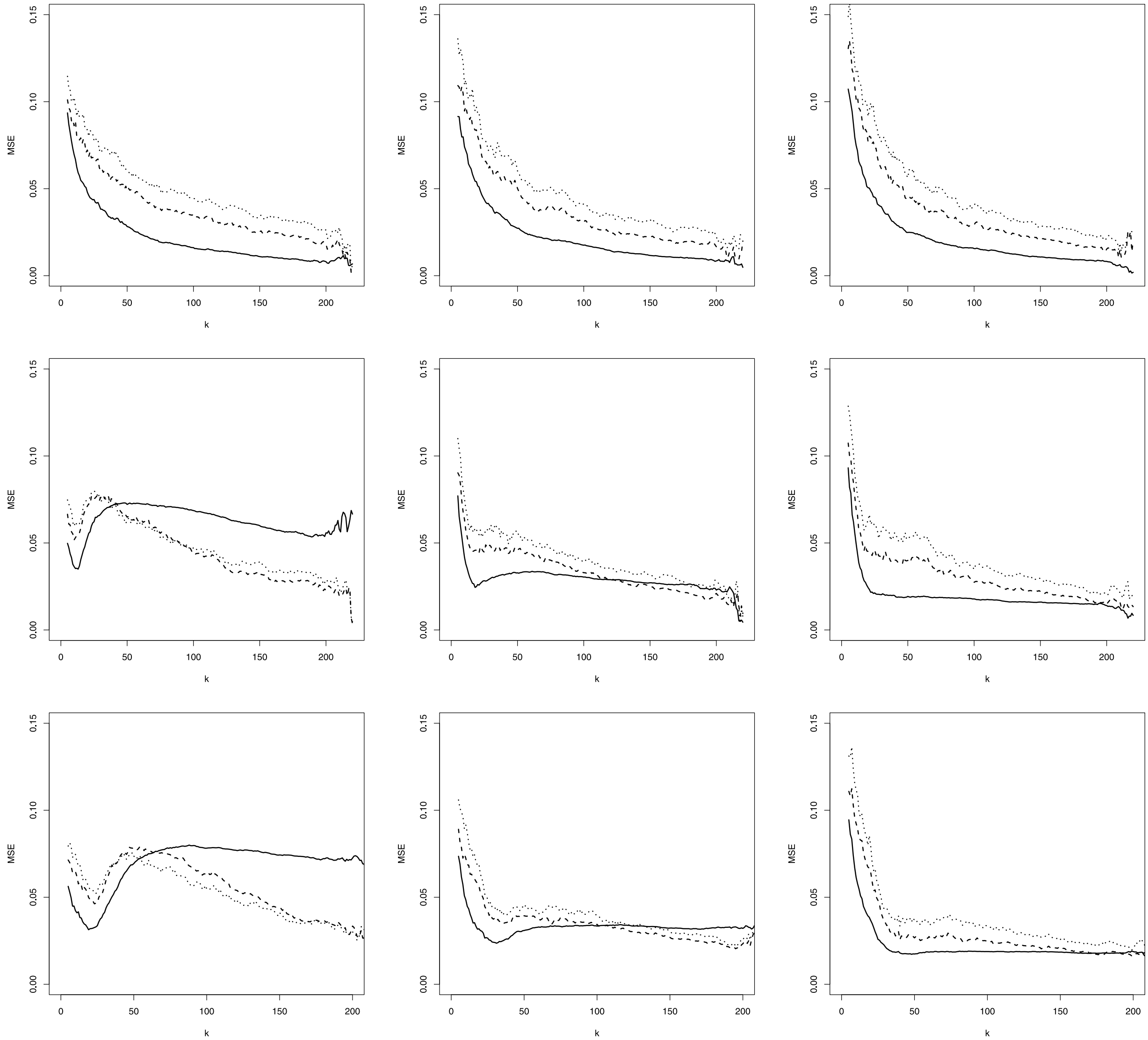


Fig. 3. BB6 simulation with (shifted) diagonal contamination. MSE of $\breve{\eta}_n(x)$ with $\alpha = 0$ (solid line), $\alpha = 0.5$ (dashed line) and $\alpha = 1$ (dotted line), as a function of k at $x = 2$ (left), $x = 3$ (middle) and $x = 4$ (right). From top to bottom: 0%, 5% and 10% of contamination.

with $\alpha = 1$, which offers the highest robustness, performs best in terms of bias. In terms of minimal MSE, using $\alpha = 0.5$ gives the best result. The advantage of $\alpha = 1$ in terms of bias is offset by its increased variance compared to $\alpha = 0.5$.

- The performance of the estimators deteriorates under increasing contamination percentages.
- For the BB6 distribution, the effect of the contamination is strongest for the smaller x values. This could be expected, as the dependence in the data is weakest at the smaller x. The dependence increases with x, and therefore at $x = 4$ the effect of contamination on the diagonal is least.
- The Farlie Gumbel Morgenstern distribution has $\eta = 0.5$, corresponding to near independence. For this distribution, contamination on the diagonal is clearly very severe.
- For the BB9 distribution, which does not satisfy our model assumptions, we still have very good estimation results, which also illustrates the robustness of our methodology with respect to violation of the model assumption. Also here we see that the effect of the contamination is biggest at the x positions where the dependence in the data is weakest.
- It appears that our estimate with $\alpha = 0.5$ gives some protection against contamination and it seems to be a safe choice whatever the framework, with or without contamination. Indeed, in case of contamination, we observe that

