Local robust estimation of Pareto-type tails with random right censoring

Goedele Dierckx, Yuri Goegebeur, Armelle Guillou

Local robust estimation of Pareto-type tails with random right censoring

Goedele Dierckx [*]
Yuri Goegebeur [†]
Armelle Guillou [‡]

July 3, 2018

Abstract. We propose a nonparametric robust estimator for the tail index of a conditional Pareto-type distribution in the presence of censoring and random covariates. The censored distribution is also of Pareto-type and the index is estimated locally within a narrow neighbourhood of the point of interest in the covariate space using the minimum density power divergence method. The main asymptotic properties of our robust estimator are derived under mild regularity conditions and its finite sample performance is illustrated on a small simulation study. A real data example is included to illustrate the practical applicability of the estimator.

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

Keywords: Pareto-type distribution; random censoring; density power divergence; local estimation.

1 Introduction

Extreme value statistics deals with modelling extreme events, that is, events that have a low frequency of occurrence but a high impact. From a theoretical point of view this implies that the interest is in quantities related to the tail of the distribution like, e.g., indices describing tail decay, extreme quantiles and small tail probabilities, rather than the centre. Estimation of such parameters is then naturally based on the largest observations in the available sample. In many practical applications one encounters also censoring, a situation where only partial information on a random variable is available, typically that it exceeds a certain value. When studying advanced age mortality one often has that some individuals of a birth cohort are still alive at

[*]KU Leuven, Faculty of Economics and Business, campus Brussel, Research Centre for Mathematical Economics, Econometrics and Statistics, Warmoesberg 26, 1000 Brussel, Belgium (email: goedele.dierckx@kuleuven.be).

[†]Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark (email: yuri.goegebeur@imada.sdu.dk).

[‡]Institut Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg et CNRS, 7 rue René Descartes, 67084 Strasbourg cedex, France (email: armelle.guillou@math.unistra.fr).

the time of follow-up, meaning that only a lower bound for their actual lifetime is available. In insurance, long developments of claims are encountered, which means that at the evaluation of a portfolio some claims might not be completely settled and thus the real payments are censored by the payments up to the time of evaluation. In the present paper we will address robust estimation of a tail heaviness parameter under a random right censoring mechanism.

More formally, let Y be the random variable of main interest and let C be the censoring random variable, with distribution functions F_Y and F_C, respectively, both being of Pareto-type, that is, with $\bullet$ denoting either Y or C, one has

$$1 - F_\bullet(y) = y^{-1/\gamma_\bullet} \ell_\bullet(y), \quad y > 0, \tag{1}$$

where $\gamma_\bullet > 0$ is the extreme value index, and $\ell_\bullet$ is a slowly varying function at infinity: $\ell_\bullet(ty)/\ell_\bullet(t) \to 1$ as $t \to \infty$ for all $y > 0$. Interest is in studying properties of the right tail of F_Y, but due to censoring one observes only $Y \wedge C$, rather than Y, together with a censoring indicator $1\!\!1_{\{Y \leq C\}}$, where $1\!\!1_{\{A\}}$ is the indicator function on the event A. Estimation is then based on a random sample of the observables $Y \wedge C$ and $1\!\!1_{\{Y \leq C\}}$, together with some correction which is needed to obtain inference for the tail of F_Y and not that of $F_{Y \wedge C}$. In this univariate context, estimation of tail parameters with random right censoring has been studied quite extensively in the extreme value literature, see for instance Beirlant *et al.* (2007), Einmahl *et al.* (2008), Worms and Worms (2014), Beirlant *et al.* (2016).

In the present paper we extend the above described setup to the case where a random covariate X is available. Model (1) can then be stated as

$$1 - F_\bullet(y|x) = y^{-1/\gamma_\bullet(x)} \ell_\bullet(y|x), \quad y > 0,$$

where $\gamma_\bullet(x) > 0$ is referred to as the conditional extreme value index and $\ell_\bullet(y|x)$ is again a slowly varying function at infinity, and interest is in estimating $\gamma_Y(x)$. Our approach is nonparametric and based on local estimation in the covariate space. Estimation of tail parameters in presence of random covariates has received quite some attention in the recent literature. Wang and Tsai (2009) followed a parametric maximum likelihood approach within the Hall subclass of Pareto-type models (Hall, 1982). Also in the framework of Pareto-type tails, Daouia *et al.* (2011) considered the nonparametric estimation of extreme conditional quantiles, and plugged these conditional quantile estimators into classical estimators for the extreme value index, such as the Hill (1975) and Pickands (1975) estimators. Goegebeur *et al.* (2014b) introduced a nonparametric and asymptotically unbiased estimator for the conditional tail index. Wang *et al.* (2012) considered the estimation of extreme conditional quantiles for Pareto-type distributions and developed a two step procedure based on quantile regression. Daouia *et al.* (2013), extended the methodology of Daouia *et al.* (2011) to the general max-domain of attraction. We also refer to Stupfler (2013) and Goegebeur *et al.* (2014a) for estimation of the conditional tail index in the general max-domain of attraction. Despite these numerous contributions to conditional extremes, the situation of censoring in regression received very little attention. Stupfler (2016) adjusted the local moment estimator introduced in Stupfler (2013) to the censoring context by diving it by the local proportion of non-censored observations. Apart from this pioneering work we are, to the best of our knowledge, not aware of other methods to deal with random covariates and censoring.

Besides censoring and random covariates we also want to develop an estimator that is robust with respect to outliers. To obtain robustness we will use the minimum density power divergence (MDPD) approach, developed by Basu *et al.* (1998). The density power divergence between density functions f and g is given by

$$\Delta_\alpha(f,g) := \begin{cases} \int_{\mathbb{R}} \left[g^{1+\alpha}(y) - \left(1+\frac{1}{\alpha}\right) g^\alpha(y) f(y) + \frac{1}{\alpha} f^{1+\alpha}(y)\right] dy, & \alpha > 0, \\ \int_{\mathbb{R}} \log \frac{f(y)}{g(y)} f(y) dy, & \alpha = 0. \end{cases} \tag{2}$$

For the purpose of estimation, f is assumed to be the true (typically unknown) density of the data, whereas g is a parametric model, depending on a parameter vector θ which is determined by minimizing the empirical version of (2). Applications of this criterion in the context of extremes can be found in, e.g., Kim and Lee (2008), Dierckx *et al.* (2013, 2014), and Escobar-Bach *et al.* (2018).

The remainder of this paper is organized as follows. In the next section we introduce the nonparametric robust estimator in the context of censorship, obtained from local fits of the extended Pareto distribution to the relative excesses over a high threshold. Then, in Section 3, we study its main asymptotic properties under some mild regularity conditions. The finite sample performance of the proposed estimator is evaluated in a small simulation study in Section 4, whereas Section 5 illustrates the applicability of the method on a real dataset. All the proofs are postponed to the Appendix.

2 Construction of the estimator

Let Y denote the response variable of interest, and C be the random right censoring time, both having a distribution depending on a random covariate X, and conditionally on X we assume Y and C to be independent random variables. We observe the random covariate X, $T := Y \wedge C$, and a censoring indicator $\mathbb{1}_{\{Y \leq C\}}$. The covariate X has density function f_X with support $S_X \subset \mathbb{R}^d$, having non-empty interior. We assume the following for the conditional distributions of Y and C given $X = x$, denoted $F_Y(.|x)$ and $F_C(.|x)$, respectively.

