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

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

[(1)] Faculty Economics and Management, Hogeschool-Universiteit Brussel, 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.

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

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

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

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

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 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 non-parametric 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.* (2012) introduced a nonparametric and asymptotically unbiased estimator for $\gamma(x)$ based on locally weighted sums of power transformed excesses over a high threshold. Recently, in Daouia *et al.* (2012), 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 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. 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.* (2012) 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 proofs of all results can be found in the appendix.

2 Estimation procedure and asymptotic properties

Let (X_i, Y_i), $i = 1, \ldots, n$, 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 $\lim_{t\to\infty} z(tx)/z(t) = x^\beta$ 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 $y^{1/\gamma(x)}\bar{F}(y; x) \to C(x) \in (0, \infty)$ as $y \to \infty$ and the function $\delta(.; x)$ 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 $|\delta| \in RV_{\rho(x)/\gamma(x)}$.

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

$$G(z; \gamma, \delta, \rho) = \begin{cases} 1 - [z(1 + \delta - \delta z^{\rho/\gamma})]^{-1/\gamma}, & z > 1, \\ 0, & z \le 1, \end{cases} \tag{3}$$

and density function

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

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

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

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.* (2012) 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 $Z_i := Y_i/u_n$, $i = 1, \ldots, n$, by means of the MDPDE criterion, adjusted to locally weighted estimation, i.e. we minimize

$$\widehat{\Delta}_\alpha(\gamma,\delta;\rho) := \frac{1}{n}\sum_{i=1}^n K_{h_n}(x-X_i)\left\{\int_1^\infty g^{1+\alpha}(z;\gamma,\delta,\rho)dz - \left(1+\frac{1}{\alpha}\right)g^\alpha(Z_i;\gamma,\delta,\rho)\right\}\mathbf{1}\{Y_i>u_n\}, \quad (4)$$

in case $\alpha > 0$ and

$$\widehat{\Delta}_0(\gamma,\delta;\rho) \quad := \quad -\frac{1}{n}\sum_{i=1}^n K_{h_n}(x-X_i)\ln g(Z_i;\gamma,\delta,\rho)\mathbf{1}\{Y_i>u_n\}, \quad (5)$$

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 $n \to \infty$, $\mathbf{1}\{A\}$ 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 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 (4) and (5) 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. 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 $\rho(x)$.

The MDPDE for $(\gamma(x), \delta(u_n;x))$ satisfies the estimating equations

$$\begin{aligned} 0 \quad &= \quad \frac{1}{n}\sum_{i=1}^n K_{h_n}(x-X_i)\mathbf{1}\{Y_i>u_n\}\int_1^\infty g^\alpha(z;\gamma,\delta,\rho)\frac{\partial g(z;\gamma,\delta,\rho)}{\partial\gamma}dz \\ &\quad -\frac{1}{n}\sum_{i=1}^n K_{h_n}(x-X_i)g^{\alpha-1}(Z_i;\gamma,\delta,\rho)\frac{\partial g(Z_i;\gamma,\delta,\rho)}{\partial\gamma}\mathbf{1}\{Y_i>u_n\}, \quad (6) \\ 0 \quad &= \quad \frac{1}{n}\sum_{i=1}^n K_{h_n}(x-X_i)\mathbf{1}\{Y_i>u_n\}\int_1^\infty g^\alpha(z;\gamma,\delta,\rho)\frac{\partial g(z;\gamma,\delta,\rho)}{\partial\delta}dz \\ &\quad -\frac{1}{n}\sum_{i=1}^n K_{h_n}(x-X_i)g^{\alpha-1}(Z_i;\gamma,\delta,\rho)\frac{\partial g(Z_i;\gamma,\delta,\rho)}{\partial\delta}\mathbf{1}\{Y_i>u_n\}. \quad (7) \end{aligned}$$

The following statistic is crucial for studying the asymptotic behavior of the estimators. Set

$\ln_+ x := \ln \max\{x, 1\}$, $x > 0$, and

$$T_n(K, s, t; x) := \frac{1}{n} \sum_{i=1}^{n} K_{h_n}(x - X_i) \left(\frac{Y_i}{u_n}\right)^s \left(\ln_+ \frac{Y_i}{u_n}\right)^t \mathbf{1}\{Y_i > u_n\}, \tag{8}$$

where $s \leq 0$ and $t \geq 0$.

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

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

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 $u_n \to \infty$ we have that

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

Now let

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

Note that $\widetilde{m}_n(K, s, t; x) := \mathbb{E}[T_n(K, s, t; x)]$. In order to obtain the asymptotic expansion of $\widetilde{m}_n(K, s, t; x)$ 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 $d(x_1, x_2)$.

Assumption $(\mathcal{B})$ *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, $x_2 \in \mathbb{R}^p$.*

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

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

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* $u_n \to \infty$ *and* $h_n \to 0$ *then*

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

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

$$\widetilde{m}_n(K,0,0;x) = \bar{F}(u_n;x)b(x)\left\{1 + O(h_n) + O(\Phi(u_n,h_n;x))\right\}, \tag{9}$$

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

$$\begin{aligned}\widetilde{m}_n(K,s,t;x) &= \gamma^t(x)\bar{F}(u_n;x)b(x)\Gamma(t+1)\left\{\frac{1}{(1-s\gamma(x))^{t+1}}\right.\\ &\quad -\frac{\delta(u_n;x)}{\gamma(x)}\left[\frac{1}{(1-s\gamma(x))^{t+1}} - \frac{1-\rho(x)}{(1-\rho(x)-s\gamma(x))^{t+1}}\right](1+o(1))\\ &\quad \left. +O(h_n) + O(\Phi(u_n,h_n;x))\right\}. \end{aligned} \tag{10}$$

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

$$\mathbb{P}_n^{(j)}(s) := r_n\left[\frac{T_n(K,s,j;x)}{\bar{F}(u_n;x)b(x)} - \mathbb{E}\left(\frac{T_n(K,s,j;x)}{\bar{F}(u_n;x)b(x)}\right)\right], \qquad j = 0,1,2,$$

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

Theorem 1 *Let* $(X_1,Y_1),\ldots,(X_n,Y_n)$ *be a sample of independent copies of the random vector* (X,Y) *where* $Y|X=x$ *satisfies* $(\mathcal{R})$, $X \sim b$, *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* $h_n \to 0$, $u_n \to \infty$, *with* $nh_n^p\bar{F}(u_n;x) \to \infty$, *then in* $\mathcal{C}^3([S,0])$

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

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

$$\mathbb{C}ov(\mathbb{P}^{(j)}(s_1),\mathbb{P}^{(k)}(s_2)) = \frac{(j+k)!\gamma^{j+k}(x)\|K\|_2^2}{[1-(s_1+s_2)\gamma(x)]^{1+j+k}}, \quad j,k = 0,1,2. \tag{11}$$

The following theorem states the existence and consistency of sequences of solutions to the estimating equations (6) and (7). 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* $(X_1,Y_1),\ldots,(X_n,Y_n)$ *be a sample of independent copies of the random vector* (X,Y) *where* $Y|X=x$ *satisfies* $(\mathcal{R})$, $X \sim b$, *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* $h_n \to 0$, $u_n \to \infty$ *with* $nh_n^p\bar{F}(u_n;x) \to \infty$, *then with probability tending to 1 there exists sequences of solutions* $(\hat{\gamma}_n(x),\hat{\delta}_n(x))$ *of the estimating equations (6) and (7), 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* $n \to \infty$.

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 $\widetilde{m}_n(K,s,j;x)$, as given in (9) and (10). Let

$$\mathbb{S}_n^{(j)}(s) := r_n \left[\frac{T_n(K,s,j;x)}{\bar{F}(u_n;x)b(x)} - \frac{j!\gamma_0^j(x)}{[1-s\gamma_0(x)]^{j+1}} \right], \quad j = 0,1,2.$$

Corollary 1 *Let $(X_1,Y_1),\ldots,(X_n,Y_n)$ be a sample of independent copies of the random vector (X,Y) where $Y|X=x$ satisfies $(\mathcal{R})$, $X \sim b$, 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 $h_n \to 0$, $u_n \to \infty$, with $nh_n^p\bar{F}(u_n;x) \to \infty$, $\sqrt{nh_n^p\bar{F}(u_n;x)}\delta(u_n;x) \to \lambda \in \mathbb{R}$, $\sqrt{nh_n^p\bar{F}(u_n;x)}h_n \to 0$, $\sqrt{nh_n^p\bar{F}(u_n;x)}\Phi(u_n,h_n;x) \to 0$, then in $\mathcal{C}^3([S,0])$*

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

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

$$\mathbb{E}[\mathbb{S}^{(j)}(s)] = -\lambda\sqrt{b(x)}j!\gamma_0^{j-1}(x) \left[\frac{1}{[1-s\gamma_0(x)]^{j+1}} - \frac{1-\rho_0(x)}{[1-\rho_0(x)-s\gamma_0(x)]^{j+1}} \right], \quad j = 0,1,2,$$

and covariance functions as given in (11).