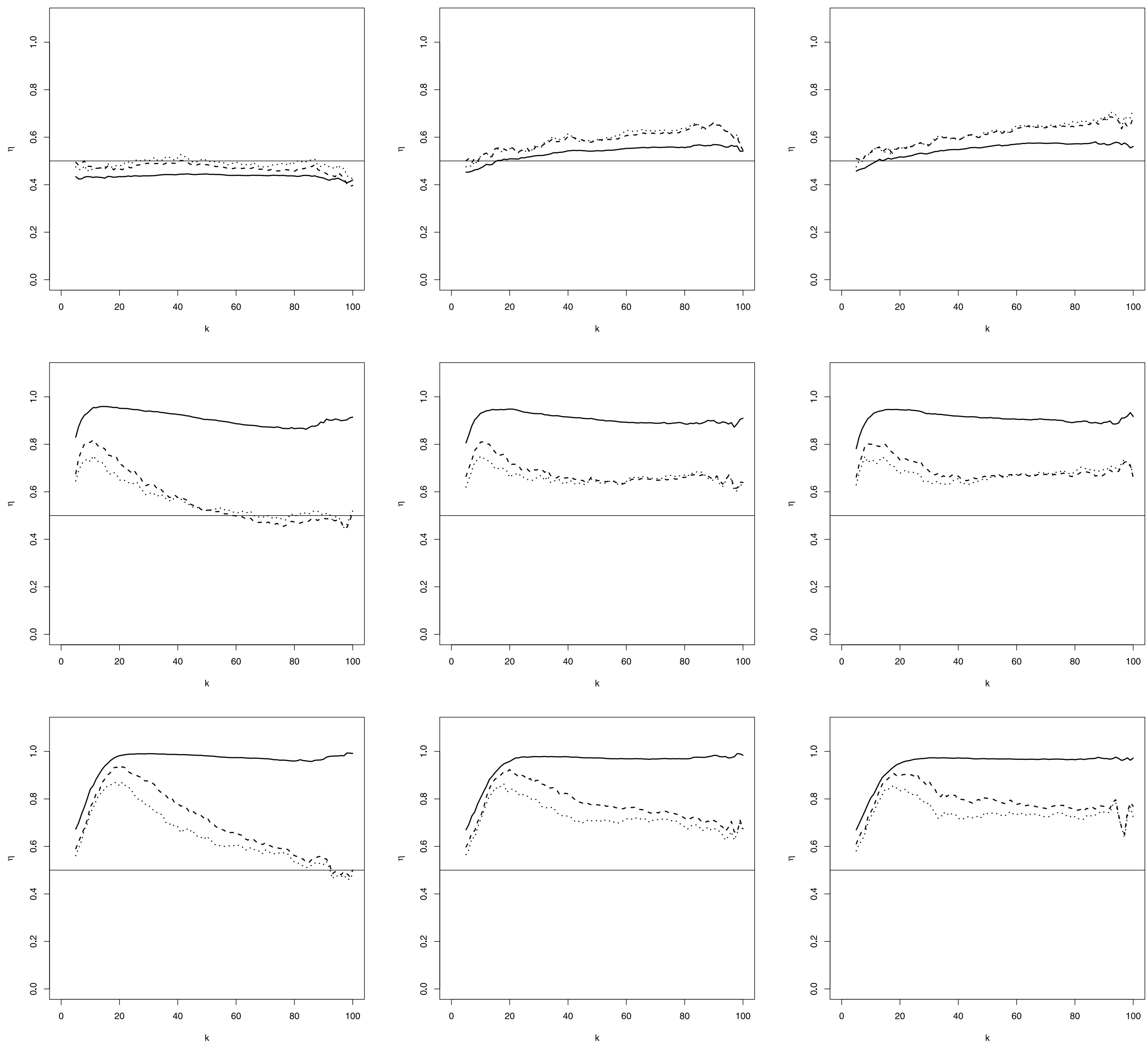


Fig. 4. FGM simulation with (shifted) diagonal contamination. Mean of $\breve{\eta}_n(x)$ with $\alpha = 0$ (solid line), $\alpha = 0.5$ (dashed line) and $\alpha = 1$ (dotted line), as a function of k at $x = -0.5$ (left), $x = 0.5$ (middle) and $x = 0.8$ (right). From top to bottom: 0%, 5% and 10% of contamination.

increasing α from 0 to 0.5 clearly improves the estimation but increasing α further to 1 does not lead to clear further improvements, since although it offers a highest robustness, the advantage of $\alpha = 1$ in terms of bias is offset by its increased variance compared to $\alpha = 0.5$. This is in line with the findings of [15] in the context without covariates.

5. A real data analysis

In this section, the proposed methodology is applied to a dataset of air pollution measurements. In environmental science, one needs to consider simultaneous high levels of several pollutants, possibly combined with high temperatures, as these may pose a major threat to human health. Estimation of the extreme dependence is thus of crucial importance in this context. 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 under consideration contains monthly maxima on, among others, temperature, and ground-level ozone, carbon monoxide and particulate matter concentrations, for the time period January 1999 to December 2013. These data are collected at stations spread over the U.S. We will estimate the extreme dependence between ground-level ozone and particulate matter concentrations, conditional on the covariates time and location, where the latter is expressed by latitude and longitude. The method is implemented with

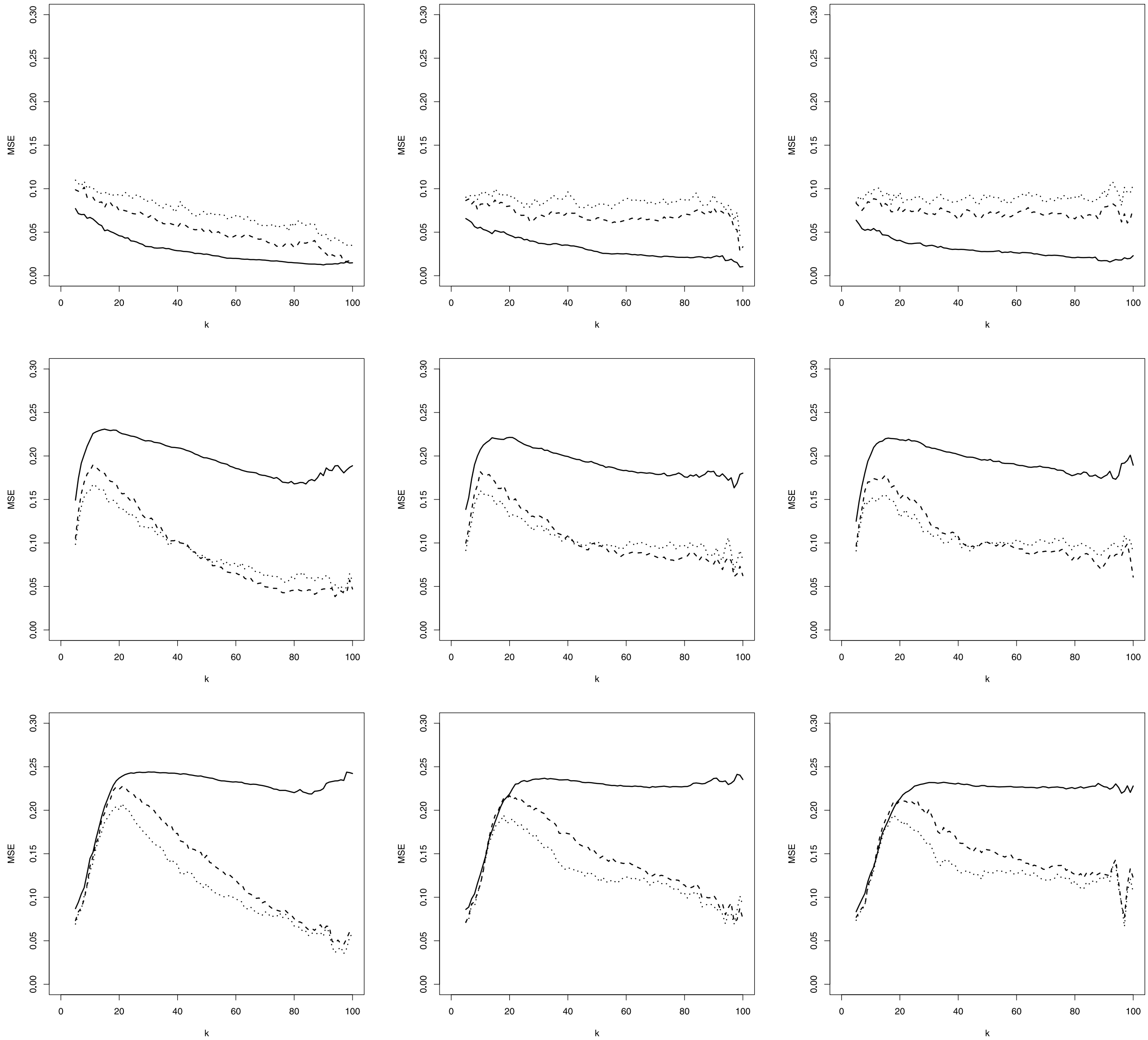


Fig. 5. FGM simulation with (shifted) diagonal contamination. MSE of $\breve{\eta}_n(x)$ with $\alpha = 0$ (solid line), $\alpha = 0.5$ (dashed line) and $\alpha = 1$ (dotted line), as a function of k at $x = -0.5$ (left), $x = 0.5$ (middle) and $x = 0.8$ (right). From top to bottom: 0%, 5% and 10% of contamination.

the same cross-validation criteria as in the simulations, though for convenience we rescaled each covariate to the interval [0, 1]. For practice we recommend rescaling the covariates in case they are measured on different scales, since by the construction of the estimator, and also in the theoretical analyses, only a single bandwidth parameter is used. As for the kernel function, we used the bi-quadratic kernel, generalized to the case $p = 3$, as follows

$$K_{h_n}(x) = \frac{1}{h_n^3} K\left(\frac{\|x\|}{h_n}\right),$$

where $x \in \mathbb{R}^3$, and $\|.\|$ denotes the Euclidean norm. In Fig. 8, we show the estimate of $\eta(x)$ with $\alpha = 0$ (solid line), $\alpha = 0.5$ (dashed line) and $\alpha = 1$ (dotted line) for the city of Los Angeles at different points in time. The reported estimate is median$\{\breve{\eta}_n(x); k = n^*/2, \ldots, n^* - 1\}$, where n^* denotes the number of observations in $B(x, h_n)$. Overall, the extreme dependence between ground-level ozone and particulate matter concentrations shows a seasonal pattern, where the dependence is typically stronger in summer than winter. For some months the estimate with $\alpha = 0$ differs noticeably from those obtained with $\alpha = 0.5$ and $\alpha = 1$, which indicates the presence of observations that are disturbing for the estimation of the dependence structure. In Fig. 9, we show the estimate $\breve{\eta}_n(x)$ with $\alpha = 0$ (solid line), $\alpha = 0.5$ (dashed line) and $\alpha = 1$ (dotted line) for months 59 (November 2003) and 100 (April 2007) as a function of k. For month 59,

