



HAL
open science

Local robust and asymptotically unbiased estimation of conditional Pareto-type tails

Goedele Dierckx, Yuri Goegebeur, Armelle Guillou

► **To cite this version:**

Goedele Dierckx, Yuri Goegebeur, Armelle Guillou. Local robust and asymptotically unbiased estimation of conditional Pareto-type tails. *Test*, 2014, 23 (2), 10.1007/s11749-013-0350-6 . hal-01311809

HAL Id: hal-01311809

<https://hal.science/hal-01311809>

Submitted on 11 May 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Local robust and asymptotically unbiased estimation of conditional Pareto-type tails

Goedele Dierckx⁽¹⁾, Yuri Goegebeur⁽²⁾ & Armelle Guillou⁽³⁾

⁽¹⁾ KU Leuven, Faculty of Economics and Business, campus Brussel, Research Centre for Mathematical Economics, Econometrics and Statistics, Warmoesberg 26, 1000 Brussel, Belgium

⁽²⁾ Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark

⁽³⁾ Institut Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg et CNRS, 7 rue René Descartes, 67084 Strasbourg cedex, France

Abstract. We introduce a nonparametric robust and asymptotically unbiased estimator for the tail index of a conditional Pareto-type response distribution in presence of random covariates. The estimator is obtained from local fits of the extended Pareto distribution to the relative excesses over a high threshold using an adjusted minimum density power divergence estimation technique. We derive the asymptotic properties of the proposed estimator under some mild regularity conditions, and also investigate its finite sample performance with a small simulation experiment. The practical applicability of the methodology is illustrated on a dataset of calcium content measurements of soil samples.

AMS Subject Classifications: 62G05, 62G20, 62G32, 62G35.

Keywords: Pareto-type distribution, tail index, bias-correction, density power divergence, local estimation.

1 Introduction

Extreme value statistics deals with drawing inferences about characteristics related to tails of distribution functions, such as indices describing tail decay, extreme quantiles, small exceedance probabilities, and measures of extremal dependence. The literature on the estimation of tail characteristics based on a sample of independent and identically distributed random variables is very elaborate. We refer to Beirlant *et al.* (2004) and de Haan and Ferreira (2006) for recent accounts of the available methodologies. However, a major statistical theme is the description of a variable of primary interest, the dependent variable, in terms of covariates, but this regression point of view on extremes has been studied much less extensively. In the present paper we will study nonparametric robust tail index estimation when the variable of interest Y , assumed to be heavy tailed, is observed simultaneously with a random covariate X .

A conditional response distribution function $F(y; x) := \mathbb{P}(Y \leq y | X = x)$ is said to be of Pareto-type if for some positive function $\gamma(x)$ we can write

$$\begin{aligned}\bar{F}(y; x) &:= 1 - F(y; x) \\ &= y^{-1/\gamma(x)} \ell_F(y; x), \quad y > 0,\end{aligned}\tag{1}$$

where ℓ_F is a slowly varying function at infinity, i.e.

$$\lim_{y \rightarrow \infty} \frac{\ell_F(\lambda y; x)}{\ell_F(y; x)} = 1, \quad \text{for all } \lambda > 0.\tag{2}$$

It is obvious that the tail heaviness of $\bar{F}(y; x)$ is governed by the tail function $\gamma(x)$, where larger values correspond with heavier tails. The Pareto-type model finds many important practical applications and has been systematically used in areas like actuarial science, finance, geology and telecommunications, to name but a few.

The estimation of $\gamma(x)$ in presence of fixed, that is nonrandom covariates, has been addressed to some extent in the recent extreme value literature, and we refer to Chapter 7 in Beirlant *et al.* (2004), and the references therein, for an overview. On the other hand, the random covariate case is much less explored. A parametric maximum likelihood approach was pursued in Wang and Tsai (2009) within the Hall subclass of Pareto-type models (Hall, 1982). Also in the framework of Pareto-type tails, Daouia *et al.* (2011) considered the nonparametric estimation of extreme conditional quantiles, and plugged these conditional quantile estimators into classical estimators for the extreme value index, such as the Hill (1975) and Pickands (1975) estimators. Goegebeur *et al.* (2013) introduced a nonparametric and asymptotically unbiased estimator for $\gamma(x)$ based on locally weighted sums of power transformed excesses over a high threshold. Wang *et al.* (2012) considered the estimation of extreme conditional quantiles for Pareto-type distributions and developed a two step procedure based on quantile regression. Recently, in Daouia *et al.* (2013), the methodology of Daouia *et al.* (2011) was extended to the general max-domain of attraction.

In the present paper we develop a nonparametric robust and asymptotically unbiased estimation procedure for the tail function $\gamma(x)$ of heavy tailed distributions when the covariates are random. The type of robustness we have in mind is robustness against outliers in the response variable. The method is based on local fits of the extended Pareto distribution to the relative excesses over a high threshold within a narrow window in the covariate space. The local fitting is performed by an adjustment of the minimum density power divergence estimation (MDPDE) criterion, originally proposed by Basu *et al.* (1998), to the locally weighted regression setting. 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 proposed model then its contribution to the estimating equations gets less weight and as such its influence on the estimation results becomes dampened. This criterion has already been used for the univariate estimation of heavy tailed distributions. For instance, Kim and Lee (2008) obtained a robust estimator for $\gamma > 0$ by fitting the strict Pareto distribution to the

largest observations in a given dataset with the MDPDE method, whereas Dierckx *et al.* (2013) used this criterion to obtain a robust and asymptotically unbiased estimator. To the best of our knowledge, its application to the nonparametric extreme value regression context is new.

Our paper is organized as follows. In the next section we introduce a nonparametric robust and asymptotically unbiased estimator, obtained from local fits of the extended Pareto distribution to the relative excesses over a high threshold, and establish its weak convergence under suitable regularity conditions. In Section 3 the finite sample performance of the proposed method is evaluated by means of a small simulation experiment. The methodology is illustrated on a dataset of calcium content measurements of soil samples in Section 4. Section 5 concludes the paper. The main arguments of the proofs are given below each result and all the details are postponed to an online appendix.

2 Estimation procedure and asymptotic properties

Basu *et al.* (1998) introduced the idea of the density power divergence for the purpose of developing a robust estimation criterion. In particular, the density power divergence between density functions f and g is given by

$$\Delta_\alpha(f, g) := \begin{cases} \int_{\mathbb{R}} [g^{1+\alpha}(y) - (1 + \frac{1}{\alpha}) g^\alpha(y)f(y) + \frac{1}{\alpha} f^{1+\alpha}(y)] dy, & \alpha > 0, \\ \int_{\mathbb{R}} \log \frac{f(y)}{g(y)} f(y) dy, & \alpha = 0. \end{cases} \quad (3)$$

For the purpose of estimation, f is assumed to be the true (typically unknown) density of the data, whereas g is a parametric model, depending on a parameter vector θ which is determined by minimizing the empirical version of (3). In the present paper we will adjust this criterion to the local estimation context with focus on estimating distributional tails.

Let (X_i, Y_i) , $i = 1, \dots, n$, be independent realizations of the random vector $(X, Y) \in \mathbb{R}^p \times \mathbb{R}_{+,0}$, where X has a distribution with joint density function b , and $\bar{F}(y; x)$ is of Pareto-type, though also satisfying the following second order condition. Denote by RV_β the class of the regularly varying functions at infinity with index β , i.e. Lebesgue measurable ultimately positive functions z satisfying $\lim_{t \rightarrow \infty} z(tx)/z(t) = x^\beta$ for all $x > 0$.

Condition (\mathcal{R}) . Let $\gamma(x) > 0$ and $\rho(x) < 0$ be constants. The conditional distribution function $F(y; x)$ is such that $y^{1/\gamma(x)} \bar{F}(y; x) \rightarrow C(x) \in (0, \infty)$ as $y \rightarrow \infty$ and the function $\delta(\cdot; x)$ defined via

$$\bar{F}(y; x) = C(x)y^{-1/\gamma(x)}(1 + \gamma(x)^{-1}\delta(y; x)),$$

is ultimately nonzero, of constant sign and $|\delta| \in RV_{\rho(x)/\gamma(x)}$.

Condition (\mathcal{R}) is not very restrictive. It is satisfied by most commonly used Pareto-type models like the Burr, Student t, F, Fréchet and log-gamma distributions. Therefore it is well accepted in the area of extreme value statistics.

Now, consider the extended Pareto distribution (Beirlant *et al.*, 2004, Beirlant *et al.*, 2009), with distribution function given by

$$G(z; \gamma, \delta, \rho) = \begin{cases} 1 - [z(1 + \delta - \delta z^{\rho/\gamma})]^{-1/\gamma}, & z > 1, \\ 0, & z \leq 1, \end{cases} \quad (4)$$

and density function

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

where $\gamma > 0$, $\rho < 0$, and $\delta > \max\{-1, \gamma/\rho\}$. It is well-known that for distribution functions satisfying (\mathcal{R}) , one can approximate the conditional distribution function of $Z := Y/u$, given that $Y > u$, where u denotes a high threshold value, by the extended Pareto distribution. Indeed, as shown in Beirlant *et al.* (2009), one has that

$$\sup_{z \geq 1} \left| \frac{\bar{F}(uz; x)}{\bar{F}(u; x)} - \bar{G}(z; \gamma(x), \delta(u; x), \rho(x)) \right| = o(\delta(u; x)), \quad \text{if } u \rightarrow \infty.$$

Clearly, based on this result, one can obtain an estimator for $\gamma(x)$ by fitting the extended Pareto distribution to the relative excesses over a high threshold. This has been pursued in the univariate context using a maximum likelihood procedure by Beirlant *et al.*, (2009), and further generalized by Dierckx *et al.* (2013) who applied the MDPDE criterion. As is well-known in extreme value statistics, by taking the second order behavior of F explicitly into account in the estimation stage one obtains asymptotically unbiased estimators for the extreme value index (see e.g. Beirlant *et al.*, 1999, Feuerverger and Hall, 1999).

