



**HAL**  
open science

## Robust conditional Weibull-type estimation

Yuri Goegebeur, Armelle Guillou, Théo Rietsch

► **To cite this version:**

Yuri Goegebeur, Armelle Guillou, Théo Rietsch. Robust conditional Weibull-type estimation. *Annals of the Institute of Statistical Mathematics*, 2015, 67 (3), 10.1007/s10463-014-0458-9 . hal-01312924

**HAL Id: hal-01312924**

**<https://hal.science/hal-01312924>**

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.

# Robust conditional Weibull-type estimation

Yuri Goegebeur<sup>(1)</sup>, Armelle Guillou<sup>(2)</sup> and Théo Rietsch<sup>(2),(3)</sup>

<sup>(1)</sup> Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark, e-mail: yuri.goegebeur@imada.sdu.dk

<sup>(2)</sup> Institut Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg et CNRS, 7 rue René Descartes, 67084 Strasbourg cedex, France, e-mail: armelle.guillou@math.unistra.fr

<sup>(3)</sup> Laboratoire des Sciences du Climat et de l'Environnement, LSCE-IPSL-CNRS, Orme des Merisiers / Bat. 701 C.E. Saclay, 91191 Gif-sur-Yvette, France, e-mail: theo.rietsch@math.unistra.fr

**Abstract.** We study nonparametric robust tail coefficient estimation when the variable of interest, assumed to be of Weibull-type, is observed simultaneously with a random covariate. In particular, we introduce a robust estimator for the tail coefficient, using the idea of the density power divergence (see Basu *et al.*, 1998) based on the relative excesses above a high threshold. The main asymptotic properties of our estimator are established under very general assumptions. The finite sample performance of the proposed procedure is evaluated by a small simulation experiment.

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

**Keywords:** Weibull-type distribution, tail coefficient, density power divergence, local estimation.

## 1 Introduction

In practical data analysis, it is not unusual to encounter outliers which may have a disturbing effect on the estimation results. In such situations, the estimates of the model according to the maximum likelihood approach are typically fairly unstable and this asks for robust methods. A special treatment of the outlying points is then required, for instance by an adequate downweighting of their influence on the estimation.

In this paper, we focus on robust procedures in order to estimate some tail parameters in an extreme value context. Such a topic has been recently studied in the literature. We can mention among others Brazauskas and Serfling (2000) and Vandewalle *et al.* (2007) for strict Pareto and Pareto-type distributions, Dupuis and Field (1998), Peng and Welsh (2001), and Juárez and Schucany (2004) for generalized extreme value or generalized Pareto distributions. In the sequel, we consider the Gumbel class, which is a rich subclass of the max-domain of attraction. Although different types of tail behavior are possible, all these distributions have in common an extreme value index equal to zero and thus differentiating them on the basis of this parameter alone is impossible. To solve this issue, we restrict our study to Weibull-type distributions for which the distribution functions have the following form:

$$\bar{F}(y) := 1 - F(y) = e^{-y^{1/\theta} \ell_F(y)}, \quad y > 0,$$

where  $\theta > 0$  and  $\ell_F$  is a slowly varying function at infinity, i.e. an ultimately positive function satisfying

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

Here  $\theta$  denotes the Weibull-tail coefficient. Different values of it allow the Weibull-type distributions to cover a large part of the Gumbel class, and hence to constitute a flexible subgroup. The estimation of this coefficient has been extensively studied in the literature (see e.g. Broniatowski (1993), Beirlant *et al.* (1995), Gardes and Girard (2005, 2008b), Diebolt *et al.* (2008), Dierckx *et al.* (2009), Goegebeur *et al.* (2010) or Goegebeur and Guillou (2011) among others) but not much attention has been paid to the regression context with covariates.

We will consider this framework of nonparametric regression estimation of conditional tails when the covariates are random. The case of random covariates is less explored in extreme value theory compared to the fixed covariates, and only few papers can indeed be mentioned: Wang and Tsai (2009) with a parametric maximum likelihood approach within the Hall subclass of Pareto-type models (Hall, 1982), Daouia *et al.* (2011) in the framework of Pareto-type distributions, and Daouia *et al.* (2013) in the general max-domain of attraction, but under rather restrictive assumptions on the underlying distribution function. Here, we consider the case of Weibull-type distributions and our approach will be based on local estimation within a narrow window around the point in the covariate space where the tail behavior of the response variable is of interest. This local fitting is performed by an adjustment of the robust minimum density power divergence (MDPD) estimation criterion, originally proposed by Basu *et al.* (1998), to the locally weighted regression setting. This criterion has already been used for robust estimation of Pareto-type distributions, see for instance Kim and Lee (2008), Dierckx *et al.* (2013a, b), but to the best of our knowledge it is new in the Weibull-type framework.

The remainder of this paper is organized as follows. In Section 2, we introduce our robust estimator of the conditional Weibull-tail coefficient and we state its main asymptotic properties. The finite sample performance of our procedure is illustrated on a small simulation study in Section 3. In Section 4 we will make some concluding remarks and discuss some possibilities for future research. The proofs of all results can be found in the Appendix.

## 2 Construction and asymptotic properties

Let  $(X_i, Y_i)$ ,  $i = 1, \dots, n$ , be independent copies of a random pair  $(X, Y) \in \mathbb{R}^p \times \mathbb{R}_+$ , where  $X$  has the density function  $f$  and the conditional survival function of  $Y$  given  $X = x$  is of Weibull-type with a tail coefficient  $\theta(x) > 0$ , that is

$$\bar{F}(y; x) = e^{-y^{1/\theta(x)} \ell_F(y; x)}, \quad y > 0. \quad (1)$$

This model can also be defined in terms of the generalized inverse of  $F$ , denoted  $Q$ , i.e.  $Q(q; x) := \inf\{y : F(y; x) \geq q\}$ ,  $0 < q < 1$ . Indeed, under (1), we have

$$Q(q; x) = (-\ln(1 - q))^{\theta(x)} \ell(-\ln(1 - q); x) \quad (2)$$

where  $\ell$  is again a slowly varying function at infinity. The function  $\theta(x)$  governs the tail behavior, with larger values indicating a slower tail decay. This function has to be adequately estimated from the data.

As is usual in an extreme value context, we base our estimation method on the relative excesses above a high threshold  $u_n$ , namely  $Z := Y/u_n$ , which admit, under model (1), the following conditional survival function:

$$\mathbb{P}\left(\frac{Y}{u_n} > t \mid Y > u_n; x\right) = \frac{\bar{F}(tu_n; x)}{\bar{F}(u_n; x)} \simeq e^{-c_n(t^{1/\theta(x)} - 1)} =: \bar{G}(t; c_n, \theta(x)) \quad \text{for } t > 1, \quad (3)$$

where  $c_n := -\ln(\bar{F}(u_n; x))$ . The approximation in (3) follows from the properties of slowly varying functions, and is valid for large values of  $u_n$ . We denote by  $g$  the density function associated to this distribution  $G$ .

The proposed estimation procedure works as follows. First we estimate  $c_n$  externally in a consistent way, cf infra. Then, we estimate  $\theta(x)$  with the MDPD criterion combined with a kernel approach, and applied to the relative excesses above  $u_n$ . More precisely, we define the MDPD estimator as the value of  $\theta$  minimizing the empirical density power divergence:

$$\widehat{\Delta}_\alpha(\theta; \widehat{c}_n) := \begin{cases} \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \mathbb{1}_{\{Y_i > u_n\}} \left\{ \int_1^\infty g^{1+\alpha}(z) dz - \left(1 + \frac{1}{\alpha}\right) g^\alpha(Y_i/u_n) \right\} & \text{for } \alpha > 0 \\ -\frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \mathbb{1}_{\{Y_i > u_n\}} \ln g(Y_i/u_n) & \text{for } \alpha = 0 \end{cases}$$

where  $\widehat{c}_n$  is a consistent estimator for  $c_n$ ,  $K_h(x) := K(x/h)/h^p$ ,  $K$  is a joint density function on  $\mathbb{R}^p$ ,  $h = h_n$  is a positive, non-random sequence of bandwidths with  $h \rightarrow 0$  if  $n \rightarrow \infty$ ,  $\mathbb{1}_{\{A\}}$  is the indicator function on the event  $A$  and  $u_n$  is a local non-random threshold sequence satisfying  $u_n \rightarrow \infty$  if  $n \rightarrow \infty$ .

Note that a joint estimation of  $\theta(x)$  and  $c_n$  with the MDPD method is practically feasible, but gives difficulties in the theoretical analysis concerning consistency and asymptotic normality, and in particular it requires the introduction of rather restrictive conditions. For this reason we opt to estimate  $c_n$  externally in a consistent way. Remark also that this density power divergence criterion is indexed by a single non-negative parameter,  $\alpha$ , that controls the trade-off between robustness and efficiency. In particular it encompasses the maximum likelihood method, corresponding to  $\alpha = 0$ , which is efficient but not robust. Increasing the value of  $\alpha$  increases the robustness and decreases the efficiency of the estimation.

The MDPD equation for  $\theta$  is thus:

$$\begin{aligned} \widehat{\Delta}'_\alpha(\theta; \widehat{c}_n) &= \frac{1+\alpha}{n} \sum_{i=1}^n K_h(x - X_i) \mathbb{1}_{\{Y_i > u_n\}} \int_1^\infty g^\alpha(z) \frac{\partial g(z)}{\partial \theta} dz \\ &\quad - \frac{1+\alpha}{n} \sum_{i=1}^n K_h(x - X_i) \mathbb{1}_{\{Y_i > u_n\}} g^{\alpha-1}\left(\frac{Y_i}{u_n}\right) \frac{\partial g(Y_i/u_n)}{\partial \theta} \\ &= \frac{1+\alpha}{n} \sum_{i=1}^n K_h(x - X_i) \mathbb{1}_{\{Y_i > u_n\}} \left\{ -\frac{\alpha e^{\widehat{c}_n(1+\alpha)} \widehat{c}_n^{\alpha\theta}}{\theta^{\alpha+1} (1+\alpha)^{2+\alpha(1-\theta)}} \left[ \theta \Psi(\alpha(1-\theta) + 1, \widehat{c}_n(1+\alpha)) \right. \right. \\ &\quad \left. \left. + (1-\theta \ln((\alpha+1)\widehat{c}_n)) \Gamma(\alpha(1-\theta) + 1, \widehat{c}_n(1+\alpha)) \right] \right\} \\ &\quad + \frac{\widehat{c}_n^\alpha}{\theta^{\alpha+1}} \frac{1+\alpha}{n} \sum_{i=1}^n K_h(x - X_i) e^{-\widehat{c}_n \alpha \left[ \left(\frac{Y_i}{u_n}\right)^{1/\theta} - 1 \right]} \left(\frac{Y_i}{u_n}\right)^{\alpha(1/\theta-1)} \mathbb{1}_{\{Y_i > u_n\}} \\ &\quad + \frac{\widehat{c}_n^\alpha}{\theta^{\alpha+2}} \frac{1+\alpha}{n} \sum_{i=1}^n K_h(x - X_i) e^{-\widehat{c}_n \alpha \left[ \left(\frac{Y_i}{u_n}\right)^{1/\theta} - 1 \right]} \left(\frac{Y_i}{u_n}\right)^{\alpha(1/\theta-1)} \ln \frac{Y_i}{u_n} \mathbb{1}_{\{Y_i > u_n\}} \\ &\quad - \frac{\widehat{c}_n^{\alpha+1}}{\theta^{\alpha+2}} \frac{1+\alpha}{n} \sum_{i=1}^n K_h(x - X_i) e^{-\widehat{c}_n \alpha \left[ \left(\frac{Y_i}{u_n}\right)^{1/\theta} - 1 \right]} \left(\frac{Y_i}{u_n}\right)^{\alpha(1/\theta-1)+1/\theta} \ln \frac{Y_i}{u_n} \mathbb{1}_{\{Y_i > u_n\}} \end{aligned} \quad (4)$$

where  $\Gamma(a, b)$  denotes the incomplete Gamma function

$$\Gamma(a, b) := \int_b^\infty z^{a-1} e^{-z} dz$$

and  $\Psi(a, b)$  its derivative with respect to the first argument

$$\Psi(a, b) := \int_b^\infty \ln z z^{a-1} e^{-z} dz.$$

In the functions  $\Gamma$  and  $\Psi$ ,  $b$  is positive whereas  $a \in \mathbb{R}$ .

The MDPD estimator for  $\theta(x)$  satisfies the estimating equation

$$\widehat{\Delta}'_\alpha(\theta; \widehat{c}_n) = 0, \quad (5)$$

where  $\widehat{c}_n := -\ln(\widehat{F}(u_n; x))$ ,

$$\widehat{F}(u_n; x) := \frac{\frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \mathbb{1}_{\{Y_i > u_n\}}}{\frac{1}{n} \sum_{i=1}^n K_h(x - X_i)},$$

is a kernel estimator for  $\bar{F}(u_n; x)$ , as considered also in Daouia *et al.* (2013) and de Wet *et al.* (2013).

In view of (4) we start by considering the following locally weighted sums of power-transformed excesses over a high threshold:

$$T_n(K, \alpha, \beta, r; x) := \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) e^{-c_n \alpha \left[ \left( \frac{Y_i}{u_n} \right)^{1/\theta(x)} - 1 \right]} \left( \frac{Y_i}{u_n} \right)^\beta \left( \ln \frac{Y_i}{u_n} \right)_+^r \mathbb{1}_{\{Y_i > u_n\}} \quad (6)$$

where  $\alpha \geq 0$ ,  $\beta \in \mathbb{R}$ ,  $r \geq 0$ , and  $(x)_+ := \max\{0, x\}$ .

To obtain the limiting behavior of (6) one has to impose some more structure on the tail of the distribution. Typically one invokes a so-called second order condition, specifying the rate of convergence of  $\ell(\lambda y; x)/\ell(y; x)$ , where  $\ell$  is the slowly varying function appearing in (2), to its limit, being one, as  $y \rightarrow \infty$ .

**Assumption ( $\mathcal{R}$ )** *There exists a constant  $\rho(x) < 0$  and a rate function  $b(\cdot; x)$  satisfying  $b(y; x) \rightarrow 0$  as  $y \rightarrow \infty$ , such that for all  $\lambda \geq 1$ , we have*

$$\ln \left( \frac{\ell(\lambda y; x)}{\ell(y; x)} \right) = b(y; x) D_{\rho(x)}(\lambda) (1 + o(1))$$

with  $D_{\rho(x)}(\lambda) := \int_1^\lambda t^{\rho(x)-1} dt$ , and where  $o(1)$  is uniform in  $\lambda \geq 1$ , as  $y \rightarrow \infty$ .

As shown in Geluk and de Haan (1987), ( $\mathcal{R}$ ) implies that  $|b(y; x)|$  is regularly varying with index  $\rho(x)$ , i.e.  $|b(\lambda y; x)|/|b(y; x)| \rightarrow \lambda^{\rho(x)}$  as  $y \rightarrow \infty$  for all  $\lambda > 0$ , so  $\rho(x)$  governs the rate of the first order convergence of  $\ell(\lambda y; x)/\ell(y; x)$  to one. If  $|\rho(x)|$  is small then the convergence is slow and the estimation of tail quantities is generally difficult. Condition ( $\mathcal{R}$ ) is well accepted in the extreme value literature, see e.g. Gardes and Girard (2008a).

As a first step in the theoretical study of estimators for  $\theta(x)$ , we consider the local behavior of the following conditional expectation:

$$m(u_n, \alpha, \beta, r; x) = \mathbb{E} \left( e^{-c_n \alpha \left[ \left( \frac{Y}{u_n} \right)^{1/\theta(x)} - 1 \right]} \left( \frac{Y}{u_n} \right)^\beta \left( \ln \frac{Y}{u_n} \right)_+^r \mathbb{1}_{\{Y > u_n\}} \middle| X = x \right).$$

**Lemma 1** *Case (i),  $\alpha = \beta = r = 0$ :*

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

*Case (ii),  $(\alpha, \beta, r) \in \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+ \setminus (0, 0, 0)$ : assume (2) and Assumption (R). We have for  $u_n \rightarrow \infty$  that*

$$\begin{aligned} m(u_n, \alpha, \beta, r; x) &= \bar{F}(u_n; x) \frac{\Gamma(1+r)}{(1+\alpha)^{1+r}} \theta^r(x) \left\{ c_n^{-r} + \frac{\theta(x)\beta}{1+\alpha} c_n^{-1} \mathbb{1}_{\{r=0\}} + \frac{r-\alpha}{1+\alpha} \frac{b(c_n; x)}{\theta(x)} c_n^{-r} \right. \\ &\quad \left. + o(b(c_n; x) c_n^{-r}) + O\left(\frac{1}{c_n^{(1+\alpha)\wedge 2}} \mathbb{1}_{\{r=0\}}\right) + O\left(\frac{1}{c_n^{(1+r+\alpha-r\varepsilon)\wedge(1+r)}} \mathbb{1}_{\{r>0\}}\right) \right\}, \end{aligned}$$

for  $\varepsilon$  sufficiently small.