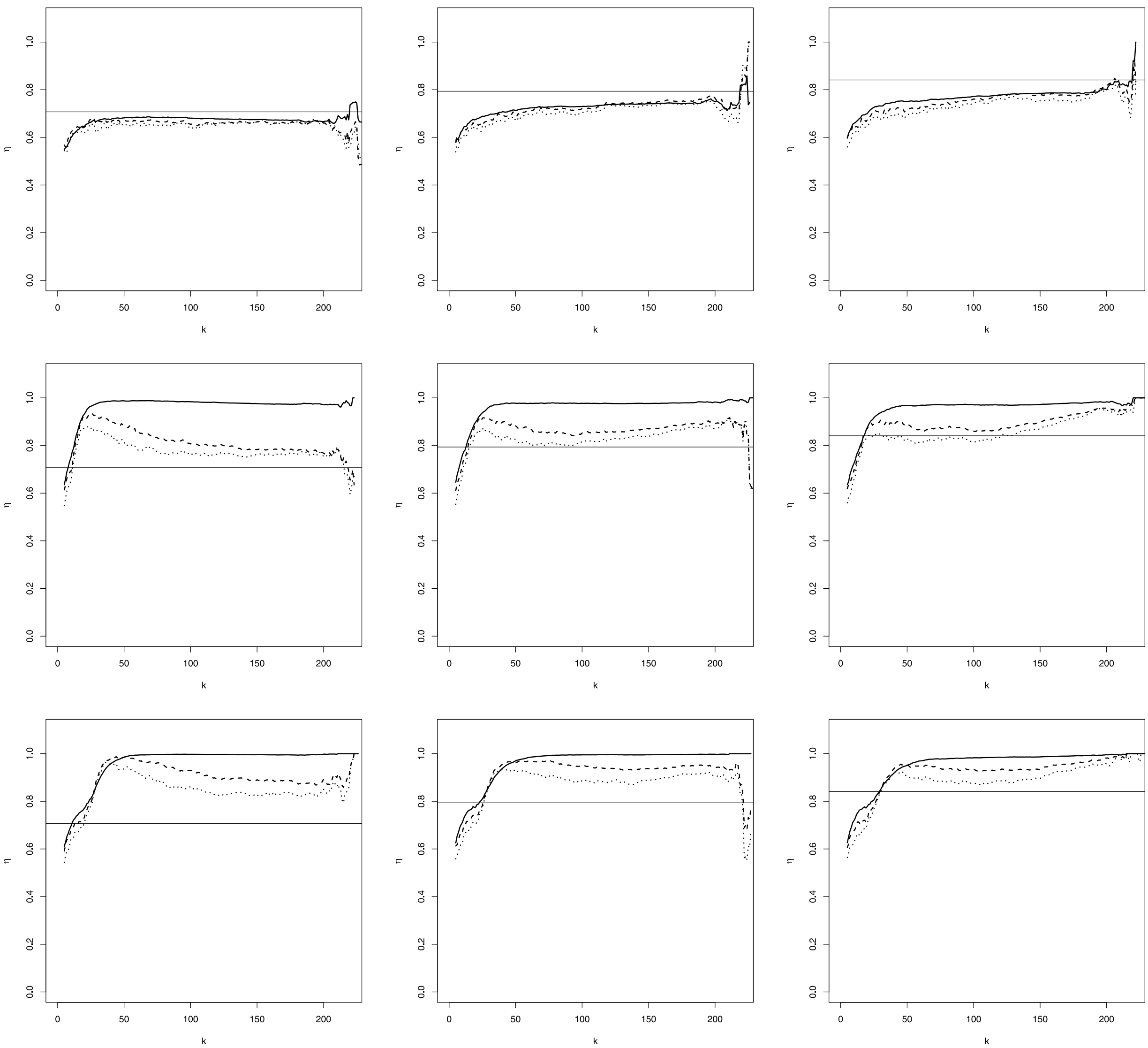


Fig. 6. BB9 simulation with (shifted) diagonal contamination. Mean of $\breve{\eta}_n(x)$ with $\alpha = 0$ (solid line), $\alpha = 0.5$ (dashed line) and $\alpha = 1$ (dotted line), as a function of k at $x = 2$ (left), $x = 3$ (middle) and $x = 4$ (right). From top to bottom: 0%, 5% and 10% of contamination.

the robust estimates show a stable pattern around $\eta(x) = 1$ for the second half of the k range, while the non-robust estimate shows nearly no stability as a function of k. On the contrary, for month 100, the robust estimates are below the non-robust estimate. Again the robust estimates show a stable horizontal pattern for the second half of the k range, which is not present for the non-robust estimate. In order to gain some further insight in the differences between the non-robust and robust estimates for these two months, we also constructed the Pareto quantile–quantile plots of the $\breve{Z}$ for which the corresponding X coordinates belong to $B(x, h_n)$, see Fig. 10. In case a Pareto-type model fits the data, the Pareto quantile–quantile plot will be linear in the largest observations with a slope reflecting η. We refer to [5], section 2.3.5, for a general discussion of Pareto quantile–quantile plots. For month 59 the quantile–quantile plot bends down for the largest ten $\breve{Z}$, which can explain why the non-robust estimate is below the robust estimate. The quantile–quantile plot for month 100 shows a clearly outlying upper observation, which apparently pulls the non-robust estimate up compared to the robust estimates. Although identification of outlying observations is not the main topic of this paper, such local Pareto quantile–quantile plots can be used in conjunction with the estimates of η obtained for different values of α to pinpoint suspect observations.