In the present context we will develop a nonparametric, robust and asymptotically unbiased estimator for $\gamma(x)$ by fitting g locally to the relative excesses $Z_i := Y_i/u_n$, $i = 1, \dots, n$, by means of the MDPDE criterion, adjusted to locally weighted estimation, i.e. we minimize

$$\hat{\Delta}_\alpha(\gamma, \delta; \rho) := \frac{1}{n} \sum_{i=1}^n K_{h_n}(x - X_i) \left\{ \int_1^\infty g^{1+\alpha}(z; \gamma, \delta, \rho) dz - \left(1 + \frac{1}{\alpha}\right) g^\alpha(Z_i; \gamma, \delta, \rho) \right\} \mathbf{1}\{Y_i > u_n\}, \quad (5)$$

in case $\alpha > 0$ and

$$\hat{\Delta}_0(\gamma, \delta; \rho) := -\frac{1}{n} \sum_{i=1}^n K_{h_n}(x - X_i) \ln g(Z_i; \gamma, \delta, \rho) \mathbf{1}\{Y_i > u_n\}, \quad (6)$$

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 $h_n \rightarrow 0$ if $n \rightarrow \infty$, $\mathbf{1}\{A\}$ is the indicator function on the event A and u_n is a local non-random threshold sequence satisfying $u_n \rightarrow \infty$ if $n \rightarrow \infty$. 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 MDPDE 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 in (5) and (6) the parameters of g are taken to be constant, i.e. not depending on X_i , which means that, in the language of local polynomial fitting, we perform a local constant estimation. Of course, the parameters γ and δ could also be replaced by polynomials, as was done e.g. in Beirlant and Goegebeur (2004) in the context of local polynomial maximum likelihood estimation of the generalized Pareto distribution, but this will make the derivations more complicated. Also note that in our approach only $\gamma(x)$ and $\delta(u_n; x)$ are estimated by the MDPDE method. The rate parameter $\rho(x)$ will either be fixed or estimated externally in a consistent way. Estimating the second order rate parameter $\rho(x)$ externally is a common approach in extreme value statistics and allows to obtain bias-corrected estimators for $\gamma(x)$ with a smaller asymptotic variance compared to those obtained with an internal estimation of $\rho(x)$; see for instance Gomes *et al.* (2007) for a discussion in the univariate framework.

The MDPD estimators of $(\gamma(x), \delta(u_n; x))$ satisfy the estimating equations

$$\begin{aligned} 0 &= \frac{1}{n} \sum_{i=1}^n K_{h_n}(x - X_i) \mathbf{1}\{Y_i > u_n\} \int_1^\infty g^\alpha(z; \gamma, \delta, \rho) \frac{\partial g(z; \gamma, \delta, \rho)}{\partial \gamma} dz \\ &\quad - \frac{1}{n} \sum_{i=1}^n K_{h_n}(x - X_i) g^{\alpha-1}(Z_i; \gamma, \delta, \rho) \frac{\partial g(Z_i; \gamma, \delta, \rho)}{\partial \gamma} \mathbf{1}\{Y_i > u_n\}, \end{aligned} \quad (7)$$

$$\begin{aligned} 0 &= \frac{1}{n} \sum_{i=1}^n K_{h_n}(x - X_i) \mathbf{1}\{Y_i > u_n\} \int_1^\infty g^\alpha(z; \gamma, \delta, \rho) \frac{\partial g(z; \gamma, \delta, \rho)}{\partial \delta} dz \\ &\quad - \frac{1}{n} \sum_{i=1}^n K_{h_n}(x - X_i) g^{\alpha-1}(Z_i; \gamma, \delta, \rho) \frac{\partial g(Z_i; \gamma, \delta, \rho)}{\partial \delta} \mathbf{1}\{Y_i > u_n\}. \end{aligned} \quad (8)$$

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

$$T_n(K, s, t; x) := \frac{1}{n} \sum_{i=1}^n K_{h_n}(x - X_i) \left(\frac{Y_i}{u_n}\right)^s \left(\ln_+ \frac{Y_i}{u_n}\right)^t \mathbf{1}\{Y_i > u_n\}, \quad (9)$$

where $s \leq 0$ and $t \geq 0$. The motivation for considering this type of statistic is that the estimating equations (7) and (8) only depend on statistics of this form. Note that $\ln_+ x$ is introduced to ensure that $(\ln_+ Y_i/u_n)^t$ is always well defined (t is nonnegative, not necessary integer).

We derive the asymptotic expansion for $\mathbb{E}[T_n(K, s, t; x)]$. First consider the conditional expectation

$$m(u_n, s, t; x) := \mathbb{E} \left[\left(\frac{Y}{u_n}\right)^s \left(\ln_+ \frac{Y}{u_n}\right)^t \mathbf{1}\{Y > u_n\} \middle| X = x \right].$$

Let Γ denote the gamma function, i.e. $\Gamma(\xi) := \int_0^\infty e^{-u} u^{\xi-1} du$, $\xi > 0$.

Lemma 1 *Case (i), $s = t = 0$:*

$$m(u_n, 0, 0; x) = \bar{F}(u_n; x).$$

Case (ii), $s < 0$ or $t > 0$: assume (\mathcal{R}) , then for $u_n \rightarrow \infty$ we have that

$$m(u_n, s, t; x) = \gamma^t(x) \bar{F}(u_n; x) \Gamma(t+1) \left\{ \frac{1}{(1-s\gamma(x))^{t+1}} - \frac{\delta(u_n; x)}{\gamma(x)} \left[\frac{1}{(1-s\gamma(x))^{t+1}} - \frac{1-\rho(x)}{(1-\rho(x)-s\gamma(x))^{t+1}} \right] (1+o(1)) \right\}.$$

Proof of Lemma 1. The result is based on a straightforward integration by parts, an application of Taylor's theorem to \bar{G} and the following slight modification of Proposition 2.3 in Beirlant *et al.* (2009)

$$\sup_{z \geq 1} z^{1/\gamma(x)} \left| \frac{\bar{F}(u_n z; x)}{\bar{F}(u_n; x)} - \bar{G}(z; \gamma(x), \delta(u_n; x), \rho(x)) \right| = o(\delta(u_n; x)), \quad u_n \rightarrow \infty.$$

See more details in the Online Appendix. □

Now let

$$\tilde{m}_n(K, s, t; x) := \mathbb{E} \left[K_{h_n}(x - X) \left(\frac{Y}{u_n} \right)^s \left(\ln_+ \frac{Y}{u_n} \right)^t \mathbf{1}\{Y > u_n\} \right].$$

Note that $\tilde{m}_n(K, s, t; x) = \mathbb{E}[T_n(K, s, t; x)]$. In order to obtain the asymptotic expansion of $\tilde{m}_n(K, s, t; x)$ we need to introduce some further conditions. For all $x_1, x_2 \in \mathbb{R}^p$, the Euclidean distance between x_1 and x_2 is denoted by $d(x_1, x_2)$.

Concerning the density function b of the covariate X we assume a Lipschitz condition.

Assumption (B) *There exists $c_b > 0$ such that $|b(x_1) - b(x_2)| \leq c_b d(x_1, x_2)$ for all $x_1, x_2 \in \mathbb{R}^p$.*

The following assumption is standard in the context of local estimation.

Assumption (K) *K is a bounded density function on \mathbb{R}^p , with support Ω included in the unit hypersphere in \mathbb{R}^p .*

Finally, we need a smoothness condition for the conditional response distribution function, when considered as a function of x . This condition will be formulated in terms of the conditional expectation $m(u_n, s, t; x)$.

Assumption (M) *The function $m(u_n, s, t; x)$ satisfies that, for $u_n \rightarrow \infty$, $h_n \rightarrow 0$, and some $S < 0$ and $T > 0$,*

$$\Phi(u_n, h_n; x) := \sup_{(s,t) \in [S,0] \times [0,T]} \sup_{z \in \Omega} \left| \frac{m(u_n, s, t; x - h_n z)}{m(u_n, s, t; x)} - 1 \right| \rightarrow 0 \text{ if } n \rightarrow \infty.$$

Lemma 2 Assume (\mathcal{R}) , (\mathcal{B}) , (\mathcal{K}) , (\mathcal{M}) and $(s, t) \in [S, 0] \times [0, T]$. For all $x \in \mathbb{R}^p$ where $b(x) > 0$ we have that if $u_n \rightarrow \infty$ and $h_n \rightarrow 0$ then

$$\tilde{m}_n(K, s, t; x) = m(u_n, s, t; x)b(x) \{1 + O(h_n) + O(\Phi(u_n, h_n; x))\}.$$

Proof of Lemma 2. A direct application of the rule of repeated expectations and straightforward calculations lead to the result. See more details in the Online Appendix. \square

By combining the result from Lemmas 1 and 2 we have that

$$\tilde{m}_n(K, 0, 0; x) = \bar{F}(u_n; x)b(x) \{1 + O(h_n) + O(\Phi(u_n, h_n; x))\}, \quad (10)$$

and, in case $(s, t) \in [S, 0] \times [0, T] \setminus (0, 0)$

$$\begin{aligned} \tilde{m}_n(K, s, t; x) &= \gamma^t(x)\bar{F}(u_n; x)b(x)\Gamma(t+1) \left\{ \frac{1}{(1-s\gamma(x))^{t+1}} \right. \\ &\quad \left. - \frac{\delta(u_n; x)}{\gamma(x)} \left[\frac{1}{(1-s\gamma(x))^{t+1}} - \frac{1-\rho(x)}{(1-\rho(x)-s\gamma(x))^{t+1}} \right] \right\} (1+o(1)) \\ &\quad + O(h_n) + O(\Phi(u_n, h_n; x)). \end{aligned} \quad (11)$$

Let $r_n := \sqrt{nh_n^p \bar{F}(u_n; x)b(x)}$, $J := \{0, 1, 2, 3\}$, and consider the empirical processes

$$\mathbb{P}_n^{(j)}(s) := r_n \left[\frac{T_n(K, s, j; x)}{\bar{F}(u_n; x)b(x)} - \mathbb{E} \left(\frac{T_n(K, s, j; x)}{\bar{F}(u_n; x)b(x)} \right) \right], \quad j \in J,$$

where $s \in [S, 0]$. In the following theorem we establish the joint convergence of these empirical processes.

Theorem 1 Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a sample of independent copies of the random vector (X, Y) where $Y|X = x$ satisfies (\mathcal{R}) , X follows a distribution with joint density function b , and assume (\mathcal{B}) , (\mathcal{K}) and (\mathcal{M}) hold. For all $x \in \mathbb{R}^p$ where $b(x) > 0$, we have that if $h_n \rightarrow 0$, $u_n \rightarrow \infty$, with $nh_n^p \bar{F}(u_n; x) \rightarrow \infty$, then in $\mathcal{C}^4([S, 0])$

$$(\mathbb{P}_n^{(0)}, \mathbb{P}_n^{(1)}, \mathbb{P}_n^{(2)}, \mathbb{P}_n^{(3)}) \rightsquigarrow (\mathbb{P}^{(0)}, \mathbb{P}^{(1)}, \mathbb{P}^{(2)}, \mathbb{P}^{(3)}), \quad \text{for } n \rightarrow \infty,$$

a zero-mean Gaussian process, with, for $s_1, s_2 \in [S, 0]$, covariance functions

$$\text{Cov}(\mathbb{P}^{(j)}(s_1), \mathbb{P}^{(k)}(s_2)) = \frac{(j+k)! \gamma^{j+k}(x) \|K\|_2^2}{[1 - (s_1 + s_2)\gamma(x)]^{1+j+k}}, \quad j, k \in J. \quad (12)$$

Proof of Theorem 1. Note that

$$\begin{aligned} \mathbb{P}_n^{(j)}(s) &= \sqrt{n} \left[\frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{h_n^p \bar{F}(u_n; x)b(x)}} K \left(\frac{x - X_i}{h_n} \right) \left(\frac{Y_i}{u_n} \right)^s \left(\ln \frac{Y_i}{u_n} \right)^j \mathbf{1}\{Y_i > u_n\} \right. \\ &\quad \left. - \mathbb{E} \left(\frac{1}{\sqrt{h_n^p \bar{F}(u_n; x)b(x)}} K \left(\frac{x - X}{h_n} \right) \left(\frac{Y}{u_n} \right)^s \left(\ln \frac{Y}{u_n} \right)^j \mathbf{1}\{Y > u_n\} \right) \right], \quad j \in J. \end{aligned}$$

As such, the empirical processes under consideration fit in the framework of Section 19.5 in van der Vaart (2007) on changing function classes. Theorem 1 then follows by using Lemmas 1 and 2 in order to check all the conditions of Theorem 19.28 in van der Vaart (2007). See more details in the Online Appendix. \square

The following theorem states the existence and consistency of sequences of solutions to the estimating equations (7) and (8). From now on we denote the true value of $\gamma(x)$ and $\rho(x)$ by $\gamma_0(x)$ and $\rho_0(x)$, respectively. In first instance we assume that $\rho_0(x)$ is known.