Theorem 3 *Let $(X_1,Y_1),\ldots,(X_n,Y_n)$ be a sample of independent copies of the random vector (X,Y) where $Y|X=x$ satisfies $(\mathcal{R})$, $X \sim b$, 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 (6) and (7), with ρ fixed at $\rho_0(x)$. For all $x \in \mathbb{R}^p$ where $b(x) > 0$, we have that if $h_n \to 0$, $u_n \to \infty$ with $nh_n^p\bar{F}(u_n;x) \to \infty$, $\sqrt{nh_n^p\bar{F}(u_n;x)}\delta(u_n;x) \to \lambda \in \mathbb{R}$, $\sqrt{nh_n^p\bar{F}(u_n;x)}h_n \to 0$, and $\sqrt{nh_n^p\bar{F}(u_n;x)}\Phi(u_n,h_n;x) \to 0$, then*

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

for $n \to \infty$, where the elements of the matrix $\mathbf{\Sigma}(\rho_0(x))$ are given by (18)-(27), and the matrices $\mathbb{B}(\rho_0(x))$ and $\mathbb{C}(\rho_0(x))$ are defined in (29) and (30), respectively.

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 $(X_1,Y_1),\ldots,(X_n,Y_n)$ 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 (6) and (7). Suppose also that $X \sim b$, 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 $h_n \to 0$, $u_n \to \infty$ with $nh_n^p\bar{F}(u_n;x) \to \infty$, when $n \to \infty$, then with probability tending to 1 there exists sequences of solutions $(\hat{\gamma}_n(x), \hat{\delta}_n(x))$ of the estimating equations (6) and (7) such that $(\hat{\gamma}_n(x), \hat{\delta}_n(x)) \xrightarrow{\mathbb{P}} (\gamma_0(x), 0)$.*

If additionally $\sqrt{nh_n^p \bar{F}(u_n;x)}\delta(u_n;x) \to \lambda \in \mathbb{R}$, $\sqrt{nh_n^p \bar{F}(u_n;x)}h_n \to 0$, *and* $\sqrt{nh_n^p \bar{F}(u_n;x)}\Phi(u_n,h_n;x) \to 0$, *then*

$$r_n \left[\begin{array}{c} \hat{\gamma}_n(x) - \gamma_0(x) \\ \hat{\delta}_n(x) \end{array} \right] \rightsquigarrow N_2(-\lambda\sqrt{b(x)}\mathbb{C}^{-1}(\tilde{\rho}(x))\mathbb{B}(\tilde{\rho}(x))\tilde{\mathbb{D}},$$
$$\mathbb{C}^{-1}(\tilde{\rho}(x))\mathbb{B}(\tilde{\rho}(x))\mathbf{\Sigma}(\tilde{\rho}(x))\mathbb{B}'(\tilde{\rho}(x))\mathbb{C}^{-1}(\tilde{\rho}(x))),$$

for $n \to \infty$, *where the elements of the vector* $\widetilde{\mathbb{D}}$ *are defined in (14), (15), (31) and (17).*

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 (6) and (7). For an example of a locally consistent estimator for $\rho(x)$ we refer to Goegebeur *et al.* (2012).

Theorem 4 *Let* $(X_1,Y_1),\ldots,(X_n,Y_n)$ *be a sample of independent copies of the random vector* (X,Y) *where* $Y|X=x$ *satisfies* $(\mathcal{R})$ *and* $X \sim b$. *The result of Theorem 2 and 3 continues to hold if* ρ *is replaced by an external consistent estimator* $\hat{\rho}_n(x)$ *in (6) and (7)*

3 Simulation study

In this section, we illustrate the finite sample behavior of our estimator $\hat{\gamma}_n$ on a small simulation study. In particular we compare our estimator with the following non-robust and biased version proposed in Goegebeur *et al.* (2012):

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

with $t=0$ and two 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 ρ we use as in Goegebeur *et al.* (2012) a Fraga Alves (2003) type estimator.
In the robust case, a first order estimator is obtained by setting $\delta = 0$, whereas a second order bias-corrected version is derived by estimating δ. In that case the value of ρ is either fixed to -1 or estimated as previously mentioned. All kernels are taken as the bi-quadratic kernel function

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

For 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 $\hat{\gamma}_n$ with $\alpha = 0$.

		$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 1$
$x_0 = 0.1$ or $x_0 = 0.9$	ρ_0 = -0.5	0.93	0.41	0.19
	ρ_0 = -1	0.94	0.47	0.25
	ρ_0 = -2	0.95	0.54	0.32
$x_0 = 0.5$	ρ_0 = -0.5	0.92	0.38	0.18
	ρ_0 = -1	0.93	0.44	0.23
	ρ_0 = -2	0.94	0.51	0.31

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

To determine an optimal value for k and h, two strategies are applied as in Goegebeur et al. (2012): an oracle strategy and a completely data driven method. For the oracle strategy, the same algorithm as in Goegebeur et al. (2012) was applied to our new estimator of γ, that is

$$(h_o, k_o) := \underset{h \in \mathcal{H}_0, k \in \mathcal{K}_0}{\arg\min} \ \Delta\left(\widehat{\gamma}(\cdot), \gamma(\cdot)\right), \tag{12}$$

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

$$\Delta^2\left(\widehat{\gamma}(\cdot), \gamma(\cdot)\right) := \frac{1}{M} \sum_{m=1}^{M} \left(\widehat{\gamma}(z_m) - \gamma(z_m)\right)^2,$$

where $z_1, \ldots, z_M$ are regularly spaced in the covariate space. Note that this method requires knowledge of the function $\gamma(x)$, which is unknown in practical situations. The minimization is performed on a grid of $h \in [0.05; 0.5]$ and of $k \in \{2, ..., m_x - 1\}$ with $M = 35$.

For the completely data driven method, the optimal bandwidth h is determined again using the leave-one-out cross-validation method of Goegebeur *et al.* (2012) and for the optimal k we proceed as follows for all x under consideration:

- we compute the estimates for $\gamma(x)$ with $k = 5, 9, 13, \ldots \lfloor m_x - 4 \rfloor$ (m_x being the number of observations in the ball $B(x, h)$);
- 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.

We simulate $N = 100$ samples of size $n = 1\,000$ from the following Burr distribution

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

where

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

In case of contamination in the response, the following distribution will be used

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

where $\tilde{F}(y;x) = 1 - \left(\frac{y}{x_c}\right)^{-0.5}, y > x_c$.
Different settings have been 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 $F(y;x)$.

Below, we only illustrate the settings 1, 2 and 4, since the settings 3 and 5 give results very similar to 2 and 4, respectively.

From Table 2, we see that the non-robust bias-corrected estimator with ρ fixed at -1 behaves best in terms of MSE. 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 γ. 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.

Figure 1: Setting 1: boxplots of $\hat{\gamma}_n$ with k_{opt} and h_{opt} 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 $\rho = -1$, row 3: bias-corrected estimator with $\rho = \hat{\rho}$.

Figure 2: Setting 1: boxplots of $\hat{\gamma}_n$ with k_{opt} and h_{opt} 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 $\rho = -1$, row 3: bias-corrected estimator with $\rho = \hat{\rho}$.

In Setting 2, the robust estimators 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 behaviour 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

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

Table 2: MSE for different estimators of γ based on 100 data sets simulated according to Setting 1.

the non-robust estimators, whereas the best results in MSE are obtained for the robust estimator with $\alpha = 0.5$ and ρ fixed at -1, although the result for ρ estimated is not much worse. Note also that in these best cases the results obtained by the data driven method are comparable to results of the oracle strategy.

Figure 3: Setting 2: boxplots of $\hat{\gamma}_n$ with k_{opt} and h_{opt} 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 $\rho = -1$, row 3: bias-corrected estimator with $\rho = \hat{\rho}$.

Figure 4: Setting 2: boxplots of $\hat{\gamma}_n$ with k_{opt} and h_{opt} 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 $\rho = -1$, row 3: bias-corrected estimator with $\rho = \hat{\rho}$.

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 ρ fixed to -1 are somewhat better than the estimators with ρ estimated and the data and oracle strategies are comparable.
In conclusion, we can conclude that in both cases (contaminated and uncontaminated) the robust bias-corrected estimator with $\alpha = 0.5$ and ρ fixed at -1 gives the best results, although using $\hat{\rho}$ is not much worse.

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

Table 3: MSE for different estimators of γ based on 100 data sets simulated according to Setting 2.

Figure 5: Setting 4: boxplots of $\hat{\gamma}_n$ with k_{opt} and h_{opt} 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 $\rho = -1$, row 3: bias-corrected estimator with $\rho = \hat{\rho}$.

Figure 6: Setting 4: boxplots of $\hat{\gamma}_n$ with k_{opt} and h_{opt} 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 $\rho = -1$, row 3: bias-corrected estimator with $\rho = \hat{\rho}$.

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

Table 4: MSE for different estimators of γ based on 100 data sets simulated according to Setting 4.