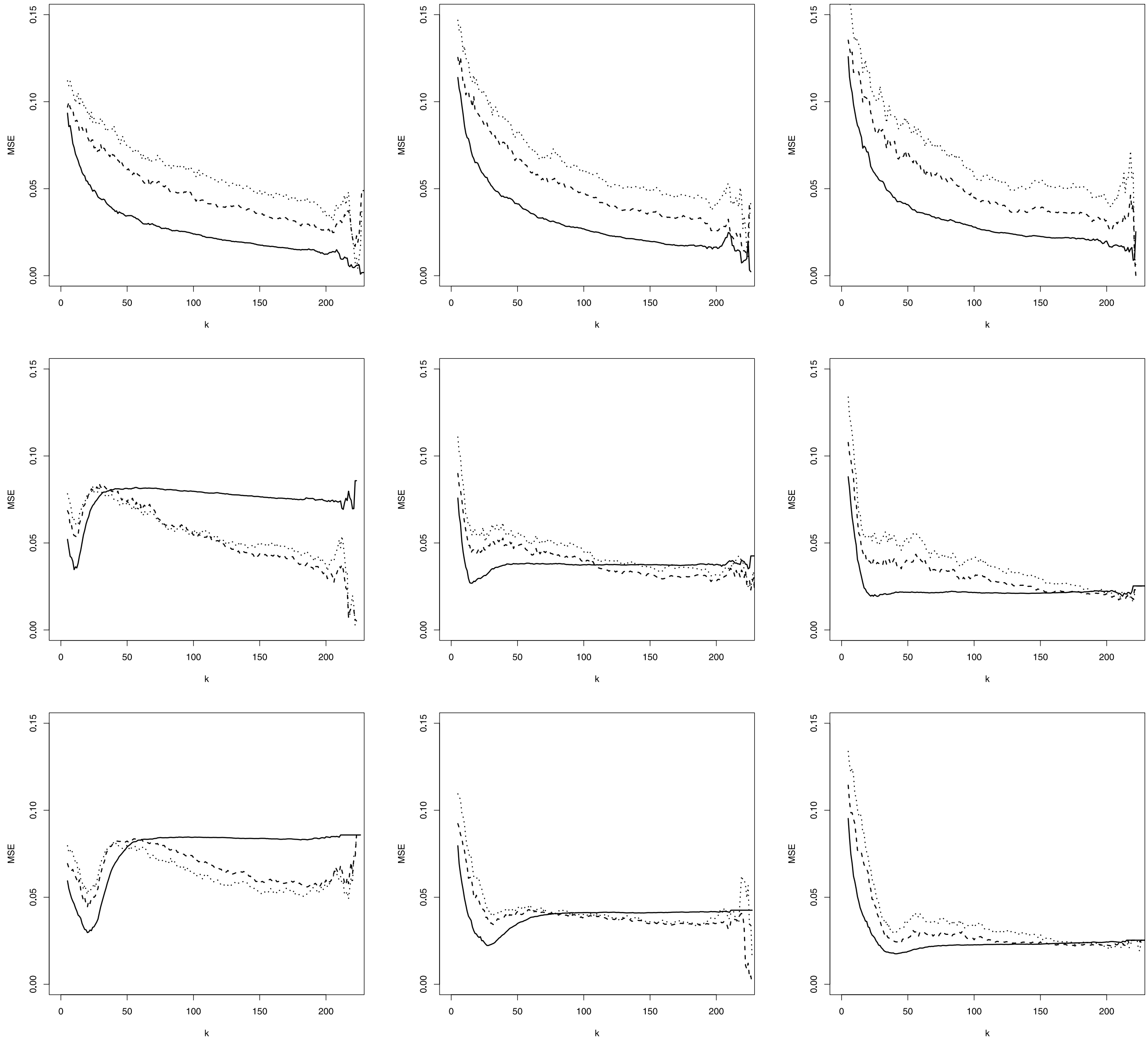


Fig. 7. BB9 simulation with (shifted) diagonal contamination. MSE of $\check{\eta}_n(x)$ with $\alpha = 0$ (solid line), $\alpha = 0.5$ (dashed line) and $\alpha = 1$ (dotted line), as a function of k at $x = 2$ (left), $x = 3$ (middle) and $x = 4$ (right). From top to bottom: 0%, 5% and 10% of contamination.

CRediT authorship contribution statement

Yuri Goegebeur: Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing. **Armelle Guillou:** Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing. **Nguyen Khanh Le Ho:** Methodology, Formal analysis, Software. **Jing Qin:** Methodology, Formal analysis, Software, Writing - original draft, Writing - review & editing.

Acknowledgments

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, France (IEA-00179). Computation/simulation for the work described in this paper was supported by the DeIC National HPC Centre, SDU, Denmark. 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.

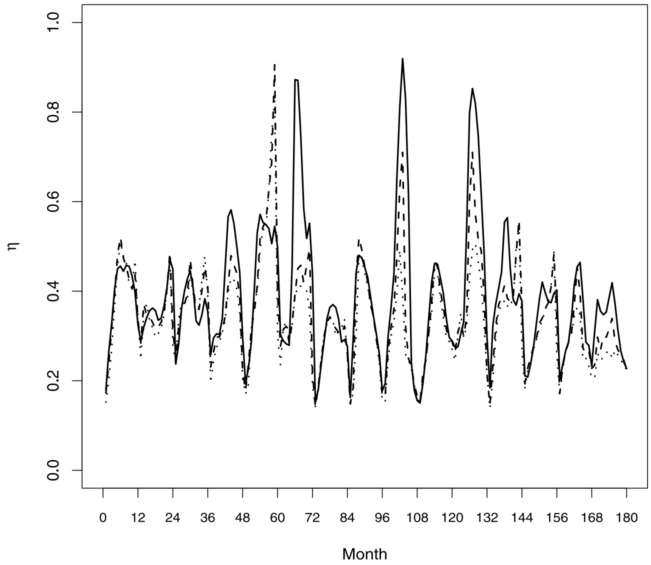


Fig. 8. Air pollution data. Time plot of $\breve{\eta}_n(x)$ with $\alpha = 0$ (solid line), $\alpha = 0.5$ (dashed line) and $\alpha = 1$ (dotted line) for the city of Los Angeles.

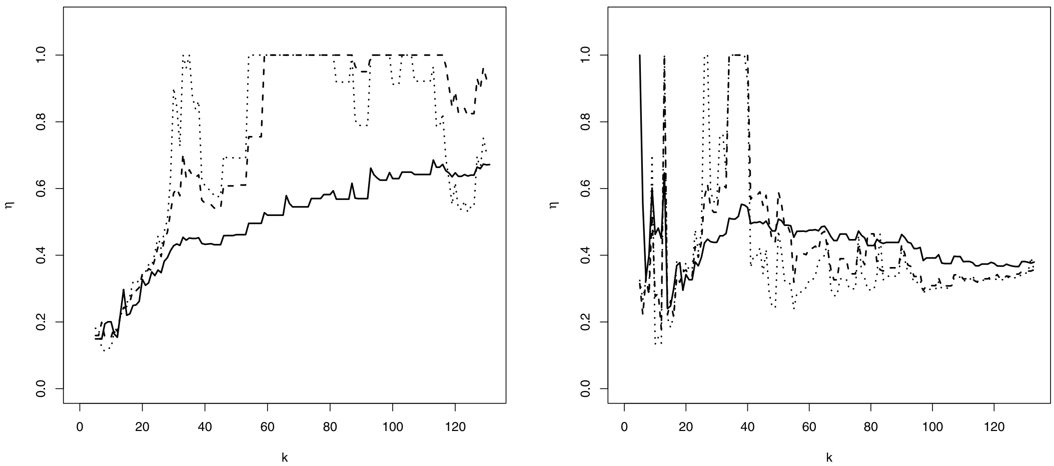


Fig. 9. Air pollution data. $\breve{\eta}_n(x)$ with $\alpha = 0$ (solid line), $\alpha = 0.5$ (dashed line) and $\alpha = 1$ (dotted line) as a function of k for months 59 (left) and 100 (right).

Appendix. Proofs

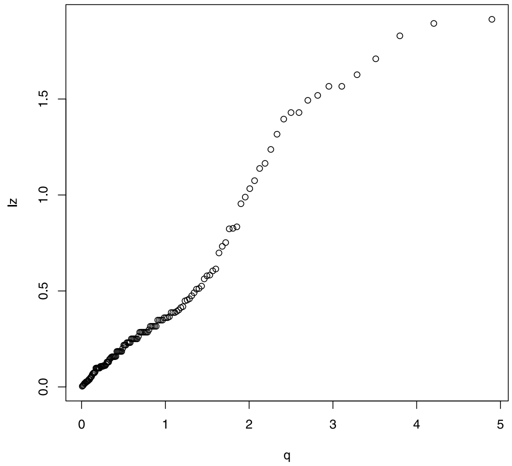
Proof of Theorem 2.1. The first step consists to show that, under our assumptions,

$$
\begin{aligned}
&E\{T_n(K,s,j|x)\} = f_X(x)\overline{F}_Z(u_n|x)\eta_0^j(x)j! \\
&\times \left(\frac{1}{\{1-s\eta_0(x)\}^{j+1}} - \frac{\delta_Z(u_n|x)}{\eta_0(x)} \left[\frac{1}{\{1-s\eta_0(x)\}^{j+1}} - \frac{1-\rho_0(x)}{\{1-\rho_0(x)-s\eta_0(x)\}^{j+1}} \right] + o(\delta_Z(u_n|x)) \right. \\
&\qquad \left. + O\left(h_n^{\delta_{f_X}\wedge\delta_C}\right) + O\left(h_n^{\delta_\eta}\ln u_n\right)\right),
\end{aligned}
$$

where the $o(\delta_Z(u_n|x))$ and $O(.)$ terms are uniform in $s \in [S, 0]$.