Assumption ($\mathcal{D}$) *The conditional survival functions of Y and C satisfy, for all $x \in S_X$, with $\bullet$ denoting either Y or C,*

$$\overline{F}_\bullet(y|x) \quad = \quad A_\bullet(x) y^{-1/\gamma_\bullet(x)} \left(1 + \frac{1}{\gamma_\bullet(x)} \delta_\bullet(y|x)\right),$$

where $A_\bullet(x) > 0$, $\gamma_\bullet(x) > 0$, and $|\delta_\bullet(.|x)|$ is normalised regularly varying with index $\rho_\bullet(x)/\gamma_\bullet(x)$, $\rho_\bullet(x) < 0$, i.e.,

$$\delta_\bullet(y|x) \quad = \quad B_\bullet(x) \exp\left(\int_1^y \frac{\varepsilon_\bullet(u|x)}{u} du\right),$$

with $B_\bullet(x) \in \mathbb{R}$ and $\varepsilon_\bullet(y|x) \to \rho_\bullet(x)/\gamma_\bullet(x)$ as $y \to \infty$. Moreover, we assume $y \to \varepsilon_\bullet(y|x)$ to be a continuous function.

Note that assumption $(\mathcal{D})$ implies that $F_Y(.|x)$ and $F_C(.|x)$ have density functions. Indeed, straightforward differentiation gives

$$f_\bullet(y|x) = \frac{A_\bullet(x)}{\gamma_\bullet(x)} y^{-1/\gamma_\bullet(x)-1}\left[1 + \left(\frac{1}{\gamma_\bullet(x)} - \varepsilon_\bullet(y|x)\right)\delta_\bullet(y|x)\right]. \tag{3}$$

The random variable T has a conditional distribution function satisfying the same properties as those of Y and C. Indeed, by straightforward computations one obtains:

$$\begin{aligned}\overline{F}_T(y|x) &= \overline{F}_Y(y|x)\overline{F}_C(y|x)\\ &= A_T(x)y^{-1/\gamma_T(x)}\left(1 + \frac{1}{\gamma_T(x)}\delta_T(y|x)\right),\end{aligned}$$

where $A_T(x) := A_Y(x)A_C(x)$, $\gamma_T(x) := \gamma_Y(x)\gamma_C(x)/(\gamma_Y(x) + \gamma_C(x))$, and

$$\delta_T(y|x) := \begin{cases} \gamma_T(x)/\gamma_Y(x)\delta_Y(y|x)(1+o(1)) & \text{if} \quad \delta_Y(y|x)/\delta_C(y|x) \to \pm\infty,\\ \gamma_T(x)/\gamma_C(x)\delta_C(y|x)(1+o(1)) & \text{if} \quad \delta_Y(y|x)/\delta_C(y|x) \to 0,\\ \gamma_T(x)\delta_Y(y|x)/\gamma_Y(x)(1+\gamma_Y(x)/(\gamma_C(x)a))(1+o(1)) & \text{if} \quad \delta_Y(y|x)/\delta_C(y|x) \to a,\end{cases}$$

where $0 < |a| < +\infty$.

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

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

and density function

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

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

$$\sup_{z\geq 1}\left|\frac{\overline{F}_\bullet(tz|x)}{\overline{F}_\bullet(t|x)} - \overline{G}(z;\gamma_\bullet(x),\delta_\bullet(t|x),\rho_\bullet(x))\right| = o(\delta_\bullet(t|x)), \qquad \text{if } t \to \infty,$$

where $\bullet$ represents Y, C or T. Clearly, based on this result, one can obtain an estimator for $\gamma_T(x)$ by fitting the extended Pareto distribution to the relative excesses over a high threshold.

Let $(T_i, X_i, \mathbb{1}_{\{Y_i \le C_i\}})$, $i = 1, \ldots, n$, be independent copies of the random vector $(T, X, \mathbb{1}_{\{Y \le C\}})$. Take $x_0 \in Int(S_X)$. We estimate $\gamma_T(x_0)$ by fitting g locally to the relative excesses $Z_i := T_i/t_n$, $i = 1, \ldots, n$, by means of the MDPD criterion, adjusted to locally weighted estimation, i.e., we minimize

$$\widehat{\Delta}_\alpha(\gamma, \delta; \rho) := \\ \frac{1}{n}\sum_{i=1}^n K_{h_n}(x_0 - X_i)\left\{\int_1^\infty g^{1+\alpha}(z;\gamma,\delta,\rho)dz - \left(1+\frac{1}{\alpha}\right)g^\alpha(Z_i;\gamma,\delta,\rho)\right\}\mathbb{1}_{\{T_i > t_n\}}, \quad (5)$$

in case $\alpha > 0$ and

$$\widehat{\Delta}_0(\gamma, \delta; \rho) := -\frac{1}{n}\sum_{i=1}^n K_{h_n}(x_0 - X_i)\ln g(Z_i;\gamma,\delta,\rho)\mathbb{1}_{\{T_i > t_n\}}, \quad (6)$$

in case $\alpha = 0$, where $K_{h_n}(x) := K(x/h_n)/h_n^d$, K is a joint density function on $\mathbb{R}^d$, h_n is a non-random sequence of bandwidths with $h_n \to 0$ if $n \to \infty$, and t_n is a local non-random threshold sequence satisfying $t_n \to \infty$ if $n \to \infty$. Note that in case $\alpha = 0$, the local empirical density power divergence criterion corresponds with a locally weighted log-likelihood function. The parameter α controls the trade-off between efficiency and robustness of the MDPD criterion: the estimator becomes more efficient but less robust as α gets closer to zero, whereas for increasing α the robustness increases and the efficiency decreases. In this paper we only estimate $\gamma_T(x_0)$ and $\delta_T(t_n|x_0)$ with the MDPD method. The parameter ρ will be fixed at some canonical value. Alternatively, one can replace ρ by an external consistent estimator. However, the estimation of ρ in a robust way is still an open problem, and moreover, using an external consistent estimator rather than a canonical value, does not, in general, improve the performance of the final MDPD estimator. For all these reasons, we only use a canonical value for the parameter ρ in the sequel.

3 Asymptotic properties

To deal with the regression context, the functions appearing in $F_Y(y|x)$ and $F_C(y|x)$ are assumed to satisfy the following Hölder conditions. Let $\|.\|$ denote some norm on $\mathbb{R}^d$.

Assumption $(\mathcal{H})$ *There exist positive constants* M_{f_X}, $M_{A_\bullet}$, $M_{\gamma_\bullet}$, $M_{B_\bullet}$, $M_{\varepsilon_\bullet}$, η_{f_X}, $\eta_{A_\bullet}$, $\eta_{\gamma_\bullet}$, $\eta_{B_\bullet}$ *and* $\eta_{\varepsilon_\bullet}$, *such that for all* $x, z \in S_X$*:*

$$\begin{aligned}
|f_X(x) - f_X(z)| &\le M_{f_X}\|x - z\|^{\eta_{f_X}}, \\
|A_\bullet(x) - A_\bullet(z)| &\le M_{A_\bullet}\|x - z\|^{\eta_{A_\bullet}}, \\
|\gamma_\bullet(x) - \gamma_\bullet(z)| &\le M_{\gamma_\bullet}\|x - z\|^{\eta_{\gamma_\bullet}}, \\
|B_\bullet(x) - B_\bullet(z)| &\le M_{B_\bullet}\|x - z\|^{\eta_{B_\bullet}}, \\
\sup_{y \ge 1}|\varepsilon_\bullet(y|x) - \varepsilon_\bullet(y|z)| &\le M_{\varepsilon_\bullet}\|x - z\|^{\eta_{\varepsilon_\bullet}}.
\end{aligned}$$

For the kernel function K we assume the following:

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

Set $\ln_+ x = \ln\max\{1, x\}$, $x > 0$, and consider, with $s \leq 0$ and $s' \geq 0$,

$$\begin{aligned} T_n^{(1)}(K, s, s'|x_0) &:= \frac{1}{n}\sum_{i=1}^n K_{h_n}(x_0 - X_i)\left(\frac{T_i}{t_n}\right)^s \left(\ln_+ \frac{T_i}{t_n}\right)^{s'} 1\!\!1_{\{T_i > t_n\}}, \\ T_n^{(2)}(K|x_0) &:= \frac{1}{n}\sum_{i=1}^n K_{h_n}(x_0 - X_i) 1\!\!1_{\{Y_i \leq C_i, T_i > t_n\}}. \end{aligned}$$

Statistics of the type $T_n^{(1)}(K, s, s'|x_0)$ are the basic building blocks for studying the asymptotic behaviour of the estimator for $\gamma_T(x_0)$. Indeed, the estimating equations resulting from (5) and (6) depend only on statistics of this type. The statistic $T_n^{(2)}(K|x_0)$ will lead to an estimator for the proportion of non-censored observations among the exceedances over t_n, which is used to correct the estimator for $\gamma_T(x_0)$ in order obtain an estimator for $\gamma_Y(x_0)$, being the quantity of main interest.