Appendix

Proof of Lemma 1

The case $(s,t) = (0,0)$ is trivial. In case $(s,t) \neq (0,0)$, we obtain, using integration by parts,

$$\begin{aligned} m(u_n,s,t;x) &= \bar{F}(u_n;x)\left\{\int_1^\infty \left[sz^{s-1}(\ln z)^t + tz^{s-1}(\ln z)^{t-1}\right]\bar{G}(z;\gamma(x),\delta(u_n;x),\rho(x))dz \right.\\ &\quad \left. + \int_1^\infty \left[sz^{s-1}(\ln z)^t + tz^{s-1}(\ln z)^{t-1}\right]\left[\frac{\bar{F}(u_n z;x)}{\bar{F}(u_n;x)} - \bar{G}(z;\gamma(x),\delta(u_n;x),\rho(x))\right]dz\right\} \\ &=: \bar{F}(u_n;x)(T_1+T_2). \end{aligned}$$

By application of Taylor's theorem to $\bar{G}$, we have that

$$\begin{aligned} T_1 &= \int_1^\infty \left[s z^{s-1} (\ln z)^t + t z^{s-1} (\ln z)^{t-1} \right] z^{-1/\gamma(x)} dz \\ &\quad - \frac{\delta(u_n;x)}{\gamma(x)} \int_1^\infty \left[s z^{s-1} (\ln z)^t + t z^{s-1} (\ln z)^{t-1} \right] z^{-1/\gamma(x)} \left[1 - z^{\rho(x)/\gamma(x)} \right] dz + o(\delta(u_n;x)) \\ &=: T_{1,1} - \frac{\delta(u_n;x)}{\gamma(x)} T_{1,2} + o(\delta(u_n;x)). \end{aligned}$$

Straightforward integration then gives

$$\begin{aligned} T_{1,1} &= \frac{\gamma^t(x)\Gamma(t+1)}{(1 - s\gamma(x))^{t+1}}, \\ T_{1,2} &= \gamma^t(x)\Gamma(t+1) \left[\frac{1}{(1 - s\gamma(x))^{t+1}} - \frac{1 - \rho(x)}{(1 - \rho(x) - s\gamma(x))^{t+1}} \right], \end{aligned}$$

and thus

$$T_1 = \gamma^t(x)\Gamma(t+1) \left\{ \frac{1}{(1 - s\gamma(x))^{t+1}} - \frac{\delta(u_n;x)}{\gamma(x)} \left[\frac{1}{(1 - s\gamma(x))^{t+1}} - \frac{1 - \rho(x)}{(1 - \rho(x) - s\gamma(x))^{t+1}} \right] (1 + o(1)) \right\}.$$

A slight modification of Proposition 2.3 in Beirlant *et al.* (2009) gives that

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

and hence $T_2 = o(\delta(u_n;x))$.

Combining the above results establishes Lemma 1.

Proof of Lemma 2

By application of the rule of repeated expectations we obtain

$$\begin{aligned} \widetilde{m}_n(K,s,t;x) &= \mathbb{E}[K_{h_n}(x - X) m(u_n,s,t;X)] \\ &= \int_\Omega K(z) m(u_n,s,t;x - h_n z) b(x - h_n z) dz, \end{aligned}$$

so, by straightforward calculations,

$$\begin{aligned} &|\widetilde{m}_n(K,s,t;x) - b(x) m(u_n,s,t;x)| \\ &\quad \leq m(u_n,s,t;x) \int_\Omega K(z) |b(x - h_n z) - b(x)| dz \\ &\quad + b(x) \int_\Omega K(z) |m(u_n,s,t;x - h_n z) - m(u_n,s,t;x)| dz \\ &\quad + \int_\Omega K(z) |b(x - h_n z) - b(x)| |m(u_n,s,t;x - h_n z) - m(u_n,s,t;x)| dz \\ &\quad =: T_3 + T_4 + T_5. \end{aligned}$$

Concerning T_3, by $(\mathcal{B})$ and $(\mathcal{K})$

$$\begin{aligned} T_3 &\leq m(u_n, s, t; x) c_b h_n \int_\Omega K(z) d(0, z) dz \\ &= m(u_n, s, t; x) b(x) O(h_n). \end{aligned}$$

The term T_4 can be analyzed by invoking $(\mathcal{M})$ and $(\mathcal{K})$ yielding

$$T_4 = m(u_n, s, t; x) b(x) O(\Phi(u_n, h_n; x)).$$

Finally, applying similar arguments to T_5 gives that $T_5 = m(u_n, s, t; x) b(x) O(h_n \Phi(u_n, h_n; x))$, and the result follows.

Proof of Theorem 1

Note that

$$\begin{aligned} \mathbb{P}_n^{(j)}(s) &= \sqrt{n} \left[\frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{h_n^p \bar{F}(u_n; x) b(x)}} K\left(\frac{x - X_i}{h_n}\right) \left(\frac{Y_i}{u_n}\right)^s \left(\ln \frac{Y_i}{u_n}\right)^j \mathbf{1}\{Y_i > u_n\} \right. \\ & \left. - \mathbb{E}\left(\frac{1}{\sqrt{h_n^p \bar{F}(u_n; x) b(x)}} K\left(\frac{x - X}{h_n}\right) \left(\frac{Y}{u_n}\right)^s \left(\ln \frac{Y}{u_n}\right)^j \mathbf{1}\{Y > u_n\} \right) \right], \quad j = 0, 1, 2. \end{aligned}$$

As such, the empirical processes under consideration fit in the framework of Section 19.5 in van der Vaart (2007) on changing function classes. Indeed, we can consider the classes $\mathcal{W}_n^{(j)} := \{w_{n,s}^{(j)}; s \in [S, 0]\}$, where

$$w_{n,s}^{(j)}(v, y) := \frac{1}{\sqrt{h_n^p \bar{F}(u_n; x) b(x)}} K\left(\frac{x - v}{h_n}\right) \left(\frac{y}{u_n}\right)^s \left(\ln \frac{y}{u_n}\right)^j \mathbf{1}\{y > u_n\}, \quad j = 0, 1, 2.$$

So, for the marginal convergence of the processes, it is sufficient to verify the conditions of Theorem 19.28 in van der Vaart (2007).

First, by Lemmas 1 and 2

$$\begin{aligned} & \mathbb{E}[w_{n,s}^{(j)}(X, Y) - w_{n,t}^{(j)}(X, Y)]^2 \\ & \quad = \frac{\|K\|_2^2}{\bar{F}(u_n; x) b(x)} \mathbb{E}\left(\frac{1}{h_n^p \|K\|_2^2} K^2\left(\frac{x - X}{h_n}\right) \left[\left(\frac{Y}{u_n}\right)^s - \left(\frac{Y}{u_n}\right)^t \right]^2 \left(\ln \frac{Y}{u_n}\right)^{2j} \mathbf{1}\{Y > u_n\} \right) \\ & \quad \leq \frac{\|K\|_2^2 (s - t)^2}{\bar{F}(u_n; x) b(x)} \mathbb{E}\left(\frac{1}{h_n^p \|K\|_2^2} K^2\left(\frac{x - X}{h_n}\right) \left(\ln \frac{Y}{u_n}\right)^{2(j+1)} \mathbf{1}\{Y > u_n\} \right) \\ & \quad = (2(j+1))! \gamma^{2(j+1)}(x) (s - t)^2 \|K\|_2^2 (1 + o(1)). \end{aligned}$$

Note that the $o(1)$ term above does not depend on s and t, and therefore

$$\begin{aligned} \sup_{|s-t| < \delta_n} \mathbb{E}[w_{n,s}^{(j)}(X, Y) - w_{n,t}^{(j)}(X, Y)]^2 &\leq (2(j+1))! \gamma^{2(j+1)}(x) \delta_n^2 \|K\|_2^2 (1 + o(1)) \\ &\to 0, \end{aligned}$$

for every sequence $\delta_n \downarrow 0$.

Next we verify the Lindeberg condition. Note that the envelope function $W_n^{(j)}$ for $\mathcal{W}_n^{(j)}$ can be taken as

$$W_n^{(j)}(v,y) = \frac{1}{\sqrt{h_n^p \bar{F}(u_n;x)b(x)}} K\left(\frac{x-v}{h_n}\right)\left(\ln \frac{y}{u_n}\right)^j \mathbf{1}\{y > u_n\}, \quad j=0,1,2.$$

Using Lemmas 1 and 2, we then have

$$\begin{aligned} \mathbb{E}\left[(W_n^{(j)}(X,Y))^2\right] &= \frac{\|K\|_2^2}{\bar{F}(u_n;x)b(x)} \mathbb{E}\left[\frac{1}{h_n^p\|K\|_2^2} K^2\left(\frac{x-X}{h_n}\right)\left(\ln \frac{Y}{u_n}\right)^{2j} \mathbf{1}\{Y > u_n\}\right] \\ &= \gamma^{2j}(x)(2j)!\|K\|_2^2(1+o(1)) = O(1), \end{aligned}$$

and, for every $\varepsilon, \alpha > 0$,

$$\begin{aligned} &\mathbb{E}\left[(W_n^{(j)}(X,Y))^2 \mathbf{1}\{W_n^{(j)}(X,Y) > \varepsilon\sqrt{n}\}\right] \\ &\quad\le \frac{1}{\varepsilon^\alpha n^{\alpha/2}} \mathbb{E}\left[(W_n^{(j)}(X,Y))^{2+\alpha}\right] \\ &\quad= \frac{\|K^{2+\alpha}\|_1}{\varepsilon^\alpha (nh_n^p \bar{F}(u_n;x)b(x))^{\alpha/2}} \frac{1}{\bar{F}(u_n;x)b(x)} \mathbb{E}\left[\frac{1}{h_n^p\|K^{2+\alpha}\|_1} K^{2+\alpha}\left(\frac{x-X}{h_n}\right)\left(\ln_+ \frac{Y}{u_n}\right)^{j(2+\alpha)} \mathbf{1}\{Y > u_n\}\right] \\ &\quad= O\left(\frac{1}{(nh_n^p\bar{F}(u_n;x))^{\alpha/2}}\right) \to 0, \end{aligned}$$

if $nh_n^p\bar{F}(u_n;x) \to \infty$, $j=0,1,2$.