Theorem 2 *Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a sample of independent copies of the random vector (X, Y) where $Y|X = x$ satisfies (\mathcal{R}) , X follows a distribution with joint density function b , and assume (\mathcal{B}) , (\mathcal{K}) and (\mathcal{M}) hold. For all $x \in \mathbb{R}^p$ where $b(x) > 0$, we have that if $h_n \rightarrow 0$, $u_n \rightarrow \infty$ with $nh_n^p \bar{F}(u_n; x) \rightarrow \infty$, then with probability tending to 1 there exists sequences of solutions $(\hat{\gamma}_n(x), \hat{\delta}_n(x))$ of the estimating equations (7) and (8), with ρ fixed at $\rho_0(x)$, such that $(\hat{\gamma}_n(x), \hat{\delta}_n(x)) \xrightarrow{\mathbb{P}} (\gamma_0(x), 0)$, as $n \rightarrow \infty$.*

Proof of Theorem 2. To prove the existence and consistency of $(\hat{\gamma}_n(x), \hat{\delta}_n(x))$ we adapt the proof of Theorem 5.1 in Chapter 6 of Lehmann and Casella (1998), where existence and consistency of solutions of the likelihood equations is established, to the MDPDE framework. Let Q_r denote the sphere centered at $(\gamma_0(x), 0)$ and radius r , and let $\hat{\Delta}_\alpha(\gamma, \delta; \rho)$ denote the density power divergence objective function. Note that r should be such that Q_r is a subset of the parameter space. First rescale $\hat{\Delta}_\alpha(\gamma, \delta; \rho)$ as $\tilde{\Delta}_\alpha(\gamma, \delta; \rho) := \hat{\Delta}_\alpha(\gamma, \delta; \rho) / (\bar{F}(u_n; x)b(x))$. We can show that for any r sufficiently small

$$\mathbb{P}_{(\gamma_0(x), 0)}(\tilde{\Delta}_\alpha(\gamma_0(x), 0; \rho_0(x)) < \tilde{\Delta}_\alpha(\gamma, \delta; \rho_0(x))) \text{ for all } (\gamma, \delta) \text{ on the surface of } Q_r \rightarrow 1.$$

We end the proof by adjusting the line of argumentation of Theorem 3.7 in Chapter 6 of Lehmann and Casella (1998). See more details in the Online Appendix. \square

In order to establish the asymptotic normality of the consistent sequence of solutions $(\hat{\gamma}_n(x), \hat{\delta}_n(x))$, we re-center the empirical processes with the leading terms of the asymptotic expansions of $\tilde{m}_n(K, s, j; x)$, as given in (10) and (11). Let

$$\mathbb{S}_n^{(j)}(s) := r_n \left[\frac{T_n(K, s, j; x)}{\bar{F}(u_n; x)b(x)} - \frac{j! \gamma_0^j(x)}{[1 - s\gamma_0(x)]^{j+1}} \right], \quad j \in J.$$

Corollary 1 *Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a sample of independent copies of the random vector (X, Y) where $Y|X = x$ satisfies (\mathcal{R}) , X follows a distribution with joint density function b , and assume (\mathcal{B}) , (\mathcal{K}) and (\mathcal{M}) hold. For all $x \in \mathbb{R}^p$ where $b(x) > 0$, we have that if $h_n \rightarrow 0$, $u_n \rightarrow \infty$, with $nh_n^p \bar{F}(u_n; x) \rightarrow \infty$, $\sqrt{nh_n^p \bar{F}(u_n; x)} \delta(u_n; x) \rightarrow \lambda \in \mathbb{R}$, $\sqrt{nh_n^p \bar{F}(u_n; x)} h_n \rightarrow 0$, $\sqrt{nh_n^p \bar{F}(u_n; x)} \Phi(u_n, h_n; x) \rightarrow 0$, then in $\mathcal{C}^4([S, 0])$*

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

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

$$\mathbb{E}[\mathbb{S}^{(j)}(s)] = -\lambda \sqrt{b(x)} j! \gamma_0^{j-1}(x) \left[\frac{1}{[1 - s\gamma_0(x)]^{j+1}} - \frac{1 - \rho_0(x)}{[1 - \rho_0(x) - s\gamma_0(x)]^{j+1}} \right], \quad j \in J,$$

and covariance functions as given in (12).

Proof of Corollary 1. We have that

$$\mathbb{S}_n^{(j)}(s) = \mathbb{P}_n^{(j)}(s) + r_n \left[\mathbb{E} \left(\frac{T_n(K, s, j; x)}{\bar{F}(u_n; x)b(x)} \right) - \frac{j! \gamma_0^j(x)}{[1 - s\gamma_0(x)]^{1+j}} \right], \quad j \in J.$$

Thus a direct application of Lemmas 1 and 2 achieves the proof. See more details in the Online Appendix. \square

Based on these results one can establish the asymptotic normality of the estimators, when suitably normalized.

Theorem 3 *Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a sample of independent copies of the random vector (X, Y) where $Y|X = x$ satisfies (\mathcal{R}) , X follows a distribution with joint density function b , and assume (\mathcal{B}) , (\mathcal{K}) and (\mathcal{M}) hold. Consider $(\hat{\gamma}_n(x), \hat{\delta}_n(x))$, a consistent sequence of estimators for $(\gamma_0(x), 0)$ satisfying (7) and (8), with ρ fixed at $\rho_0(x)$. For all $x \in \mathbb{R}^p$ where $b(x) > 0$, we have that if $h_n \rightarrow 0$, $u_n \rightarrow \infty$ with $nh_n^p \bar{F}(u_n; x) \rightarrow \infty$, $\sqrt{nh_n^p \bar{F}(u_n; x)} \delta(u_n; x) \rightarrow \lambda \in \mathbb{R}$, $\sqrt{nh_n^p \bar{F}(u_n; x)} h_n \rightarrow 0$, and $\sqrt{nh_n^p \bar{F}(u_n; x)} \Phi(u_n, h_n; x) \rightarrow 0$, then*

$$r_n \begin{bmatrix} \hat{\gamma}_n(x) - \gamma_0(x) \\ \hat{\delta}_n(x) - \delta(u_n; x) \end{bmatrix} \rightsquigarrow N_2(\mathbf{0}, \mathbb{C}^{-1}(\rho_0(x)) \mathbb{B}(\rho_0(x)) \boldsymbol{\Sigma}(\rho_0(x)) \mathbb{B}'(\rho_0(x)) \mathbb{C}^{-1}(\rho_0(x))),$$

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

$$\mathbb{B}(\rho_0(x)) := \gamma_0^{-\alpha-2}(x) \begin{bmatrix} -\frac{\alpha \gamma_0(x)(1+\gamma_0(x))}{[1+\alpha(1+\gamma_0(x))]^2} & \gamma_0(x) & 0 & -1 \\ -\frac{\alpha \gamma_0(x) \rho_0(x)(1+\gamma_0(x))}{[1+\alpha(1+\gamma_0(x))][1-\rho_0(x)+\alpha(1+\gamma_0(x))]} & \gamma_0(x) & -\gamma_0(x)(1-\rho_0(x)) & 0 \end{bmatrix},$$

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

$$\begin{aligned}
\sigma_{11}(\rho_0(x)) &:= \|K\|_2^2 \\
\sigma_{21}(\rho_0(x)) &:= \frac{\|K\|_2^2}{1 + \alpha(1 + \gamma_0(x))} \\
\sigma_{22}(\rho_0(x)) &:= \frac{\|K\|_2^2}{1 + 2\alpha(1 + \gamma_0(x))} \\
\sigma_{31}(\rho_0(x)) &:= \frac{\|K\|_2^2}{1 - \rho_0(x) + \alpha(1 + \gamma_0(x))} \\
\sigma_{32}(\rho_0(x)) &:= \frac{\|K\|_2^2}{1 - \rho_0(x) + 2\alpha(1 + \gamma_0(x))} \\
\sigma_{33}(\rho_0(x)) &:= \frac{\|K\|_2^2}{1 - 2\rho_0(x) + 2\alpha(1 + \gamma_0(x))} \\
\sigma_{41}(\rho_0(x)) &:= \frac{\gamma_0(x)\|K\|_2^2}{[1 + \alpha(1 + \gamma_0(x))]^2} \\
\sigma_{42}(\rho_0(x)) &:= \frac{\gamma_0(x)\|K\|_2^2}{[1 + 2\alpha(1 + \gamma_0(x))]^2} \\
\sigma_{43}(\rho_0(x)) &:= \frac{\gamma_0(x)\|K\|_2^2}{[1 - \rho_0(x) + 2\alpha(1 + \gamma_0(x))]^2} \\
\sigma_{44}(\rho_0(x)) &:= \frac{2\gamma_0^2(x)\|K\|_2^2}{[1 + 2\alpha(1 + \gamma_0(x))]^3}
\end{aligned}$$