As a first step we establish the asymptotic expansions of $\mathbb{E}(T_n^{(1)}(K, s, s'|x))$ and $\mathbb{E}(T_n^{(2)}(K|x))$. Let $\eta := \eta_{\gamma_Y} \wedge \eta_{\gamma_C} \wedge \eta_{\varepsilon_Y} \wedge \eta_{\varepsilon_C}$.

Theorem 1 *Assume $(\mathcal{D})$, $(\mathcal{H})$, $(\mathcal{K})$ and $x_0 \in Int(S_X)$ with $f_X(x_0) > 0$. Then if $t_n \to \infty$ and $h_n \to 0$ as $n \to \infty$ in such a way that $h_n^\eta \ln t_n \to 0$, we have the following asymptotic expansions*

$$\begin{aligned} \mathbb{E}(T_n^{(1)}(K, s, s'|x_0)) &= \overline{F}_T(t_n|x_0) f_X(x_0)\gamma_T^{s'}(x_0)\Gamma(s'+1)\left\{\frac{1}{(1 - s\gamma_T(x_0))^{s'+1}}\right. \\ &\quad -\frac{\delta_T(t_n|x_0)}{\gamma_T(x_0)}\left[\frac{1}{(1 - s\gamma_T(x_0))^{s'+1}} - \frac{1 - \rho_T(x_0)}{(1 - \rho_T(x_0) - s\gamma_T(x_0))^{s'+1}}\right](1 + o(1)) \\ &\quad \left. + O(h_n^{\eta_{f_X} \wedge \eta_{A_Y} \wedge \eta_{A_C}}) + O(h_n^{\eta_{\gamma_Y} \wedge \eta_{\gamma_C}} \ln t_n)\right\}, \end{aligned}$$

where the $o(1)$ and $O(.)$ terms are uniform in $(s, s') \in [S, 0] \times [0, S']$, with $S < 0$ and $S' > 0$, and

$$\begin{aligned} \mathbb{E}(T_n^{(2)}(K|x_0)) &= \overline{F}_T(t_n|x_0) f_X(x_0)\frac{\gamma_T(x_0)}{\gamma_Y(x_0)}\left\{1 + \delta_C(t_n|x_0)\frac{\gamma_T(x_0)\rho_C(x_0)}{\gamma_C(x_0)(\gamma_C(x_0) - \gamma_T(x_0)\rho_C(x_0))}(1 + o(1))\right. \\ &\quad -\delta_Y(t_n|x_0)\frac{\rho_Y(x_0)(\gamma_Y(x_0) - \gamma_T(x_0))}{\gamma_Y(x_0)(\gamma_Y(x_0) - \gamma_T(x_0)\rho_Y(x_0))}(1 + o(1)) \\ &\quad \left. + O(h_n^{\eta_{f_X} \wedge \eta_{A_Y} \wedge \eta_{A_C}}) + O(h_n^{\eta_{\gamma_Y} \wedge \eta_{\gamma_C}} \ln t_n)\right\}. \end{aligned}$$

We now turn to establishing the joint convergence of $T_n^{(1)}(K, s, s'|x_0)$ for several values of (s, s')

and $T_n^{(2)}(K|x_0)$. Let

$$\mathbb{T}_n := \frac{1}{\overline{F}_T(t_n|x_0)f_X(x_0)}\begin{bmatrix} T_n^{(1)}(K,s_1,s_1'|x_0) \\ \vdots \\ T_n^{(1)}(K,s_J,s_J'|x_0) \\ T_n^{(2)}(K|x_0) \end{bmatrix},$$

$r_n := \sqrt{nh^d\overline{F}_T(t_n|x_0)f_X(x_0)}$, and '$\rightsquigarrow$' denoting convergence in distribution.

Theorem 2 *Assume ($\mathcal{D}$), ($\mathcal{H}$), ($\mathcal{K}$) and $x_0 \in Int(S_X)$ with $f_X(x_0) > 0$. Then if $t_n \to \infty$ and $h_n \to 0$ as $n \to \infty$ in such a way that $h_n^\eta \ln t_n \to 0$ and $r_n \to \infty$, then*

$$r_n(\mathbb{T}_n - \mathbb{E}(\mathbb{T}_n)) \rightsquigarrow N(0,\Sigma),$$

with

$$\begin{aligned}
\Sigma_{j,k} &:= \frac{\gamma_T^{s_j'+s_k'}(x_0)\Gamma(s_j'+s_k'+1)\|K\|_2^2}{(1-(s_j+s_k)\gamma_T(x_0))^{s_j'+s_k'+1}}, \qquad j,k\in\{1,\ldots,J\},\\
\Sigma_{J+1,J+1} &:= \frac{\gamma_T(x_0)\|K\|_2^2}{\gamma_Y(x_0)},\\
\Sigma_{J+1,j} &:= \frac{\gamma_T^{s_j'+1}(x_0)\Gamma(s_j'+1)\|K\|_2^2}{\gamma_Y(x_0)(1-s_j\gamma_T(x_0))^{s_j'+1}}, \qquad j\in\{1,\ldots,J\}.
\end{aligned}$$

We now establish the joint weak convergence of the statistics $T_n^{(1)}(K,s,j|x_0)$ as stochastic processes in $s \in [S,0]$, with $j = 0,1,2,3$, and $T_n^{(2)}(K|x_0)$. To this aim, let

$$\begin{aligned}
\mathbb{S}_n^{(j)} &:= \left\{ r_n\left[\frac{T_n^{(1)}(K,s,j|x_0)}{\overline{F}(t_n|x_0)f_X(x_0)} - \frac{j!\gamma_T^j(x_0)}{(1-s\gamma_T(x_0))^{j+1}},\right]; s\in[S,0]\right\}, \qquad j\in\{0,1,2,3\},\\
\mathbb{S}_n^{(4)} &:= r_n\left[\frac{T_n^{(2)}(K|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)} - \frac{\gamma_T(x_0)}{\gamma_Y(x_0)}\right].
\end{aligned}$$