Thirdly, we verify the condition on the bracketing integrals $J_{[\,]}(\delta_n, \mathcal{W}_n^{(j)}, L_2(\mathbb{P}))$, $j=0,1,2$, in Theorem 19.28 of van der Vaart (2007). We have that

$$\begin{aligned} |w_{n,s}^{(j)}(v,y) - w_{n,t}^{(j)}(v,y)| &\le \frac{|s-t|}{\sqrt{h_n^p\bar{F}(u_n;x)b(x)}} K\left(\frac{x-v}{h_n}\right)\left(\ln \frac{y}{u_n}\right)^{j+1} \mathbf{1}\{y > u_n\}, \\ &=: |s-t| w^{(j)}(v,y). \end{aligned}$$

Note

$$\mathbb{E}\left[\left(w^{(j)}(X,Y)\right)^2\right] = \gamma^{2(j+1)}(x)(2(j+1))!\|K\|_2^2(1+o(1)), \quad j=0,1,2.$$

So that the condition on $J_{[\,]}(\delta_n, \mathcal{W}_n^{(j)}, L_2(\mathbb{P}))$, $j=0,1,2$, is easy to verify using the result of Example 19.7 in van der Vaart (2007) on parametric function classes.

Finally, we comment on the pointwise convergence of the covariance functions on $[S,0]^2$. For $(s_1,s_2)\in[S,0]^2$ we have that

$$\begin{aligned}
&\mathbb{C}ov(\mathbb{P}_n^{(j)}(s_1),\mathbb{P}_n^{(j)}(s_2))\\
&= \mathbb{C}ov(w_{n,s_1}^{(j)}(X,Y),w_{n,s_2}^{(j)}(X,Y))\\
&= \frac{\|K\|_2^2}{\bar F(u_n;x)b(x)}\mathbb{E}\left[\frac{1}{h_n^p\|K\|_2^2}K^2\left(\frac{x-X}{h_n}\right)\left(\frac{Y}{u_n}\right)^{s_1+s_2}\left(\ln\frac{Y}{u_n}\right)^{2j}\mathbf{1}\{Y>u_n\}\right]\\
&\quad -\frac{h_n^p}{\bar F(u_n;x)b(x)}\mathbb{E}\left[K_{h_n}(x-X)\left(\frac{Y}{u_n}\right)^{s_1}\left(\ln\frac{Y}{u_n}\right)^{j}\mathbf{1}\{Y>u_n\}\right]\times\\
&\quad \mathbb{E}\left[K_{h_n}(x-X)\left(\frac{Y}{u_n}\right)^{s_2}\left(\ln\frac{Y}{u_n}\right)^{j}\mathbf{1}\{Y>u_n\}\right]\\
&\to \frac{\gamma^{2j}(x)(2j)!\|K\|_2^2}{[1-(s_1+s_2)\gamma(x)]^{1+2j}},\quad n\to\infty;\quad j=0,1,2.
\end{aligned}$$

The joint convergence of the empirical processes follows then from the fact that the coordinate classes being Donsker is equivalent to the union of the coordinate classes being Donsker, see van der Vaart p. 270. The pointwise convergence of the covariances between the processes $\mathbb{P}_n^{(j)}$, $j=0,1,2$, can be established along the same line of arguments as above.

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 $\widehat\Delta_\alpha(\gamma,\delta;\rho)$ denote the density power divergence objective function. Note that r should be such that Q_r is a subset of the parameter space. First we rescale $\widehat\Delta_\alpha(\gamma,\delta;\rho)$ as $\widetilde\Delta_\alpha(\gamma,\delta;\rho):=\widehat\Delta_\alpha(\gamma,\delta;\rho)/(\bar F(u_n;x)b(x))$, and we show that for any r sufficiently small

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

Let $f_s(\gamma,\delta;\rho_0(x))$, $s=1,2$, denote the derivatives of $\widetilde\Delta_\alpha(\gamma,\delta;\rho_0(x))$ with respect to γ and δ, respectively, without the common scale factor $1+\alpha$. Similarly, f_{st} and f_{stu}, $s,t,u=1,2$, denote the second and third order derivatives, respectively (again apart from the common scaling by $1+\alpha$).

By Taylor's theorem

$$
\begin{aligned}
&\widetilde{\Delta}_\alpha(\gamma,\delta;\rho_0(x)) - \widetilde{\Delta}_\alpha(\gamma_0(x),0;\rho_0(x)) \\
&= (1+\alpha)\left\{f_1(\gamma_0(x),0;\rho_0(x))(\gamma-\gamma_0(x)) + f_2(\gamma_0(x),0;\rho_0(x))\delta\right. \\
&\quad +\frac{1}{2}\left[f_{11}(\gamma_0(x),0;\rho_0(x))(\gamma-\gamma_0(x))^2 + f_{22}(\gamma_0(x),0;\rho_0(x))\delta^2 + 2f_{12}(\gamma_0(x),0;\rho_0(x))(\gamma-\gamma_0(x))\delta\right] \\
&\quad +\frac{1}{6}\left[f_{111}(\tilde{\gamma},\tilde{\delta};\rho_0(x))(\gamma-\gamma_0(x))^3 + f_{222}(\tilde{\gamma},\tilde{\delta};\rho_0(x))\delta^3 + 3f_{112}(\tilde{\gamma},\tilde{\delta};\rho_0(x))(\gamma-\gamma_0(x))^2\delta\right. \\
&\quad \left.\left.+3f_{122}(\tilde{\gamma},\tilde{\delta};\rho_0(x))(\gamma-\gamma_0(x))\delta^2\right]\right\} \qquad (13) \\
&=: (1+\alpha)\{S_1+S_2+S_3\},
\end{aligned}
$$

where $(\tilde{\gamma},\tilde{\delta})$ is a point on the line segment connecting (γ,δ) and $(\gamma_0(x),0)$. After some tedious, but straightforward derivations one obtains

$$
\begin{aligned}
&f_1(\gamma_0(x),0;\rho_0(x)) \\
&= \gamma_0^{-\alpha-2}(x)\left[-\frac{\alpha\gamma_0(x)(1+\gamma_0(x))}{[1+\alpha(1+\gamma_0(x))]^2}\frac{T_n(K,0,0;x)}{\bar{F}(u_n;x)b(x)}\right. \\
&\quad \left.+\gamma_0(x)\frac{T_n(K,-\alpha(1+\gamma_0(x))/\gamma_0(x),0;x)}{\bar{F}(u_n;x)b(x)} - \frac{T_n(K,-\alpha(1+\gamma_0(x))/\gamma_0(x),1;x)}{\bar{F}(u_n;x)b(x)}\right], \\
&f_2(\gamma_0(x),0;\rho_0(x)) \\
&= \gamma_0^{-\alpha-1}(x)\left[-\frac{\alpha\rho_0(x)(1+\gamma_0(x))}{[1+\alpha(1+\gamma_0(x))][1-\rho_0(x)+\alpha(1+\gamma_0(x))]}\frac{T_n(K,0,0;x)}{\bar{F}(u_n;x)b(x)}\right. \\
&\quad \left.+\frac{T_n(K,-\alpha(1+\gamma_0(x))/\gamma_0(x),0;x)}{\bar{F}(u_n;x)b(x)} -(1-\rho_0(x))\frac{T_n(K,-(\alpha(1+\gamma_0(x))-\rho_0(x))/\gamma_0(x),0;x)}{\bar{F}(u_n;x)b(x)}\right].
\end{aligned}
$$

By using the results of Lemmas 1, 2 and Theorem 1, we have that $f_1(\gamma_0(x),0;\rho_0(x)) \stackrel{\mathbb{P}}{\to} 0$ and $f_2(\gamma_0(x),0;\rho_0(x)) \stackrel{\mathbb{P}}{\to} 0$, so, for any given $r>0$ we have that $|f_1(\gamma_0(x),0;\rho_0(x))| < r^2$ and $|f_2(\gamma_0(x),0;\rho_0(x))| < r^2$ with probability tending to 1, and hence, on Q_r, $|S_1| < 2r^3$ with probability tending to 1.