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

$$\begin{aligned}
C_{11}(\rho_0(x)) &:= \gamma_0^{-\alpha-2}(x) \frac{1 + \alpha^2(1 + \gamma_0(x))^2}{[1 + \alpha(1 + \gamma_0(x))]^3} \\
C_{21}(\rho_0(x)) &:= \gamma_0^{-\alpha-2}(x) \frac{\rho_0(x)(1 - \rho_0(x))[1 + \alpha(1 + \gamma_0(x)) + \alpha^2(1 + \gamma_0(x))^2] + \alpha^3\rho_0(x)(1 + \gamma_0(x))^3}{[1 + \alpha(1 + \gamma_0(x))]^2[1 - \rho_0(x) + \alpha(1 + \gamma_0(x))]^2} \\
C_{22}(\rho_0(x)) &:= \gamma_0^{-\alpha-2}(x) \frac{(1 - \rho_0(x))\rho_0^2(x) + \alpha\rho_0^2(x)(1 + \gamma_0(x))[\alpha(1 + \gamma_0(x)) - \rho_0(x)]}{[1 + \alpha(1 + \gamma_0(x))][1 - \rho_0(x) + \alpha(1 + \gamma_0(x))][1 - 2\rho_0(x) + \alpha(1 + \gamma_0(x))]}
\end{aligned}$$

Proof of Theorem 3. Let $f_s(\gamma, \delta; \rho_0(x))$, $s = 1, 2$, denote the derivatives of $\tilde{\Delta}_\alpha(\gamma, \delta; \rho_0(x))$ with respect to γ and δ , respectively, without the common scale factor $1 + \alpha$. First we establish the joint limiting distribution of the random terms appearing in $f_s(\gamma_0(x), 0; \rho_0(x))$, $s = 1, 2$, when appropriately normalized. To this aim, let

$$\mathbb{T}_n := \frac{1}{\bar{F}(u_n; x)b(x)} \begin{bmatrix} T_n(K, 0, 0; x) \\ T_n(K, -\alpha(1 + \gamma_0(x))/\gamma_0(x), 0; x) \\ T_n(K, -(\alpha(1 + \gamma_0(x)) - \rho_0(x))/\gamma_0(x), 0; x) \\ T_n(K, -\alpha(1 + \gamma_0(x))/\gamma_0(x), 1; x) \end{bmatrix},$$

and

$$\tilde{\mathbb{T}} := \begin{bmatrix} 1 \\ \frac{1}{1+\alpha(1+\gamma_0(x))} \\ \frac{1}{1-\rho_0(x)+\alpha(1+\gamma_0(x))} \\ \frac{\gamma_0(x)}{[1+\alpha(1+\gamma_0(x))]^2} \end{bmatrix}.$$

An application of Corollary 1 leads to

$$r_n[\mathbb{T}_n - \tilde{\mathbb{T}}] \rightsquigarrow N_4(\lambda\sqrt{b(x)}\mathbb{D}, \boldsymbol{\Sigma}(\rho_0(x))),$$

where \mathbb{D} is a (4×1) vector with elements given by

$$\begin{aligned} D_1 &:= 0 \\ D_2 &:= -\frac{\alpha\rho_0(x)(1+\gamma_0(x))}{\gamma_0(x)[1+\alpha(1+\gamma_0(x))][1-\rho_0(x)+\alpha(1+\gamma_0(x))]} \\ D_3 &:= -\frac{\rho_0(x)[\alpha(1+\gamma_0(x))-\rho_0(x)]}{\gamma_0(x)[1-\rho_0(x)+\alpha(1+\gamma_0(x))][1-2\rho_0(x)+\alpha(1+\gamma_0(x))]} \\ D_4 &:= \frac{\rho_0(x)(1-\rho_0(x))-\alpha^2\rho_0(x)(1+\gamma_0(x))^2}{[1+\alpha(1+\gamma_0(x))]^2[1-\rho_0(x)+\alpha(1+\gamma_0(x))]^2}. \end{aligned}$$

Then by applying a Taylor series expansion of the estimating equations $f_s(\hat{\gamma}_n(x), \hat{\delta}_n(x); \rho_0(x)) = 0$, $s = 1, 2$, around $(\gamma_0(x), 0)$, Theorem 3 follows. See more details in the Online Appendix. \square

Note that the expected value of the limiting random vector in Theorem 3 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 estimator when the parameter ρ is fixed at some value $\tilde{\rho}(x) < 0$, possibly misspecified.

Proposition 1 *Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a sample of independent copies of the random vector (X, Y) where $Y|X = x$ satisfies (\mathcal{R}) and assume the parameter ρ is fixed at $\tilde{\rho}(x)$ in (7) and (8). Suppose also that X follows a distribution with joint density function b , and assume (\mathcal{B}) , (\mathcal{M}) and (\mathcal{K}) hold. For all $x \in \mathbb{R}^p$ where $b(x) > 0$, we have that if $h_n \rightarrow 0$, $u_n \rightarrow \infty$ with $nh_n^p \bar{F}(u_n; x) \rightarrow \infty$, when $n \rightarrow \infty$, then with probability tending to 1 there exists sequences of solutions $(\hat{\gamma}_n(x), \hat{\delta}_n(x))$ of the estimating equations (7) and (8) such that $(\hat{\gamma}_n(x), \hat{\delta}_n(x)) \xrightarrow{\mathbb{P}} (\gamma_0(x), 0)$. If additionally $\sqrt{nh_n^p \bar{F}(u_n; x)} \delta(u_n; x) \rightarrow \lambda \in \mathbb{R}$, $\sqrt{nh_n^p \bar{F}(u_n; x)} h_n \rightarrow 0$, and $\sqrt{nh_n^p \bar{F}(u_n; x)} \Phi(u_n, h_n; x) \rightarrow 0$, then*

$$\begin{aligned} r_n \begin{bmatrix} \hat{\gamma}_n(x) - \gamma_0(x) \\ \hat{\delta}_n(x) \end{bmatrix} &\rightsquigarrow N_2(-\lambda\sqrt{b(x)}\mathbb{C}^{-1}(\tilde{\rho}(x))\mathbb{B}(\tilde{\rho}(x))\tilde{\mathbb{D}}, \\ &\mathbb{C}^{-1}(\tilde{\rho}(x))\mathbb{B}(\tilde{\rho}(x))\boldsymbol{\Sigma}(\tilde{\rho}(x))\mathbb{B}'(\tilde{\rho}(x))\mathbb{C}^{-1}(\tilde{\rho}(x))), \end{aligned}$$

for $n \rightarrow \infty$, where the elements of the vector $\tilde{\mathbb{D}}$ are the following

$$\begin{aligned}\tilde{D}_1 &:= D_1 \\ \tilde{D}_2 &:= D_2 \\ \tilde{D}_3 &:= -\frac{[\alpha(1 + \gamma_0(x)) - \tilde{\rho}(x)]\rho_0(x)}{\gamma_0(x)[1 - \tilde{\rho}(x) + \alpha(1 + \gamma_0(x))][1 - \rho_0(x) - \tilde{\rho}(x) + \alpha(1 + \gamma_0(x))]} \\ \tilde{D}_4 &:= D_4.\end{aligned}$$

Proof of Proposition 1. The arguments needed to prove the consistency and asymptotic normality are the same as those used in the proofs of Theorems 2 and 3. See more details in the Online Appendix. \square

Note that, as expected, by a misspecification of ρ at some value $\tilde{\rho}(x)$, one loses the bias-correcting effect of taking the second order structure of F into account in the estimation. However, the variance expression remains the same as in Theorem 3, but with $\rho_0(x)$ replaced by $\tilde{\rho}(x)$.

Finally, we examine the asymptotic behavior of $(\hat{\gamma}_n(x), \hat{\delta}_n(x))$ in the case where ρ is replaced by an external consistent estimator $\hat{\rho}_n(x)$ in (7) and (8). For an example of a locally consistent estimator for $\rho(x)$ we refer to Goegebeur *et al.* (2013), where the estimator proposed by Fraga Alves *et al.* (2003) was adjusted to the local estimation context.

Theorem 4 *The result of Theorems 2 and 3 continues to hold if ρ is replaced by an external consistent estimator $\hat{\rho}_n(x)$ in (7) and (8).*

Proof of Theorem 4. Again the proof is similar to that of Theorems 2 and 3. See more details in the Online Appendix. \square

3 Simulation study

In this section, we illustrate the finite sample behavior of our estimator $\hat{\gamma}_n(x)$ with a small simulation study. In particular we compare our estimator with the following non-robust and biased version proposed in Goegebeur *et al.* (2013):

$$\hat{\gamma}_n^{(2)}(x, t, K, K) := \frac{1}{t+1} \frac{\sum_{i=1}^n K_h(x - X_i) (\ln Y_i - \ln u_n)_+^{t+1} \mathbf{1}\{Y_i > u_n\}}{\sum_{i=1}^n K_h(x - X_i) (\ln Y_i - \ln u_n)_+^t \mathbf{1}\{Y_i > u_n\}}$$

with $t = 0$ and two (non-robust) bias-corrected versions of the form

$$\hat{\gamma}_n^{(2)}(x, \beta) := \beta \hat{\gamma}_n^{(2)}(x, 0, K, K) + (1 - \beta) \hat{\gamma}_n^{(2)}(x, 1, K, K)$$

with $\beta = -1$ and $\beta = 1/\hat{\rho}(x)$. To estimate $\rho(x)$ we use the consistent estimator proposed in Goegebeur *et al.* (2013). In the robust case, a first order estimator is obtained by setting $\delta = 0$ in (4), whereas a second order bias-corrected version is derived by estimating γ and δ

		$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 1$
$\gamma_0 = 0.2$	$\rho_0 = -0.5$	0.93	0.42	0.20
	$\rho_0 = -1$	0.94	0.48	0.26
	$\rho_0 = -2$	0.95	0.54	0.33
$\gamma_0 = 0.4$	$\rho_0 = -0.5$	0.91	0.36	0.17
	$\rho_0 = -1$	0.92	0.42	0.22
	$\rho_0 = -2$	0.93	0.49	0.29

Table 1: Asymptotic relative efficiency of $\hat{\gamma}_n(x)$ with $\alpha = 0.1, 0.5$ and 1 relative to $\hat{\gamma}_n(x)$ with $\alpha = 0$.

jointly. In that case the value of $\rho(x)$ is either fixed to -1 or estimated consistently as previously mentioned. All kernels are taken as the bi-quadratic kernel function

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

Concerning α , the values $\alpha = 0.1$ and $\alpha = 0.5$ are considered. According to Table 1, higher values of α are not appropriate, due to a low asymptotic relative efficiency compared to $\hat{\gamma}_n(x)$ with $\alpha = 0$. The values $\gamma_0 = 0.2$ and $\gamma_0 = 0.4$ are representative for the range of the function $\gamma(x)$ considered in the simulation experiment. This decrease in asymptotic relative efficiency for increasing α is not specific for our estimator. Indeed, it was also observed in the univariate extreme value framework by Kim and Lee (2008) for the strict Pareto distribution and by Juárez and Schucany (2004) for the generalized Pareto distribution. We also refer to the original Basu *et al.* (1998) paper, where the asymptotic relative efficiencies for a range of common models like normal, Poisson and exponential were examined, and where a similar decrease was observed. In view of this decrease of efficiency, when using the MDPDE method it is generally recommended to keep α rather small, say in the range $0 \leq \alpha \leq 1$. For what concerns our method, in absence of outliers we have thus that $\alpha = 0$ (maximum likelihood estimation) is optimal in the class of asymptotically unbiased density power divergence estimators (since it leads to the smallest asymptotic variance). However, as we will see in the simulation experiments, if one keeps α small then one does not lose much. The decrease in asymptotic efficiency compared to the maximum likelihood estimator is the price to pay for the increased robustness obtained with the MDPD estimator for increasing values of α . Concerning the robustness, we refer to Dierckx *et al.* (2013), where the influence functions for the MDPD estimators of the extended Pareto distribution are derived in the univariate framework. These influence functions are bounded when considered as a function of the point of contamination when α is not too close to zero. Similar findings can be expected to hold also here.