Theorem 3 *Under the assumptions of Theorem 2 and assuming additionally that $r_n\delta_T(t_n|x_0) \to \lambda \in \mathbb{R}$, $r_nh^{\eta_{f_X}\wedge\eta_{A_Y}\wedge\eta_{A_C}} \to 0$ and $r_nh^{\eta_{\gamma_Y}\wedge\eta_{\gamma_C}}\ln t_n \to 0$ as $n \to \infty$, we have*

$$(\mathbb{S}_n^{(0)},\mathbb{S}_n^{(1)},\mathbb{S}_n^{(2)},\mathbb{S}_n^{(3)},\mathbb{S}_n^{(4)}) \rightsquigarrow (\mathbb{S}^{(0)},\mathbb{S}^{(1)},\mathbb{S}^{(2)},\mathbb{S}^{(3)},\mathbb{S}^{(4)}),$$

where $\mathbb{S}^{(0)}$, $\mathbb{S}^{(1)}$, $\mathbb{S}^{(2)}$ and $\mathbb{S}^{(3)}$ are tight Gaussian processes on $[S,0]$ and $\mathbb{S}_n^{(4)}$ is a Gaussian random variable, where, with $s \in [S,0]$,

$$\mathbb{E}(\mathbb{S}^{(j)}(s)) = -\lambda j!\gamma_T^{j-1}(x_0)\left[\frac{1}{(1-s\gamma_T(x_0))^{j+1}} - \frac{1-\rho_T(x_0)}{(1-\rho_T(x_0)-s\gamma_T(x_0))^{j+1}}\right],\ j\in\{0,1,2,3\},$$

$$\mathbb{E}(\mathbb{S}^{(4)}) = \lambda\times\begin{cases} -\frac{\rho_Y(x_0)(\gamma_Y(x_0)-\gamma_T(x_0))}{\gamma_Y(x_0)(\gamma_Y(x_0)-\gamma_T(x_0)\rho_Y(x_0))}, & \text{if } \delta_Y(y|x)/\delta_C(y|x)\to\pm\infty,\\ \frac{\gamma_T(x_0)\rho_C(x_0)}{\gamma_Y(x_0)(\gamma_C(x_0)-\gamma_T(x_0)\rho_C(x_0))}, & \text{if } \delta_Y(y|x)/\delta_C(y|x)\to 0,\\ \frac{a\gamma_C(x_0)}{a\gamma_C(x_0)+\gamma_Y(x_0)}\left[\frac{\gamma_T(x_0)\rho_C(x_0)}{a\gamma_C(x_0)(\gamma_C(x_0)-\gamma_T(x_0)\rho_C(x_0))}\right. & \\ \qquad\left. - \frac{\rho_Y(x_0)(\gamma_Y(x_0)-\gamma_T(x_0))}{\gamma_Y(x_0)(\gamma_Y(x_0)-\gamma_T(x_0)\rho_Y(x_0))}\right] & \text{if } \delta_Y(y|x)/\delta_C(y|x)\to a, \end{cases} \tag{7}$$

and, with $s_1, s_2 \in [S, 0]$,

$$\begin{aligned}
\mathbb{C}ov(\mathbb{S}^{(j)}(s_1), \mathbb{S}^{(k)}(s_2)) &= \frac{(j+k)!\gamma_T^{j+k}(x_0)\|K\|_2^2}{(1-(s_1+s_2)\gamma_T(x_0))^{j+k+1}},\ j,k \in \{0,1,2,3\}, \\
\mathbb{C}ov(\mathbb{S}^{(j)}(s), \mathbb{S}^{(4)}) &= \frac{j!\gamma_T^{j+1}(x_0)\|K\|_2^2}{\gamma_Y(x_0)(1-s\gamma_T(x_0))^{j+1}},\ j \in \{0,1,2,3\}, \\
\mathbb{V}ar(\mathbb{S}^{(4)}) &= \frac{\gamma_T(x_0)\|K\|_2^2}{\gamma_Y(x_0)}.
\end{aligned}$$

For the sequel, we denote by $\gamma_T^{(0)}(x_0)$ and $\delta_T^{(0)}(t_n|x_0)$ the true value of $\gamma_T(x_0)$ and $\delta_T(t_n|x_0)$, respectively. Let $\widehat{\gamma}_{T,n}(x_0|\tilde{\rho})$ and $\widehat{\delta}_{T,n}(x_0|\tilde{\rho})$ be the corresponding MDPD estimators, obtained from solving the following estimating equations, with the second order parameter ρ fixed at the value $\tilde{\rho}$:

$$\begin{aligned}
0 &= \frac{1}{n}\sum_{i=1}^n K_{h_n}(x_0 - X_i)\mathbb{1}_{\{T_i>t_n\}} \int_1^\infty g^\alpha(z;\gamma,\delta,\tilde{\rho})\frac{\partial g(z;\gamma,\delta,\tilde{\rho})}{\partial\gamma}dz \\
&\quad -\frac{1}{n}\sum_{i=1}^n K_{h_n}(x_0 - X_i)g^{\alpha-1}(Z_i;\gamma,\delta,\tilde{\rho})\frac{\partial g(Z_i;\gamma,\delta,\tilde{\rho})}{\partial\gamma}\mathbb{1}_{\{T_i>t_n\}}, \qquad (8) \\
0 &= \frac{1}{n}\sum_{i=1}^n K_{h_n}(x_0 - X_i)\mathbb{1}_{\{T_i>t_n\}} \int_1^\infty g^\alpha(z;\gamma,\delta,\tilde{\rho})\frac{\partial g(z;\gamma,\delta,\tilde{\rho})}{\partial\delta}dz \\
&\quad -\frac{1}{n}\sum_{i=1}^n K_{h_n}(x_0 - X_i)g^{\alpha-1}(Z_i;\gamma,\delta,\tilde{\rho})\frac{\partial g(Z_i;\gamma,\delta,\tilde{\rho})}{\partial\delta}\mathbb{1}_{\{T_i>t_n\}}. \qquad (9)
\end{aligned}$$

Define

$$\widehat{p}_n(x_0) := \frac{T_n^{(2)}(K|x_0)}{T_n^{(1)}(K,0,0|x_0)}.$$

From Theorem 1 it is clear that $\widehat{p}_n(x_0)$ estimates $\gamma_T^{(0)}(x_0)/\gamma_Y^{(0)}(x_0)$, which motivates $\widehat{\gamma}_{Y,n}(x_0|\tilde{\rho}) := \widehat{\gamma}_{T,n}(x_0|\tilde{\rho})/\widehat{p}_n(x_0)$ as estimator for $\gamma_Y^{(0)}(x_0)$. The consistency of this estimator is formalised in the next theorem. Let '$\xrightarrow{P}$' denote convergence in probability.

Theorem 4 *Under the assumptions of Theorem 2 we have, for* $n \to \infty$,

$$(\widehat{\gamma}_{T,n}(x_0|\tilde{\rho}), \widehat{\delta}_{T,n}(x_0|\tilde{\rho}), \widehat{p}_n(x_0)) \xrightarrow{P} (\gamma_T^{(0)}(x_0), 0, \gamma_T^{(0)}(x_0)/\gamma_Y^{(0)}(x_0)),$$

and hence $\widehat{\gamma}_{Y,n}(x_0|\tilde{\rho}) \xrightarrow{P} \gamma_Y^{(0)}(x_0)$.

Finally, we obtain the asymptotic normality of $\widehat{\gamma}_{Y,n}(x_0;\tilde{\rho})$, when properly normalised.

Theorem 5 *Under the assumptions of Theorem 3 we have for $n \to \infty$*

$$r_n(\widehat{\gamma}_{Y,n}(x_0|\tilde{\rho}) - \gamma_Y^{(0)}(x_0)) \rightsquigarrow N(\lambda L^T \check{\Delta}^{-1}(\tilde{\rho})B(\tilde{\rho})\mu(\tilde{\rho}), L^T \check{\Delta}^{-1}(\tilde{\rho})B(\tilde{\rho})\Sigma(\tilde{\rho})B(\tilde{\rho})^T \check{\Delta}^{-1}(\tilde{\rho})L),$$

where the precise definitions of L, $\check{\Delta}(\tilde{\rho})$, $B(\tilde{\rho})$, $\mu(\tilde{\rho})$ and $\Sigma(\tilde{\rho})$ are given in the proof of the theorem in the appendix.