To obtain this result for the case $j > 0$, use the following decomposition

$$
\begin{aligned}
E\{T_n(K,s,j|x)\} &= f_X(x)\int_1^\infty p(z)\overline{F}_Z(u_n z|x)dz \\
&\quad + \int_{S_K} K(v)\int_1^\infty p(z)\overline{F}_Z(u_n z|x)dz\{f_X(x-h_n v) - f_X(x)\}dv
\end{aligned}
$$



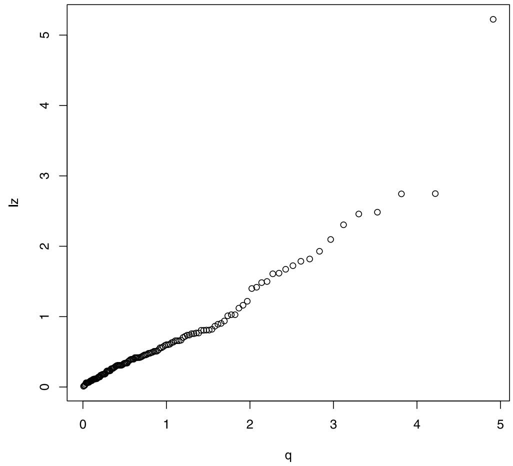


Fig. 10. Air pollution data. Local Pareto quantile–quantile plots for months 59 (left) and 100 (right).

$$+f_X(x)\int_{S_K}K(v)\int_1^\infty p(z)\{\overline{F}_Z(u_nz|x-h_nv)-\overline{F}_Z(u_nz|x)\}dz\,dv$$
$$+\int_{S_K}K(v)\int_1^\infty p(z)\{\overline{F}_Z(u_nz|x-h_nv)-\overline{F}_Z(u_nz|x)\}dz\,\{f_X(x-h_nv)-f_X(x)\}dv,$$

where $p(z):=sz^{s-1}(\ln z)^j+jz^{s-1}(\ln z)^{j-1}$. Each term can be treated using our Hölder-type conditions, which imply in particular that, for n large enough, $z\geq u_n$, and some constants M_1,M_2,M_3

$$\left|\frac{\overline{F}_Z(z|x-h_nv)}{\overline{F}_Z(z|x)}-1\right|\leq M_1\left\{h_n^{\delta_C}+z^{M_2h_n^{\delta_\eta}}h_n^{\delta_\eta}\ln z+|\delta_Z(z|x)|h_n^{\delta_A}+|\delta_Z(z|x)|z^{M_3h_n^{\delta_\varepsilon}}h_n^{\delta_\varepsilon}\ln z\right\} \tag{12}$$

combined with a slight modification of Proposition 2.3 in [6] which ensures that

$$\sup_{z\geq 1}z^{\frac{1}{\eta(x)}}\left|\frac{\overline{F}_Z(u_nz|x)}{\overline{F}_Z(u_n|x)}-\overline{G}\{z;\eta(x),\delta_Z(u_n|x),\rho(x)\}\right|=o(|\delta_Z(u_n|x)|)\text{ as }u_n\to\infty.$$

In case $j=0$ we obtain

$$E\{T_n(K,s,0|x)\}=\int_{S_K}K(v)\int_1^\infty p(z)\overline{F}_Z(u_nz|x-h_nv)dzf_X(x-h_nv)dv$$
$$+\int_{S_K}K(v)\overline{F}_Z(u_n|x-h_nv)f_X(x-h_nv)dv,$$

where $p(z)=sz^{s-1}$. Both terms can be analysed with decompositions similar to the ones used for the case $j>0$.

Then we can follow the lines of proofs of Theorem 1 and Corollary 1 in [13] in order to achieve the proof of Theorem 2.1.

Proof of Theorem 3.1. First remark that

$$\sqrt{\frac{nh_n^p}{\overline{F}_Z(u_n|x)f_X(x)}}\left\{\breve{T}_n(K,s,j|x)-T_n(K,s,j|x)\right\}$$
$$=\sqrt{\frac{nh_n^p}{\overline{F}_Z(u_n|x)f_X(x)}}\left[\breve{T}_n(K,s,j|x)-T_n(K,s,j|x)-E\left\{\breve{T}_n(K,s,j|x)-T_n(K,s,j|x)\right\}\right]$$
$$+\sqrt{\frac{nh_n^p}{\overline{F}_Z(u_n|x)f_X(x)}}E\left\{\breve{T}_n(K,s,j|x)-T_n(K,s,j|x)\right\}$$
$$=:R_{n,1}+R_{n,2}.$$

We will study the two terms $R_{n,1}$ and $R_{n,2}$ separately. First, we start with the term $R_{n,1}$. Define for any $s<0$ with $j\in J:=\{0,1,2,3\}$ or $(s,j)=(0,0)$

$$g_{\xi,n}^{(s,j)}(y_1,y_2,v):=\sqrt{\frac{h_n^p}{\overline{F}_Z(u_n|x)f_X(x)}}K_{h_n}(x-v)q_{\xi,n}^{(s,j)}(y_1,y_2,v)$$

with

$$q_{\xi,n}^{(s,j)}(y_1,y_2,v) := \left\{\frac{Z_\xi(y_1,y_2,v)}{u_n}\right\}^s \left\{\ln \frac{Z_\xi(y_1,y_2,v)}{u_n}\right\}^j \mathbb{1}_{\{Z_\xi(y_1,y_2,v)>u_n\}},$$

$$Z_\xi(y_1,y_2,v) := \min\left\{\frac{1}{|1-\xi_1(y_1,y_2,v)|}, \frac{1}{|1-\xi_2(y_1,y_2,v)|}\right\},$$

and measurable $\xi \in H := \{\xi = (\xi_1,\xi_2); \xi : \mathbb{R}\times\mathbb{R}\times S_X \to \mathbb{R}^2\}$.

For convenience, denote $\xi_n = (F_{n,1}, F_{n,2})$ and $\xi_0 = (F_1, F_2)$. According to Lemma 3.1, $r_n^{-1}|\xi_n - \xi_0|$ converges in probability towards the null function $H_0 = \{0\}$ in H, endowed with the norm $\|\xi\|_H := \|\xi_1\|_\infty + \|\xi_2\|_\infty$ for any $\xi \in H$. Consider now the class

$$\mathcal{E}_n^{(s,j)}(b) := \{g_{\xi_0+r_n\xi,n}^{(s,j)} - g_{\xi_0,n}^{(s,j)} : \xi \in H, \|\xi\|_H \le b\},$$

with envelope function $G_n^{(s,j)}(b)$. Our aim is to apply Theorem 2.3 in [32]. To reach this goal, we need to introduce some notations. Let P denote the law of the vector $(Y^{(1)}, Y^{(2)}, X)$ and define the expectation of any real-valued measurable function f under P by $Pf = \int f dP$.

We have now to show the two following results:

Assertion 1. For any $s<0$ with $j \in J$ or $(s,j)=(0,0)$, we have

$$\sqrt{n} P G_n^{(s,j)}(b_n) \longrightarrow 0 \text{ for all } b_n \to 0,$$

and

Assertion 2. For any $s<0$ with $j \in J$ or $(s,j)=(0,0)$, we have

$$P\left[\{G_n^{(s,j)}(b)\}^2\right] \longrightarrow 0.$$

Proof of Assertion 1. We start to consider the case where $s<0$ with $j\in J$. As a first step we derive an envelope function for $\mathcal{E}_n^{(s,j)}(b_n)$. We have