We now turn to the derivation of the asymptotic expansion for the unconditional moment. Let

$$m_n(K, \alpha, \beta, r; x) := \mathbb{E} \left[ K_h(x - X) e^{-c_n \alpha \left[ \left( \frac{Y}{u_n} \right)^{1/\theta(x)} - 1 \right]} \left( \frac{Y}{u_n} \right)^\beta \left( \ln \frac{Y}{u_n} \right)_+^r \mathbb{1}_{\{Y > u_n\}} \right].$$

Note that since  $T_n(K, \alpha, \beta, r; x)$  is an average of independent and identically distributed (i.i.d.) terms we also have that  $m_n(K, \alpha, \beta, r; x) = \mathbb{E}(T_n(K, \alpha, \beta, r; x))$ .

We introduce the following further conditions. Let  $\|\cdot\|$  be a norm on  $\mathbb{R}^p$ .

**Assumption (F)** There exists  $M_f > 0$  and  $\eta_f > 0$  such that  $|f(x) - f(x')| \leq M_f \|x - x'\|^{\eta_f}$  for all  $x, x' \in \mathbb{R}^p$ .

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

Finally we introduce a condition on the oscillation of the response distribution in a neighborhood of the point  $x$  where the estimation will take place. This condition is formulated in terms of the conditional excess function:

**Assumption (M)** The conditional excess function  $m(u_n, \alpha, \beta, r; x)$  satisfies for  $u_n \rightarrow \infty$ ,  $h \rightarrow 0$ , and some  $\bar{\alpha} > 0$ ,  $R > 0$ ,  $\xi > 0$  that

$$\Phi_n(x) := \sup_{\alpha \in [0, \bar{\alpha}]} \sup_{\beta \in [\alpha/\theta(x) - \xi, \alpha/\theta(x) + \xi]} \sup_{r \in [0, R]} \sup_{z \in \Omega} \left| \frac{m(u_n, \alpha, \beta, r; x - hz)}{m(u_n, \alpha, \beta, r; x)} - 1 \right| \rightarrow 0 \text{ as } n \rightarrow \infty.$$

The following lemma gives then the asymptotic expansion of  $m_n(K, \alpha, \beta, r; x)$ .

**Lemma 2** *Assume (2), (R), (F), (K) and (M). For all  $x \in \mathbb{R}^p$  where  $f(x) > 0$ , we have that if  $u_n \rightarrow \infty$  and  $h \rightarrow 0$  as  $n \rightarrow \infty$  then*

$$m_n(K, \alpha, \beta, r; x) = m(u_n, \alpha, \beta, r; x) f(x) (1 + O(h^{\eta_f}) + O(\Phi_n(x))).$$

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 (2) and  $X \sim f$ . If in addition to the previous assumptions one also has that  $nh^p \bar{F}(u_n; x) \rightarrow \infty$  as  $n \rightarrow \infty$  then

$$\tilde{T}_n(K, \alpha, \beta, r; x) := \frac{c_n^r T_n(K, \alpha, \beta, r; x)}{\bar{F}(u_n; x) f(x)} \xrightarrow{\mathbb{P}} \frac{\theta^r(x) \Gamma(1+r)}{(1+\alpha)^{1+r}} \text{ as } n \rightarrow \infty.$$

The consistency of  $\widehat{F}(u_n; x)$  follows now easily from Lemma 2.

**Corollary 1** *Let  $(X_1, Y_1), \dots, (X_n, Y_n)$  be i.i.d. random vectors and assume  $(\mathcal{F})$ ,  $(\mathcal{K})$  and  $(\mathcal{M})$ . For all  $x \in \mathbb{R}^p$  where  $f(x) > 0$  we have that if  $h \rightarrow 0$ ,  $u_n \rightarrow \infty$  with  $nh^p \overline{F}(u_n; x) \rightarrow \infty$  as  $n \rightarrow \infty$ , then*

$$\frac{\widehat{F}(u_n; x)}{\overline{F}(u_n; x)} \xrightarrow{\mathbb{P}} 1.$$

Note that this result holds for a general conditional survival function  $\overline{F}(y; x)$ , where  $X \sim f$ , i.e. the assumption of conditional Weibull-type behavior, and hence  $(\mathcal{R})$ , is not needed. In our context we have then  $\widehat{c}_n - c_n = -\ln \widehat{F}(u_n; x) / \overline{F}(u_n; x) \xrightarrow{\mathbb{P}} 0$ , by a straightforward application of the continuous mapping theorem.

The following theorem states the existence and consistency of sequences of solutions to the estimating equation (5). From now on we denote the true value of  $\theta(x)$  by  $\theta_0(x)$ .

**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 (2),  $X \sim f$ , and assume  $(\mathcal{R})$ ,  $(\mathcal{F})$ ,  $(\mathcal{K})$  and  $(\mathcal{M})$  hold. For all  $x \in \mathbb{R}^p$  where  $f(x) > 0$ , we have that if  $h_n \rightarrow 0$ ,  $u_n \rightarrow \infty$  with  $nh_n^p \overline{F}(u_n; x) \rightarrow \infty$ , then with probability tending to 1 there exists sequences of solutions  $(\widehat{\theta}_n(x))_{n \in \mathbb{N}}$  of the estimating equation (5), such that  $\widehat{\theta}_n(x) \xrightarrow{\mathbb{P}} \theta_0(x)$ , as  $n \rightarrow \infty$ .*

We now derive the limiting distribution of a vector of statistics of the form (6), when properly normalized. This result will form the basis for proving the asymptotic normality of the MDPD estimator. Let

$$\mathbb{T}'_n := (\widetilde{T}_n(K_1, \alpha_1, \beta_1, r_1; x), \dots, \widetilde{T}_n(K_J, \alpha_J, \beta_J, r_J; x))$$

for some positive integer  $J$  and let  $\Sigma$  be a  $(J \times J)$  covariance matrix with elements

$$\sigma_{j,k} := \frac{\theta_0^{r_j+r_k}(x) \|K_j K_k\|_1 \Gamma(1+r_j+r_k)}{(1+\alpha_j+\alpha_k)^{1+r_j+r_k}}.$$

**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 (2) and  $X \sim f$ , and assume  $(\mathcal{R})$ ,  $(\mathcal{F})$ ,  $(\mathcal{M})$  hold and kernel functions  $K_1, \dots, K_J$  satisfying  $(\mathcal{K})$ . For all  $x \in \mathbb{R}^p$  where  $f(x) > 0$ , we have that if  $h \rightarrow 0$ ,  $u_n \rightarrow \infty$  for  $n \rightarrow \infty$ , with  $nh^p \overline{F}(u_n; x) \rightarrow \infty$ , then*

$$\sqrt{nh^p \overline{F}(u_n; x)} [\mathbb{T}_n - \mathbb{E}(\mathbb{T}_n)] \rightsquigarrow N_J \left( 0, \frac{1}{f(x)} \Sigma \right).$$

With the result of Theorem 2, we can now establish the asymptotic normality of  $\widehat{\theta}_n(x)$ , when properly 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 (2),  $X \sim f$ , and assume  $(\mathcal{R})$ ,  $(\mathcal{F})$ ,  $(\mathcal{K})$  and  $(\mathcal{M})$  hold. Consider  $(\widehat{\theta}_n(x))_{n \in \mathbb{N}}$ , a consistent sequence of estimators for  $\theta_0(x)$  satisfying (5). For all  $x \in \mathbb{R}^p$  where  $f(x) > 0$ , we have that if  $h \rightarrow 0$  and  $u_n \rightarrow \infty$  as  $n \rightarrow \infty$  with  $nh^p \overline{F}(u_n; x) \rightarrow \infty$ ,  $\sqrt{nh^p \overline{F}(u_n; x)} b(c_n; x) \rightarrow \lambda \in \mathbb{R}$ ,  $\sqrt{nh^p \overline{F}(u_n; x)} h_n^{\eta_f} \rightarrow 0$ ,  $\sqrt{nh^p \overline{F}(u_n; x)} \Phi_n(x) \rightarrow 0$ , and  $\sqrt{nh^p \overline{F}(u_n; x)} / c_n^{(1+\alpha-\varepsilon) \wedge 1} \rightarrow 0$  (for some small  $\varepsilon > 0$ ) then*

$$\begin{aligned} & \sqrt{nh^p \overline{F}(u_n; x)} (\widehat{\theta}_n(x) - \theta_0(x)) \\ & \rightsquigarrow N \left( \lambda, \frac{\theta_0^2(x) (1+\alpha)^2 \|K\|_2^2}{(1+\alpha^2)^2 (1+2\alpha)^3} (1+4\alpha+9\alpha^2+14\alpha^3+13\alpha^4+8\alpha^5+4\alpha^6) \right). \end{aligned}$$

Note that the mean of the limiting distribution in Theorem 3 depends only on  $\lambda$  and not on  $\alpha$  nor on the weight function  $K$  or parameters related to the distribution of  $Y$  given  $X = x$ . This is in line with the usual asymptotic normality result in the univariate case, see e.g. Girard (2004), Gardes and Girard (2008b) and Goegebeur *et al.* (2010). The asymptotic variance in Theorem 3 is increasing in  $\alpha$ , which reflects the decreasing efficiency of the MDPD estimation method when  $\alpha$  increases. The maximum likelihood estimator, corresponding to  $\alpha = 0$ , has an asymptotic variance equal to  $\theta_0^2(x)\|K\|_2^2$ , which, apart from the factor  $\|K\|_2^2$ , coincides with the asymptotic variance of the Hill type estimator proposed by Girard (2004) in the univariate context. In Figure 1 we show the asymptotic standard deviation as a function of  $\alpha$  when  $\theta_0(x) = 1$  and  $K(u) = 0.5 \mathbb{1}_{u \in [-1,1]}$ .

### 3 Simulation results

The aim of this section is to illustrate the efficiency of our robust estimation method on a small simulation study. As is clear from the above discussion, the computation of the estimator requires a selection for the bandwidth parameter  $h$  and the threshold  $u_n$ . We select the threshold as usual in extreme value theory, that is, we take the  $(k + 1)$  largest response observation in the ball  $B(x, h)$  for any fixed value of  $x$ . We propose a data driven method to determine  $(h, k)$  and we compare it with a theoretical one, called Oracle strategy, which requires the knowledge of the function  $\theta(x)$ . These two methods are similar to those used in Goegebeur *et al.* (2013a) and are briefly recalled as follows:

**The Oracle strategy**, proposed in Gardes *et al.* (2010) consists in selecting  $(h, k)$  satisfying

$$(h_0, k_0) := \underset{h, k}{\operatorname{argmin}} \Delta \left( \widehat{\theta}(\cdot), \theta(\cdot) \right), \quad (7)$$

where

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

and  $z_1, \dots, z_M$  are points regularly spread in the covariate space. The retained value  $\widehat{\theta}(x)$  is the one corresponding to this pair  $(h_0, k_0)$ .

**The data driven method** does not require any prior knowledge about  $\theta(x)$  and thus can be directly applied to a real data set. First, the method involves the selection of an optimal bandwidth  $h$ . To this aim, we can use the following cross validation criterion introduced by Yao (1999), implemented by Gannoun *et al.* (2002), and studied by Daouia *et al.* (2011) and Daouia *et al.* (2013) in an extreme value context:

$$h_{cv} := \underset{h \in \mathcal{H}}{\operatorname{argmin}} \sum_{i=1}^n \sum_{j=1}^n \left( \mathbb{1}_{\{Y_i \leq Y_j\}} - \widetilde{F}_{n,-i}(Y_j | X_i) \right)^2, \quad (8)$$

where  $\mathcal{H}$  is a grid of values for  $h$  and

$$\widetilde{F}_{n,-i}(y|x) := \frac{\sum_{j=1, j \neq i}^n K_h(x - X_j) \mathbb{1}_{\{Y_j \leq y\}}}{\sum_{j=1, j \neq i}^n K_h(x - X_j)}.$$

Once the bandwidth  $h$  has been chosen, we compute  $\widehat{\theta}(x)$  for each  $k = 5, \dots, k_{max}$  where  $k_{max}$  is an appropriate maximum value of the number of exceedances. The retained value  $\widehat{\theta}$  is the median of these estimates of  $\theta$ .



In the simulations, we use the following smooth and symmetric function

$$\theta(x) = \frac{1}{2} \left( \frac{1}{10} + \sin(\pi x) \right) \left( \frac{11}{10} - \frac{1}{2} \exp \left\{ -64 \left( x - \frac{1}{2} \right)^2 \right\} \right)$$

proposed by Daouia *et al.* (2011), though originally in the context of Pareto-type tails, and the estimators are based on the bi-quadratic kernel function

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

The function  $\theta(x)$  is differentiable and has several stationary points. As such, it is more challenging than a monotone function. We assume that  $X$  is uniformly distributed on  $[0, 1]$  and we consider four different settings for the conditional distribution of  $Y$  given  $X = x$ :

- The strict Weibull distribution  $\mathcal{W}(\xi(x), \lambda)$ ,

$$1 - F(y; x) = e^{-\lambda y^{\xi(x)}}, \quad y > 0; \xi(x), \lambda > 0,$$

for which  $\theta(x) = 1/\xi(x)$  and  $\rho(x) = -\infty$ . We consider the case  $\lambda = 1$ .

- The extended Weibull distribution  $\mathcal{EW}(\xi(x), \lambda)$  (Kluppelberg and Villaseñor, 1993),

$$1 - F(y; x) = r(y) e^{-y^{\xi(x)}},$$

where  $\xi(x) > 0$  and  $r(y)$  is a regularly varying function at infinity with index  $\lambda$ . Here  $\theta(x) = 1/\xi(x)$  and  $\rho(x) = -1$ . We choose  $r(y) = 1/y$ , so  $\lambda = -1$ .

- The perturbed Weibull distribution  $\widetilde{\mathcal{W}}(\xi(x), \lambda)$  (Dierckx *et al.*, 2009),

$$1 - F(y; x) = e^{-y^{\xi(x)}(C + Dy^\lambda)}, \quad \xi(x) > 0, \lambda < 0, C > 0, D \in \mathbb{R},$$

having  $\theta(x) = 1/\xi(x)$  and  $\rho(x) = \lambda\theta(x)$ . In this case we use  $\lambda = -5$ ,  $C = 1$  and  $D = -1$ .

- A contaminated distribution with  $F_\varepsilon(y; x) = (1 - \varepsilon)F(y; x) + \varepsilon\check{F}(y; x)$  where the distribution function  $F$  is one of the three above mentioned distributions and for the contaminating distribution function  $\check{F}$  we consider

- a shifted strict Weibull distribution, i.e. it has distribution function  $\check{F}(y; x) = 1 - e^{-(y^\beta - y_c^\beta)}$ ,  $y > y_c$ . We choose  $\beta = 4/3$ , two different values for  $\varepsilon$ , namely 0.005 and 0.01, and  $y_c = 1.2$  times the 95% quantile of the uncontaminated distribution  $F$ ;
- a shifted strict Pareto distribution, with distribution function  $\check{F}(y; x) = 1 - (y/y_c)^{-1/\gamma}$ ,  $y > y_c$ . We choose  $\gamma = 1/4$ , the same fractions of contamination  $\varepsilon$  and two values for the shift parameter  $y_c$ :  $y_c = 1.2$  and 1.35 times the 95% quantile of the uncontaminated distribution  $F$ .

Note that the contamination from a shifted strict Pareto distribution is more severe than that from a shifted strict Weibull. Indeed, the strict Pareto distribution is in the max-domain of attraction of the Fréchet distribution, for which  $\gamma > 0$ , and consisting of heavy-tailed distributions, whereas the strict Weibull distribution is in fact in the same class as the main distribution function  $F$ , and thus having  $\gamma = 0$ , though shifted. In the simulation experiment we also considered other values for  $\beta$ ,  $\gamma$  and  $y_c$ , but those lead to similar results and therefore we do not include these here for brevity.

Each time  $N = 100$  samples of size  $n = 1000$  are generated. Both methods are implemented on  $M = 37$  values of  $x$ , equally spaced in  $[0, 1]$ , namely  $\{0.05, 0.075, \dots, 0.925, 0.95\}$ . In all the settings, the minimization (7) is performed on a grid  $\mathcal{H} = \{0.05, 0.075, \dots, 0.15\}$  and  $k_{max} = 25$ .