We now focus on the second order derivatives appearing in S_2. Again, by tedious calculus one obtains

$$
\begin{aligned}
&f_{11}(\gamma_0(x),0;\rho_0(x))\\
&= \gamma_0^{-\alpha-2}(x)\Bigg[\left(\frac{\alpha+2}{1+\alpha(1+\gamma_0(x))}-\frac{2\alpha+4}{[1+\alpha(1+\gamma_0(x))]^2}+\frac{2\alpha+2}{[1+\alpha(1+\gamma_0(x))]^3}\right)\frac{T_n(K,0,0;x)}{\bar F(u_n;x)b(x)}\\
&\quad -(\alpha+1)\frac{T_n(K,-\alpha(1+\gamma_0(x))/\gamma_0(x),0;x)}{\bar F(u_n;x)b(x)}+\frac{2\alpha+2}{\gamma_0(x)}\frac{T_n(K,-\alpha(1+\gamma_0(x))/\gamma_0(x),1;x)}{\bar F(u_n;x)b(x)}\\
&\quad -\frac{\alpha}{\gamma_0^2(x)}\frac{T_n(K,-\alpha(1+\gamma_0(x))/\gamma_0(x),2;x)}{\bar F(u_n;x)b(x)}\Bigg],\\
&f_{12}(\gamma_0(x),0;\rho_0(x))\\
&= \gamma_0^{-\alpha-2}(x)\Bigg[\Bigg(\frac{1+\alpha(2+\alpha)(1+\gamma_0(x))}{[1+\alpha(1+\gamma_0(x))]^2}\\
&\quad -\frac{(1-\rho_0(x))^2-\alpha[\rho_0(x)(1-\rho_0(x))-2(1+\gamma_0(x))(1-\rho_0(x))]+\alpha^2(1+\gamma_0(x))(1-\rho_0(x))}{[1-\rho_0(x)+\alpha(1+\gamma_0(x))]^2}\Bigg)\\
&\quad \times\frac{T_n(K,0,0;x)}{\bar F(u_n;x)b(x)}-(1+\alpha)\frac{T_n(K,-\alpha(1+\gamma_0(x))/\gamma_0(x),0;x)}{\bar F(u_n;x)b(x)}\\
&\quad +(\alpha+1)(1-\rho_0(x))\frac{T_n(K,-(\alpha(1+\gamma_0(x))-\rho_0(x))/\gamma_0(x),0;x)}{\bar F(u_n;x)b(x)}\\
&\quad +\frac{\alpha}{\gamma_0(x)}\frac{T_n(K,-\alpha(1+\gamma_0(x))/\gamma_0(x),1;x)}{\bar F(u_n;x)b(x)}\\
&\quad -\frac{(\alpha-\rho_0(x))(1-\rho_0(x))}{\gamma_0(x)}\frac{T_n(K,-(\alpha(1+\gamma_0(x))-\rho_0(x))/\gamma_0(x),1;x)}{\bar F(u_n;x)b(x)}\Bigg],\\
&f_{22}(\gamma_0(x),0;\rho_0(x))\\
&= \gamma_0^{-\alpha-2}(x)\Bigg[\Bigg(\frac{1+\alpha+\gamma_0(x)}{1+\alpha(1+\gamma_0(x))}-\frac{2(1-\rho_0(x))(1+\gamma_0(x)+\alpha)}{1-\rho_0(x)+\alpha(1+\gamma_0(x))}\\
&\quad +\frac{(1+\gamma_0(x))(1-2\rho_0(x))+\alpha(1-\rho_0(x))^2}{1-2\rho_0(x)+\alpha(1+\gamma_0(x))}\Bigg)\frac{T_n(K,0,0;x)}{\bar F(u_n;x)b(x)}\\
&\quad -(\alpha+\gamma_0(x))\frac{T_n(K,-\alpha(1+\gamma_0(x))/\gamma_0(x),0;x)}{\bar F(u_n;x)b(x)}\\
&\quad +2(1-\rho_0(x))(\alpha+\gamma_0(x))\frac{T_n(K,-(\alpha(1+\gamma_0(x))-\rho_0(x))/\gamma_0(x),0;x)}{\bar F(u_n;x)b(x)}\\
&\quad -[(1+\gamma_0(x))(1-2\rho_0(x))+(\alpha-1)(1-\rho_0(x))^2]\frac{T_n(K,-(\alpha(1+\gamma_0(x))-2\rho_0(x))/\gamma_0(x),0;x)}{\bar F(u_n;x)b(x)}\Bigg].
\end{aligned}
$$

Now, let $f^*_{st}(\gamma_0(x),0;\rho_0(x))$ denote the limits of the random terms $f_{st}(\gamma_0(x),0;\rho_0(x))$, $s,t=1,2$.

These can be obtained from the results of Lemmas 1, 2 and Theorem 1, and are given by

$$
\begin{aligned}
&f^*_{11}(\gamma_0(x), 0; \rho_0(x)) = \gamma_0^{-\alpha-2}(x)\, \frac{1+\alpha^2(1+\gamma_0(x))^2}{[1+\alpha(1+\gamma_0(x))]^3},\\
&f^*_{12}(\gamma_0(x), 0; \rho_0(x))\\
&\quad = \gamma_0^{-\alpha-2}(x)\, \frac{\rho_0(x)(1-\rho_0(x))[1+\alpha(1+\gamma_0(x))+\alpha^2(1+\gamma_0(x))^2]+\alpha^3\rho_0(x)(1+\gamma_0(x))^3}{[1+\alpha(1+\gamma_0(x))]^2[1-\rho_0(x)+\alpha(1+\gamma_0(x))]^2},\\
&f^*_{22}(\gamma_0(x), 0; \rho_0(x))\\
&\quad = \gamma_0^{-\alpha-2}(x)\, \frac{(1-\rho_0(x))\rho_0^2(x)+\alpha\rho_0^2(x)(1+\gamma_0(x))[\alpha(1+\gamma_0(x))-\rho_0(x)]}{[1+\alpha(1+\gamma_0(x))][1-\rho_0(x)+\alpha(1+\gamma_0(x))][1-2\rho_0(x)+\alpha(1+\gamma_0(x))]}.
\end{aligned}
$$

Now, write

$$
\begin{aligned}
2S_2 \;=\;& f^*_{11}(\gamma_0(x), 0; \rho_0(x))(\gamma-\gamma_0(x))^2 + f^*_{22}(\gamma_0(x), 0; \rho_0(x))\delta^2 + 2f^*_{12}(\gamma_0(x), 0; \rho_0(x))(\gamma-\gamma_0(x))\delta\\
&+[f_{11}(\gamma_0(x), 0; \rho_0(x)) - f^*_{11}(\gamma_0(x), 0; \rho_0(x))](\gamma-\gamma_0(x))^2\\
&+[f_{22}(\gamma_0(x), 0; \rho_0(x)) - f^*_{22}(\gamma_0(x), 0; \rho_0(x))]\delta^2\\
&+2[f_{12}(\gamma_0(x), 0; \rho_0(x)) - f^*_{12}(\gamma_0(x), 0; \rho_0(x))](\gamma-\gamma_0(x))\delta.
\end{aligned}
$$

Note that the first three terms are in fact a nonrandom positive definite quadratic form in $(\gamma - \gamma_0(x))$ and δ. This can be verified analytically, but the result is not included in the paper. By the spectral decomposition this quadratic form can be rewritten as $\lambda_1\xi_1^2 + \lambda_2\xi_2^2$, where $0 < \lambda_1 \leq \lambda_2$ are the eigenvalues and ξ_1 and ξ_2 are orthogonal transformations of $(\gamma - \gamma_0(x))$ and δ. Note that in this new coordinate system Q_r becomes $\xi_1^2 + \xi_2^2 = r^2$. Thus, for the quadratic form we have that $\lambda_1\xi_1^2 + \lambda_2\xi_2^2 \geq \lambda_1(\xi_1^2 + \xi_2^2) = \lambda_1 r^2$. For the random part of S_2 we know from Lemmas 1, 2 and Theorem 1 that $f_{st}(\gamma_0(x), 0; \rho_0(x)) \xrightarrow{\mathbb{P}} f^*_{st}(\gamma_0(x), 0; \rho_0(x))$, $s, t = 1, 2$, and thus in absolute value the random part is less than $4r^3$ with probability tending to 1. Overall, we have that there exists $c > 0$ and $r_0 > 0$ such that for $r < r_0$

$$S_2 > cr^2$$

with probability tending to 1.

For the term S_3, one can show that $|f_{stu}(\gamma, \delta; \rho_0(x))| \leq M_{stu}(\boldsymbol{V})$, where $\boldsymbol{V} := [(X_1, Y_1), \ldots, (X_n, Y_n)]$, for $(\gamma, \delta) \in Q_r$, with $M_{stu}(\boldsymbol{V}) \xrightarrow{\mathbb{P}} m_{stu}$, $s, t, u = 1, 2$, which is bounded. The derivations are straightforward, and are for brevity omitted from the paper. Thus, with probability tending to 1, $|f_{stu}(\tilde{\gamma}, \tilde{\delta}; \rho_0(x))| < 2m_{stu}$, and hence $|S_3| < er^3$ on Q_r, where

$$e := \frac{1}{3}\sum_{s=1}^{2}\sum_{t=1}^{2}\sum_{u=1}^{2} m_{stu}.$$

Combining the above we find that with probability tending to 1,

$$\min(S_1 + S_2 + S_3) > cr^2 - (2+e)r^3,$$

where the minimum is over (γ, δ) on the surface of Q_r. Clearly, the right-hand side of the above inequality is positive if $r < c/(2+e)$.