The estimator $\widehat{\gamma}_{Y,n}(x_0;\tilde{\rho})$ depends on $\widehat{\gamma}_{T,n}(x_0|\tilde{\rho})$ and $\widehat{p}_n(x_0)$. In Dierckx *et al.* (2014) it was shown that when $\tilde{\rho}$ is correctly specified, then $\widehat{\gamma}_{T,n}(x_0|\tilde{\rho})$ is asymptotically unbiased in the sense that the mean of the normal limiting distribution is zero. In case $\tilde{\rho}$ is mis-specified then the mean of the limiting distribution of $\widehat{\gamma}_{T,n}(x_0|\tilde{\rho})$ is not necessarily zero, but, being a second order estimator, it continues to perform well compared to estimators that are not corrected for bias, as observed in Dierckx *et al.* (2014). Also $\widehat{p}_n(x_0)$ (after proper normalisation) has a limiting distribution with a mean that is not necessarily equal to zero, as is common in extreme value statistics. At the theoretical level one can thus expect a non-zero mean of the limiting distribution of $\widehat{\gamma}_{Y,n}(x_0;\tilde{\rho})$, but despite this, the proposed estimator performs well in practice. Also, it can tolerate outliers and high proportions of censoring, as is illustrated in the simulation experiment.

4 A simulation study

In this section, we illustrate the performance of the proposed MDPD estimator $\widehat{\gamma}_{Y,n}(x_0|\tilde{\rho})$ with a small simulation study. To this aim, we need first to choose the function K and the canonical value $\tilde{\rho}$. Then, we have to select the bandwidth parameter h_n and the threshold t_n. Concerning α, three values will be tried: $\alpha = 0$ corresponding to a local maximum likelihood estimator, $\alpha = 0.1$ and $\alpha = 0.5$.

Regarding the value $\tilde{\rho}$, according to Beirlant *et al.* (2016) in the uncontaminated framework, $\tilde{\rho} = -0.5$ is a suitable value which allows to stabilize the estimators adapted to censorship as a function of k, the number of top order statistics used in the estimation. Thus this value will be also used in our context. As for the kernel function K, we take the bi-quadratic function

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

For the selection of h_n and t_n, we proceed as follows. For each dataset, an optimal bandwidth $h_{n,o}$ is selected using the leave-one-out cross validation method, already used in the extreme value literature, see, e.g., Daouia *et al.* (2011, 2013) and Goegebeur *et al.* (2014a). Once this optimal bandwidth is selected an optimal value for t_n has to be determined for every x_0. As usual in extreme value statistics, t_n is selected as the $(k+1)$-th largest response observation in the ball $B(x_0, h_n)$, where the optimal k-value is obtained for all x_0 under consideration by the following algorithm:

- We compute $\widehat{\gamma}_{T,n}(x_0|\tilde{\rho})$ with $k = 5, 9, 13, \ldots, m_{x_0}-4$ (m_{x_0} being the number of observations in the ball $B(x_0, h_{n,o})$) by minimizing (5) or (6) . The minimization is carried out with the numerical minimization procedure described in Byrd *et al.* (1995) (R function `optim`, with `method = "L-BFGS-B"`). This method is a quasi Newton method adapted to allow for the constraint $\gamma_T(x_0) > 0$. Also, in order to avoid unstable estimates at some values of k we added some smoothness condition linking estimates at subsequent values of k: $|\widehat{\delta}_{T,n}(x_0|\tilde{\rho})|$ is at most 5 percent larger for k compared to $k+1$;

- we deduce $\widehat{\gamma}_{Y,n}(x_0|\tilde{\rho})$;
- we split the range of k into several blocks of same size 40;
- we calculate the standard deviation of the estimates of $\gamma_Y(x_0)$ in each block;
- the block with minimal standard deviation determines the $k_{n,o}(x_0)$ to be used.

Note that in this procedure, h_n and k are selected separately. One could also determine the two parameters simultaneously, as was attempted in, e.g., Daouia *et al.* (2013). However, as reported in that paper, this is not without problems, and in practice it does not perform better than with a separate selection.

We simulated $N = 100$ samples of size $n = 1\,000$, with $X \sim U(0,1)$ and $Y|X = x$ is generated from the following Burr distribution

$$1 - F_Y(y;x) = \left(1 + y^{-\rho(x)/\gamma_Y(x)}\right)^{1/\rho(x)}, \qquad y > 0.$$

We set here

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

This function γ_Y was also used in Daouia *et al.* (2011) and Goegebeur *et al.* (2014b).

In case of censoring, the data are censored using a Burr distribution also with $\rho(x) = -1$, but with an index $\gamma_C(x)$ set at two values, 0.75 and 0.5, respectively.

On top of the censoring, contamination is added in the response variable according to the following mixture of distribution functions:

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

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

The following settings were considered:

- Setting 1: censoring with $\gamma_C(x) = 0.75$

  uncontaminated situation;

  $\epsilon = 0.01$, x_c= 1.2 times the 99.99% quantile of $F_T(y;x)$;

  $\epsilon = 0.05$, x_c= 1.2 times the 99.99% quantile of $F_T(y;x)$;

- Setting 2: censoring with $\gamma_C(x) = 0.5$

  uncontaminated situation;

  $\epsilon = 0.01$, x_c= 1.2 times the 99.99% quantile of $F_T(y;x)$;

  $\epsilon = 0.05$, x_c= 1.2 times the 99.99% quantile of $F_T(y;x)$.

In the simulation experiment, h_n was selected from the range 0.1 until 0.2

Figure 1 and Table 1 display the results for Setting 1. In particular, in Figure 1, we show the boxplots of $\widehat{\gamma}_{Y,n}(x_0|\tilde{\rho})$ for the three values of α (corresponding to the three columns of the figure) and the three percentages of contamination (corresponding to the three rows of the figure). The full line represents the true value of γ_Y. In the uncontaminated situation (first row), the non-robust estimator ($\alpha = 0$) and the robust estimators (with $\alpha = 0.1$ and $\alpha = 0.5$) perform similarly and capture the sine behaviour of the function γ_Y quite well, although the variance becomes slightly larger as α increases, as expected. Indeed in case of no-contamination, maximum likelihood approaches (corresponding to $\alpha = 0$) are efficient. On the contrary, in case of contamination (rows 2 and 3), a larger value of α is required in order to capture the sine behaviour of the function γ_Y without too much bias. Indeed, in that context, a small value of α (0 or 0.1) leads to estimators $\widehat{\gamma}_{Y,n}(x_0|\tilde{\rho})$ with a considerable bias and variance. Also, as expected, increasing the percentage of contamination deteriorates the estimation.

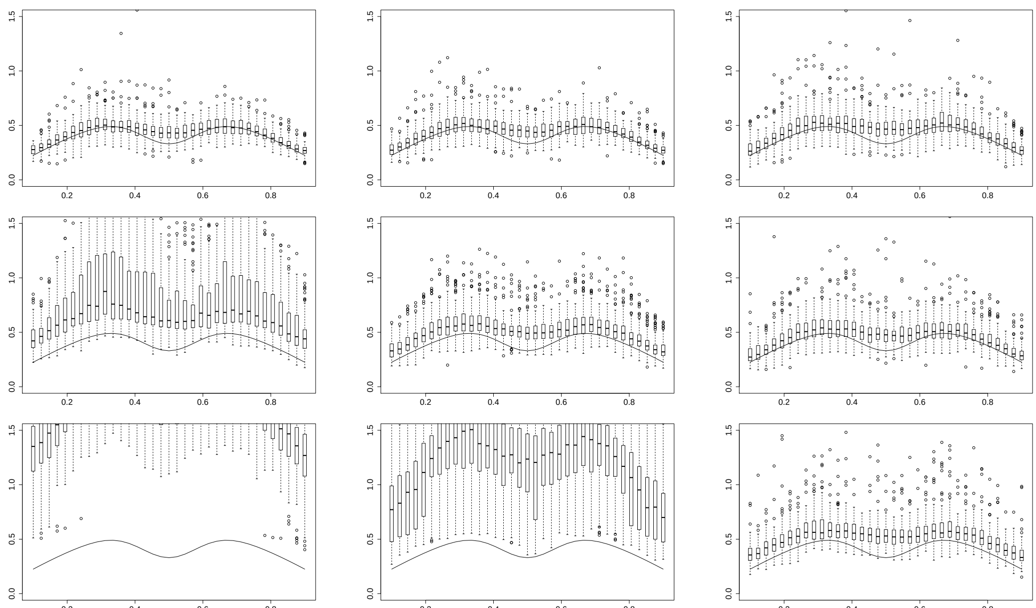

Figure 1: Setting 1: boxplots of $\widehat{\gamma}_{Y,n}(x_0|\tilde{\rho})$, from the left to the right: $\alpha = 0$, $\alpha = 0.1$ and $\alpha = 0.5$; from the top to the bottom: no-contamination, $\epsilon = 0.01$ and $\epsilon = 0.05$. The full line represents the true function γ_Y.