$$\begin{aligned}
\left|q_{\xi_0+r_n\xi,n}^{(s,j)} - q_{\xi_0,n}^{(s,j)}\right| &= \bigg|\int_1^\infty p(a)\mathbb{1}_{(u_n<u_na<Z_{\xi_0+r_n\xi})}\mathbb{1}_{(Z_{\xi_0+r_n\xi}>u_n)}da - \int_1^\infty p(a)\mathbb{1}_{(u_n<u_na<Z_{\xi_0})}\mathbb{1}_{(Z_{\xi_0}>u_n)}da \\
&\quad + \mathbb{1}_{(j=0)}\mathbb{1}_{(Z_{\xi_0+r_n\xi}>u_n)} - \mathbb{1}_{(j=0)}\mathbb{1}_{(Z_{\xi_0}>u_n)}\bigg| \\
&\le \int_1^\infty |p(a)| \left|\mathbb{1}_{(u_n<u_na<Z_{\xi_0+r_n\xi})}\mathbb{1}_{(Z_{\xi_0+r_n\xi}>u_n)} - \mathbb{1}_{(u_n<u_na<Z_{\xi_0})}\mathbb{1}_{(Z_{\xi_0}>u_n)}\right| da \\
&\quad + \mathbb{1}_{(j=0)}\left|\mathbb{1}_{(Z_{\xi_0+r_n\xi}>u_n)} - \mathbb{1}_{(Z_{\xi_0}>u_n)}\right| \\
&\le \int_1^\infty |p(a)| \left|\mathbb{1}_{(u_n<u_na<Z_{\xi_0})} - \mathbb{1}_{(u_n<u_na<Z_{\xi_0+r_n\xi})}\right| \mathbb{1}_{(Z_{\xi_0}>u_n)}da \\
&\quad + \int_1^\infty |p(a)|\, \mathbb{1}_{(u_n<u_na<Z_{\xi_0+r_n\xi})}\left|\mathbb{1}_{(Z_{\xi_0+r_n\xi}>u_n)} - \mathbb{1}_{(Z_{\xi_0}>u_n)}\right| da + \mathbb{1}_{(j=0)}\left|\mathbb{1}_{(Z_{\xi_0+r_n\xi}>u_n)} - \mathbb{1}_{(Z_{\xi_0}>u_n)}\right| \\
&\le \int_1^\infty |p(a)|\, \mathbb{1}_{\{\min(Z_{\xi_0},Z_{\xi_0+r_n\xi})\le u_na\le \max(Z_{\xi_0},Z_{\xi_0+r_n\xi})\}}\mathbb{1}_{(Z_{\xi_0}>u_n)}\, da \\
&\quad + \int_1^\infty |p(a)|\, \mathbb{1}_{(u_n<u_na<Z_{\xi_0+r_n\xi})}\mathbb{1}_{\{\min(Z_{\xi_0},Z_{\xi_0+r_n\xi})\le u_n\le \max(Z_{\xi_0},Z_{\xi_0+r_n\xi})\}}\, da \\
&\quad + \mathbb{1}_{(j=0)}\mathbb{1}_{\{\min(Z_{\xi_0},Z_{\xi_0+r_n\xi})\le u_n\le \max(Z_{\xi_0},Z_{\xi_0+r_n\xi})\}}.
\end{aligned} \tag{13}$$

Remark now that

$$\begin{aligned}
&\left(u_n\, a \in [\min(Z_{\xi_0}, Z_{\xi_0+r_n\xi}); \max(Z_{\xi_0}, Z_{\xi_0+r_n\xi})]\right) \\
&= \Bigg(u_n\, a \in \Bigg[\min\left\{\min\left(\frac{1}{|1-F_1-r_n\xi_1|}, \frac{1}{|1-F_2-r_n\xi_2|}\right), \min\left(\frac{1}{1-F_1}, \frac{1}{1-F_2}\right)\right\}; \\
&\qquad \max\left\{\min\left(\frac{1}{|1-F_1-r_n\xi_1|}, \frac{1}{|1-F_2-r_n\xi_2|}\right), \min\left(\frac{1}{1-F_1}, \frac{1}{1-F_2}\right)\right\}\Bigg]\Bigg) \\
&= \Bigg(\frac{1}{u_n\, a} \in [\min\{\max(|1-F_1-r_n\xi_1|, |1-F_2-r_n\xi_2|), \max(1-F_1, 1-F_2)\}; \\
&\qquad \max\{\max(|1-F_1-r_n\xi_1|, |1-F_2-r_n\xi_2|), \max(1-F_1, 1-F_2)\}]\Bigg) \\
&\subset \left(\frac{1}{u_n\, a} \in [\min(|1-F_1-r_n\xi_1|, 1-F_1), \max(|1-F_1-r_n\xi_1|, 1-F_1)]\right)
\end{aligned}$$

$$\begin{aligned}&\cup\left(\frac{1}{u_n\,a}\in\left[\min\left(|1-F_2-r_n\xi_2|,\,1-F_2\right),\,\max\left(|1-F_2-r_n\xi_2|,\,1-F_2\right)\right]\right)\\ &\subset\left(\frac{1}{u_n\,a}\in[1-F_1-r_n\,b_n,\,1-F_1+r_n\,b_n]\right)\cup\left(\frac{1}{u_n\,a}\in[1-F_2-r_n\,b_n,\,1-F_2+r_n\,b_n]\right)\\ &=: A_{n,1}(a)\cup A_{n,2}(a).\end{aligned}\tag{14}$$

Also,

$$\begin{aligned}\mathbb{1}_{(Z_{\xi_0+r_n\xi}>u_na)} &= \mathbb{1}_{(|1-F_1-r_n\xi_1|<\frac{1}{u_na},|1-F_2-r_n\xi_2|<\frac{1}{u_na})}\leq\mathbb{1}_{(-\frac{1}{u_na}-r_nb_n<1-F_1<\frac{1}{u_na}+r_nb_n,-\frac{1}{u_na}-r_nb_n<1-F_2<\frac{1}{u_na}+r_nb_n)}\\ &\leq \mathbb{1}_{(1-F_1<\frac{1}{u_na}+r_nb_n,1-F_2<\frac{1}{u_na}+r_nb_n)},\end{aligned}$$

and, taking into account that

$$\mathbb{1}_{(Z_{\xi_0}>u_n)}=\mathbb{1}_{(1-F_1<\frac{1}{u_n},1-F_2<\frac{1}{u_n})},$$

we obtain

$$\mathbb{1}_{(Z_{\xi_0+r_n\xi}>u_na)}\leq\mathbb{1}_{\{1-F_1<\frac{1}{u_n}\left(\frac{1}{a}+r_nu_nb_n\right),1-F_2<\frac{1}{u_n}\left(\frac{1}{a}+r_nu_nb_n\right)\}}=\mathbb{1}_{(Z_{\xi_0}>\frac{u_n}{\frac{1}{a}+r_nu_nb_n})}.\tag{15}$$

Thus, combining (13)–(15), we obtain the following envelope for $\mathcal{E}_n^{(s,j)}(b_n)$:

$$\begin{aligned}G_n^{(s,j)}(b_n) &:= \sqrt{\frac{h_n^p}{\overline{F}_Z(u_n|x)f_X(x)}}K_{h_n}(x-.)\left[\int_1^\infty|p(a)|\,\mathbb{1}_{\{A_{n,1}(a)\cup A_{n,2}(a)\}}\mathbb{1}_{(Z_{\xi_0}>u_n)}\,da\right.\\ &\left.+\int_1^\infty|p(a)|\,\mathbb{1}_{\left(Z_{\xi_0}>\frac{u_n}{\frac{1}{a}+r_nu_nb_n}\right)}\mathbb{1}_{\{A_{n,1}(1)\cup A_{n,2}(1)\}}\,da+\mathbb{1}_{(j=0)}\mathbb{1}_{\{A_{n,1}(1)\cup A_{n,2}(1)\}}\right],\end{aligned}\tag{16}$$