As indicators of efficiency we compute the bias, together with the mean squared error and the standard error

$$\text{Bias}(\hat{\theta}(\cdot)) := \frac{1}{MN} \sum_{m=1}^M \sum_{i=1}^N \left| \hat{\theta}^{(i)}(z_m) - \theta(z_m) \right|, \quad \text{MSE}(\hat{\theta}(\cdot)) := \frac{1}{MN} \sum_{m=1}^M \sum_{i=1}^N \left[ \hat{\theta}^{(i)}(z_m) - \theta(z_m) \right]^2,$$

$$\text{Sd}(\hat{\theta}(\cdot)) := \sqrt{\frac{1}{M} \sum_{m=1}^M \frac{1}{N-1} \sum_{i=1}^N \left( \hat{\theta}^{(i)}(z_m) - \bar{\hat{\theta}}(z_m) \right)^2},$$

where  $\bar{\hat{\theta}}(z_m) = \frac{1}{N} \sum_{i=1}^N \hat{\theta}^{(i)}(z_m)$  and  $\hat{\theta}^{(i)}(z_m)$  is the estimate of  $\theta(z_m)$  obtained with the  $i^{\text{th}}$  sample and our estimator  $\hat{\theta}(x)$  is evaluated at points  $z_1, \dots, z_M$  regularly spaced in  $[0, 1]$ .

The boxplots based on the 100 simulations in the uncontaminated case are given in Figure 2 for the Oracle approach and Figure 3 for the data driven method. Overall, both methods perform quite well, with, as expected, a better performance with the Oracle approach, which is also confirmed by the bias, the standard deviation and the MSE given in Table 1. In addition, we can observe that all these indicators slightly increase with  $\alpha$  whatever the method used. The increase in standard deviation when  $\alpha$  increases is in line with the decreasing efficiency of the MDPD method for increasing  $\alpha$ , as mentioned before.

Now we consider setting 4 where the distribution is contaminated. Figures 4 and 5 summarize the results for  $\varepsilon = 0.005$  with the Oracle and data driven approach, respectively, when the contamination is generated from the shifted strict Weibull distribution. The corresponding results for the strict Pareto contamination are shown in Figures 6 and 7 when  $y_c = 1.2$  times the 95% quantile of the uncontaminated distribution and in Figures 8 and 9 when 1.2 is replaced by 1.35. Again the Oracle strategy outperforms the data driven method, but the latter is still acceptable and fits quite well the curve of the function  $\theta$ . Note also that if we increase the value of  $y_c$  the fit deteriorates since the contamination becomes heavier. Similar comments can be made in terms of bias, standard deviation and MSE, reported in Tables 2 and 3, for  $\varepsilon = 0.005$  and  $\varepsilon = 0.01$ , respectively, in case of strict Weibull contamination, and in Tables 4 and 5 in case of strict Pareto contamination with  $y_c = 1.2$  times the 95% quantile of the uncontaminated distribution and Tables 6 and 7 when 1.2 is replaced by 1.35. Indeed, based on these tables we can draw the following conclusions:

- The Oracle strategy clearly outperforms the data driven method, and, as expected, the estimates obtained with both methods deteriorate in terms of bias, MSE and standard deviation when the proportion of contamination  $\varepsilon$  increases.
- Unlike the uncontaminated cases where the smallest value of  $\alpha$ , here 0.1, gave the best results, we see now that a larger value of  $\alpha$  is needed in order to deal with the contamination. For the Oracle method the results in the tables suggest to take  $\alpha = 0.25$  or  $\alpha = 0.5$ , while for the data driven method a larger value of  $\alpha$  seems to be needed, say  $\alpha = 1$  or  $\alpha = 2$ .
- The results for the contamination from the shifted strict Weibull and strict Pareto distributions are similar, with bias, MSE and standard deviation being of the same order of magnitude in case where  $y_c = 1.2$  times the 95% quantile of the uncontaminated distribution. This can possibly be explained by the fact that if we compare the mean of the first two moments of the contaminated

samples in both cases, they are quite similar, despite the heavier tail in case of contamination by a strict Pareto distribution. However, if we increase the contamination by increasing the value of  $y_c$ , the results deteriorate a bit while still remaining acceptable.

Summarized, we can state that the proposed MDPD estimator for the tail coefficient of Weibull-type distributions is a promising new estimator. Indeed, on uncontaminated data the estimation results do not indicate the typical bias-issues that occur in practice when estimating Weibull-type tails, as experienced in, among others, Gardes and Girard (2008b) and Goegebeur *et al.* (2010) in the univariate context. Also in case of contamination the method continues to work well, despite the fact that the contamination was quite severe in terms of shift and tail heaviness of the contaminating distribution.

## 4 Conclusion and future work

In this paper we introduced a nonparametric robust estimator for the conditional Weibull-tail coefficient, which is obtained by the local fitting of an approximation to the distribution of the relative excesses over a high threshold by the MDPD method. The method showed a good performance on simulated data, both in situations with and without outliers, and does not seem to suffer from the typical bias issues that are encountered when fitting Weibull-type tails. Our paper is a first contribution to robust estimation of conditional Weibull-type tails with random covariates, and therefore we focused initially on the study of local asymptotic properties, which is already challenging, but it can also be of interest to study uniform or stochastic process properties of the proposed method. So far, uniform or stochastic process properties of estimators of the conditional tail-index with random covariates are seldom considered in the extreme value literature. In the framework of local estimation of conditional Pareto-type tails with random covariates - which is generally much better explored than conditional Weibull-type tails - there are to the best of our knowledge only two contributions on uniform convergence available: Gardes and Stupfler (2013), who study the uniform weak consistency of a smoothed Hill estimator, and Goegebeur *et al.* (2013b) where the uniform almost sure convergence of a local Hill-type estimator is established. These papers are highly technical in nature even though they consider in fact only estimators which are in structure quite simple - much simpler than e.g. the MDPD estimator considered in the present paper which is not explicitly available. At this stage it remains uncertain whether in the setting of conditional Pareto-type tails with random covariates one can obtain a weak limit for the properly normalised conditional tail index estimator considered as a stochastic process. In future work we intend to establish the uniform consistency of nonparametric estimators for conditional Weibull-type tails, as well as to examine their properties when considered as stochastic processes. In first instance we will though focus on estimators that are explicitly available and thus have a simpler structure than the MDPD estimator considered in the present paper.

## Appendix

### Proof of Lemma 1

The case  $\alpha = \beta = r = 0$  is trivial, so we only consider case (ii). Let  $p_n := F(u_n; x)$ . Remark that

$$\begin{aligned}
m(u_n, \alpha, \beta, r; x) &= \mathbb{E} \left( e^{-c_n \alpha \left[ \left( \frac{Q(U; x)}{Q(p_n; x)} \right)^{1/\theta(x)} - 1 \right]} \left( \frac{Q(U; x)}{Q(p_n; x)} \right)^\beta \left( \ln \frac{Q(U; x)}{Q(p_n; x)} \right)_+^r \mathbb{1}_{\{Q(U; x) > Q(p_n; x)\}} \right) \\
&= \int_{p_n}^{\tilde{p}_n} e^{-c_n \alpha \left[ \left( \frac{Q(u; x)}{Q(p_n; x)} \right)^{1/\theta(x)} - 1 \right]} \left( \frac{Q(u; x)}{Q(p_n; x)} \right)^\beta \left( \ln \frac{Q(u; x)}{Q(p_n; x)} \right)^r du \\
&\quad + \int_{\tilde{p}_n}^1 e^{-c_n \alpha \left[ \left( \frac{Q(u; x)}{Q(p_n; x)} \right)^{1/\theta(x)} - 1 \right]} \left( \frac{Q(u; x)}{Q(p_n; x)} \right)^\beta \left( \ln \frac{Q(u; x)}{Q(p_n; x)} \right)^r du \\
&=: m^{(1)}(u_n, \alpha, \beta, r; x) + m^{(2)}(u_n, \alpha, \beta, r; x),
\end{aligned}$$

where  $U$  is a uniform  $[0, 1]$  random variable and  $\tilde{p}_n := 1 - \frac{1-p_n}{\ln \frac{1}{1-p_n}}$ .

We will study the two terms separately. First remark that

$$\frac{Q(u; x)}{Q(p_n; x)} = \left( 1 + \frac{-\ln \frac{1-u}{1-p_n}}{-\ln(1-p_n)} \right)^{\theta(x)} \frac{\ell \left( \left( 1 + \frac{-\ln \frac{1-u}{1-p_n}}{-\ln(1-p_n)} \right) (-\ln(1-p_n)); x \right)}{\ell(-\ln(1-p_n); x)}. \quad (9)$$

Thus by the change of variable  $z = \frac{1-u}{1-p_n}$ , Assumption  $(\mathcal{R})$  and the bound  $\frac{\rho(x)-1}{2} z^2 \leq D_{\rho(x)}(1+z) - z \leq 0$ , for  $z \geq 0$ , we deduce that

$$\begin{aligned}
m^{(1)}(u_n, \alpha, \beta, r; x) &= (1-p_n) \int_{\frac{1-\tilde{p}_n}{1-p_n}}^1 e^{-c_n \alpha \left[ \left( 1 + \frac{-\ln z}{c_n} \right) (1 + b(c_n; x) D_{\rho(x)} \left( 1 + \frac{-\ln z}{c_n} \right) (1 + o(1))) \right]^{1/\theta(x)} - 1} \\
&\quad \times \left( 1 + \frac{-\ln z}{c_n} \right)^{\theta(x)\beta} \left( 1 + b(c_n; x) D_{\rho(x)} \left( 1 + \frac{-\ln z}{c_n} \right) (1 + o(1)) \right)^\beta \\
&\quad \times \left( \ln \left[ \left( 1 + \frac{-\ln z}{c_n} \right)^{\theta(x)} \left( 1 + b(c_n; x) D_{\rho(x)} \left( 1 + \frac{-\ln z}{c_n} \right) (1 + o(1)) \right) \right] \right)^r dz \\
&= (1-p_n) \int_{\frac{1-\tilde{p}_n}{1-p_n}}^1 z^\alpha \left[ \theta^r(x) \left( \frac{-\ln z}{c_n} \right)^r + \theta^r(x) \left( \theta(x)\beta - \frac{r}{2} \right) \left( \frac{-\ln z}{c_n} \right)^{r+1} \right. \\
&\quad \left. + r\theta^{r-1}(x) \left( \frac{-\ln z}{c_n} \right)^r b(c_n; x)(1 + o(1)) - \alpha\theta^{r-1}(x) \frac{(-\ln z)^{r+1}}{c_n^r} b(c_n; x)(1 + o(1)) \right. \\
&\quad \left. + O \left( \left( \frac{-\ln z}{c_n} \right)^{r+2} \right) \right] dz.
\end{aligned}$$

Now remark that

$$\int_{\frac{1-\tilde{p}_n}{1-p_n}}^1 z^\alpha (-\ln z)^r dz = \frac{1}{(1+\alpha)^{r+1}} \{ \Gamma(r+1) - \Gamma(r+1, (1+\alpha) \ln(1+c_n)) \}.$$

Thus

$$\begin{aligned}
m^{(1)}(u_n, \alpha, \beta, r; x) &= (1 - p_n) \frac{\Gamma(1+r)}{(1+\alpha)^{1+r}} \theta^r(x) \left\{ c_n^{-r} + \frac{\theta(x)\beta}{1+\alpha} c_n^{-1} \mathbb{1}_{\{r=0\}} + \frac{r-\alpha}{1+\alpha} \frac{b(c_n; x)}{\theta(x)} c_n^{-r} \right. \\
&\quad - c_n^{-1-\alpha} \mathbb{1}_{\{r=0\}} + o(b(c_n; x)c_n^{-r}) + O\left(\frac{\ln c_n}{c_n^{2+\alpha}} \mathbb{1}_{\{r=0\}}\right) + O\left(\frac{1}{c_n^2} \mathbb{1}_{\{r=0\}}\right) \\
&\quad \left. + O\left(\frac{(\ln c_n)^r}{c_n^{1+r+\alpha}} \mathbb{1}_{\{r>0\}}\right) + O\left(\frac{1}{c_n^{1+r}} \mathbb{1}_{\{r>0\}}\right) \right\}.
\end{aligned}$$

Now, concerning the  $m^{(2)}(u_n, \alpha, \beta, r; x)$  term, using the monotonicity of  $Q$  and of the exponential function leads to the inequality

$$\begin{aligned}
m^{(2)}(u_n, \alpha, \beta, r; x) &\leq e^{-c_n \alpha \left[ \left( \frac{Q(\tilde{p}_n; x)}{Q(p_n; x)} \right)^{1/\theta(x)} - 1 \right]} \int_{\tilde{p}_n}^1 \left( \frac{Q(u; x)}{Q(p_n; x)} \right)^\beta \left( \ln \frac{Q(u; x)}{Q(p_n; x)} \right)^r du \\
&=: T_1 \times T_2.
\end{aligned}$$

Clearly, using (9), Assumption  $(\mathcal{R})$  and the bound for  $D_{\rho(x)}(1 + \cdot)$ , we have

$$\left( \frac{Q(\tilde{p}_n; x)}{Q(p_n; x)} \right)^{1/\theta(x)} = 1 + \frac{b(c_n; x)}{\theta(x)} \frac{\ln(1+c_n)}{c_n} (1 + o(1)) + \frac{\ln(1+c_n)}{c_n}.$$

This implies that

$$T_1 = e^{-\alpha \ln(1+c_n)} e^{-\frac{\alpha}{\theta(x)} b(c_n; x) \ln(1+c_n) (1+o(1))} = c_n^{-\alpha} (1 + o(1))$$

since  $\rho(x) < 0$ .

Now, concerning the term  $T_2$ , using the tail quantile function  $U(y; x) := Q\left(1 - \frac{1}{y}; x\right)$ ,  $y > 1$ , combined with the change of variables  $z = \frac{1-p_n}{1-u}$ , we deduce that

$$\begin{aligned}
T_2 &= (1 - p_n) \left( \frac{a\left(\frac{1}{1-p_n}; x\right)}{U\left(\frac{1}{1-p_n}; x\right)} \right)^r \int_{1+c_n}^\infty \left[ 1 + \frac{a\left(\frac{1}{1-p_n}; x\right)}{U\left(\frac{1}{1-p_n}; x\right)} \frac{U\left(\frac{z}{1-p_n}; x\right) - U\left(\frac{1}{1-p_n}; x\right)}{a\left(\frac{1}{1-p_n}; x\right)} \right]^\beta \frac{1}{z^2} \\
&\quad \times \left( \frac{\ln U\left(\frac{z}{1-p_n}; x\right) - \ln U\left(\frac{1}{1-p_n}; x\right)}{a\left(\frac{1}{1-p_n}; x\right) / U\left(\frac{1}{1-p_n}; x\right)} \right)^r dz,
\end{aligned}$$

where  $a$  is the positive function that appears in the max-domain of attraction condition

$$\frac{U(tx) - U(t)}{a(t)} \rightarrow \ln x, \text{ as } t \rightarrow \infty, \text{ for all } x > 0.$$

We have to study two cases depending on the sign of  $\beta$ .

**First case:**  $\beta \leq 0$ . Using the fact that  $U(\cdot)$  is an increasing function combined with Corollary B.2.10 in de Haan and Ferreira (2006, p. 376), we deduce that for  $p_n$  sufficiently large and  $\varepsilon$  sufficiently small that

$$T_2 \leq (1 - p_n) \left( \frac{a\left(\frac{1}{1-p_n}; x\right)}{U\left(\frac{1}{1-p_n}; x\right)} \right)^r O(c_n^{r\varepsilon-1}) = O\left(\frac{1-p_n}{c_n^{1+r-r\varepsilon}}\right),$$

where we have also used that

$$\frac{a\left(\frac{1}{1-p_n}; x\right)}{U\left(\frac{1}{1-p_n}; x\right)} = O\left(\frac{1}{c_n}\right), \text{ as } p_n \uparrow 1,$$

see e.g. the proof of Lemma 1 in de Wet *et al.* (2013).