We complete the graphical representations by an indicator of efficiency, called MSE in Table 1, obtained by computing the average over the 100 simulated datasets of

$$\frac{1}{M}\sum_{m=1}^{M}\left(\widehat{\gamma}_{Y,n}(z_m|\tilde{\rho})-\gamma_Y(z_m)\right)^2,$$

where $z_1,\ldots,z_M$ are regularly spaced in the covariate space. Here M is set at the value 35. Table 1 corroborates the conclusions derived from Figure 1, namely that the value $\alpha=0.5$ is necessary in case of contamination to obtain good estimates of the true value $\gamma_Y(x_0)$ and this choice does not deteriorate too much the estimation in case of no-contamination.

% of contamination	non robust ($\alpha=0$)	$\alpha=0.1$	$\alpha=0.5$
0%	0.011	0.011	0.020
1%	0.195	0.031	0.021
5%	2.090	0.818	0.038

Table 1: Setting 1: MSE for $\widehat{\gamma}_{Y,n}(x_0|\tilde{\rho})$.

Figure 2 and Table 2 are constructed similarly as Figure 1 and Table 1 but in case of Setting 2. The conclusions are exactly the same, although the estimation of $\gamma_Y(x_0)$ is more difficult in Figure 2 than in Figure 1. This can be explained by the fact that the tail of the censoring distribution is less heavy in Setting 2 than in Setting 1 leading to more extremes censored, as illustrated in Figure 3 where the theoretical asymptotic proportion of censoring, namely $\gamma_Y(x)/(\gamma_Y(x)+\gamma_C(x))$, is plotted. As a result, MSE values in Table 2 are somewhat larger than in Table 1.

% of contamination	non robust ($\alpha=0$)	$\alpha=0.1$	$\alpha=0.5$
0%	0.012	0.015	0.022
1%	0.279	0.036	0.027
5%	2.718	0.707	0.043

Table 2: Setting 2: MSE for $\widehat{\gamma}_{Y,n}(x_0|\tilde{\rho})$.

5 Practical example

In this section, we illustrate our methodology with the Australian AIDS survival dataset before 1 July 1991, coming from Dr P.J. Solomon and the Australian National Centre in HIV Epidemiology and Clinical Research; see Venables and Ripley (2002). This dataset `aids2` is available in the R package `MASS`, and has recently been considered several times in the extreme value literature, in particular by Einmahl *et al.* (2008) but without taking care of the covariates, and by Ndao *et al.* (2014), Stupfler (2016) and Goegebeur *et al.* (2018) in a regression context. However, up to now, the question of whether this dataset contains outliers or not, has not yet been addressed.

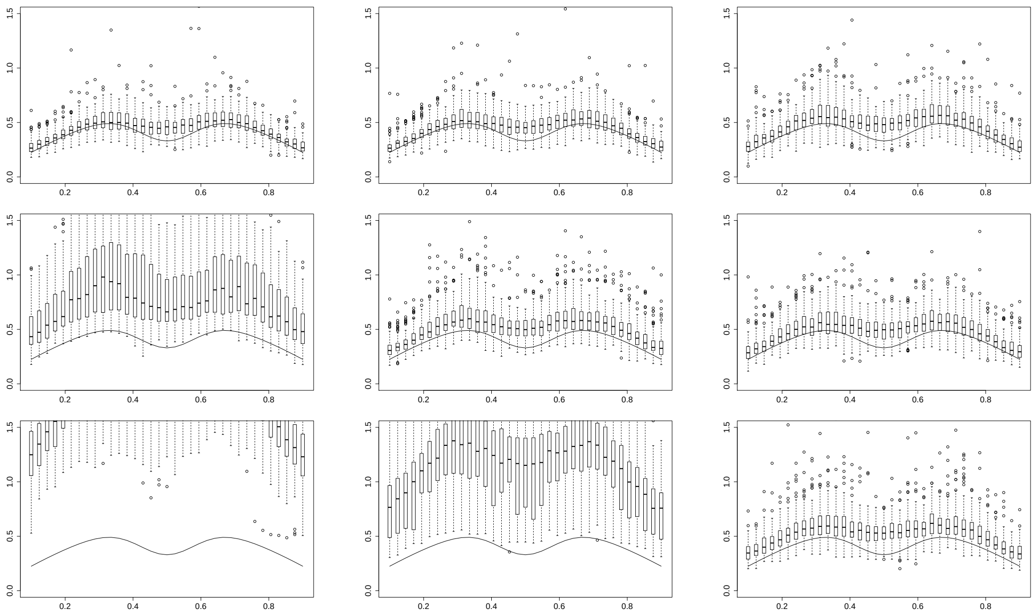

Figure 2: Setting 2: boxplots of $\widehat{\gamma}_{Y,n}(x_0|\tilde{\rho})$, from the left to the right: $\alpha = 0$, $\alpha = 0.1$ and $\alpha = 0.5$; from the top to the bottom: no-contamination, $\epsilon = 0.01$ and $\epsilon = 0.05$. The full line represents the true function γ_Y.

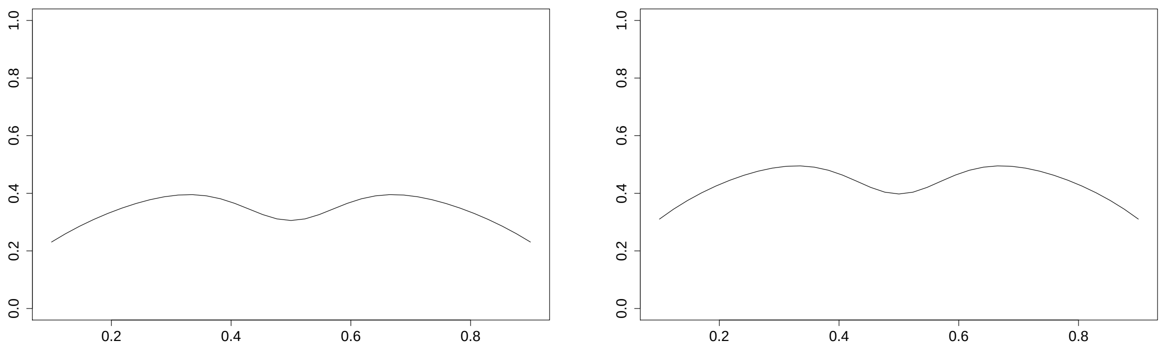

Figure 3: Theoretical asymptotic proportion of censoring for Setting 1 (left panel) and Setting 2 (right panel).

The information on each patient includes gender, date of diagnosis, date of death or end of observation and an indicator as to which of the two is the case. The dataset contains 2843 patients, of which 1761 died, the other survival times are right censored. Since this dataset contains only 89 women, only the 2754 male patients are considered. Our methodology was applied in order to estimate the conditional extreme value index of the survival time Y of a patient conditionally on his age at the time of the diagnosis.