To complete the proof of the existence and consistency we adjust the line of argumentation of Theorem 3.7 in Chapter 6 of Lehmann and Casella (1998). For $r > 0$, small enough that Q_r is a subset of the parameter space, consider

$$S_n(r) := \{\boldsymbol{v} : \widetilde{\Delta}_\alpha(\gamma_0(x), 0; \rho_0(x)) < \widetilde{\Delta}_\alpha(\gamma, \delta; \rho_0(x)) \text{ for all } (\gamma, \delta) \text{ on the surface of } Q_r\}.$$

From the above we have that $\mathbb{P}_{(\gamma_0(x),0)}(S_n(r)) \to 1$ for any such r, and hence there exists a sequence $r_n^* \downarrow 0$ such that $\mathbb{P}_{(\gamma_0(x),0)}(S_n(r_n^*)) \to 1$ as $n \to \infty$. By the differentiability of $\widetilde{\Delta}_\alpha(\gamma, \delta; \rho_0(x))$ we have that $\boldsymbol{v} \in S_n(r_n^*)$ implies that there exists a point $(\hat{\gamma}_n(r_n^*), \hat{\delta}_n(r_n^*)) \in Q_{r_n^*}$ for which $\widetilde{\Delta}_\alpha(\gamma, \delta; \rho_0(x))$ attains a local minimum, and thus $f_s(\hat{\gamma}_n(r_n^*), \hat{\delta}_n(r_n^*); \rho_0(x)) = 0$, $s = 1, 2$. Now let $(\hat{\gamma}_n^*(x), \hat{\delta}_n^*(x)) := (\hat{\gamma}_n(r_n^*), \hat{\delta}_n(r_n^*))$ for $\boldsymbol{v} \in S_n(r_n^*)$ and arbitrary otherwise. Clearly

$$\mathbb{P}_{(\gamma_0(x),0)}(f_1(\hat{\gamma}_n^*(x), \hat{\delta}_n^*(x); \rho_0(x)) = 0, f_2(\hat{\gamma}_n^*(x), \hat{\delta}_n^*(x); \rho_0(x)) = 0) \geq \mathbb{P}_{(\gamma_0(x),0)}(S_n(r_n^*)) \to 1,$$

as $n \to \infty$. Thus with probability tending to 1 there exists a sequence of solutions to the estimating equations (6) and (7). Also, for any fixed $r > 0$ and n sufficiently large

$$\begin{aligned} \mathbb{P}_{(\gamma_0(x),0)}(d((\hat{\gamma}_n^*(x), \hat{\delta}_n^*(x)), (\gamma_0(x), 0)) < r) &\geq \mathbb{P}_{(\gamma_0(x),0)}(d((\hat{\gamma}_n^*(x), \hat{\delta}_n^*(x)), (\gamma_0(x), 0)) < r_n^*) \\ &\geq \mathbb{P}_{(\gamma_0(x),0)}(S_n(r_n^*)) \to 1, \end{aligned}$$

which establishes the consistency of the sequence $(\hat{\gamma}_n^*(x), \hat{\delta}_n^*(x))$.

Proof of Corollary 1

We have that

$$\mathbb{S}_n^{(j)}(s) = \mathbb{P}_n^{(j)}(s) + r_n \left[\mathbb{E}\left(\frac{T_n(K, s, j; x)}{\bar{F}(u_n; x) b(x)} \right) - \frac{j! \gamma_0^j(x)}{[1 - s\gamma_0(x)]^{1+j}} \right], \quad j = 0, 1, 2.$$

From Lemmas 1 and 2

$$\begin{aligned} & r_n \left[\mathbb{E}\left(\frac{T_n(K, s, j; x)}{\bar{F}(u_n; x) b(x)} \right) - \frac{j! \gamma_0^j(x)}{[1 - s\gamma_0(x)]^{1+j}} \right] \\ & \quad = -\lambda \sqrt{b(x)} j! \gamma_0^{j-1}(x) \left[\frac{1}{[1 - s\gamma_0(x)]^{j+1}} - \frac{1 - \rho_0(x)}{[1 - \rho_0(x) - s\gamma_0(x)]^{j+1}} \right] + o(1), \quad j = 0, 1, 2, \end{aligned}$$

where the $o(1)$ terms are uniform in $s \in [S, 0]$.

Proof of Theorem 3

To start 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. Let

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

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

and set $\overline{\mathbb{A}}_n(\rho_0(x)) := r_n[\mathbb{T}_n - \widetilde{\mathbb{T}}]$. Thus, from Corollary 1, we get that

$$\overline{\mathbb{A}}_n(\rho_0(x)) \rightsquigarrow N_4(\lambda\sqrt{b(x)}\mathbb{D}, \mathbf{\Sigma}(\rho_0(x))),$$

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

$$D_1 := 0, \tag{14}$$

$$D_2 := -\frac{\alpha\rho_0(x)(1+\gamma_0(x))}{\gamma_0(x)[1+\alpha(1+\gamma_0(x))][1-\rho_0(x)+\alpha(1+\gamma_0(x))]}, \tag{15}$$

$$D_3 := -\frac{\rho_0(x)[\alpha(1+\gamma_0(x))-\rho_0(x)]}{\gamma_0(x)[1-\rho_0(x)+\alpha(1+\gamma_0(x))][1-2\rho_0(x)+\alpha(1+\gamma_0(x))]}, \tag{16}$$

$$D_4 := \frac{\rho_0(x)(1-\rho_0(x))-\alpha^2\rho_0(x)(1+\gamma_0(x))^2}{[1+\alpha(1+\gamma_0(x))]^2[1-\rho_0(x)+\alpha(1+\gamma_0(x))]^2}, \tag{17}$$

and $\mathbf{\Sigma}(\rho_0(x))$ a symmetric (4×4) matrix with elements

$$\sigma_{11}(\rho_0(x)) := \|K\|_2^2, \tag{18}$$

$$\sigma_{21}(\rho_0(x)) := \frac{\|K\|_2^2}{1+\alpha(1+\gamma_0(x))}, \tag{19}$$

$$\sigma_{22}(\rho_0(x)) := \frac{\|K\|_2^2}{1+2\alpha(1+\gamma_0(x))}, \tag{20}$$

$$\sigma_{31}(\rho_0(x)) := \frac{\|K\|_2^2}{1-\rho_0(x)+\alpha(1+\gamma_0(x))}, \tag{21}$$

$$\sigma_{32}(\rho_0(x)) := \frac{\|K\|_2^2}{1-\rho_0(x)+2\alpha(1+\gamma_0(x))}, \tag{22}$$

$$\sigma_{33}(\rho_0(x)) := \frac{\|K\|_2^2}{1-2\rho_0(x)+2\alpha(1+\gamma_0(x))}, \tag{23}$$

$$\sigma_{41}(\rho_0(x)) := \frac{\gamma_0(x)\|K\|_2^2}{[1+\alpha(1+\gamma_0(x))]^2}, \tag{24}$$

$$\sigma_{42}(\rho_0(x)) := \frac{\gamma_0(x)\|K\|_2^2}{[1+2\alpha(1+\gamma_0(x))]^2}, \tag{25}$$

$$\sigma_{43}(\rho_0(x)) := \frac{\gamma_0(x)\|K\|_2^2}{[1-\rho_0(x)+2\alpha(1+\gamma_0(x))]^2}, \tag{26}$$

$$\sigma_{44}(\rho_0(x)) := \frac{2\gamma_0^2(x)\|K\|_2^2}{[1+2\alpha(1+\gamma_0(x))]^3}. \tag{27}$$

Now, apply a Taylor series expansion of the estimating equations $f_1(\hat{\gamma}_n(x), \hat{\delta}_n(x); \rho_0(x)) = 0$ and

$f_2(\hat{\gamma}_n(x), \hat{\delta}_n(x); \rho_0(x)) = 0$ around $(\gamma_0(x), 0)$. This gives

$$
\begin{aligned}
0 &= f_1(\gamma_0(x), 0; \rho_0(x)) + f_{11}(\gamma_0(x), 0; \rho_0(x))(\hat{\gamma}_n(x) - \gamma_0(x)) + f_{12}(\gamma_0(x), 0; \rho_0(x))\hat{\delta}_n(x) \\
&\quad + \frac{1}{2}\Big\{f_{111}(\breve{\gamma}_n(x), \breve{\delta}_n(x); \rho_0(x))(\hat{\gamma}_n(x) - \gamma_0(x))^2 + f_{122}(\breve{\gamma}_n(x), \breve{\delta}_n(x); \rho_0(x))\hat{\delta}_n^2(x) \\
&\quad + 2f_{112}(\breve{\gamma}_n(x), \breve{\delta}_n(x); \rho_0(x))(\hat{\gamma}_n(x) - \gamma_0(x))\hat{\delta}_n(x)\Big\}, \\
0 &= f_2(\gamma_0(x), 0; \rho_0(x)) + f_{21}(\gamma_0(x), 0; \rho_0(x))(\hat{\gamma}_n(x) - \gamma_0(x)) + f_{22}(\gamma_0(x), 0; \rho_0(x))\hat{\delta}_n(x) \\
&\quad + \frac{1}{2}\Big\{f_{211}(\breve{\gamma}_n(x), \breve{\delta}_n(x); \rho_0(x))(\hat{\gamma}_n(x) - \gamma_0(x))^2 + f_{222}(\breve{\gamma}_n(x), \breve{\delta}_n(x); \rho_0(x))\hat{\delta}_n^2(x) \\
&\quad + 2f_{122}(\breve{\gamma}_n(x), \breve{\delta}_n(x); \rho_0(x))(\hat{\gamma}_n(x) - \gamma_0(x))\hat{\delta}_n(x)\Big\},
\end{aligned}
$$