**Second case:**  $\beta > 0$ . Using again Corollary B.2.10 in de Haan and Ferreira (2006, p. 376) we have for  $p_n$  sufficiently large,  $\delta$  and  $\tilde{\delta}$  positive constants, and  $\varepsilon$  and  $\tilde{\varepsilon}$  sufficiently small that

$$\begin{aligned} T_2 &\leq (1-p_n)\delta^r \left(\frac{a\left(\frac{1}{1-p_n}; x\right)}{U\left(\frac{1}{1-p_n}; x\right)}\right)^{r+\beta} \tilde{\delta}^\beta \left[1 + \frac{U\left(\frac{1}{1-p_n}; x\right)}{a\left(\frac{1}{1-p_n}; x\right)} \frac{1}{\tilde{\delta}(1+c_n)^{\tilde{\varepsilon}}}\right]^\beta \int_{1+c_n}^\infty z^{\beta\tilde{\varepsilon}+r\varepsilon-2} dz \\ &= (1-p_n)\delta^r \frac{1}{(1+c_n)^{\tilde{\varepsilon}\beta}} \left(\frac{a\left(\frac{1}{1-p_n}; x\right)}{U\left(\frac{1}{1-p_n}; x\right)}\right)^r \left[1 + \frac{a\left(\frac{1}{1-p_n}; x\right)}{U\left(\frac{1}{1-p_n}; x\right)} \tilde{\delta}(1+c_n)^{\tilde{\varepsilon}}\right]^\beta \int_{1+c_n}^\infty z^{\beta\tilde{\varepsilon}+r\varepsilon-2} dz \\ &= O\left(\frac{1-p_n}{c_n^{1+r-r\varepsilon}}\right). \end{aligned}$$

Finally

$$m^{(2)}(u_n, \alpha, \beta, r; x) = O\left(\frac{1-p_n}{c_n^{1+r+\alpha-r\varepsilon}}\right).$$

Combining all these results together leads to Lemma 1. □

## Proof of Lemma 2

From the rule of repeated expectations we have that

$$m_n(K, \alpha, \beta, r; x) = \mathbb{E}(K_h(x-X)m(u_n, \alpha, \beta, r; X)).$$

Straightforward operations give

$$\begin{aligned} m_n(K, \alpha, \beta, r; x) &= \int_{\Omega} K(z)m(u_n, \alpha, \beta, r; x-hz)f(x-hz)dz \\ &= m(u_n, \alpha, \beta, r; x)f(x) + m(u_n, \alpha, \beta, r; x) \int_{\Omega} K(z)(f(x-hz) - f(x))dz \\ &\quad + f(x) \int_{\Omega} K(z)(m(u_n, \alpha, \beta, r; x-hz) - m(u_n, \alpha, \beta, r; x))dz \\ &\quad + \int_{\Omega} K(z)(m(u_n, \alpha, \beta, r; x-hz) - m(u_n, \alpha, \beta, r; x))(f(x-hz) - f(x))dz \\ &=: m(u_n, \alpha, \beta, r; x)f(x) + T_3 + T_4 + T_5. \end{aligned}$$

We now analyze each of the terms separately. By  $(\mathcal{F})$  and  $(\mathcal{K})$  we have that

$$\begin{aligned} |T_3| &\leq m(u_n, \alpha, \beta, r; x)M_f \int_{\Omega} K(z)\|hz\|^{\eta_f} dz \\ &= O(m(u_n, \alpha, \beta, r; x)h^{\eta_f}), \end{aligned}$$

and, by  $(\mathcal{M})$  and  $(\mathcal{K})$ ,

$$\begin{aligned} |T_4| &\leq f(x)m(u_n, \alpha, \beta, r; x) \int_{\Omega} K(z) \left| \frac{m(u_n, \alpha, \beta, r; x - hz)}{m(u_n, \alpha, \beta, r; x)} - 1 \right| dz \\ &= O(m(u_n, \alpha, \beta, r; x)\Phi_n(x)). \end{aligned}$$

Using similar arguments one obtains  $T_5 = O(m(u_n, \alpha, \beta, r; x)h^{\eta_f}\Phi_n(x))$ . This proves the statement about the unconditional expectation.

For what concerns the convergence in probability, we have already from the first part of the proof that

$$\mathbb{E} \left( \tilde{T}_n(K, \alpha, \beta, r; x) \right) = \frac{\theta^r(x)\Gamma(1+r)}{(1+\alpha)^{1+r}} (1 + o(1)).$$

Also, again by using the result from the first part of the proof

$$\begin{aligned} \text{Var} \left( \tilde{T}_n(K, \alpha, \beta, r; x) \right) &= \frac{c_n^{2r} \text{Var} \left( K_h(x - X) e^{-c_n \alpha \left[ \left( \frac{Y}{u_n} \right)^{1/\theta(x)} - 1 \right]} \left( \frac{Y}{u_n} \right)^\beta \left( \ln \frac{Y}{u_n} \right)_+^r \mathbb{1}_{\{Y > u_n\}} \right)}{n(\bar{F}(u_n; x)f(x))^2} \\ &= \frac{\theta^{2r}(x) \|K\|_2^2 \Gamma(1+2r)}{(1+2\alpha)^{1+2r} n h^p \bar{F}(u_n; x) f(x)} (1 + o(1)). \end{aligned}$$

Thus

$$\text{Var} \left( \tilde{T}_n(K, \alpha, \beta, r; x) \right) \rightarrow 0$$

under the assumptions of the lemma, and the convergence in probability follows.  $\square$

## Proof of Corollary 1

First note that

$$\hat{f}_n(x) := \frac{1}{n} \sum_{i=1}^n K_h(x - X_i),$$

is a classical kernel density estimator for  $f$ . As shown in Parzen (1962), if  $nh^p \rightarrow \infty$ , then for all  $x \in \mathbb{R}^p$  where  $f(x) > 0$  one has that  $\hat{f}_n(x) \xrightarrow{\mathbb{P}} f(x)$ . The result follows then by noting that

$$\frac{\hat{F}(u_n; x)}{\hat{f}_n(x)} = \frac{f(x)}{\hat{f}_n(x)} \tilde{T}_n(K, 0, 0, 0; x).$$

$\square$

## Proof of Theorem 1

To prove the theorem we will adjust the arguments used to prove existence and consistency of solutions of the likelihood estimating equation, see e.g. Theorem 3.7 and Theorem 5.1 in Chapter 6 of Lehmann and Casella (1998), to the MDPD framework. Rescale the objective function  $\hat{\Delta}_\alpha(\theta; \hat{c}_n)$  as

$$\tilde{\Delta}_\alpha(\theta; \hat{c}_n) := \frac{\hat{\Delta}_\alpha(\theta; \hat{c}_n)}{\bar{F}(u_n; x) f(x) c_n^\alpha}.$$

First, we will show that

$$\mathbb{P}_{\theta_0(x)}(\tilde{\Delta}_\alpha(\theta_0(x); \hat{c}_n) < \tilde{\Delta}_\alpha(\theta; \hat{c}_n)) \rightarrow 1 \quad (10)$$

as  $n \rightarrow \infty$ , for any  $\theta$  sufficiently close to  $\theta_0(x)$ .

By Taylor's theorem

$$\begin{aligned} \tilde{\Delta}_\alpha(\theta; \hat{c}_n) - \tilde{\Delta}_\alpha(\theta_0(x); \hat{c}_n) &= \tilde{\Delta}'_\alpha(\theta_0(x); \hat{c}_n)(\theta - \theta_0(x)) + \frac{1}{2}\tilde{\Delta}''_\alpha(\theta_0(x); \hat{c}_n)(\theta - \theta_0(x))^2 \\ &\quad + \frac{1}{6}\tilde{\Delta}'''_\alpha(\tilde{\theta}; \hat{c}_n)(\theta - \theta_0(x))^3, \end{aligned}$$

where  $\tilde{\theta}$  is a value between  $\theta$  and  $\theta_0(x)$ . The term  $\tilde{\Delta}'_\alpha(\theta_0(x); \hat{c}_n)$  can be obtained from (4). Write  $\tilde{\Delta}'_\alpha(\theta_0(x); \hat{c}_n) =: R_1 + R_2 + R_3 - R_4$ . For analyzing the term  $R_1$ , we use the recursive relationships

$$\begin{aligned} \Gamma(a, b) &= e^{-b}b^{a-1} + (a-1)\Gamma(a-1, b), \\ \Psi(a, b) &= e^{-b}b^{a-1} \ln b + (a-1)\Psi(a-1, b) + \Gamma(a-1, b), \end{aligned}$$

Lemma 2, and the consistency of  $\hat{F}(u_n; x)$ , giving

$$R_1 \xrightarrow{\mathbb{P}} -\frac{\alpha}{\theta_0^{\alpha+1}(x)(1+\alpha)}.$$

For  $R_2$  we rearrange the terms to obtain

$$\begin{aligned} R_2 &= \frac{1+\alpha}{\theta_0^{\alpha+1}(x)} (1 + o_{\mathbb{P}}(1)) \left\{ \frac{T_n(K, \alpha, \alpha(1/\theta_0(x) - 1), 0; x)}{\bar{F}(u_n; x)f(x)} \right. \\ &\quad \left. + \frac{\frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \left[ e^{-\hat{c}_n \alpha \left[ \left( \frac{Y_i}{u_n} \right)^{1/\theta_0(x)} - 1 \right]} - e^{-c_n \alpha \left[ \left( \frac{Y_i}{u_n} \right)^{1/\theta_0(x)} - 1 \right]} \right] \left( \frac{Y_i}{u_n} \right)^{\alpha(1/\theta_0(x) - 1)} \mathbb{1}_{\{Y_i > u_n\}}}{\bar{F}(u_n; x)f(x)} \right\} \\ &=: \frac{1+\alpha}{\theta_0^{\alpha+1}(x)} (R_{2,1} + R_{2,2})(1 + o_{\mathbb{P}}(1)). \end{aligned}$$

By Lemma 2 we have that  $R_{2,1} \xrightarrow{\mathbb{P}} (1+\alpha)^{-1}$ . For the term  $R_{2,2}$ , we use the mean value theorem to obtain, with  $\tilde{c}_n$  being a random value between  $c_n$  and  $\hat{c}_n$ ,

$$\begin{aligned} R_{2,2} &= \alpha \ln \frac{\hat{F}(u_n; x)}{\bar{F}(u_n; x)} \left[ \frac{\frac{1}{n} \sum_{i=1}^n K_h(x - X_i) e^{-\tilde{c}_n \alpha \left[ \left( \frac{Y_i}{u_n} \right)^{1/\theta_0(x)} - 1 \right]} \left( \frac{Y_i}{u_n} \right)^{\alpha(1/\theta_0(x) - 1) + 1/\theta_0(x)} \mathbb{1}_{\{Y_i > u_n\}}}{\bar{F}(u_n; x)f(x)} \right. \\ &\quad \left. - \frac{\frac{1}{n} \sum_{i=1}^n K_h(x - X_i) e^{-\tilde{c}_n \alpha \left[ \left( \frac{Y_i}{u_n} \right)^{1/\theta_0(x)} - 1 \right]} \left( \frac{Y_i}{u_n} \right)^{\alpha(1/\theta_0(x) - 1)} \mathbb{1}_{\{Y_i > u_n\}}}{\bar{F}(u_n; x)f(x)} \right] \\ &=: \alpha \ln \frac{\hat{F}(u_n; x)}{\bar{F}(u_n; x)} (R_{2,2,1} - R_{2,2,2}), \end{aligned}$$



which can be easily bounded as follows

$$\begin{aligned}
R_{2,2,1} &\leq \frac{\frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \left(\frac{Y_i}{u_n}\right)^{\alpha(1/\theta_0(x)-1)+1/\theta_0(x)} \mathbb{1}_{\{Y_i > u_n\}}}{\widehat{F}(u_n; x) f(x)} = O_{\mathbb{P}}(1), \\
R_{2,2,2} &\leq \frac{\frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \left(\frac{Y_i}{u_n}\right)^{\alpha(1/\theta_0(x)-1)} \mathbb{1}_{\{Y_i > u_n\}}}{\widehat{F}(u_n; x) f(x)} = O_{\mathbb{P}}(1),
\end{aligned}$$

and therefore, by the consistency of  $\widehat{F}(u_n; x)$ , the convergence  $R_{2,2} \xrightarrow{\mathbb{P}} 0$  follows. Combining all results gives

$$R_2 \xrightarrow{\mathbb{P}} \frac{1}{\theta_0^{\alpha+1}(x)}.$$

The terms  $R_3$  and  $R_4$  can be analyzed in an analogous way and yield

$$R_3 \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad R_4 \xrightarrow{\mathbb{P}} \frac{1}{\theta_0^{\alpha+1}(x)(1+\alpha)}.$$

Thus  $\widetilde{\Delta}'_{\alpha}(\theta_0(x); \widehat{c}_n) \xrightarrow{\mathbb{P}} 0$ . Let  $|\theta - \theta_0(x)| = r$ ,  $r > 0$ . With probability tending to 1 we have that

$$\left| \widetilde{\Delta}'_{\alpha}(\theta_0(x); \widehat{c}_n)(\theta - \theta_0(x)) \right| < r^3.$$

We now turn to the analysis of  $\widetilde{\Delta}''_{\alpha}(\theta_0(x); \widehat{c}_n)$ . Let

$$\phi(a, b) := \int_b^{\infty} \ln^2 z z^{a-1} e^{-z} dz,$$

and

$$\widehat{T}_n(K, \alpha, \beta, r; x) := \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) e^{-\widehat{c}_n \alpha \left[ \left(\frac{Y_i}{u_n}\right)^{1/\theta_0(x)} - 1 \right]} \left(\frac{Y_i}{u_n}\right)^{\beta} \left(\ln \frac{Y_i}{u_n}\right)_+^r \mathbb{1}_{\{Y_i > u_n\}}.$$

Note that the function  $\phi(a, b)$  satisfies the recursive relationship

$$\phi(a, b) = e^{-b} b^{a-1} \ln^2 b + (a-1)\phi(a-1, b) + 2\Psi(a-1, b). \quad (11)$$

After tedious calculations one obtains the following expression for  $\tilde{\Delta}''_{\alpha}(\theta_0(x); \hat{c}_n)$ :

$$\begin{aligned}
\tilde{\Delta}''_{\alpha}(\theta_0(x); \hat{c}_n) = & \frac{T_n(K, 0, 0, 0; x)}{\overline{F}(u_n; x)f(x)} \frac{e^{\hat{c}_n(1+\alpha)} \hat{c}_n^{\alpha\theta_0(x)}}{\theta_0^{\alpha+2}(x)(1+\alpha)^{1+\alpha(1-\theta_0(x))} c_n^{\alpha}} \\
& \times \{ \alpha(1+\alpha)\Gamma(\alpha(1-\theta_0(x))+1, \hat{c}_n(1+\alpha)) + 2\alpha^2\theta_0(x)\Psi(\alpha(1-\theta_0(x))+1, \hat{c}_n(1+\alpha)) \\
& + \alpha^2\theta_0^2(x)\phi(\alpha(1-\theta_0(x))+1, \hat{c}_n(1+\alpha)) \\
& - 2\alpha^2\theta_0(x)\ln(\hat{c}_n(1+\alpha))[\Gamma(\alpha(1-\theta_0(x))+1, \hat{c}_n(1+\alpha)) + \theta_0(x)\Psi(\alpha(1-\theta_0(x))+1, \hat{c}_n(1+\alpha))] \\
& + \alpha^2\theta_0^2(x)\ln^2(\hat{c}_n(1+\alpha))\Gamma(\alpha(1-\theta_0(x))+1, \hat{c}_n(1+\alpha)) \} \\
& - \frac{(\alpha+1)^2 \hat{c}_n^{\alpha} \widehat{T}_n(K, \alpha, \alpha(1/\theta_0(x)-1), 0; x)}{\theta_0^{\alpha+2}(x)c_n^{\alpha}} \frac{\widehat{T}_n(K, \alpha, \alpha(1/\theta_0(x)-1), 1; x)}{\overline{F}(u_n; x)f(x)} \\
& + \frac{2(\alpha+1)^2 \hat{c}_n^{\alpha+1} \widehat{T}_n(K, \alpha, \alpha(1/\theta_0(x)-1) + 1/\theta_0(x), 1; x)}{\theta_0^{\alpha+3}(x)c_n^{\alpha}} \frac{\widehat{T}_n(K, \alpha, \alpha(1/\theta_0(x)-1) + 1/\theta_0(x), 2; x)}{\overline{F}(u_n; x)f(x)} \\
& - \frac{\alpha(1+\alpha)\hat{c}_n^{\alpha} \widehat{T}_n(K, \alpha, \alpha(1/\theta_0(x)-1), 2; x)}{\theta_0^{\alpha+4}(x)c_n^{\alpha}} \frac{\widehat{T}_n(K, \alpha, \alpha(1/\theta_0(x)-1) + 1/\theta_0(x), 2; x)}{\overline{F}(u_n; x)f(x)} \\
& + \frac{(1+2\alpha)(1+\alpha)\hat{c}_n^{\alpha+1} \widehat{T}_n(K, \alpha, \alpha(1/\theta_0(x)-1) + 1/\theta_0(x), 2; x)}{\theta_0^{\alpha+4}(x)c_n^{\alpha}} \frac{\widehat{T}_n(K, \alpha, \alpha(1/\theta_0(x)-1) + 2/\theta_0(x), 2; x)}{\overline{F}(u_n; x)f(x)}.
\end{aligned}$$

By a line of argumentation similar to that used for  $\tilde{\Delta}'_{\alpha}(\theta_0(x); \hat{c}_n)$  and also using (11) one obtains that under the conditions of the theorem

$$\tilde{\Delta}''_{\alpha}(\theta_0(x); \hat{c}_n) \xrightarrow{\mathbb{P}} \frac{1+\alpha^2}{\theta_0^{\alpha+2}(x)(1+\alpha)^2}. \quad (12)$$

Write

$$\begin{aligned}
\frac{1}{2} \tilde{\Delta}''_{\alpha}(\theta_0(x); \hat{c}_n)(\theta - \theta_0(x))^2 &= \frac{1+\alpha^2}{2\theta_0^{\alpha+2}(x)(1+\alpha)^2} (\theta - \theta_0(x))^2 \\
&+ \frac{1}{2} \left( \tilde{\Delta}''_{\alpha}(\theta_0(x); \hat{c}_n) - \frac{1+\alpha^2}{\theta_0^{\alpha+2}(x)(1+\alpha)^2} \right) (\theta - \theta_0(x))^2.
\end{aligned}$$

The random part in the right-hand side of the above display is in absolute value less than  $r^3$  with probability tending to 1. There exist thus a  $\delta_1 > 0$  and an  $r_0 > 0$  such that for  $r < r_0$

$$\frac{1}{2} \tilde{\Delta}''_{\alpha}(\theta_0(x); \hat{c}_n)(\theta - \theta_0(x))^2 > \delta_1 r^2$$

with probability tending to 1.