For the practical implementation of our estimators we have to determine the bandwidth parameter h_n and the threshold u_n , where we take, as usual in extreme value statistics, the latter as the $(k + 1)$ -th largest response observation in the ball $B(x, h_n)$. To determine an optimal value for k and h_n , two strategies are applied as in Goegebeur *et al.* (2013): an oracle strategy and a completely data driven method; see also Daouia *et al.* (2011, 2013). For the oracle strategy, the

same algorithm as in Goegebeur *et al.* (2013) was applied to our new estimator of γ , that is

$$(h_{n,o}, k_o) := \arg \min_{h_n \in \mathcal{H}_0, k \in \mathcal{K}_0} \Psi(\hat{\gamma}(\cdot), \gamma(\cdot)), \quad (13)$$

where \mathcal{H}_0 and \mathcal{K}_0 are grids of values of h_n and k , respectively, and

$$\Psi^2(\hat{\gamma}(\cdot), \gamma(\cdot)) := \frac{1}{M} \sum_{m=1}^M (\hat{\gamma}(z_m) - \gamma(z_m))^2,$$

where z_1, \dots, z_M are regularly spaced in the covariate space. Note that this method requires knowledge of the function $\gamma(x)$, which is unknown in practical situations.

For the completely data driven method, the optimal bandwidth h_n is determined using the leave-one-out cross-validation method of Daouia *et al.* (2011) and for the optimal k we proceed as follows for all x under consideration:

- we compute the estimates for $\gamma(x)$ with $k = 5, 9, 13, \dots, m_x - 4$ (m_x being the number of observations in the ball $B(x, h_n)$);
- we split the range of k into several blocks of same size;
- we calculate the standard deviation of the estimates for $\gamma(x)$ in each block;
- the block with minimal standard deviation determines the k to be used.

Note that in this implementation h_n and k are selected separately. One could also pursue a simultaneous determination of these two parameters, as was attempted in e.g. Daouia *et al.* (2013). However, as reported in that paper, the simultaneous selection is not without problems, and in practice it does not seem to perform any better than the separate selection of the tuning parameters.

Practically, the parameter estimates are determined by minimizing the objective function given in (5) or (6), depending on the value of α , with a program written in the R language. In this, the adaptive quadrature numerical integration procedure as described in Piessens *et al.* (1983) (implemented in the R function `integrate`) is used to determine the integral in the objective function. The minimization is carried out with the numerical minimization procedure described in Byrd *et al.* (1995) (R function `optim`, with `method = "L-BFGS-B"`). This method is a quasi Newton method adapted to allow for the constraint $\gamma > 0$.

We simulate $N = 100$ samples of size $n = 1000$, with $X \sim U(0, 1)$ and $Y|X = x$ is generated from the following Burr distribution

$$1 - F(y; x) = \left(1 + y^{-\rho(x)/\gamma(x)}\right)^{1/\rho(x)}, \quad y > 0.$$

This model clearly satisfies condition (\mathcal{R}) . In fact, the above Burr model is parameterized in such a way that $\gamma(x)$ and $\rho(x)$ correspond with those appearing in (\mathcal{R}) . We set here

$$\gamma(x) = 0.5(0.1 + \sin(\pi x))(1.1 - 0.5 \exp(-64(x - 0.5)^2)) \text{ and } \rho(x) = -1.$$

This function for $\gamma(x)$ was also used in the simulation experiments of Daouia *et al.* (2011) and Goegebeur *et al.* (2013). In case of contamination in the response variable, the following distribution function will be used

$$F_\epsilon(y; x) = (1 - \epsilon)F(y; x) + \epsilon\tilde{F}(y; x)$$

where $\tilde{F}(y; x) = 1 - \left(\frac{y}{x_c}\right)^{-0.5}$, $y > x_c$, and $\epsilon \in (0, 1)$ is the fraction of contamination.

The following settings were considered

- Setting 1: uncontaminated situation;
- Setting 2: $\epsilon = 0.01$, $x_c = 1.2$ times the 99.99% quantile of $F(y; x)$;
- Setting 3: $\epsilon = 0.01$, $x_c = 2$ times the 99.99% quantile of $F(y; x)$;
- Setting 4: $\epsilon = 0.05$, $x_c = 1.2$ times the 99.99% quantile of $F(y; x)$;
- Setting 5: $\epsilon = 0.05$, $x_c = 2$ times the 99.99% quantile of $F(y; x)$.

In the paper we only illustrate the Settings 1, 2 and 4, since the Settings 3 and 5 give results very similar to 2 and 4, respectively. In the oracle method, the minimization is performed on a grid of $h_n \in [0.05; 0.5]$ and of $k \in \{2, \dots, m_x - 1\}$ with $M = 35$.

In the uncontaminated situation (Setting 1), from Table 2, we see that the non-robust bias-corrected estimator with $\rho(x)$ fixed at -1 behaves best in terms of mean squared error (MSE, computed as the average of Ψ^2 over the 100 simulated datasets). However, in general the robust estimators are competitive compared to the corresponding non-robust ones. The MSE increases slightly when α increases, mainly due to the larger variance in the estimation of $\gamma(x)$. For the robust estimators, as well as for the non-robust ones, the bias-corrected estimators outperform the biased ones in terms of MSE. Further, when the data driven method is applied, the MSE is usually at least twice the MSE obtained using the oracle strategy. The difference is largest for the biased estimators. Thus unsurprisingly, the robust biased estimator using the data driven method behaves the worst in terms of MSE. In Figures 1 and 2 we show the boxplots of the different estimators for $\gamma(x)$ under consideration for the oracle and data driven strategies, respectively. In all figures, the dashed line represents the true function $\gamma(x)$.

In Setting 2, the non-robust estimators clearly suffer from the contamination. In the first column of Figures 3 and 4, one can observe the large biases and variances, especially for the bias-corrected estimators. In particular, the sinus behavior of γ as a function of x is not captured very well by the non-robust estimators. This is confirmed in Table 3, where we can observe that the MSE is largest for the non-robust estimators, whereas the best results in terms of MSE are obtained for the robust estimator with $\alpha = 0.5$ and $\rho(x)$ fixed at -1, although the result for $\rho(x)$ estimated is not much worse. Note also that in these best cases the results obtained by the data driven method are comparable to those of the oracle strategy.

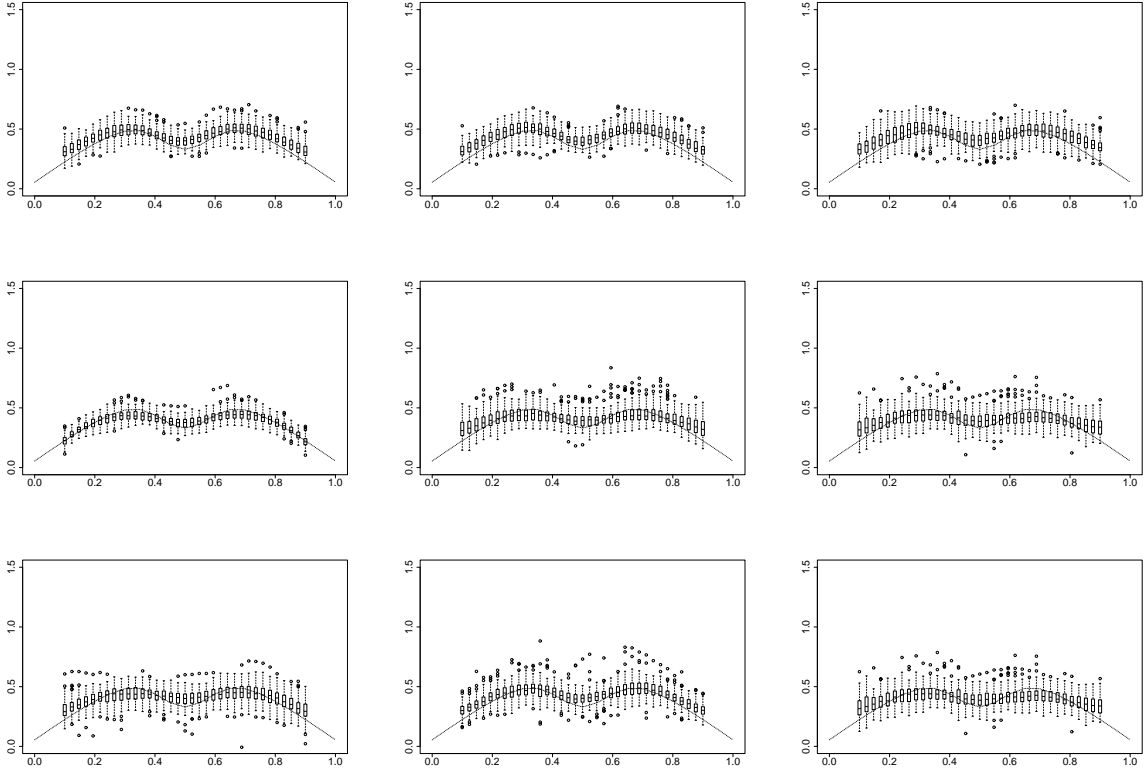


Figure 1: Setting 1: boxplots of $\hat{\gamma}_n(x)$ with k_o and $h_{n,o}$ determined using the oracle strategy; column 1: non-robust estimators, column 2: robust estimators with $\alpha = 0.1$, column 3: robust estimators with $\alpha = 0.5$, row 1: biased estimator, row 2: bias-corrected estimator with $\rho(x) = -1$, row 3: bias-corrected estimator with $\rho(x) = \hat{\rho}(x)$.

Non Robust/Robust	Estimator	Oracle strategy	Data driven method
non robust	biased	0.006	0.019
non robust	bias-corrected $\rho(x) = -1$	0.003	0.006
non robust	bias-corrected $\rho(x) = \hat{\rho}(x)$	0.007	0.006
robust $\alpha = 0.1$	biased	0.006	0.025
robust $\alpha = 0.1$	bias-corrected $\rho(x) = -1$	0.007	0.011
robust $\alpha = 0.1$	bias-corrected $\rho(x) = \hat{\rho}(x)$	0.006	0.007
robust $\alpha = 0.5$	biased	0.008	0.055
robust $\alpha = 0.5$	bias-corrected $\rho(x) = -1$	0.007	0.017
robust $\alpha = 0.5$	bias-corrected $\rho(x) = \hat{\rho}(x)$	0.007	0.019

Table 2: MSE for different estimators of $\gamma(x)$ based on 100 datasets simulated according to Setting 1.