In Figure 4, we show the scatterplot of the original dataset (left) and dataset after rescaling the age into $(0, 1)$ (right). The censored observations are indicated by crosses and the uncensored ones by circles. We can see some points which are away from the main cloud, and thus the question is whether these points might be considered as outliers or not. This question is of interest because as outlined in the simulation section, if the answer is yes, using the estimator with $\alpha = 0$ could have an adverse effect on the estimation of the extreme value index γ. To answer this question, we plot in Figure 5 our estimator $\widehat{\gamma}_{Y,n}(x_0| - 0.5)$ with two different values of α: $\alpha = 0$ (full line), $\alpha = 0.5$ (dotted line). Note that we have not included $\alpha = 0.1$ in order to have a visually clearer figure. On the contrary, we have added the estimator $\widehat{\gamma}_{T,n}(x_0| - 0.5)$ (dashed line) to outline the effect of censoring. As is clear from Figure 5, there exists a substantial difference between the estimates $\widehat{\gamma}_{Y,n}(x_0| - 0.5)$ and $\widehat{\gamma}_{T,n}(x_0| - 0.5)$, due to the high percentage of censoring in the dataset. This outlines the importance of taking the censoring into account. However, the different values for α do not change too much the estimation of γ, with estimates almost always overlapping. This indicates that this dataset does not contain disturbing outliers.

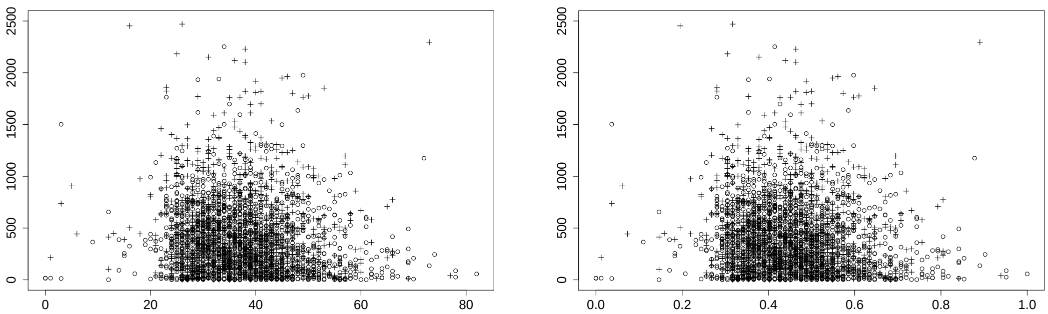

Figure 4: Aids dataset: survival time as a function of age (left) and of aged rescaled into $(0, 1)$ (right).

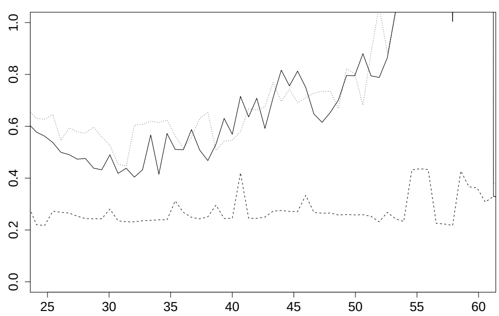

Figure 5: Aids dataset: $\widehat{\gamma}_{Y,n}(x_0| - 0.5)$ with two different values of α: $\alpha = 0$ (full line), $\alpha = 0.5$ (dotted line) and $\widehat{\gamma}_{T,n}(x_0| - 0.5)$ (dashed line) as a function of age.

To reinforce this idea, we now contaminate the dataset by adding 15 pairs represented with triangles in the scatterplot in the left panel of Figure 6, and whose coordinates are given in

Table 3. Then, the estimators $\widehat{\gamma}_{T,n}(x_0|-0.5)$ and $\widehat{\gamma}_{Y,n}(x_0|-0.5)$ with the two same values of α are computed again in the right panel of Figure 6. The first estimator is slightly higher than previously. Concerning $\widehat{\gamma}_{Y,n}(x_0|-0.5)$, we can see, this time, a notable difference between the estimates with $\alpha = 0$ and $\alpha = 0.5$, especially for the covariate range 40 till 60 years, which is precisely the age range where the contamination was located. This highlights the presence of contamination in the dataset, and the importance of taking it into account in the estimation of the extreme value index.

age	time	age	time	age	time
46	4693.527	47	4551.934	51	4982.919
45	4104.4	44	4191.676	47	4143.882
43	4699.267	46	4864.752	50	4449.309
51	4911.656	47	4047.968	43	4576.016
51	4707.304	52	4905.974	49	4833.587

Table 3: Age and survival time of the outliers added to the aids dataset.

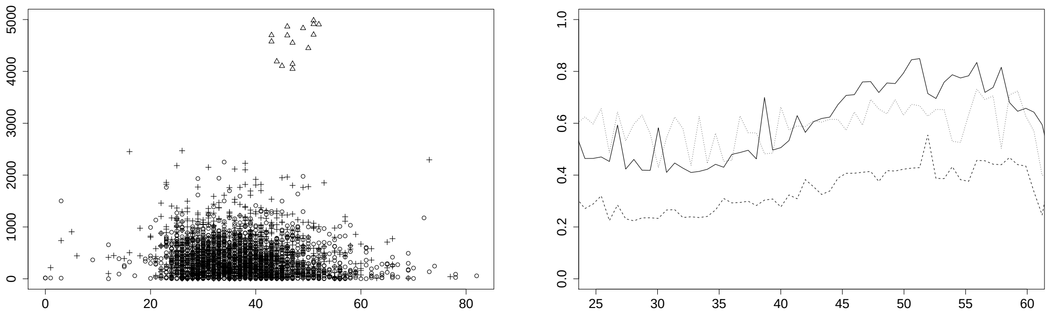

Figure 6: Aids dataset: with outliers represented with triangles (left panel); $\widehat{\gamma}_{Y,n}(x_0|-0.5)$ with two different values of α: $\alpha = 0$ (full line), $\alpha = 0.5$ (dotted line) and $\widehat{\gamma}_{T,n}(x_0|-0.5)$ (dashed line) as a function of age (right panel).

Appendix

Proof of Theorem 1

We focus on deriving the asymptotic expansion for $\mathbb{E}(T_n^{(2)}(K|x_0))$. Then, $\mathbb{E}(T_n^{(1)}(K, s, s'|x_0))$ can be handled similarly, combined with some ideas from Dierckx *et al.* (2014). Note that Dierckx *et al.* (2014) also considered the statistic $T_n^{(1)}(K, s, s'|x_0)$, though it was analysed under their high level assumption called $(\mathcal{M})$, which is avoided in the present paper, and this allows us to obtain a more precise statement of the remainder terms using the Hölder exponents from condition $(\mathcal{H})$.

We have

$$\begin{aligned}
\mathbb{E}(T_n^{(2)}(K|x_0)) &= \mathbb{E}\left(K_{h_n}(x_0-X)\mathbb{1}_{\{Y\leq C, T>t_n\}}\right)\\
&= \mathbb{E}(K_{h_n}(x_0-X)\mathbb{E}(\mathbb{1}_{\{Y\leq C, T>t_n\}}|X))\\
&= \mathbb{E}\left[K_{h_n}(x_0-X)\int_{t_n}^{\infty} f_Y(y|X)\overline{F}_C(y|X)dy\right]\\
&= \int_{\mathbb{R}^d} K_{h_n}(x_0-u)\int_{t_n}^{\infty} f_Y(y|u)\overline{F}_C(y|u)dy f_X(u)du\\
&= \int_{S_K} K(z)\int_{t_n}^{\infty} f_Y(y|x_0-h_nz)\overline{F}_C(y|x_0-h_nz)dy f_X(x_0-h_nz)dz.
\end{aligned}$$