with

$$\begin{aligned}\sqrt{n}PG_n^{(s,j)}(b_n) &= \sqrt{\frac{nh_n^p}{\overline{F}_Z(u_n|x)f_X(x)}}\left\{E\left(K_{h_n}(x-X)\int_1^\infty|p(a)|\,E\left[\mathbb{1}_{\{A_{n,1}(a)\cup A_{n,2}(a)\}}\mathbb{1}_{(Z_{\xi_0}>u_n)}\middle|X\right]\,da\right)\right.\\ &+E\left(K_{h_n}(x-X)\int_1^\infty|p(a)|\,E\left[\mathbb{1}_{\left(Z_{\xi_0}>\frac{u_n}{\frac{1}{a}+r_nu_nb_n}\right)}\mathbb{1}_{\{A_{n,1}(1)\cup A_{n,2}(1)\}}\middle|X\right]\,da\right)\\ &\left.+\mathbb{1}_{(j=0)}E\left(K_{h_n}(x-X)E\left[\mathbb{1}_{\{A_{n,1}(1)\cup A_{n,2}(1)\}}\middle|X\right]\right)\right\}\\ &=\sqrt{\frac{nh_n^p}{\overline{F}_Z(u_n|x)f_X(x)}}\left(\int_{S_K}K(v)\int_1^\infty|p(a)|\,E\left[\mathbb{1}_{\{A_{n,1}(a)\cup A_{n,2}(a)\}}\mathbb{1}_{(Z_{\xi_0}>u_n)}\middle|X=x-h_nv\right]\,daf_X(x-h_nv)dv\right.\\ &+\int_{S_K}K(v)\int_1^\infty|p(a)|\,E\left[\mathbb{1}_{\left(Z_{\xi_0}>\frac{u_n}{\frac{1}{a}+r_nu_nb_n}\right)}\mathbb{1}_{\{A_{n,1}(1)\cup A_{n,2}(1)\}}\middle|X=x-h_nv\right]\,daf_X(x-h_nv)dv\\ &\left.+\mathbb{1}_{(j=0)}\int_{S_K}K(v)E\left[\mathbb{1}_{\{A_{n,1}(1)\cup A_{n,2}(1)\}}\middle|X=x-h_nv\right]f_X(x-h_nv)dv\right)=:\sqrt{\frac{nh_n^p}{\overline{F}_Z(u_n|x)f_X(x)}}\left(T_1+T_2+T_3\right).\end{aligned}$$

Consider T_1. By the Cauchy–Schwarz inequality

$$T_1\leq\int_{S_K}K(v)\int_1^\infty|p(a)|\sqrt{\Pr\{A_{n,1}(a)\cup A_{n,2}(a)|X=x-h_nv\}\overline{F}_Z(u_n|x-h_nv)}daf_X(x-h_nv)dv.$$

The sub-additivity of probability measures and some straightforward calculations give then

$$\begin{aligned}\Pr\{A_{n,1}(a)\cup A_{n,2}(a)|X=x-h_nv\} &\leq \Pr\{A_{n,1}(a)|X=x-h_nv\}+\Pr\{A_{n,2}(a)|X=x-h_nv\}\\ &=\int_0^1\mathbb{1}_{(\frac{1}{u_n\,a}\in[z-r_nb_n,z+r_nb_n])}dz+\int_0^1\mathbb{1}_{(\frac{1}{u_n\,a}\in[z-r_nb_n,z+r_nb_n])}dz\leq 2r_nb_n+2r_nb_n=4r_nb_n.\end{aligned}\tag{17}$$

Thus

$$T_1\leq 2\sqrt{r_nb_n\overline{F}_Z(u_n|x)}\int_1^\infty|p(a)|\,da\int_{S_K}K(v)\sqrt{\frac{\overline{F}_Z(u_n|x-h_nv)}{\overline{F}_Z(u_n|x)}}f_X(x-h_nv)dv,$$

and hence, by (12) and the Hölder continuity of f_X, we have $T_1=O\left(\sqrt{r_n\overline{F}_Z(u_n|x)}\right)$.

As for T_2 use again the Cauchy–Schwarz inequality and (17) to obtain

$$T_2 \leq 2\sqrt{r_n b_n} \int_1^\infty |p(a)|\, da \int_{S_K} K(v) \sqrt{\overline{F}_Z\left(\frac{u_n}{1+r_n u_n b_n}\middle| x - h_n v\right)} f_X(x-h_n v) dv$$

$$= 2\sqrt{r_n b_n \overline{F}_Z\left(\frac{u_n}{1+r_n u_n b_n}\middle| x\right)} \int_1^\infty |p(a)|\, da \int_{S_K} K(v) \sqrt{\frac{\overline{F}_Z\left(\frac{u_n}{1+r_n u_n b_n}\middle| x-h_n v\right)}{\overline{F}_Z\left(\frac{u_n}{1+r_n u_n b_n}\middle| x\right)}} f_X(x-h_n v) dv.$$

Note that under our assumptions, $r_n u_n \to 0$, as $n \to \infty$. Thus using (12) and the fact that

$$\frac{\overline{F}_Z\left(\frac{u_n}{1+r_n u_n b_n}\middle| x\right)}{\overline{F}_Z(u_n|x)} \to 1, \tag{18}$$

we have that $T_2 = O\left(\sqrt{r_n \overline{F}_Z(u_n|x)}\right)$.

By similar arguments, we get $T_3 = O(r_n) = o\left(\sqrt{r_n \overline{F}_Z(u_n|x)}\right)$ under our assumptions (10) and (11). Combining the above

$$\sqrt{n} P G_n^{(s,j)}(b_n) = O\left(\sqrt{n h_n^p r_n}\right).$$

Now, we move to the case $(s,j) = (0,0)$ and use a similar proof. In that case, using (17), we have

$$\sqrt{n} P G_n^{(0,0)}(b_n) = \sqrt{\frac{n h_n^p}{\overline{F}_Z(u_n|x) f_X(x)}} \int_{S_K} K(v) \Pr\left\{A_{n,1}(1) \cup A_{n,2}(1) \middle| X = x - h_n v\right\} f_X(x-h_n v) dv$$

$$\leq 4 r_n b_n \sqrt{\frac{n h_n^p}{\overline{F}_Z(u_n|x) f_X(x)}} \int_{S_K} K(v) f_X(x-h_n v) dv = O\left(\sqrt{\frac{n h_n^p}{\overline{F}_Z(u_n|x)}}\, r_n\, b_n\right) = o\left(\sqrt{n h_n^p r_n}\right)$$

which achieves the proof of Assertion 1 in case $(s,j) = (0,0)$.

Proof of Assertion 2. Again, we start to look at the case $s < 0$ and $j \in J$. From (16) and straightforward bounds, we deduce that

$$\left\{G_n^{(s,j)}(b)\right\}^2 \leq \frac{h_n^p}{\overline{F}_Z(u_n|x) f_X(x)} K_{h_n}^2(x-\cdot) \left[\left\{\int_1^\infty |p(a)| da\right\} \int_1^\infty |p(a)| \mathbb{1}_{\left(Z_{\xi_0} > \frac{u_n}{\frac{1}{a} + r_n u_n b}\right)} \mathbb{1}_{\{A_{n,1}(1) \cup A_{n,2}(1)\}} da \right.$$
$$\left. + 3\left\{\int_1^\infty |p(a)| da\right\} \int_1^\infty |p(a)| \mathbb{1}_{\{A_{n,1}(a) \cup A_{n,2}(a)\}} \mathbb{1}_{(Z_{\xi_0} > u_n)} da + \left\{1 + 4\int_1^\infty |p(a)| da\right\} \mathbb{1}_{\{A_{n,1}(1) \cup A_{n,2}(1)\}}\right].$$

Since $\int_1^\infty |p(a)| da < \infty$, using again the Cauchy–Schwarz inequality combined with (17), we deduce that

$$P\left[\left\{G_n^{(s,j)}(b)\right\}^2\right] \leq \frac{C\sqrt{r_n}}{\overline{F}_Z(u_n|x) f_X(x)} \int_{S_K} K^2(v) \int_1^\infty |p(a)| \sqrt{\overline{F}_Z\left(\frac{u_n}{\frac{1}{a} + r_n u_n b}\middle| x - h_n v\right)}\, da\, f_X(x-h_n v)\, dv$$
$$+ \frac{C\sqrt{r_n}}{\overline{F}_Z(u_n|x) f_X(x)} \int_{S_K} K^2(v) \int_1^\infty |p(a)| \sqrt{\overline{F}_Z(u_n|x-h_n v)}\, da\, f_X(x-h_n v)\, dv$$
$$+ \frac{C r_n}{\overline{F}_Z(u_n|x) f_X(x)} \int_{S_K} K^2(v) f_X(x-h_n v)\, dv,$$

where C is a constant which can change from one line to each other.