For the third order derivative, one can show that  $|\tilde{\Delta}'''_{\alpha}(\theta; \hat{c}_n)| \leq M(\mathbf{V})$ , where  $\mathbf{V} := [(X_1, Y_1), \dots, (X_n, Y_n)]$ , for  $\theta \in (\theta_0(x) - r, \theta_0(x) + r)$ , with  $M(\mathbf{V}) \xrightarrow{\mathbb{P}} M$ , which is bounded. The derivation is straightforward but lengthy and is therefore omitted from the paper. We can thus conclude that with probability tending to 1

$$\frac{1}{6} |\tilde{\Delta}'''_{\alpha}(\tilde{\theta}; \hat{c}_n)(\theta - \theta_0(x))^3| < \frac{1}{3} Mr^3.$$

Overall, we have that with probability tending to 1

$$\tilde{\Delta}_\alpha(\theta; \hat{c}_n) - \tilde{\Delta}_\alpha(\theta_0(x); \hat{c}_n) > \delta_1 r^2 - (1 + M/3)r^3,$$

which is positive if  $r < \delta_1/(1 + M/3)$ , and thus (10) follows.

To complete the proof we adjust the line of argumentation of Theorem 3.7 in Chapter 6 of Lehmann and Casella (1998). Let  $\delta > 0$  be such that  $\theta_0(x) - \delta > 0$ , and define

$$S_n(\delta) = \left\{ \mathbf{v} : \tilde{\Delta}_\alpha(\theta_0(x); \hat{c}_n) < \tilde{\Delta}_\alpha(\theta_0(x) - \delta; \hat{c}_n) \text{ and } \tilde{\Delta}_\alpha(\theta_0(x); \hat{c}_n) < \tilde{\Delta}_\alpha(\theta_0(x) + \delta; \hat{c}_n) \right\}.$$

For any  $\mathbf{v} \in S_n(\delta)$ , since  $\tilde{\Delta}_\alpha(\theta; \hat{c}_n)$  is differentiable with respect to  $\theta$ , we have that there exists a  $\hat{\theta}_{n,\delta}(x) \in (\theta_0(x) - \delta, \theta_0(x) + \delta)$  where  $\tilde{\Delta}_\alpha(\theta; \hat{c}_n)$  achieves a local minimum, so  $\tilde{\Delta}'_\alpha(\hat{\theta}_{n,\delta}(x); \hat{c}_n) = 0$ . By the first part of the proof of the theorem,  $\mathbb{P}_{\theta_0(x)}(S_n(\delta)) \rightarrow 1$  for any  $\delta$  small enough, and hence there exists a sequence  $\delta_n \downarrow 0$  such that  $\mathbb{P}_{\theta_0(x)}(S_n(\delta_n)) \rightarrow 1$  as  $n \rightarrow \infty$ . Now let  $\hat{\theta}_n^*(x) = \hat{\theta}_{n,\delta_n}(x)$  if  $\mathbf{v} \in S_n(\delta_n)$  and arbitrary otherwise. Since  $\mathbf{v} \in S_n(\delta_n)$  implies  $\tilde{\Delta}'_\alpha(\hat{\theta}_n^*(x); \hat{c}_n) = 0$  we have that

$$\mathbb{P}_{\theta_0(x)}(\tilde{\Delta}'_\alpha(\hat{\theta}_n^*(x); \hat{c}_n) = 0) \geq \mathbb{P}_{\theta_0(x)}(S_n(\delta_n)) \rightarrow 1,$$

as  $n \rightarrow \infty$ , which establishes the existence part. For the consistency of the solution sequence, note that for any fixed  $\delta > 0$  and  $n$  large enough

$$\mathbb{P}_{\theta_0(x)}(|\hat{\theta}_n^*(x) - \theta_0(x)| < \delta) \geq \mathbb{P}_{\theta_0(x)}(|\hat{\theta}_n^*(x) - \theta_0(x)| < \delta_n) \geq \mathbb{P}_{\theta_0(x)}(S_n(\delta_n)) \rightarrow 1,$$

as  $n \rightarrow \infty$ , whence the consistency of the estimator sequence.  $\square$

## Proof of Theorem 2

Let  $r_n := \sqrt{nh^p \bar{F}(u_n; x)}$ . To prove the theorem we will make use of the Cramér-Wold device (see e.g. Severini, 2005, p. 337) according to which it is sufficient to show that

$$\Lambda_n := \xi' r_n [\mathbb{T}_n - \mathbb{E}(\mathbb{T}_n)] \rightsquigarrow N\left(0, \frac{1}{f(x)} \xi' \Sigma \xi\right),$$

for all  $\xi \in \mathbb{R}^J$ .

Take an arbitrary  $\xi \in \mathbb{R}^J$ . A straightforward rearrangement of terms leads to

$$\begin{aligned} \Lambda_n &= \sum_{i=1}^n \sqrt{\frac{h^p}{n\bar{F}(u_n; x)}} \frac{1}{f(x)} \left[ \sum_{j=1}^J \xi_j c_n^{r_j} K_{j,h}(x - X_i) e^{-c_n \alpha_j \left[\left(\frac{Y_i}{u_n}\right)^{1/\theta_0(x)} - 1\right]} \left(\frac{Y_i}{u_n}\right)^{\beta_j} \left(\ln \frac{Y_i}{u_n}\right)_+^{r_j} \mathbb{1}_{\{Y_i > u_n\}} \right. \\ &\quad \left. - \mathbb{E} \left( \sum_{j=1}^J \xi_j c_n^{r_j} K_{j,h}(x - X_i) e^{-c_n \alpha_j \left[\left(\frac{Y_i}{u_n}\right)^{1/\theta_0(x)} - 1\right]} \left(\frac{Y_i}{u_n}\right)^{\beta_j} \left(\ln \frac{Y_i}{u_n}\right)_+^{r_j} \mathbb{1}_{\{Y_i > u_n\}} \right) \right] \\ &=: \sum_{i=1}^n W_i. \end{aligned}$$

By the model assumptions,  $W_1, \dots, W_n$  are i.i.d. random variables, and therefore  $\text{Var}(\Lambda_n) = n \text{Var}(W_1)$ . We have

$$\text{Var}(W_1) = \frac{h^p}{n\bar{F}(u_n; x) f^2(x)} \sum_{j=1}^J \sum_{k=1}^J \xi_j \xi_k c_n^{r_j + r_k} \mathbb{C}_{j,k},$$

with

$$\begin{aligned} \mathbb{C}_{j,k} &:= \mathbb{E} \left[ K_{j,h}(x - X_1) K_{k,h}(x - X_1) e^{-c_n(\alpha_j + \alpha_k)} \left[ \left( \frac{Y_1}{u_n} \right)^{1/\theta_0(x)} - 1 \right] \left( \frac{Y_1}{u_n} \right)^{\beta_j + \beta_k} \left( \ln \frac{Y_1}{u_n} \right)_+^{r_j + r_k} \mathbb{1}_{\{Y_1 > u_n\}} \right] \\ &\quad - \mathbb{E} \left[ K_{j,h}(x - X_1) e^{-c_n \alpha_j} \left[ \left( \frac{Y_1}{u_n} \right)^{1/\theta_0(x)} - 1 \right] \left( \frac{Y_1}{u_n} \right)^{\beta_j} \left( \ln \frac{Y_1}{u_n} \right)_+^{r_j} \mathbb{1}_{\{Y_1 > u_n\}} \right] \times \\ &\quad \mathbb{E} \left[ K_{k,h}(x - X_1) e^{-c_n \alpha_k} \left[ \left( \frac{Y_1}{u_n} \right)^{1/\theta_0(x)} - 1 \right] \left( \frac{Y_1}{u_n} \right)^{\beta_k} \left( \ln \frac{Y_1}{u_n} \right)_+^{r_k} \mathbb{1}_{\{Y_1 > u_n\}} \right]. \end{aligned}$$

By using the results of Lemmas 1 and 2 we have then

$$\mathbb{C}_{j,k} = \frac{\bar{F}(u_n; x) f(x) \|K_j K_k\|_1 \Gamma(1 + r_j + r_k) \theta_0^{r_j + r_k}(x)}{h^p c_n^{r_j + r_k} (1 + \alpha_j + \alpha_k)^{1 + r_j + r_k}} (1 + o(1)),$$

which gives that  $\text{Var}(\Lambda_n) = 1/f(x) \xi' \Sigma \xi (1 + o(1))$ . To establish the convergence in distribution to a normal random variable we have to verify the Lyapounov condition for triangular arrays of random variables (Billingsley, 1995, p. 362). In the present context this simplifies to verifying that  $n\mathbb{E}|W_1|^3 \rightarrow 0$ . We have

$$\begin{aligned} \mathbb{E}|W_1|^3 &\leq \left( \frac{h^p}{n\bar{F}(u_n; x)} \right)^{3/2} \frac{1}{f^3(x)} \times \\ &\quad \left\{ \mathbb{E} \left[ \left( \sum_{j=1}^J |\xi_j| c_n^{r_j} K_{j,h}(x - X_1) e^{-c_n \alpha_j} \left[ \left( \frac{Y_1}{u_n} \right)^{1/\theta_0(x)} - 1 \right] \left( \frac{Y_1}{u_n} \right)^{\beta_j} \left( \ln \frac{Y_1}{u_n} \right)_+^{r_j} \mathbb{1}_{\{Y_1 > u_n\}} \right)^3 \right] \right. \\ &\quad + 3\mathbb{E} \left[ \left( \sum_{j=1}^J |\xi_j| c_n^{r_j} K_{j,h}(x - X_1) e^{-c_n \alpha_j} \left[ \left( \frac{Y_1}{u_n} \right)^{1/\theta_0(x)} - 1 \right] \left( \frac{Y_1}{u_n} \right)^{\beta_j} \left( \ln \frac{Y_1}{u_n} \right)_+^{r_j} \mathbb{1}_{\{Y_1 > u_n\}} \right)^2 \right] \\ &\quad \times \mathbb{E} \left[ \sum_{j=1}^J |\xi_j| c_n^{r_j} K_{j,h}(x - X_1) e^{-c_n \alpha_j} \left[ \left( \frac{Y_1}{u_n} \right)^{1/\theta_0(x)} - 1 \right] \left( \frac{Y_1}{u_n} \right)^{\beta_j} \left( \ln \frac{Y_1}{u_n} \right)_+^{r_j} \mathbb{1}_{\{Y_1 > u_n\}} \right] \\ &\quad \left. + 4 \left[ \mathbb{E} \left( \sum_{j=1}^J |\xi_j| c_n^{r_j} K_{j,h}(x - X_1) e^{-c_n \alpha_j} \left[ \left( \frac{Y_1}{u_n} \right)^{1/\theta_0(x)} - 1 \right] \left( \frac{Y_1}{u_n} \right)^{\beta_j} \left( \ln \frac{Y_1}{u_n} \right)_+^{r_j} \mathbb{1}_{\{Y_1 > u_n\}} \right)^3 \right] \right\}. \end{aligned}$$

Again by using Lemmas 1 and 2 we obtain that

$$\mathbb{E}|W_1|^3 = O \left( \left( n \sqrt{nh^p \bar{F}(u_n; x)} \right)^{-1} \right),$$

and hence,  $n\mathbb{E}|W_1|^3 \rightarrow 0$ . □

### Proof of Theorem 3

Apply a Taylor series expansion to the estimating equation  $\tilde{\Delta}'_\alpha(\hat{\theta}_n(x); \hat{c}_n) = 0$  around  $\theta_0(x)$ . This gives

$$0 = \tilde{\Delta}'_\alpha(\theta_0(x); \hat{c}_n) + \tilde{\Delta}''_\alpha(\theta_0(x); \hat{c}_n)(\hat{\theta}_n(x) - \theta_0(x)) + \frac{1}{2} \tilde{\Delta}'''_\alpha(\tilde{\theta}_n(x); \hat{c}_n)(\hat{\theta}_n(x) - \theta_0(x))^2$$

where  $\tilde{\theta}_n(x)$  is a random value between  $\hat{\theta}_n(x)$  and  $\theta_0(x)$ . A straightforward rearrangement of the terms leads then to

$$\begin{aligned} r_n(\hat{\theta}_n(x) - \theta_0(x)) &= -\frac{1}{\tilde{\Delta}_\alpha''(\theta_0(x); \hat{c}_n) + \frac{1}{2}\tilde{\Delta}_\alpha'''(\tilde{\theta}_n(x); \hat{c}_n)(\hat{\theta}_n(x) - \theta_0(x))} r_n \tilde{\Delta}'_\alpha(\theta_0(x); \hat{c}_n) \\ &= -\frac{\theta_0^{\alpha+2}(x)(1+\alpha)^2}{1+\alpha^2} r_n \tilde{\Delta}'_\alpha(\theta_0(x); \hat{c}_n)(1 + o_{\mathbb{P}}(1)) \end{aligned} \quad (13)$$

by (12), the consistency of  $\hat{\theta}_n(x)$  and the boundedness of the third derivative. Another application of Taylor's theorem gives

$$r_n \tilde{\Delta}'_\alpha(\theta_0(x); \hat{c}_n) = r_n \tilde{\Delta}'_\alpha(\theta_0(x); c_n) - \left. \frac{\partial}{\partial \tilde{c}_n} \tilde{\Delta}'_\alpha(\theta_0(x); \tilde{c}_n) \right|_{\tilde{c}_n} r_n \ln \frac{\hat{F}(u_n; x)}{F(u_n; x)}$$

with  $\tilde{c}_n$  being a random value between  $\hat{c}_n$  and  $c_n$ . Direct computations allow us to prove that, under our assumptions, and using the second part of Lemma 2 and arguments similar to those used in the proof of Theorem 1, we have

$$\left. \frac{\partial}{\partial \tilde{c}_n} \tilde{\Delta}'_\alpha(\theta_0(x); \tilde{c}_n) \right|_{\tilde{c}_n} = o_{\mathbb{P}}(1).$$

In addition, by Theorem 2 in de Wet *et al.* (2013), we deduce that

$$\begin{aligned} r_n \tilde{\Delta}'_\alpha(\theta_0(x); \hat{c}_n) &= r_n \tilde{\Delta}'_\alpha(\theta_0(x); c_n) + o_{\mathbb{P}}(1) \\ &= -\frac{\alpha}{\theta_0^{\alpha+1}(x)(1+\alpha)} r_n \left[ \tilde{T}_n(K, 0, 0, 0; x) - 1 \right] \\ &\quad + \frac{1+\alpha}{\theta_0^{\alpha+1}(x)} r_n \left[ \tilde{T}_n(K, \alpha, \alpha(1/\theta_0(x) - 1), 0; x) - \frac{1}{1+\alpha} \right] \\ &\quad - \frac{1+\alpha}{\theta_0^{\alpha+2}(x)} r_n \left[ \tilde{T}_n(K, \alpha, \alpha(1/\theta_0(x) - 1) + 1/\theta_0(x), 1; x) - \frac{\theta_0(x)}{(1+\alpha)^2} \right] + o_{\mathbb{P}}(1). \end{aligned} \quad (14)$$

Finally, combining (13) and (14) with Theorem 2 and the delta-method, Theorem 3 follows.  $\square$

## Acknowledgement

The authors are very grateful to the referee for her/his 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. and Jones, M.C. (1998). Robust and efficient estimation by minimizing a density power divergence. *Biometrika* **85**, 549–559.
- [2] Beirlant, J., Broniatowski, M., Teugels, J.L. and Vynckier, P. (1995). The mean residual life function at great age: applications to tail estimation. *J. Statist. Plann. Inference* **45**, 21–48.
- [3] Billingsley, P. (1995). *Probability and Measure*. Wiley series in Probability and Mathematical Statistics.