where $(\breve{\gamma}_n(x), \breve{\delta}_n(x))$ is a point on the line segment connecting $(\hat{\gamma}_n(x), \hat{\delta}_n(x))$ and $(\gamma_0(x), 0)$. A straightforward rearrangement gives a set of random equations where interest is in $r_n(\hat{\gamma}_n(x) - \gamma_0(x))$ and $r_n\hat{\delta}_n(x)$:

$$
-r_n \begin{bmatrix} f_1(\gamma_0(x), 0; \rho_0(x)) \\ f_2(\gamma_0(x), 0; \rho_0(x)) \end{bmatrix} = \begin{bmatrix} \tilde{f}_{11}(\gamma_0(x), 0; \rho_0(x)) & \tilde{f}_{12}(\gamma_0(x), 0; \rho_0(x)) \\ \tilde{f}_{12}(\gamma_0(x), 0; \rho_0(x)) & \tilde{f}_{22}(\gamma_0(x), 0; \rho_0(x)) \end{bmatrix} \begin{bmatrix} r_n(\hat{\gamma}_n(x) - \gamma_0(x)) \\ r_n\hat{\delta}_n(x) \end{bmatrix}, \quad (28)
$$

where

$$
\begin{aligned}
\tilde{f}_{11}(\gamma_0(x), 0; \rho_0(x)) &:= f_{11}(\gamma_0(x), 0; \rho_0(x)) + \frac{1}{2}\Big[f_{111}(\breve{\gamma}_n(x), \breve{\delta}_n(x); \rho_0(x))(\hat{\gamma}_n(x) - \gamma_0(x)) \\
&\quad + f_{112}(\breve{\gamma}_n(x), \breve{\delta}_n(x); \rho_0(x))\hat{\delta}_n(x)\Big], \\
\tilde{f}_{12}(\gamma_0(x), 0; \rho_0(x)) &:= f_{12}(\gamma_0(x), 0; \rho_0(x)) + \frac{1}{2}\Big[f_{122}(\breve{\gamma}_n(x), \breve{\delta}_n(x); \rho_0(x))\hat{\delta}_n(x) \\
&\quad + f_{112}(\breve{\gamma}_n(x), \breve{\delta}_n(x); \rho_0(x))(\hat{\gamma}_n(x) - \gamma_0(x))\Big], \\
\tilde{f}_{22}(\gamma_0(x), 0; \rho_0(x)) &:= f_{22}(\gamma_0(x), 0; \rho_0(x)) + \frac{1}{2}\Big[f_{222}(\breve{\gamma}_n(x), \breve{\delta}_n(x); \rho_0(x))\hat{\delta}_n(x) \\
&\quad + f_{122}(\breve{\gamma}_n(x), \breve{\delta}_n(x); \rho_0(x))(\hat{\gamma}_n(x) - \gamma_0(x))\Big].
\end{aligned}
$$

Now, introduce

$$
\mathbb{B}(\rho_0(x)) := \gamma_0^{-\alpha-2}(x) \begin{bmatrix} b_{11}(\rho_0(x)) & \gamma_0(x) & 0 & -1 \\ b_{21}(\rho_0(x)) & \gamma_0(x) & -\gamma_0(x)(1-\rho_0(x)) & 0 \end{bmatrix}, \quad (29)
$$

with

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

so that

$$
r_n \begin{bmatrix} f_1(\gamma_0(x), 0; \rho_0(x)) \\ f_2(\gamma_0(x), 0; \rho_0(x)) \end{bmatrix} = \mathbb{B}(\rho_0(x))\overline{\mathbb{A}}_n(\rho_0(x)),
$$

leading to the weak convergence

$$r_n \left[\begin{array}{c} f_1(\gamma_0(x), 0; \rho_0(x)) \\ f_2(\gamma_0(x), 0; \rho_0(x)) \end{array} \right] \rightsquigarrow N_2(\lambda\sqrt{b(x)}\mathbb{B}(\rho_0(x))\mathbb{D}, \mathbb{B}(\rho_0(x))\boldsymbol{\Sigma}(\rho_0(x))\mathbb{B}'(\rho_0(x))).$$

Concerning the terms $\tilde{f}_{st}(\gamma_0(x), 0; \rho_0(x))$, $s, t = 1, 2$, we have by Lemmas 1 and 2, Theorem 1, the consistency of $(\hat{\gamma}_n(x), \hat{\delta}_n(x))$ and because $|f_{stu}(\gamma, \delta; \rho_0(x))| \leq M_{stu}(\boldsymbol{V})$, in some open neighborhood of $(\gamma_0(x), 0)$, with $M_{stu}(\boldsymbol{V}) = O_{\mathbb{P}}(1)$, $s, t, u = 1, 2$, that $\tilde{f}_{st}(\gamma_0(x), 0; \rho_0(x)) \stackrel{\mathbb{P}}{\to} f^*_{st}(\gamma_0(x), 0; \rho_0(x))$, $s, t = 1, 2$. Let

$$\mathbb{C}(\rho_0(x)) := \left[\begin{array}{cc} f^*_{11}(\gamma_0(x), 0; \rho_0(x)) & f^*_{12}(\gamma_0(x), 0; \rho_0(x)) \\ f^*_{12}(\gamma_0(x), 0; \rho_0(x)) & f^*_{22}(\gamma_0(x), 0; \rho_0(x)) \end{array} \right]. \tag{30}$$

From the proof of the consistency, we know that $\mathbb{C}(\rho_0(x))$ is a positive definite matrix, and thus invertible. Then, according to Lemma 5.2 in Chapter 6 of Lehmann and Casella (1998), for the solution of the system of equations (28), we have the following convergence

$$\begin{aligned} r_n \left[\begin{array}{c} \hat{\gamma}_n(x) - \gamma_0(x) \\ \hat{\delta}_n(x) \end{array} \right] \quad \rightsquigarrow \quad & N_2(-\lambda\sqrt{b(x)}\mathbb{C}^{-1}(\rho_0(x))\mathbb{B}(\rho_0(x))\mathbb{D}, \\ & \mathbb{C}^{-1}(\rho_0(x))\mathbb{B}(\rho_0(x))\boldsymbol{\Sigma}(\rho_0(x))\mathbb{B}'(\rho_0(x))\mathbb{C}^{-1}(\rho_0(x))). \end{aligned}$$

After tedious calculations one can show that $-\mathbb{C}^{-1}(\rho_0(x))\mathbb{B}(\rho_0(x))\mathbb{D} = [0, 1]'$. Taking into account that $r_n\delta(u_n; x) \to \lambda\sqrt{b(x)}$, the theorem follows.

Proof of Proposition 1

The arguments needed to prove the consistency and asymptotic normality are the same as those used in the proofs of Theorem 2 and 3, and therefore we limit ourselves to giving some comments to the main ideas. Concerning the consistency one works with $\widetilde{\Delta}_\alpha(\gamma, \delta; \tilde{\rho}(x))$ and its derivatives. Again by Lemmas 1, 2 and Theorem 1 we have that $f_s(\gamma_0(x), 0; \tilde{\rho}(x)) \stackrel{\mathbb{P}}{\to} 0$, $s = 1, 2$, and that $f_{st}(\gamma_0(x), 0; \tilde{\rho}(x)) \stackrel{\mathbb{P}}{\to} f^*_{st}(\gamma_0(x), 0; \tilde{\rho}(x))$, $s, t = 1, 2$, leading to the results for S_1 and S_2. Also for the third order derivatives we can use the same arguments. This establishes the existence and the consistency. To prove the asymptotic normality one uses the same line of argumentation as in Theorem 3, with $\rho_0(x)$ replaced by $\tilde{\rho}(x)$ in $\overline{\mathbb{A}}_{k,n}(\rho_0(x))$, $\boldsymbol{\Sigma}(\rho_0(x))$, $\mathbb{B}(\rho_0(x))$ and $\mathbb{C}(\rho_0(x))$, and replacing the vector $\mathbb{D}$ by $\widetilde{\mathbb{D}}$, having as elements $\widetilde{D}_1 := D_1$, $\widetilde{D}_2 := D_2$, $\widetilde{D}_4 := D_4$ and

$$\widetilde{D}_3 := -\frac{[\alpha(1 + \gamma_0(x)) - \tilde{\rho}(x)]\rho_0(x)}{\gamma_0(x)[1 - \tilde{\rho}(x) + \alpha(1 + \gamma_0(x))][1 - \rho_0(x) - \tilde{\rho}(x) + \alpha(1 + \gamma_0(x))]}. \tag{31}$$

Proof of Theorem 4

The proof of Theorem 4 is similar to that of Theorems 2 and 3, and therefore we only give the big lines of argument.