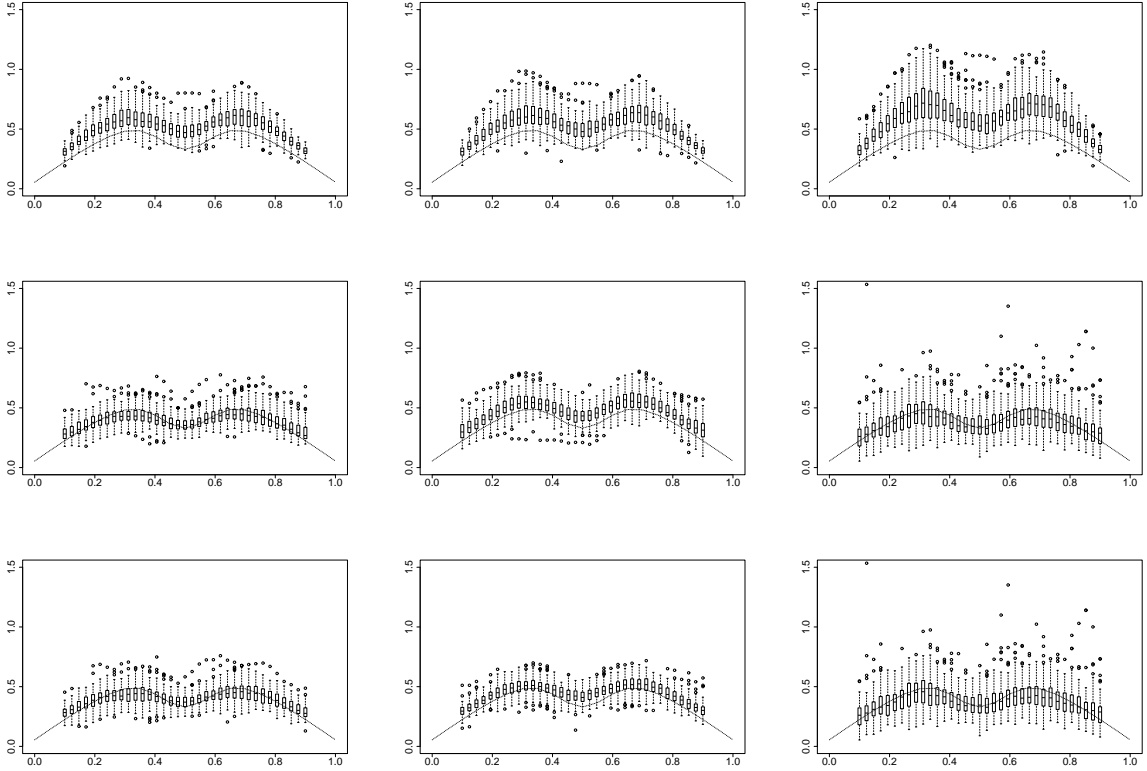


Figure 2: Setting 1: boxplots of $\hat{\gamma}_n(x)$ with k_o and $h_{n,o}$ determined using the data driven method; column 1: non-robust estimators, column 2: robust estimators with $\alpha = 0.1$, column 3: robust estimators with $\alpha = 0.5$, row 1: biased estimator, row 2: bias-corrected estimator with $\rho(x) = -1$, row 3: bias-corrected estimator with $\rho(x) = \hat{\rho}(x)$.

Non Robust/Robust	Estimator	Oracle strategy	Data driven method
non robust	biased	0.053	0.069
non robust	bias-corrected $\rho(x) = -1$	0.291	0.977
non robust	bias-corrected $\rho(x) = \hat{\rho}(x)$	0.447	0.470
robust $\alpha = 0.1$	biased	0.020	0.039
robust $\alpha = 0.1$	bias-corrected $\rho(x) = -1$	0.011	0.025
robust $\alpha = 0.1$	bias-corrected $\rho(x) = \hat{\rho}(x)$	0.014	0.023
robust $\alpha = 0.5$	biased	0.012	0.060
robust $\alpha = 0.5$	bias-corrected $\rho(x) = -1$	0.007	0.009
robust $\alpha = 0.5$	bias-corrected $\rho(x) = \hat{\rho}(x)$	0.009	0.012

Table 3: MSE for different estimators of $\gamma(x)$ based on 100 datasets simulated according to Setting 2.

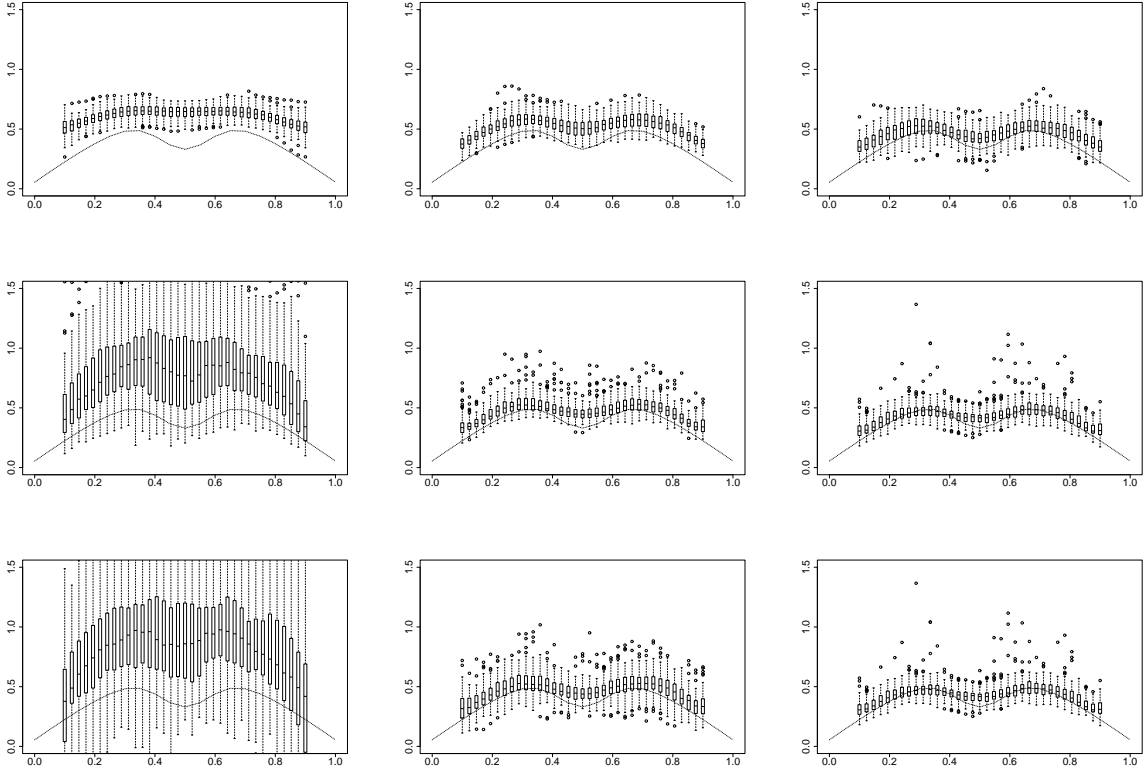


Figure 3: Setting 2: boxplots of $\hat{\gamma}_n(x)$ with k_o and $h_{n,o}$ determined using the oracle strategy; column 1: non-robust estimators, column 2: robust estimators with $\alpha = 0.1$, column 3: robust estimators with $\alpha = 0.5$, row 1: biased estimator, row 2: bias-corrected estimator with $\rho(x) = -1$, row 3: bias-corrected estimator with $\rho(x) = \hat{\rho}(x)$.

When the contamination percentage is increased from 1 to 5%, the results for the non-robust estimators are appalling, especially when the bias-corrected versions are applied. Whereas the robust estimator with $\alpha = 0.1$ could somewhat withstand 1% contamination, this is no longer true for 5% contamination. Although behaving better than the corresponding non-robust estimators, the estimators show a considerable bias and variance. The bias-corrected robust estimators with $\alpha = 0.5$ behave now the best by far. In these cases, the estimators with $\rho(x)$ fixed to -1 are somewhat better than the estimators with $\rho(x)$ estimated, and the data and oracle strategies are comparable.

In conclusion, we can say that in both cases (contaminated and uncontaminated) the robust bias-corrected estimator with $\alpha = 0.5$ and $\rho(x)$ fixed at -1 has a very attractive behavior. Also, using the consistent estimator $\hat{\rho}(x)$ for $\rho(x)$ in the estimation of $\hat{\gamma}_n(x)$ is not much worse than using the true value of this parameter. The selection of the tuning parameters h_n and k is challenging, though the proposed methods give a good practical performance. Note that the selection of h_n

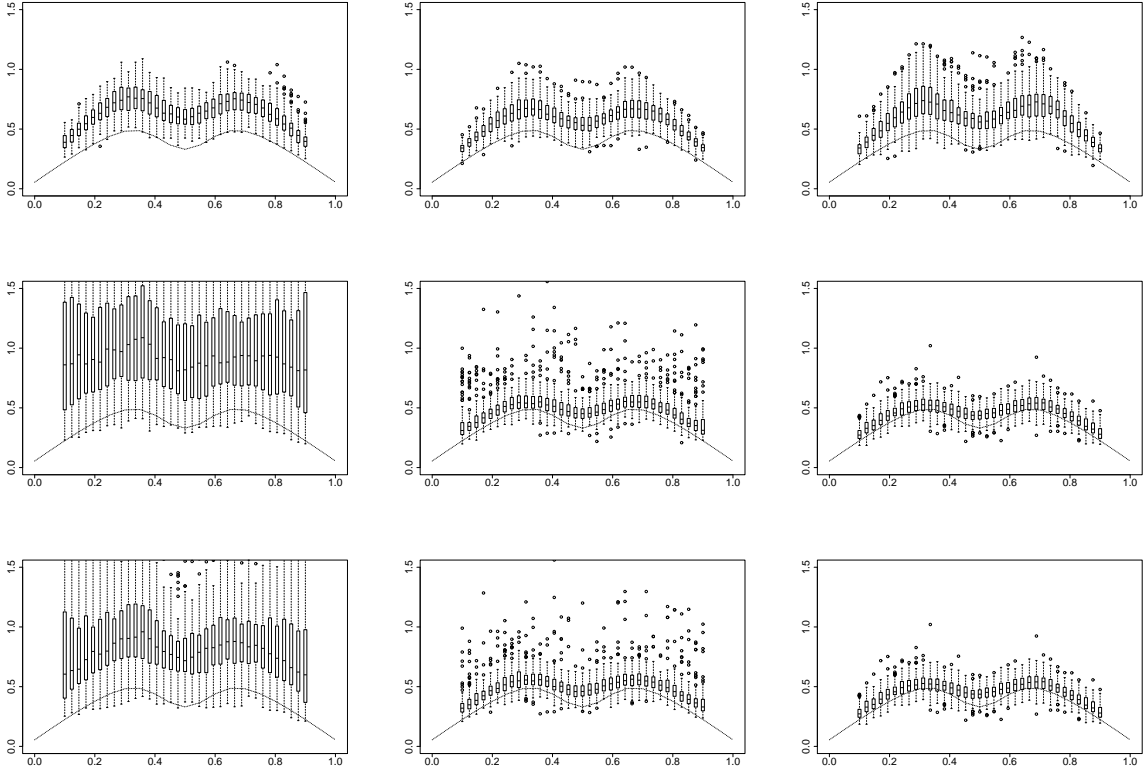


Figure 4: Setting 2: boxplots of $\hat{\gamma}_n(x)$ with k_o and $h_{n,o}$ determined using the data strategy; column 1: non-robust estimators, column 2: robust estimators with $\alpha = 0.1$, column 3: robust estimators with $\alpha = 0.5$, row 1: biased estimator, row 2: bias-corrected estimator with $\rho(x) = -1$, row 3: bias-corrected estimator with $\rho(x) = \hat{\rho}(x)$.