- [4] Brazauskas, V. and Serfling, R. (2000). Robust estimation of tail parameters for two-parameter Pareto and exponential models via generalized quantile statistics. *Extremes* **3**, 231–249.
- [5] Broniatowski, M. (1993). On the estimation of the Weibull tail coefficient. *J. Statist. Plann. Inference* **35**, 349–366.
- [6] Daouia, A., Gardes, L. and Girard, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli*, to appear.
- [7] Daouia, A., Gardes, L., Girard, S. and Lekina, A. (2011). Kernel estimators of extreme level curves. *Test* **20**, 311–333.
- [8] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer.
- [9] de Wet, T., Goegebeur, Y., Guillou, A. and Osmann, M. (2013). Kernel regression with Weibull-type tails. Submitted.
- [10] Diebolt, J., Gardes, L., Girard, S. and Guillou, A. (2008). Bias-reduced estimators of the Weibull tail-coefficient. *Test* **17**, 311–331.
- [11] Dierckx, G., Beirlant, J., De Waal, D. and Guillou, A. (2009). A new estimation method for Weibull-type tails based on the mean excess function. *J. Statist. Plann. Inference* **139**, 1905–1920.
- [12] Dierckx, G., Goegebeur, Y. and Guillou, A. (2013a). An asymptotically unbiased minimum density power divergence estimator for the Pareto-tail index. *J. Multivariate Anal.*, to appear.
- [13] Dierckx, G., Goegebeur, Y. and Guillou, A. (2013b). Local robust and asymptotically unbiased estimation of conditional Pareto-type tails. Submitted.
- [14] Dupuis, D. and Field, C. (1998). Robust estimation of extremes. *Canad. J. Statist.* **26**, 119–215.
- [15] Gannoun, A., Girard, S., Guinot, C. and Saracco, J. (2002). Reference ranges based on nonparametric quantile regression. *Stat. Med.* **21**, 3119–3135.
- [16] Gardes, L. and Girard, S. (2005). Estimating extreme quantiles of Weibull tail distributions. *Comm. Statist. Theory Methods* **34**, 1065–1080.
- [17] Gardes, L. and Girard, S. (2008a). A moving window approach for nonparametric estimation of the conditional tail index. *J. Multivariate Anal.* **99**, 2368–2388.
- [18] Gardes, L. and Girard, S. (2008b). Estimation of the Weibull-tail coefficient with linear combination of upper order statistics. *J. Statist. Plann. Inference* **138**, 1416–1427.
- [19] Gardes, L., Girard, S. and Lekina, A. (2010). Functional nonparametric estimation of conditional extreme quantiles. *J. Multivariate Anal.* **101**, 419–433.
- [20] Gardes, L. and Stupfler, G. (2013). Estimation of the conditional tail index using a smoothed local Hill estimator. *Extremes*, to appear.
- [21] Geluk, J.L. and de Haan, L. (1987). *Regular Variation, Extensions and Tauberian Theorems*. CWI Tract 40, Center for Mathematics and Computer Science, Amsterdam.
- [22] Girard, S. (2004). A Hill type estimator of the Weibull tail coefficient. *Comm. Statist. Theory Methods* **33**, 205–234.
- [23] Goegebeur, Y., Beirlant, J. and de Wet, T. (2010). Generalized kernel estimators for the Weibull tail coefficient. *Comm. Statist. Theory Methods* **39**, 3695–3716.

- [24] Goegebeur, Y. and Guillou, A. (2011). A weighted mean excess function approach to the estimation of Weibull-type tails. *Test* **20**, 138–162.
- [25] Goegebeur, Y., Guillou, A. and Schorgen, A. (2013a). Nonparametric regression estimation of conditional tails – the random covariate case. To appear in *Statistics*.
- [26] Goegebeur, Y., Guillou, A. and Stupfler, G. (2013b). Uniform asymptotic properties of a nonparametric regression estimator of conditional tails. Submitted. Available at <http://hal.archives-ouvertes.fr/hal-00794724>.
- [27] Hall, P. (1982). On some simple estimates of an exponent of regular variation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **44**, 37–42.
- [28] Juárez, S. and Schucany, W. (2004). Robust and efficient estimation for the generalized Pareto distribution. *Extremes* **7**, 237–251.
- [29] Kim, M. and Lee, S. (2008). Estimation of a tail index based on minimum density power divergence. *J. Multivariate Anal.* **99**, 2453–2471.
- [30] Klüppelberg, C. and Villaseñor, J.A. (1993). Estimation of distribution tails - a semiparametric approach. *Blätter der Deutschen Gesellschaft für Versicherungsmathematik* **21**, 213–235.
- [31] Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation*. Springer.
- [32] Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33**, 1065–1076.
- [33] Peng, L. and Welsh, A. (2001). Robust estimation of the generalized Pareto distribution. *Extremes* **4**, 53–65.
- [34] Severini, T. (2005). *Elements of Distribution Theory*. Cambridge Series in Statistical and Probabilistic Mathematics.
- [35] Vandewalle, B., Beirlant, J., Christmann, A. and Hubert, M. (2007). A robust estimator for the tail index of Pareto-type distributions. *Comput. Statist. Data Anal.* **51**, 6252–6268.
- [36] Wang, H. and Tsai, C.L. (2009). Tail index regression. *J. Amer. Statist. Assoc.* **104**, 1233–1240.
- [37] Yao, Q. (1999). Conditional predictive regions for stochastic processes. Technical report, University of Kent at Canterbury.

Distribution	Value of $\alpha$	Oracle			Data Driven		
		Bias	MSE	Sd	Bias	MSE	Sd
$\mathcal{W}(1/\theta(x), 1)$	0.1	0.051	0.004	0.065	0.095	0.014	0.097
	0.25	0.053	0.005	0.067	0.095	0.014	0.099
	0.5	0.059	0.006	0.076	0.096	0.015	0.104
	1	0.070	0.008	0.088	0.098	0.016	0.109
	2	0.079	0.010	0.100	0.101	0.017	0.111
$\mathcal{EW}(1/\theta(x), -1)$	0.1	0.053	0.005	0.065	0.089	0.013	0.100
	0.25	0.055	0.005	0.068	0.089	0.013	0.102
	0.5	0.060	0.006	0.075	0.091	0.014	0.107
	1	0.068	0.007	0.085	0.093	0.014	0.110
	2	0.068	0.007	0.085	0.093	0.014	0.110
$\widetilde{\mathcal{W}}(1/\theta(x), -5)$	0.1	0.062	0.006	0.062	0.077	0.009	0.084
	0.25	0.063	0.006	0.066	0.077	0.009	0.084
	0.5	0.069	0.007	0.074	0.079	0.010	0.086
	1	0.075	0.009	0.086	0.083	0.010	0.090
	2	0.082	0.010	0.092	0.084	0.011	0.091

Table 1: Bias, mean squared error and standard deviation of the MDPD estimator for the three distributions, the two approaches and different values of  $\alpha$  in the case where there is no contamination.



Distribution $F$	Value of $\alpha$	Oracle			Data Driven		
		Bias	MSE	Sd	Bias	MSE	Sd
$\mathcal{W}(1/\theta(x), 1)$	0.1	0.068	0.008	0.086	0.173	0.057	0.170
	0.25	0.062	0.007	0.079	0.148	0.041	0.151
	0.5	0.064	0.007	0.081	0.130	0.031	0.138
	1	0.072	0.009	0.091	0.126	0.028	0.133
	2	0.080	0.011	0.102	0.130	0.029	0.132
$\mathcal{EW}(1/\theta(x), -1)$	0.1	0.069	0.008	0.089	0.179	0.058	0.175
	0.25	0.062	0.006	0.079	0.150	0.040	0.154
	0.5	0.064	0.007	0.081	0.130	0.029	0.141
	1	0.071	0.008	0.089	0.124	0.026	0.135
	2	0.076	0.009	0.095	0.125	0.026	0.134
$\widetilde{\mathcal{W}}(1/\theta(x), -5)$	0.1	0.074	0.008	0.082	0.142	0.044	0.175
	0.25	0.069	0.008	0.077	0.117	0.029	0.151
	0.5	0.074	0.009	0.082	0.102	0.020	0.132
	1	0.079	0.010	0.090	0.098	0.017	0.121
	2	0.083	0.011	0.095	0.097	0.016	0.119

Table 2: Bias, mean squared error and standard deviation of the MDPD estimator for the three distributions, the two approaches and different values of  $\alpha$  in the shifted strict Weibull contaminated case with  $\varepsilon = 0.005$  and  $\beta = 4/3$ .

Distribution $F$	Value of $\alpha$	Oracle			Data Driven		
		Bias	MSE	Sd	Bias	MSE	Sd
$\mathcal{W}(1/\theta(x), 1)$	0.1	0.082	0.012	0.099	0.257	0.117	0.208
	0.25	0.069	0.008	0.087	0.228	0.094	0.200
	0.5	0.068	0.008	0.087	0.193	0.069	0.182
	1	0.076	0.010	0.097	0.177	0.057	0.167
	2	0.084	0.012	0.107	0.178	0.056	0.162
$\mathcal{EW}(1/\theta(x), -1)$	0.1	0.078	0.011	0.100	0.251	0.108	0.214
	0.25	0.067	0.008	0.087	0.209	0.078	0.197
	0.5	0.067	0.008	0.086	0.171	0.054	0.177
	1	0.074	0.009	0.093	0.157	0.044	0.160
	2	0.080	0.010	0.100	0.157	0.043	0.154
$\widetilde{\mathcal{W}}(1/\theta(x), -5)$	0.1	0.083	0.011	0.097	0.198	0.078	0.204
	0.25	0.074	0.009	0.086	0.160	0.055	0.190
	0.5	0.075	0.009	0.086	0.130	0.036	0.168
	1	0.080	0.010	0.093	0.118	0.027	0.149
	2	0.084	0.011	0.099	0.114	0.025	0.141

Table 3: Bias, mean squared error and standard deviation of the MDPD estimator for the three distributions, the two approaches and different values of  $\alpha$  in the shifted strict Weibull contaminated case with  $\varepsilon = 0.01$  and  $\beta = 4/3$ .

Distribution $F$	Value of $\alpha$	Oracle			Data Driven		
		Bias	MSE	Sd	Bias	MSE	Sd
$\mathcal{W}(1/\theta(x), 1)$	0.1	0.066	0.008	0.076	0.147	0.043	0.152
	0.25	0.062	0.006	0.069	0.125	0.028	0.125
	0.5	0.066	0.007	0.075	0.114	0.023	0.118
	1	0.075	0.009	0.084	0.114	0.023	0.119
	2	0.083	0.011	0.093	0.117	0.025	0.121
$\mathcal{EW}(1/\theta(x), -1)$	0.1	0.069	0.008	0.076	0.142	0.039	0.146
	0.25	0.063	0.006	0.069	0.123	0.029	0.130
	0.5	0.067	0.007	0.072	0.111	0.023	0.121
	1	0.072	0.008	0.078	0.108	0.021	0.118
	2	0.083	0.011	0.093	0.109	0.021	0.117
$\widetilde{\mathcal{W}}(1/\theta(x), -5)$	0.1	0.078	0.010	0.078	0.129	0.037	0.162
	0.25	0.071	0.008	0.069	0.103	0.020	0.120
	0.5	0.076	0.009	0.073	0.097	0.016	0.107
	1	0.083	0.010	0.081	0.097	0.016	0.104
	2	0.088	0.012	0.086	0.096	0.015	0.104

Table 4: Bias, mean squared error and standard deviation of the MDPD estimator for the three distributions, the two approaches and different values of  $\alpha$  in the shifted strict Pareto contaminated case with  $\varepsilon = 0.005$ ,  $\gamma = 1/4$ ,  $y_c = 1.2$  times the 95% quantile of the uncontaminated distribution.

Distribution $F$	Value of $\alpha$	Oracle			Data Driven		
		Bias	MSE	Sd	Bias	MSE	Sd
$\mathcal{W}(1/\theta(x), 1)$	0.1	0.078	0.009	0.071	0.196	0.071	0.185
	0.25	0.079	0.009	0.071	0.154	0.045	0.157
	0.5	0.083	0.011	0.077	0.129	0.032	0.141
	1	0.090	0.012	0.085	0.123	0.028	0.134
	2	0.096	0.014	0.090	0.124	0.028	0.132
$\mathcal{EW}(1/\theta(x), -1)$	0.1	0.078	0.010	0.085	0.187	0.068	0.178
	0.25	0.072	0.008	0.075	0.152	0.042	0.147
	0.5	0.072	0.008	0.075	0.132	0.030	0.133
	1	0.079	0.010	0.082	0.124	0.027	0.126
	2	0.085	0.011	0.089	0.124	0.027	0.124
$\widetilde{\mathcal{W}}(1/\theta(x), -5)$	0.1	0.086	0.011	0.086	0.167	0.055	0.180
	0.25	0.075	0.009	0.076	0.127	0.030	0.142
	0.5	0.078	0.009	0.078	0.110	0.021	0.122
	1	0.083	0.011	0.084	0.104	0.018	0.115
	2	0.088	0.012	0.089	0.102	0.018	0.113

Table 5: Bias, mean squared error and standard deviation of the MDPD estimator for the three distributions, the two approaches and different values of  $\alpha$  in the shifted strict Pareto contaminated case with  $\varepsilon = 0.01$ ,  $\gamma = 1/4$ ,  $y_c = 1.2$  times the 95% quantile of the uncontaminated distribution.

Distribution $F$	Value of $\alpha$	Oracle			Data Driven		
		Bias	MSE	Sd	Bias	MSE	Sd
$\mathcal{W}(1/\theta(x), 1)$	0.1	0.073	0.009	0.083	0.171	0.058	0.190
	0.25	0.066	0.007	0.073	0.141	0.040	0.163
	0.5	0.067	0.007	0.073	0.123	0.030	0.148
	1	0.074	0.009	0.081	0.118	0.026	0.138
	2	0.083	0.011	0.090	0.121	0.026	0.136
$\mathcal{EW}(1/\theta(x), -1)$	0.1	0.074	0.009	0.082	0.173	0.061	0.196
	0.25	0.068	0.007	0.073	0.145	0.041	0.168
	0.5	0.071	0.008	0.075	0.130	0.031	0.152
	1	0.076	0.009	0.082	0.124	0.027	0.145
	2	0.082	0.011	0.089	0.123	0.027	0.143
$\widetilde{\mathcal{W}}(1/\theta(x), -5)$	0.1	0.083	0.011	0.084	0.145	0.048	0.198
	0.25	0.073	0.008	0.073	0.119	0.030	0.163
	0.5	0.075	0.009	0.076	0.103	0.020	0.138
	1	0.082	0.010	0.083	0.099	0.018	0.130
	2	0.088	0.012	0.088	0.098	0.017	0.128

Table 6: Bias, mean squared error and standard deviation of the MDPD estimator for the three distributions, the two approaches and different values of  $\alpha$  in the shifted strict Pareto contaminated case with  $\varepsilon = 0.005$ ,  $\gamma = 1/4$ ,  $y_c = 1.35$  times the 95% quantile of the uncontaminated distribution.