Concerning the existence and consistency of $(\hat{\gamma}(x), \hat{\delta}_n(x))$ as estimators for $(\gamma_0(x), 0)$, we have that by the consistency of $\hat{\rho}_n(x)$ and conditioning on the event $\hat{\rho}_n(x) \in (\rho_0(x) - \varepsilon, \rho_0(x) + \varepsilon)$ for some $\varepsilon > 0$, it is sufficient to show that

$$\begin{aligned}&\mathbb{P}_{(\gamma_0(x),0)}(\widetilde{\Delta}_\alpha(\gamma_0(x), 0; \hat{\rho}_n(x)) < \widetilde{\Delta}_\alpha(\gamma, \delta; \hat{\rho}_n(x)) \\ &\qquad \text{for all } (\gamma, \delta) \text{ on the surface of } Q_r \mid \hat{\rho}_n(x) \in (\rho_0(x) - \varepsilon, \rho_0(x) + \varepsilon)) \to 1.\end{aligned}$$

First make a Taylor series expansion as in (13), though now with $\rho_0(x)$ replaced by $\hat{\rho}_n(x)$.

Assume that $(-(\alpha(1+\gamma_0(x)) - (\rho_0(x) - \varepsilon))/\gamma_0(x), -(\alpha(1+\gamma_0(x)) - (\rho_0(x) + \varepsilon))/\gamma_0(x)) \in [S, 0]$. Concerning S_1, we have that $f_1(\gamma_0(x), 0; \hat{\rho}_n(x))$ does not depend on $\hat{\rho}_n(x)$ and therefore obviously $f_1(\gamma_0, 0; \hat{\rho}_n(x)) \xrightarrow{\mathbb{P}} 0$, whereas for $f_2(\gamma_0, 0; \hat{\rho}_n(x))$ we write

$$\begin{aligned}
&f_2(\gamma_0(x), 0; \hat{\rho}_n(x)) \\
&= \gamma_0^{-\alpha-1}(x)\Bigg[-\frac{\alpha\hat{\rho}_n(x)(1+\gamma_0(x))}{[1+\alpha(1+\gamma_0(x))][1-\hat{\rho}_n(x)+\alpha(1+\gamma_0(x))]}\frac{T_n(K,0,0;x)}{\bar{F}(u_n;x)b(x)} \\
&\quad + \frac{T_n(K, -\alpha(1+\gamma_0(x))/\gamma_0(x), 0; x)}{\bar{F}(u_n;x)b(x)} \\
&\quad -(1-\hat{\rho}_n(x))\left(\frac{T_n(K, -(\alpha(1+\gamma_0(x)) - \rho_0(x))/\gamma_0(x), 0; x)}{\bar{F}(u_n;x)b(x)} - \frac{1}{1-\rho_0(x)+\alpha(1+\gamma_0(x))}\right) \\
&\quad -(1-\hat{\rho}_n(x))\left(\frac{T_n(K, -(\alpha(1+\gamma_0(x)) - \hat{\rho}_n(x))/\gamma_0(x), 0; x)}{\bar{F}(u_n;x)b(x)} - \frac{T_n(K, -(\alpha(1+\gamma_0(x)) - \rho_0(x))/\gamma_0(x), 0; x)}{\bar{F}(u_n;x)b(x)}\right) \\
&\quad -\frac{1-\hat{\rho}_n(x)}{1-\rho_0(x)+\alpha(1+\gamma_0(x))}\Bigg] \\
&=: \gamma_0^{\alpha-1}(x)\left[T_1+T_2+T_3+T_4+T_5\right].
\end{aligned}$$

Now use Lemmas 1, 2 and Theorem 1 to obtain

$$\begin{aligned}
T_1 &\xrightarrow{\mathbb{P}} -\frac{\alpha\rho_0(x)(1+\gamma_0(x))}{[1+\alpha(1+\gamma_0(x))][1-\rho_0(x)+\alpha(1+\gamma_0(x))]}, \\
T_2 &\xrightarrow{\mathbb{P}} \frac{1}{1+\alpha(1+\gamma_0(x))}, \\
T_3 &\xrightarrow{\mathbb{P}} 0, \\
|T_4| &\leq \frac{1-\hat{\rho}_n(x)}{\gamma_0(x)}\,|\hat{\rho}_n(x)-\rho_0(x)|\,\frac{T_n(K,0,1;x)}{\bar{F}(u_n;x)b(x)} = o_{\mathbb{P}}(1), \\
T_5 &\xrightarrow{\mathbb{P}} -\frac{1-\rho_0(x)}{1-\rho_0(x)+\alpha(1+\gamma_0(x))}.
\end{aligned}$$

Combining these results give that $f_2(\gamma_0, 0; \hat{\rho}_n(x)) \xrightarrow{\mathbb{P}} 0$.

For S_2, write

$$\begin{aligned}2S_2 &= f_{11}^*(\gamma_0(x),0;\rho_0(x))(\gamma-\gamma_0(x))^2 + f_{22}^*(\gamma_0(x),0;\rho_0(x))\delta^2 \\ &\quad +2f_{12}^*(\gamma_0(x),0;\rho_0(x))(\gamma-\gamma_0(x))\delta \\ &\quad +[f_{11}(\gamma_0(x),0;\hat\rho_n(x)) - f_{11}^*(\gamma_0(x),0;\rho_0(x))](\gamma-\gamma_0(x))^2 \\ &\quad +[f_{22}(\gamma_0(x),0;\hat\rho_n(x)) - f_{22}^*(\gamma_0(x),0;\rho_0(x))]\delta^2 \\ &\quad +2[f_{12}(\gamma_0(x),0;\hat\rho_n(x)) - f_{12}^*(\gamma_0(x),0;\rho_0(x))](\gamma-\gamma_0(x))\delta.\end{aligned}$$

By arguments similar to those used above when treating $f_2(\gamma_0(x),0;\hat\rho_n(x))$, we have that $f_{st}(\gamma_0(x),0;\hat\rho_n(x)) \xrightarrow{\mathbb{P}} f_{st}^*(\gamma_0(x),0;\rho_0(x))$, $s,t=1,2$, and hence we can proceed as in the proof of Theorem 2. Finally, conditionally on $\hat\rho_n(x)\in(\rho_0(x)-\varepsilon,\rho_0(x)+\varepsilon)$, also the argument for the third order derivatives holds and the proof for the existence and consistency can be completed in the same way as in the proof of Theorem 2.

The proof of asymptotic normality proceeds along the lines of Theorem 3. To start we make a Taylor series expansion of the estimating equations, leading to (28), though with $\rho_0(x)$ replaced by $\hat\rho_n(x)$. Since $\mathbb{P}(\hat\rho_n(x)\in(\rho_0(x)-\varepsilon,\rho_0(x)+\varepsilon))\to 1$, we have that (by an appropriate choice of S in Corollary 1)

$$\overline{\mathbb{A}}_{k,n}(\hat\rho_n(x)) \rightsquigarrow N_4(\lambda\sqrt{b(x)}\mathbb{D}, \mathbf{\Sigma}(\rho_0(x))),$$

and hence

$$\begin{aligned} r_n\left[\begin{array}{c} f_1(\gamma_0(x),0;\hat\rho_n(x)) \\ f_2(\gamma_0(x),0;\hat\rho_n(x))\end{array}\right] &= \mathbb{B}(\hat\rho_n(x))\overline{\mathbb{A}}_{k,n}(\hat\rho_n(x)) \\ &\rightsquigarrow N_2(\lambda\sqrt{b(x)}\mathbb{B}(\rho_0(x))\mathbb{D}, \mathbb{B}(\rho_0(x))\mathbf{\Sigma}(\rho_0(x))\mathbb{B}'(\rho_0(x))).\end{aligned}$$

The rest of the proof is identical to that of Theorem 3.

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.

[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.

[6] Daouia, A., Gardes, L., Girard, S., 2012. On kernel smoothing for extremal quantile regression. Bernoulli, to appear.

[7] Daouia, A., Gardes, L., Girard, S., Lekina, A., 2011. Kernel estimators of extreme level curves. Test, 20, 311–333.

[8] de Haan, L., Ferreira, A., 2006. *Extreme Value Theory: An Introduction.* Springer.

[9] Dierckx, G., Goegebeur, Y., Guillou, A., 2012. An asymptotically unbiased minimum density power divergence estimator for the Pareto-tail index. Submitted.

[10] Feuerverger, A., Hall, P., 1999. Estimating a tail exponent by modelling departure from a Pareto distribution. The Annals of Statistics, 27, 760–781.

[11] 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.

[12] Goegebeur, Y., Guillou, A., Schorgen, A., 2012. Nonparametric regression estimation of conditional tails - the random covariate case. Submitted.

[13] Hall, P., 1982. On some simple estimates of an exponent of regular variation. Journal of the Royal Statistical Society Series B, 44, 37–42.

[14] Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. Annals of Statistics, 3, 1163–1174.

[15] Kim, M., Lee, S., 2008. Estimation of a tail index based on minimum density power divergence. Journal of Multivariate Analysis, 99, 2453–2471.

[16] Lehmann, E.L., Casella, G., 1998. *Theory of point estimation.* Springer.

[17] Pickands, J., 1975. Statistical inference using extreme order statistics. Annals of Statistics, 3, 119–131.

[18] van der Vaart, A.W., 2007. *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistic Mathematics.

[19] Wang, H., Tsai, C.L., 2009. Tail index regression. Journal of the American Statistical Association, 104, 1233–1240.