Non robust/Robust	Estimator	Oracle strategy	Data driven method
non robust	biased	0.368	0.419
non robust	bias-corrected $\rho(x) = -1$	1.312	7.508
non robust	bias-corrected $\rho(x) = \hat{\rho}(x)$	2.752	18244.6
robust $\alpha = 0.1$	biased	0.124	0.159
robust $\alpha = 0.1$	bias-corrected $\rho(x) = -1$	0.197	0.676
robust $\alpha = 0.1$	bias-corrected $\rho(x) = \hat{\rho}(x)$	0.240	0.668
robust $\alpha = 0.5$	biased	0.036	0.091
robust $\alpha = 0.5$	bias-corrected $\rho(x) = -1$	0.013	0.017
robust $\alpha = 0.5$	bias-corrected $\rho(x) = \hat{\rho}(x)$	0.023	0.020

Table 4: MSE for different estimators of $\gamma(x)$ based on 100 datasets simulated according to Setting 4.

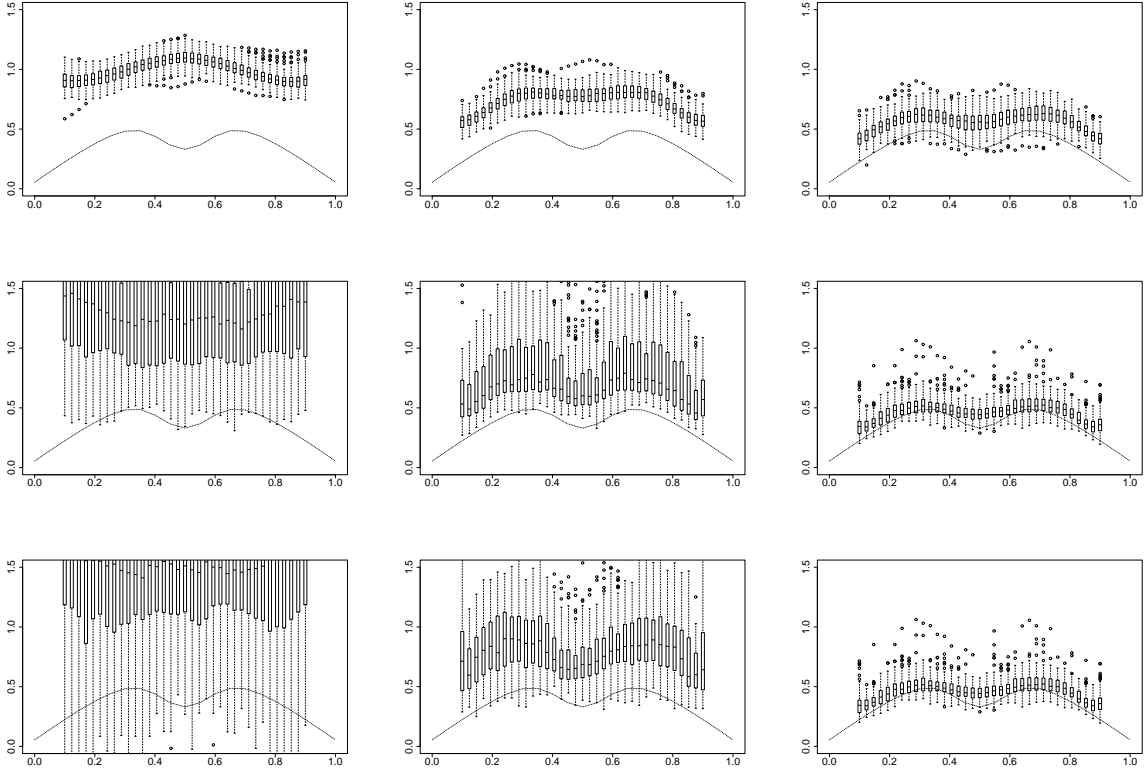


Figure 5: Setting 4: boxplots of $\hat{\gamma}_n(x)$ with k_o and $h_{n,o}$ determined using the oracle strategy; column 1: non-robust estimators, column 2: robust estimators with $\alpha = 0.1$, column 3: robust estimators with $\alpha = 0.5$, row 1: biased estimator, row 2: bias-corrected estimator with $\rho(x) = -1$, row 3: bias-corrected estimator with $\rho(x) = \hat{\rho}(x)$.

and k (or some other threshold related parameter) is not specific for our paper, but a recurrent issue when one applies extreme value methods nonparametrically by local estimation. Indeed, we refer to Beirlant and Goegebeur (2004) for an early contribution to local estimation of the GPD, and more recently to Daouia *et al.* (2011) and Daouia *et al.* (2013) for examples of kernel estimators. For the considered tuning parameter selection criteria, we can say that the data driven method gives for the above suggested robust bias-corrected γ estimator a MSE that is quite close to the one obtained with the oracle strategy, indicating the general good performance of the method. Also, both methods capture the sine behavior of the γ function quite well. The selection of the tuning parameters h_n and k (or u_n), can also be addressed from a theoretical point of view, though this is highly technical and outside the scope of this paper, which mainly focused on the theoretical asymptotic properties of the estimators for $\gamma(x)$.

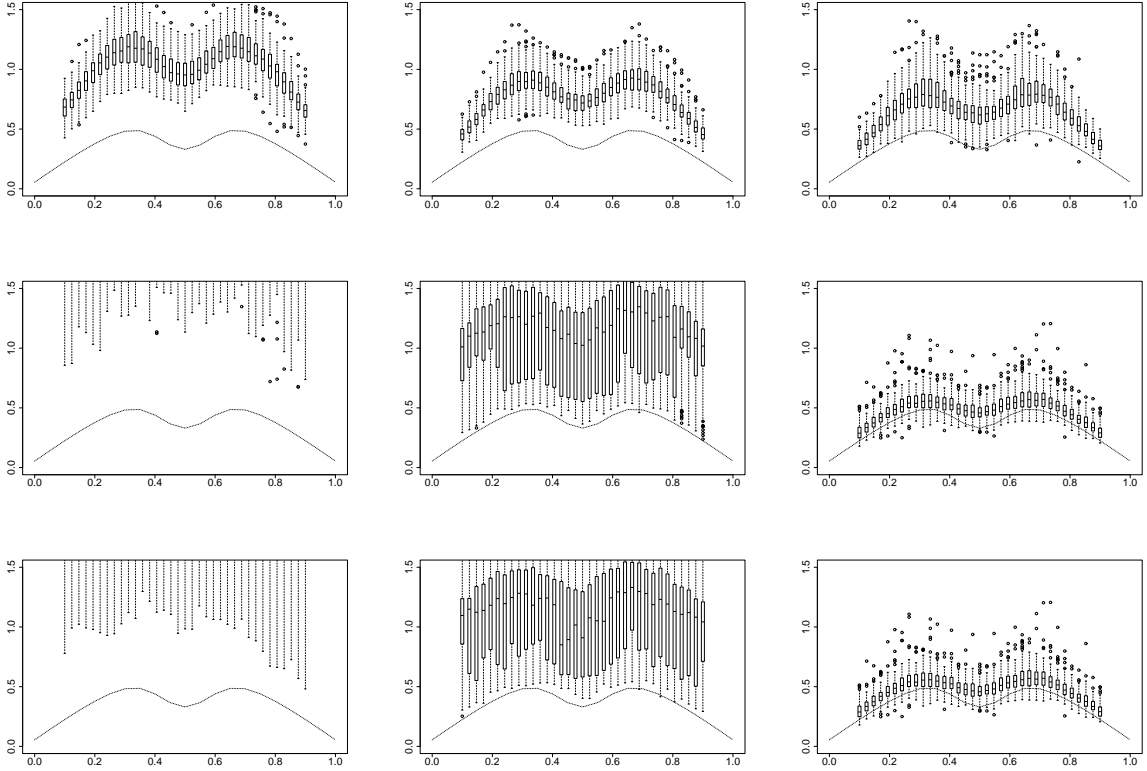


Figure 6: Setting 4: boxplots of $\hat{\gamma}_n(x)$ with k_o and $h_{n,o}$ determined using the data driven method; column 1: non-robust estimators, column 2: robust estimators with $\alpha = 0.1$, column 3: robust estimators with $\alpha = 0.5$, row 1: biased estimator, row 2: bias-corrected estimator with $\rho(x) = -1$, row 3: bias-corrected estimator with $\rho(x) = \hat{\rho}(x)$.

4 Case study: Condroz soil data

In agriculture, a new concept of crop management, called precision farming has emerged. This development permits within-field variation of crop techniques, for instance the adjustment of fertilizer inputs on the basis of soil sampling and soil analysis. As a consequence of precision farming there is an increased demand for soil data, and laboratories are now burdened with large datasets. Inevitably there is concern about data quality and outliers, since outliers may heavily influence estimation results in data analysis. Robust estimation methods have therefore become the necessary tools in order to provide high quality information.