Distribution $F$	Value of $\alpha$	Oracle			Data Driven		
		Bias	MSE	Sd	Bias	MSE	Sd
$\mathcal{W}(1/\theta(x), 1)$	0.1	0.089	0.014	0.098	0.226	0.093	0.218
	0.25	0.077	0.010	0.085	0.186	0.066	0.194
	0.5	0.076	0.009	0.082	0.154	0.046	0.173
	1	0.081	0.010	0.086	0.140	0.037	0.156
	2	0.087	0.012	0.094	0.141	0.036	0.151
$\mathcal{EW}(1/\theta(x), -1)$	0.1	0.089	0.013	0.099	0.237	0.104	0.233
	0.25	0.080	0.010	0.085	0.195	0.072	0.206
	0.5	0.078	0.010	0.082	0.163	0.051	0.184
	1	0.083	0.011	0.086	0.148	0.041	0.169
	2	0.088	0.012	0.092	0.145	0.039	0.163
$\widetilde{\mathcal{W}}(1/\theta(x), -5)$	0.1	0.089	0.013	0.094	0.183	0.067	0.215
	0.25	0.082	0.010	0.083	0.143	0.039	0.177
	0.5	0.084	0.011	0.085	0.120	0.026	0.157
	1	0.086	0.011	0.088	0.111	0.022	0.145
	2	0.091	0.013	0.093	0.108	0.020	0.139

Table 7: Bias, mean squared error and standard deviation of the MDPD estimator for the three distributions, the two approaches and different values of  $\alpha$  in the shifted strict Pareto contaminated case with  $\varepsilon = 0.01$ ,  $\gamma = 1/4$ ,  $y_c = 1.35$  times the 95% quantile of the uncontaminated distribution.

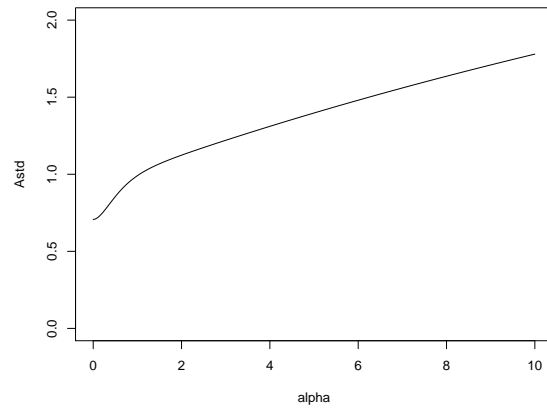


Figure 1: Asymptotic standard deviation of the MDPD estimator for  $\theta_0(x)$  as a function of  $\alpha$  when  $\theta_0(x) = 1$  and  $K(u) = 0.5 \mathbb{1}_{u \in [-1,1]}$ .

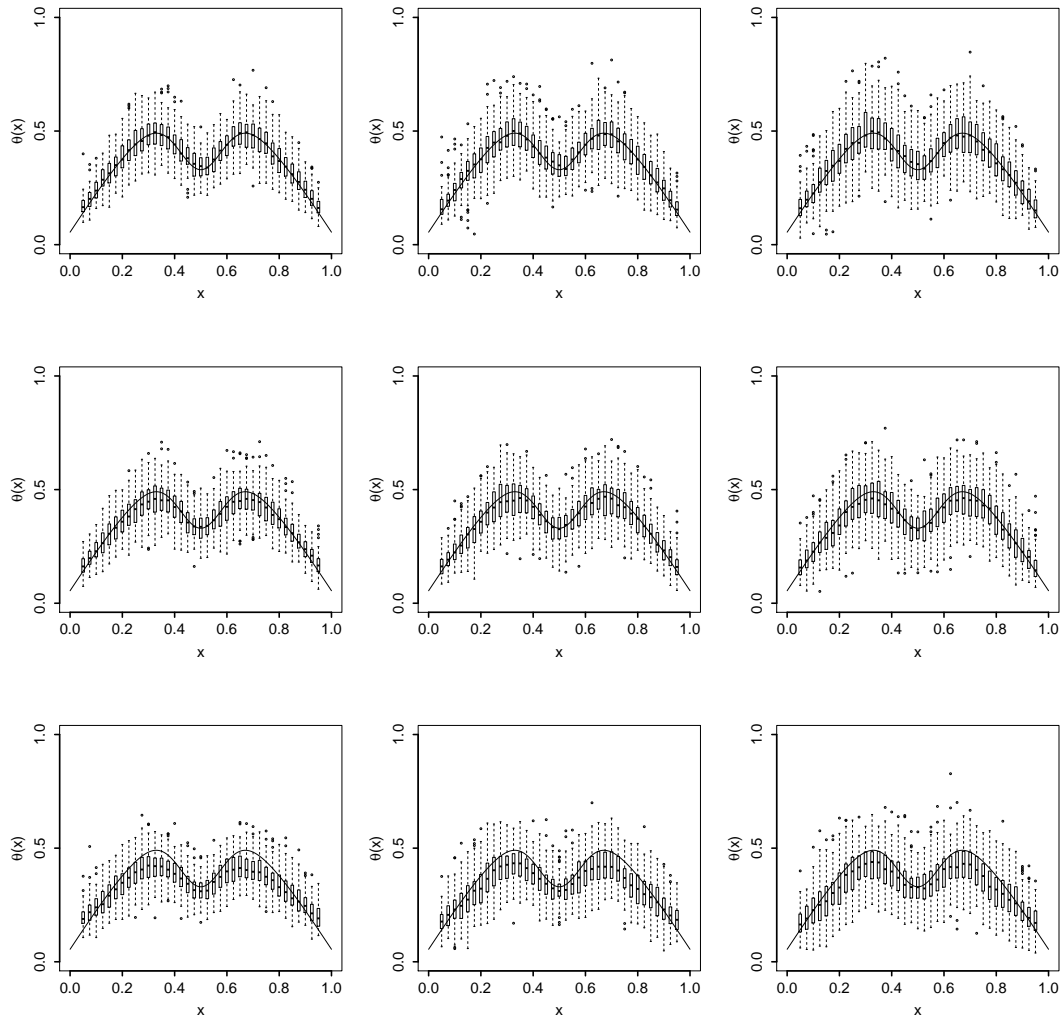


Figure 2: Results for the Oracle method for the three distributions and different values of  $\alpha$ . Row 1: Strict Weibull, row 2: Extended Weibull, row 3: Perturbed Weibull; Column 1:  $\alpha = 0.1$ , column 2:  $\alpha = 0.5$ , column 3:  $\alpha = 1$ .

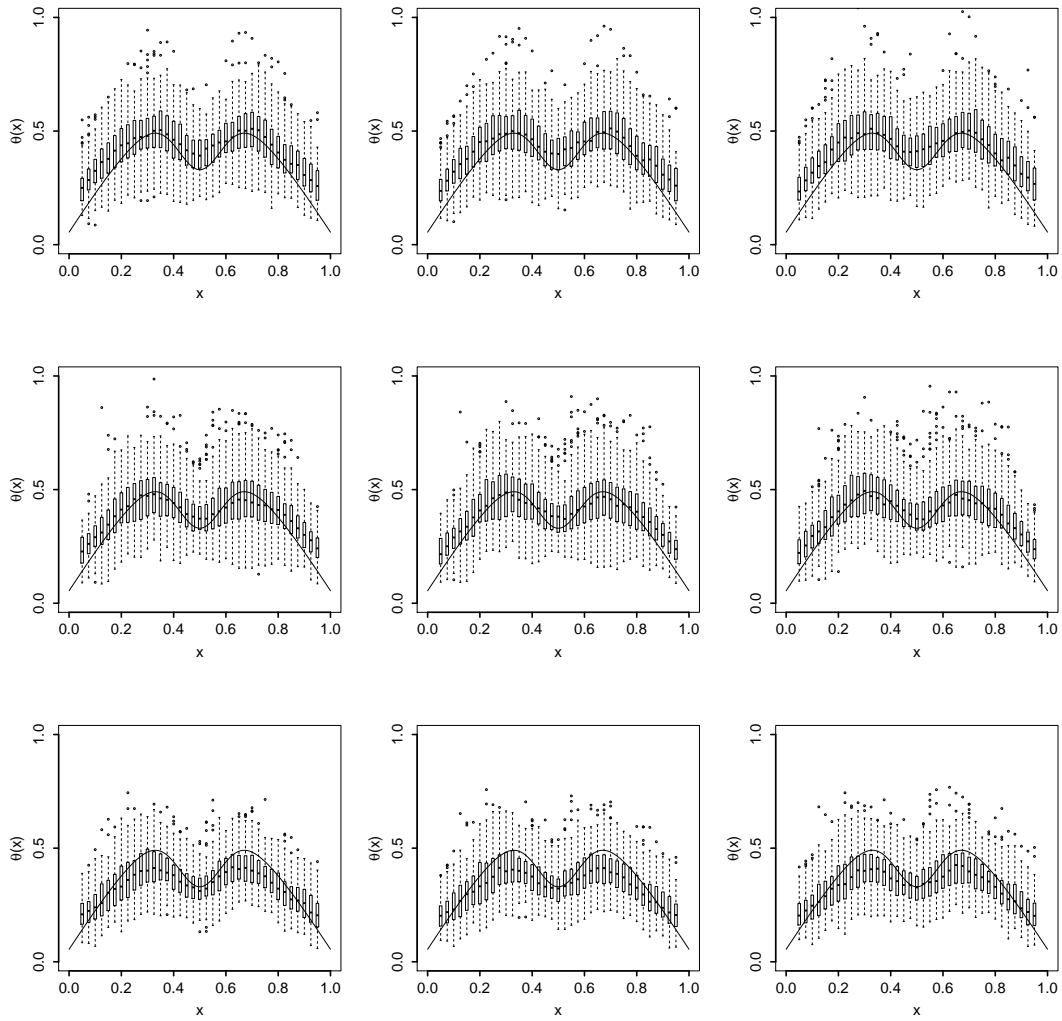


Figure 3: Results for the data driven strategy for the three distributions and different values of  $\alpha$ . Row 1: Strict Weibull, row 2: Extended Weibull, row 3: Perturbed Weibull; Column 1:  $\alpha = 0.1$ , column 2:  $\alpha = 0.5$ , column 3:  $\alpha = 1$ .

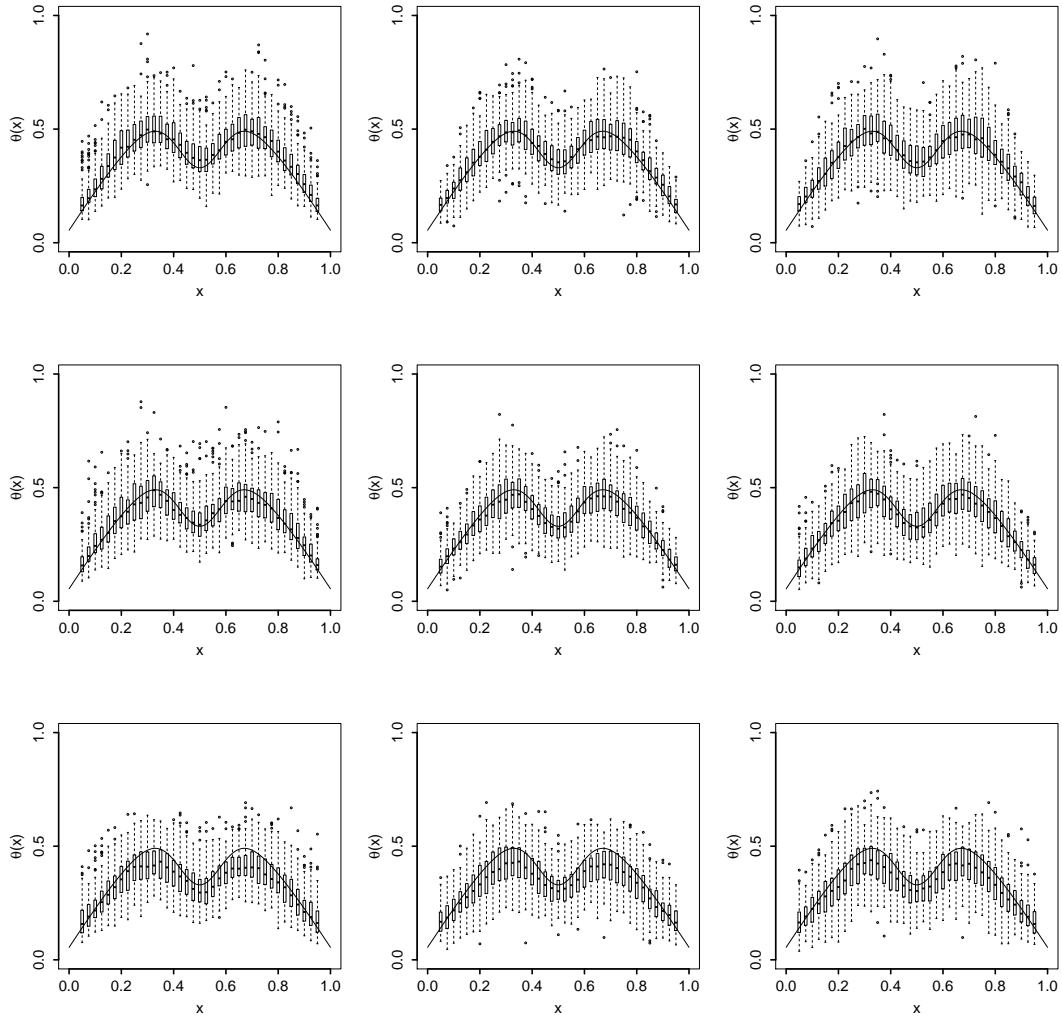


Figure 4: Results for the Oracle method for the three distributions with shifted strict Weibull contamination,  $\varepsilon = 0.005$ ,  $\beta = 4/3$  and different values of  $\alpha$ . Row 1: contaminated Strict Weibull, row 2: contaminated Extended Weibull, row 3: contaminated Perturbed Weibull; Column 1:  $\alpha = 0.1$ , column 2:  $\alpha = 0.5$ , column 3:  $\alpha = 1$ .

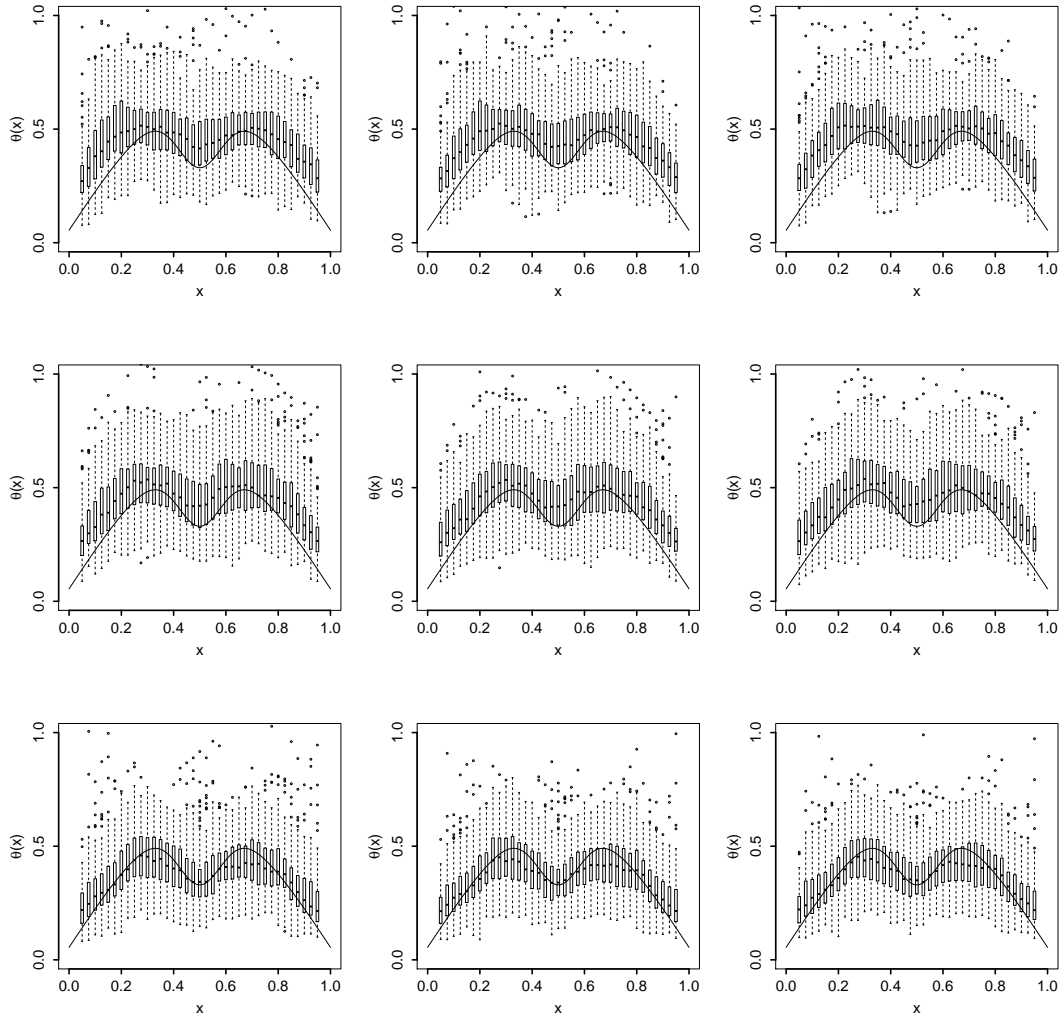


Figure 5: Results for the data driven strategy for the three distributions with shifted strict Weibull contamination,  $\varepsilon = 0.005$ ,  $\beta = 4/3$  and different values of  $\alpha$ . Row 1: contaminated Strict Weibull, row 2: contaminated Extended Weibull, row 3: contaminated Perturbed Weibull; Column 1:  $\alpha = 0.1$ , column 2:  $\alpha = 0.5$ , column 3:  $\alpha = 1$ .