Finally, combining (12) with (18), we deduce that

$$P\left[\left\{G_n^{(s,j)}(b)\right\}^2\right] = O\left(\sqrt{\frac{r_n}{\overline{F}_Z(u_n|x)}}\right).$$

The case $(s,j) = (0,0)$ can be dealt with similarly and leads to

$$P\left[\left\{G_n^{(0,0)}(b)\right\}^2\right] = O\left(\frac{r_n}{\overline{F}_Z(u_n|x)}\right).$$

This achieves the proof of Assertion 2.

Combining Assertions 1 and 2 with Theorem 2.3 in [32] yields that $R_{n,1} = o_P(1)$.

Now, it remains to study the term $R_{n,2}$. To this aim, note that, for n large,

$$
\begin{aligned}
|R_{n,2}| &\leq \sqrt{\frac{nh_n^p}{\overline{F}_Z(u_n|x)f_X(x)}} E\left|\breve{T}_n(K,s,j|x) - T_n(K,s,j|x)\right| \\
&\leq \sqrt{n} E\left|g^{(s,j)}_{\xi_n,n}(Y^{(1)},Y^{(2)},X) - g^{(s,j)}_{\xi_0,n}(Y^{(1)},Y^{(2)},X)\right| \leq \sqrt{n} P G^{(s,j)}_n(b),
\end{aligned}
$$

since $\xi_n \in \xi_0 + r_n\mathcal{B}(0,b)$ where $\mathcal{B}(0,b) := \{\xi : \|\xi\|_H \leq b\}$ (where we use the Skorohod representation). According to the proof of Assertion 1, since $b_n \to 0$ can be replaced by any fixed value b without changing the conclusion, we have $R_{n,2} = o(1)$.

Combining the results for $R_{n,1}$ and $R_{n,2}$ achieves the proof of Theorem 3.1.

References

[1] P. Barbe, A.-L. Fougères, C. Genest, On the tail behavior of sums of dependent risks, Astin Bull. 36 (2006) 361–373.
[2] A. Basu, I.R. Harris, N.L. Hjort, M.C. Jones, Robust and efficient estimation by minimizing a density power divergence, Biometrika 85 (1998) 549–559.
[3] J. Beirlant, G. Dierckx, Y. Goegebeur, G. Matthys, Tail index estimation and an exponential regression model, Extremes 2 (1999) 177–200.
[4] J. Beirlant, G. Dierckx, A. Guillou, Bias-reduced estimators for bivariate tail modelling, Insurance Math. Econom. 49 (2011) 18–26.
[5] J. Beirlant, Y. Goegebeur, J. Segers, J. Teugels, Statistics of Extremes – Theory and Applications, Wiley, 2004.
[6] J. Beirlant, E. Joossens, J. Segers, Second-order refined peaks-over-threshold modelling for heavy-tailed distributions, J. Statist. Plann. Inference 139 (2009) 2800–2815.
[7] J. Beirlant, B. Vandewalle, Some comments on the estimation of a dependence index in bivariate extreme value statistics, Statist. Probab. Lett. 60 (2002) 265–278.
[8] A. Daouia, L. Gardes, S. Girard, On kernel smoothing for extremal quantile regression, Bernoulli 19 (2013) 2557–2589.
[9] A. Daouia, L. Gardes, S. Girard, A. Lekina, Kernel estimators of extreme level curves, TEST 20 (2011) 311–333.
[10] L. de Haan, A. Ferreira, Extreme Value Theory: An Introduction, Springer, 2006.
[11] R. Dell'Aquila, P. Embrechts, Extremes and robustness: A contradiction? Financ. Mark. Portf. Manag. 20 (2006) 103–118.
[12] G. Dierckx, Y. Goegebeur, A. Guillou, An asymptotically unbiased minimum density power divergence estimator for the Pareto-tail index, J. Multivariate Anal. 121 (2013) 70–86.
[13] G. Dierckx, Y. Goegebeur, A. Guillou, Local robust and asymptotically unbiased estimation of conditional Pareto type-tails, TEST 23 (2014) 330–355.
[14] G. Draisma, H. Drees, A. Ferreira, L. de Haan, Bivariate tail estimation: dependence in asymptotic independence, Bernoulli 10 (2004) 251–280.
[15] C. Dutang, Y. Goegebeur, A. Guillou, Robust and unbiased estimation of the coefficient of tail dependence, Insurance Math. Econom. 57 (2014) 46–57.
[16] C. Dutang, Y. Goegebeur, A. Guillou, Robust and bias-corrected estimation of extreme failures sets, Sankhya A 78 (2016) 52–86.
[17] M. Escobar-Bach, Y. Goegebeur, A. Guillou, Local robust estimation of the Pickands dependence function, Ann. Statist. 46 (2018) 2806–2843.
[18] A. Feuerverger, P. Hall, Estimating a tail exponent by modelling departure from a Pareto distribution, Ann. Statist. 27 (1999) 760–781.
[19] A. Ghosh, Divergence based robust estimation of the tail index through an exponential regression model, Stat. Methods Appl. 26 (2017) 181–213.
[20] E. Giné, A. Guillou, Rates of strong uniform consistency for multivariate kernel density estimators, Ann. Inst. Henri Poincaré Probab. Stat. 38 (2002) 907–921.
[21] E. Giné, V. Koltchinskii, J. Zinn, Weighted uniform consistency of kernel density estimators, Ann. Probab. 32 (2004) 2570–2605.
[22] Y. Goegebeur, A. Guillou, T. Rietsch, Robust conditional Weibull-type estimation, Ann. Inst. Statist. Math. 67 (2015) 479–514.
[23] M.I. Gomes, M.J. Martins, Bias reduction and explicit estimation of the extreme value index, J. Statist. Plann. Inference 124 (2004) 361–378.
[24] P. Hall, On some simple estimates of an exponent of regular variation, J. R. Stat. Soc. Ser. B Stat. Methodol. 44 (1982) 37–42.
[25] H. Joe, Multivariate Models and Dependence Concepts, Chapman and Hall, London, 1997.
[26] S.F. Juárez, W.R. Schucany, Robust and efficient estimation for the generalized Pareto distribution, Extremes 7 (2004) 237–251.
[27] M. Kim, S. Lee, Estimation of a tail index based on minimum density power divergence, J. Multivariate Anal. 99 (2008) 2453–2471.
[28] A.W. Ledford, J.A. Tawn, Modelling dependence within joint tail regions, J. R. Stat. Soc. Ser. B Stat. Methodol. 59 (1997) 475–499.
[29] D. Nolan, D. Pollard, U-processes: rates of convergence, Ann. Statist. 15 (1987) 780–799.
[30] L. Peng, Estimation of the coefficient of tail dependence in bivariate extremes, Statist. Probab. Lett. 43 (1999) 399–409.
[31] L. Peng, A practical way for estimating tail dependence functions, Statist. Sinica 20 (2010) 365–378.
[32] A.W. van der Vaart, J.A. Wellner, Empirical processes indexed by estimated functions, in: Asymptotics: Particles, Processes and Inverse Problems, in: IMS Lecture Notes Monogr. Ser., vol. 55, 2007, pp. 234–252.
[33] B. Vandewalle, J. Beirlant, A. Christmann, M. Hubert, A robust estimator for the tail index of Pareto-type distributions, Comput. Statist. Data Anal. 51 (2007) 6252–6268.
[34] B. Vandewalle, J. Beirlant, M. Hubert, A robust estimator of the tail index based on an exponential regression model, in: M. Hubert, G. Pison, A. Struyf, S. Van Aelst (Eds.), Theory and Applications of Recent Robust Methods, Springer Basel AG, 2004.
[35] Q. Yao, Conditional Predictive Regions for Stochastic Processes, Technical Report, University of Kent at Canterbury, 1999.