In this section we will illustrate our robust estimation method on soil data from the Condroz region in Belgium. The Condroz soil database is centralized at the Unit of Biometry, Data Management and Agrometeorology of the Agricultural Research Centre of Gembloux, and contains information about the chemical composition (pHKCL, K, Mg, Ca, ...) of soil samples originat-

ing from the Condroz region. We refer to Laroche and Oger (1999) and Goegebeur *et al.* (2005) for further details about the database. In particular, we will focus on the estimation of the tail index of the calcium content distribution conditional on the soil pH level using the data from one of the communities in the Condroz. The dataset under consideration has 1505 observations. In Figure 7 (a) we show the scatterplot of calcium content versus pH level of the soil samples. The scatterplot clearly indicates an overall positive association between the two variables. Also, some extreme observations for calcium content are present, especially at the larger pH levels. In panels (b), (c) and (d) of Figure 7 we show local (i.e. for some neighborhoods in the covariate space) Pareto quantile plots of the calcium content of soil samples. For $\text{pH} \in [5.5, 5.7]$ (panel (b)) the Pareto quantile plot becomes linear in the largest observations indicating a good fit of the conditional Pareto-type model (1). The same can be said for $\text{pH} \in [6.5, 6.7]$ (panel (c)) and $\text{pH} \in [7, 7.2]$ (panel (d)) though there seem to be some observations that are outlying with respect to the conditional Pareto-type model, especially at the larger values of pH. These outlying observations clearly require that the tail index $\gamma(\text{pH})$ is estimated in a robust way. The MDPD estimator for $\gamma(\text{pH})$ is obtained with the completely data-driven method as described in the simulation section, where for the optimal selection of k we used blocks of size 10 and $h_n \in \{0.05, 0.10, \dots, 0.50\}$. The cross-validation criterion leads to a bandwidth $h_{n,o} = 0.15$. In Figure 8 we show the bias-corrected MDPD estimator for $\gamma(\text{pH})$ with $\rho(\text{pH})$ fixed at -1 and $\alpha = 0$ (solid line), $\alpha = 0.1$ (dashed line) and $\alpha = 0.5$ (dotted line). For pH values where the data do not contain outliers for the variable calcium content, the estimates obtained with the three values of α are as expected very similar. For $\text{pH} \in [7, 7.5]$ we see that the estimates for $\gamma(\text{pH})$ corresponding with $\alpha = 0$ and $\alpha = 0.1$ are clearly influenced by the observations that are outlying with respect to the conditional Pareto-type model, leading to elevated estimates for $\gamma(\text{pH})$ compared to the more robust estimator obtained with $\alpha = 0.5$.

5 Conclusion

We considered the robust and asymptotically unbiased estimation of the tail index of a conditional Pareto-type response distribution in the presence of random covariates. The asymptotic properties of our estimator were established together with its performance on simulated and real datasets. The theoretical study of the selection of the tuning parameters h_n and u_n , and the robust asymptotically unbiased estimation of extreme conditional quantiles will be the subjects of forthcoming studies.

Acknowledgement

The authors are very grateful to the two anonymous referees for their very constructive comments on the paper. The suggestions have definitely improved the presentation of the material.

References

- [1] 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.

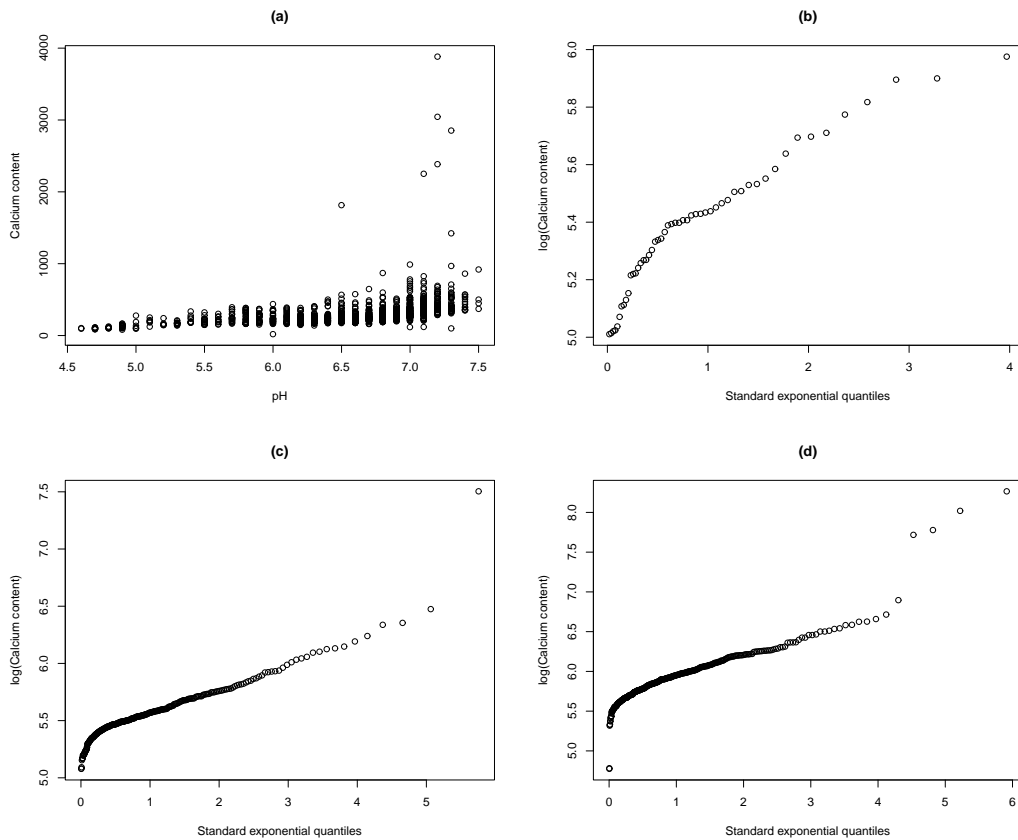


Figure 7: Condroz dataset: (a) Scatterplot of calcium content versus pH, (b) Pareto quantile plot of the calcium content for soil samples with $\text{pH} \in [5.5, 5.7]$, (c) Pareto quantile plot of the calcium content for soil samples with $\text{pH} \in [6.5, 6.7]$, (d) Pareto quantile plot of the calcium content for soil samples with $\text{pH} \in [7, 7.2]$.

- [2] Beirlant, J., Dierckx, G., Goegebeur, Y., Matthys, G., 1999. Tail index estimation and an exponential regression model. *Extremes*, 2, 177–200.
- [3] Beirlant, J., Goegebeur, Y., 2004. Local polynomial maximum likelihood estimation for Pareto-type distributions. *Journal of Multivariate Analysis*, 89, 97–118.
- [4] Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2004. *Statistics of Extremes – Theory and Applications*. Wiley Series in Probability and Statistics.
- [5] 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.

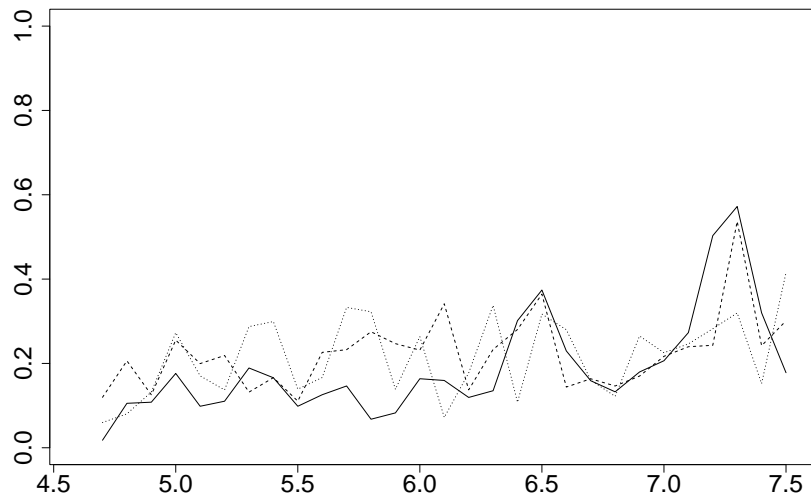


Figure 8: Condroz dataset: $\hat{\gamma}_n(pH)$ as a function of pH for $\alpha = 0$ (solid line), $\alpha = 0.1$ (dashed line) and $\alpha = 0.5$ (dotted line).

- [6] Byrd, R.H., Lu, P., Nocedal, J., Zhu, C., 1995. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16, 1190–1208.
- [7] Daouia, A., Gardes, L., Girard, S., 2013. On kernel smoothing for extremal quantile regression. *Bernoulli*, to appear.
- [8] Daouia, A., Gardes, L., Girard, S., Lekina, A., 2011. Kernel estimators of extreme level curves. *Test*, 20, 311–333.
- [9] de Haan, L., Ferreira, A., 2006. *Extreme Value Theory: An Introduction*. Springer.
- [10] Dierckx, G., Goegebeur, Y., Guillou, A., 2013. An asymptotically unbiased minimum density power divergence estimator for the Pareto-tail index. *Journal of Multivariate Analysis*, to appear.
- [11] Feuerverger, A., Hall, P., 1999. Estimating a tail exponent by modelling departure from a Pareto distribution. *Annals of Statistics*, 27, 760–781.
- [12] Fraga Alves, M.I., Gomes, M.I., de Haan, L., 2003. A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica* 60, 193–213.
- [13] Goegebeur, Y., Guillou, A., Schorgen, A., 2013. Nonparametric regression estimation of conditional tails - the random covariate case. *Statistics* (to appear), DOI:10.1080/02331888.2013.800064.

- [14] Goegebeur, Y., Planchon, V., Beirlant, J., Oger, R., 2005. Quality assessment of pedochemical data using extreme value methodology. *Journal of Applied Science*, 5, 1092–1102.
- [15] Gomes, M.I., Martins, M.J., Neves, M., 2007. Improving second order reduced bias extreme value index estimation. *REVSTAT*, 5, 177–207.
- [16] Hall, P., 1982. On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society Series B*, 44, 37–42.
- [17] Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3, 1163–1174.
- [18] Juárez, S.F., Schucany, W.R., 2004. Robust and efficient estimation for the generalized Pareto distribution. *Extremes*, 7, 237–251.
- [19] Kim, M., Lee, S., 2008. Estimation of a tail index based on minimum density power divergence. *Journal of Multivariate Analysis*, 99, 2453–2471.
- [20] Laroche, J., Oger, R., 1999. Base de données sols. Première synthèse. Faculté Universitaire des Sciences Agronomique de Gembloux. Unité de Géopédologie, Requasud asbl, Gembloux.
- [21] Lehmann, E.L., Casella, G., 1998. *Theory of Point Estimation*. Springer.
- [22] Pickands, J., 1975. Statistical inference using extreme order statistics. *Annals of Statistics*, 3, 119–131.
- [23] Piessens, R., de Doncker-Kapenga, E., Uberhuber, C., Kahaner, D., 1983. *Quadpack: a Subroutine Package for Automatic Integration*. Springer Verlag.
- [24] van der Vaart, A.W., 2007. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics.
- [25] Wang, H., Li, D., He, X., 2012. Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of the American Statistical Association*, 107, 1453–1464.
- [26] Wang, H., Tsai, C.L., 2009. Tail index regression. *Journal of the American Statistical Association*, 104, 1233–1240.