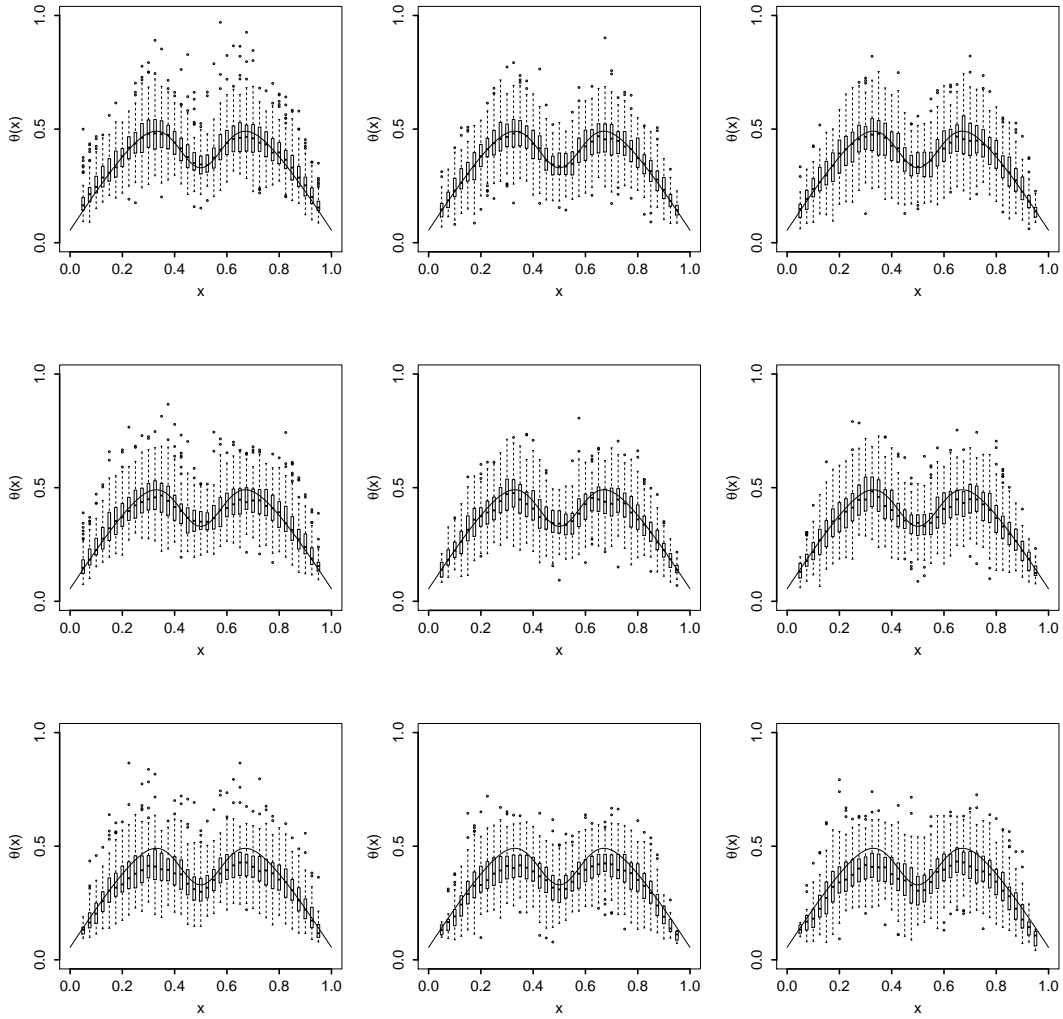


Figure 6: Results for the Oracle method for the three distributions with shifted strict Pareto contamination,  $\varepsilon = 0.005$ ,  $\gamma = 1/4$ ,  $y_c = 1.2$  times the 95% quantile of the uncontaminated distribution and different values of  $\alpha$ . Row 1: contaminated Strict Weibull, row 2: contaminated Extended Weibull, row 3: contaminated Perturbed Weibull; Column 1:  $\alpha = 0.1$ , column 2:  $\alpha = 0.5$ , column 3:  $\alpha = 1$ .

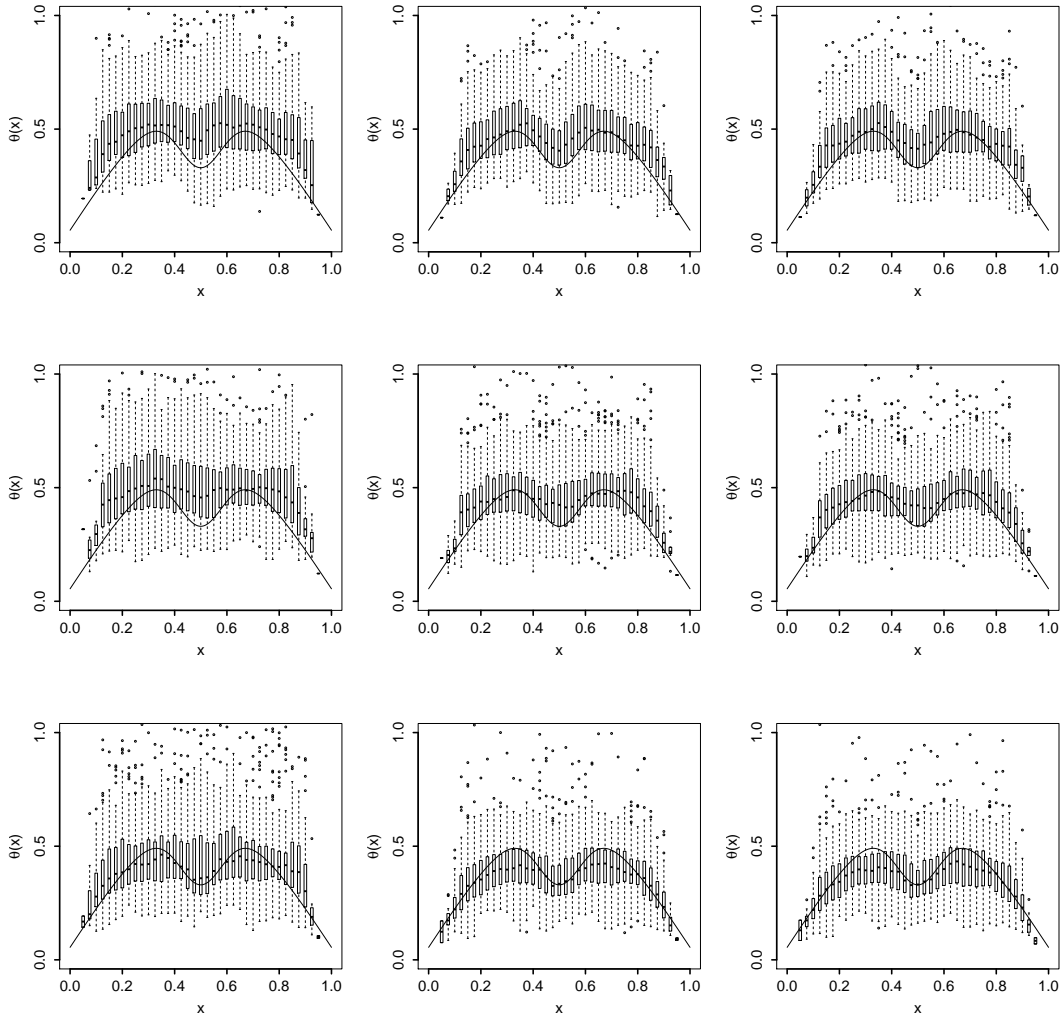


Figure 7: Results for the data driven strategy for the three distributions with shifted strict Pareto contamination,  $\varepsilon = 0.005$ ,  $\gamma = 1/4$ ,  $y_c = 1.2$  times the 95% quantile of the uncontaminated distribution and different values of  $\alpha$ . Row 1: contaminated Strict Weibull, row 2: contaminated Extended Weibull, row 3: contaminated Perturbed Weibull; Column 1:  $\alpha = 0.1$ , column 2:  $\alpha = 0.5$ , column 3:  $\alpha = 1$ .

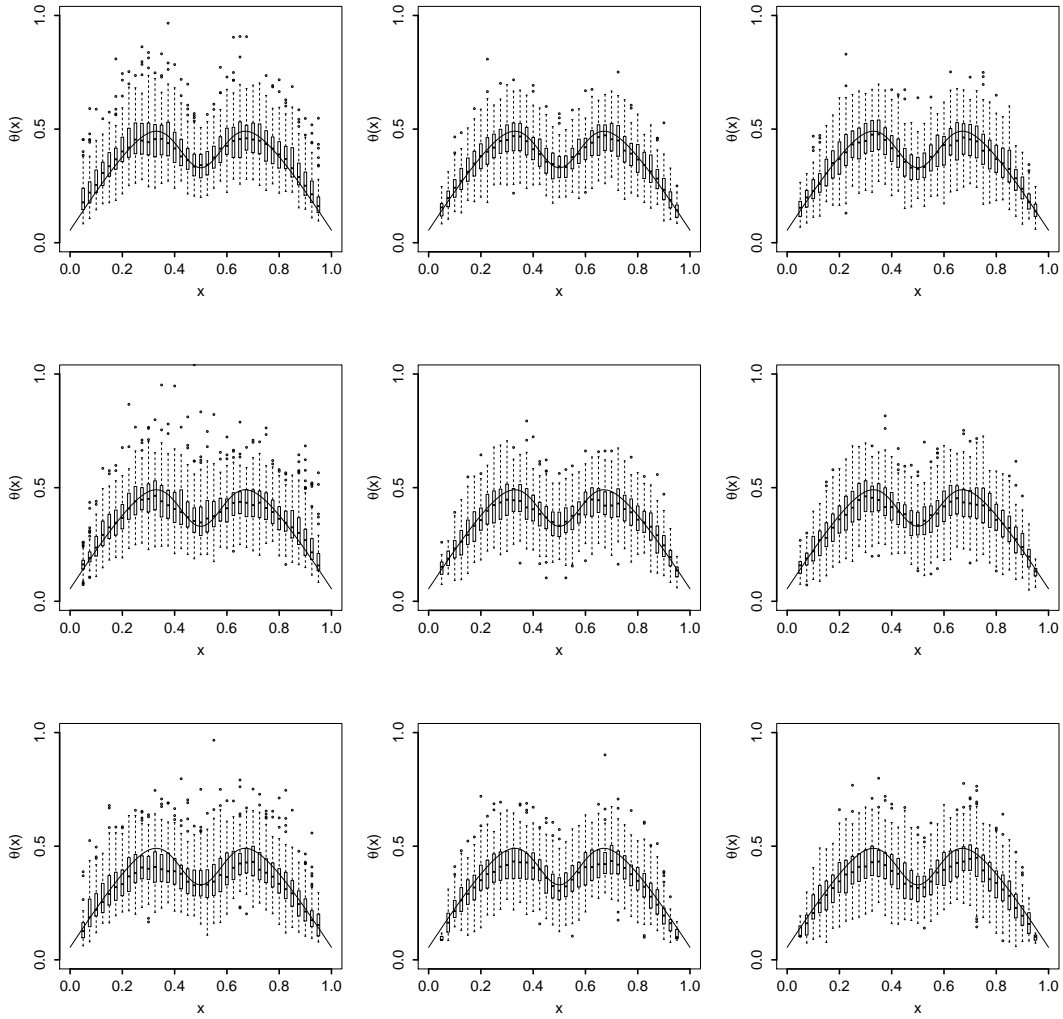


Figure 8: Results for the Oracle method for the three distributions with shifted strict Pareto contamination,  $\varepsilon = 0.005$ ,  $\gamma = 1/4$ ,  $y_c = 1.35$  times the 95% quantile of the uncontaminated distribution and different values of  $\alpha$ . Row 1: contaminated Strict Weibull, row 2: contaminated Extended Weibull, row 3: contaminated Perturbed Weibull; Column 1:  $\alpha = 0.1$ , column 2:  $\alpha = 0.5$ , column 3:  $\alpha = 1$ .

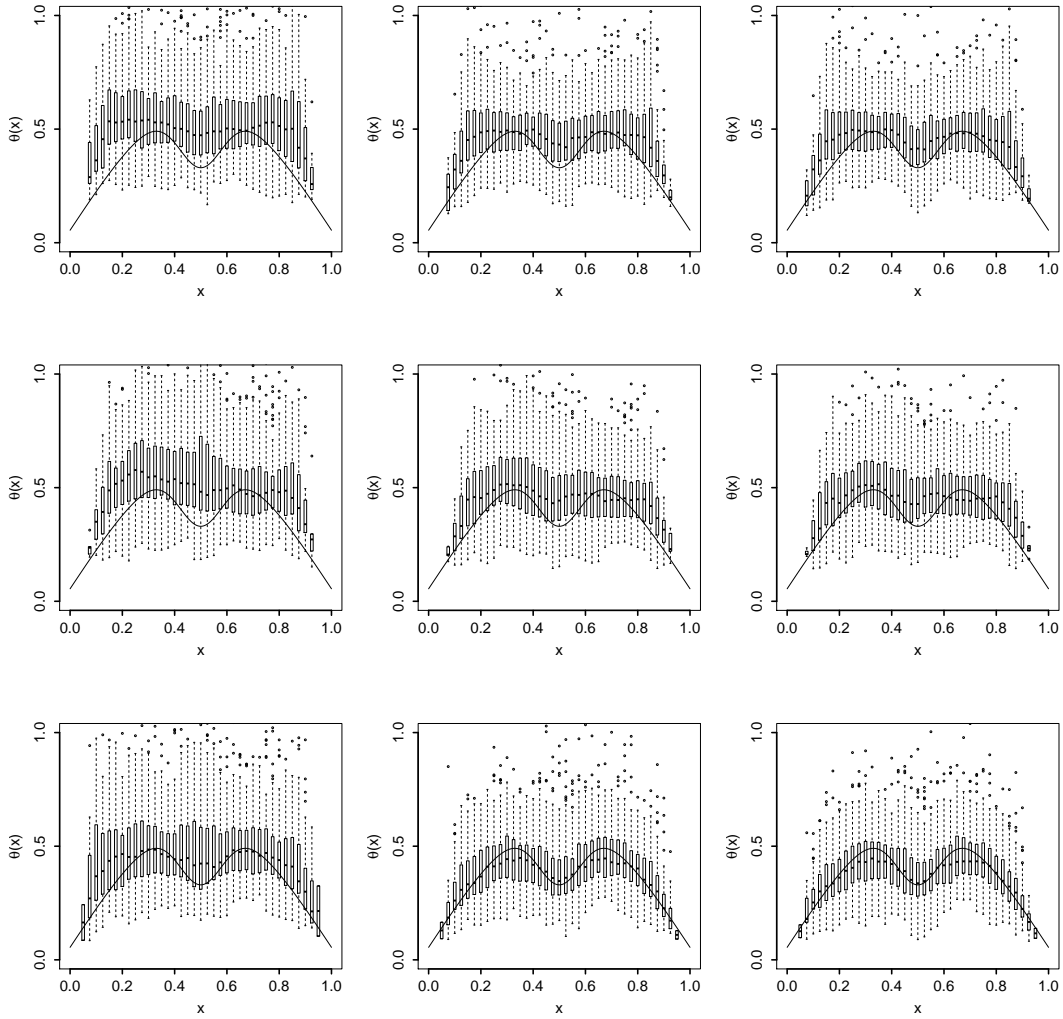


Figure 9: Results for the data driven strategy for the three distributions with shifted strict Pareto contamination,  $\varepsilon = 0.005$ ,  $\gamma = 1/4$ ,  $y_c = 1.35$  times the 95% quantile of the uncontaminated distribution and different values of  $\alpha$ . Row 1: contaminated Strict Weibull, row 2: contaminated Extended Weibull, row 3: contaminated Perturbed Weibull; Column 1:  $\alpha = 0.1$ , column 2:  $\alpha = 0.5$ , column 3:  $\alpha = 1$ .