In view of the various Hölder conditions, the latter is further decomposed as

$$\begin{aligned}
&\mathbb{E}(T_n^{(2)}(K|x_0)) =\\
&\quad f_X(x_0)\int_{t_n}^{\infty} f_Y(y|x_0)\overline{F}_C(y|x_0)dy\\
&\quad +\int_{t_n}^{\infty} f_Y(y|x_0)\overline{F}_C(y|x_0)dy\int_{S_K} K(z)(f_X(x_0-h_nz)-f_X(x_0))dz\\
&\quad +f_X(x_0)\int_{S_K} K(z)\int_{t_n}^{\infty} f_Y(y|x_0)(\overline{F}_C(y|x_0-h_nz)-\overline{F}_C(y|x_0))dydz\\
&\quad +\int_{S_K} K(z)\int_{t_n}^{\infty} f_Y(y|x_0)(\overline{F}_C(y|x_0-h_nz)-\overline{F}_C(y|x_0))dy(f_X(x_0-h_nz)-f_X(x_0))dz\\
&\quad +f_X(x_0)\int_{S_K} K(z)\int_{t_n}^{\infty}(f_Y(y|x_0-h_nz)-f_Y(y|x_0))\overline{F}_C(y|x_0))dydz\\
&\quad +\int_{S_K} K(z)\int_{t_n}^{\infty}(f_Y(y|x_0-h_nz)-f_Y(y|x_0))\overline{F}_C(y|x_0)dy(f_X(x_0-h_nz)-f_X(x_0))dz\\
&\quad +f_X(x_0)\int_{S_K} K(z)\int_{t_n}^{\infty}(f_Y(y|x_0-h_nz)-f_Y(y|x_0))(\overline{F}_C(y|x_0-h_nz)-\overline{F}_C(y|x_0))dydz\\
&\quad +\int_{S_K} K(z)\int_{t_n}^{\infty}(f_Y(y|x_0-h_nz)-f_Y(y|x_0))(\overline{F}_C(y|x_0-h_nz)-\overline{F}_C(y|x_0))dy(f_X(x_0-h_nz)-f_X(x_0))dz\\
&\quad =: T_1+\cdots+T_8.
\end{aligned}$$

Concerning T_1 we have

$$T_1 = t_n f_Y(t_n|x_0)\overline{F}_C(t_n|x_0)f_X(x_0)\int_1^{\infty}\frac{f_Y(t_nz|x_0)}{f_Y(t_n|x_0)}\frac{\overline{F}_C(t_nz|x_0)}{\overline{F}_C(t_n|x_0)}dz.$$

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

$$\sup_{z\geq 1} z^{1/\gamma_\bullet(x)}\left|\frac{\overline{F}_\bullet(t_nz|x_0)}{\overline{F}_\bullet(t_n|x_0)}-\overline{G}(z;\gamma_\bullet(x_0),\delta_\bullet(t_n|x_0),\rho_\bullet(x_0))\right| = o(\delta_\bullet(t_n|x_0)), \qquad t_n\to\infty.$$

This leads to the decomposition

$$\begin{aligned} T_1 &= t_n f_Y(t_n|x_0)\overline{F}_C(t_n|x_0)f_X(x_0)\left[\int_1^\infty \frac{f_Y(t_nz|x_0)}{f_Y(t_n|x_0)}\overline{G}(z;\gamma_C(x_0),\delta_C(t_n|x_0),\rho_C(x_0))dz \right. \\ &\quad \left. + \int_1^\infty \frac{f_Y(t_nz|x_0)}{f_Y(t_n|x_0)}\left(\frac{\overline{F}_C(t_nz|x_0)}{\overline{F}_C(t_n|x_0)} - \overline{G}(z;\gamma_C(x_0),\delta_C(t_n|x_0),\rho_C(x_0))\right)dz\right] \\ &=: t_n f_Y(t_n|x_0)\overline{F}_C(t_n|x_0)f_X(x_0)(T_{1,1}+T_{1,2}). \end{aligned}$$

From (3) we can write

$$\begin{aligned} T_{1,1} &= \int_1^\infty z^{-1/\gamma_Y(x_0)-1}\overline{G}(z;\gamma_C(x_0),\delta_C(t_n|x_0),\rho_C(x_0))dz \\ &+\frac{1}{1+\left(\frac{1}{\gamma_Y(x_0)}-\varepsilon_Y(t_n|x_0)\right)\delta_Y(t_n|x_0)}\Bigg[\\ &\frac{\delta_Y(t_n|x_0)}{\gamma_Y(x_0)}\int_1^\infty z^{-1/\gamma_Y(x_0)-1}(z^{\rho_Y(x_0)/\gamma_Y(x_0)}-1)\overline{G}(z;\gamma_C(x_0),\delta_C(t_n|x_0),\rho_C(x_0))dz \\ &+\frac{\delta_Y(t_n|x_0)}{\gamma_Y(x_0)}\int_1^\infty z^{-1/\gamma_Y(x_0)-1}\left(\frac{\delta_Y(t_nz|x_0)}{\delta_Y(t_n|x_0)}-z^{\rho_Y(x_0)/\gamma_Y(x_0)}\right)\overline{G}(z;\gamma_C(x_0),\delta_C(t_n|x_0),\rho_C(x_0))dz \\ &-\varepsilon_Y(t_n|x_0)\delta_Y(t_n|x_0)\int_1^\infty z^{-1/\gamma_Y(x_0)-1}\left(\frac{\varepsilon_Y(t_nz|x_0)}{\varepsilon_Y(t_n|x_0)}-1\right)\frac{\delta_Y(t_nz|x_0)}{\delta_Y(t_n|x_0)}\overline{G}(z;\gamma_C(x_0),\delta_C(t_n|x_0),\rho_C(x_0))dz \\ &-\varepsilon_Y(t_n|x_0)\delta_Y(t_n|x_0)\int_1^\infty z^{-1/\gamma_Y(x_0)-1}\left(\frac{\delta_Y(t_nz|x_0)}{\delta_Y(t_n|x_0)}-1\right)\overline{G}(z;\gamma_C(x_0),\delta_C(t_n|x_0),\rho_C(x_0))dz\Bigg] \\ &=: T_{1,1,1}+\frac{1}{1+\left(\frac{1}{\gamma_Y(x_0)}-\varepsilon_Y(t_n|x_0)\right)\delta_Y(t_n|x_0)}(T_{1,1,2}+\cdots+T_{1,1,5}). \end{aligned}$$

In order to deal with these integrals, the following expansion of the extended Pareto distribution is useful

$$\overline{G}(z;\gamma_\bullet(x_0),\delta_\bullet(t_n|x_0),\rho_\bullet(x_0)) = z^{-1/\gamma_\bullet(x_0)}\left(1-\frac{\delta_\bullet(t_n|x_0)}{\gamma_\bullet(x_0)}(1-z^{\rho_\bullet(x_0)/\gamma_\bullet(x_0)})+O(\delta_\bullet^2(t_n|x_0))\right),$$

where $O(\delta_\bullet^2(t_n|x_0))$ is uniform in $z\geq 1$.

A straightforward calculation gives then

$$\begin{aligned} T_{1,1,1} &= \gamma_T(x_0)+\delta_C(t_n|x_0)\frac{\gamma_T^2(x_0)\rho_C(x_0)}{\gamma_C(x_0)(\gamma_C(x_0)-\gamma_T(x_0)\rho_C(x_0))}+O(\delta_C^2(t_n|x_0)), \\ T_{1,1,2} &= \delta_Y(t_n|x_0)\frac{\gamma_T^2(x_0)\rho_Y(x_0)}{\gamma_Y(x_0)(\gamma_Y(x_0)-\gamma_T(x_0)\rho_Y(x_0))}+O(\delta_Y(t_n|x_0)\delta_C(t_n|x_0)). \end{aligned}$$

For $T_{1,1,3}$ we use Proposition B.1.10 in de Haan and Ferreira (2006), see also Drees (1998). Thus, for $\varepsilon>0$ and $0<\delta<1/\gamma_T(x_0)-\rho_Y(x_0)/\gamma_Y(x_0)$ arbitrary, and n sufficiently large, we have

$$|T_{1,1,3}| \leq \varepsilon\frac{|\delta_Y(t_n|x_0)|}{\gamma_Y(x_0)}\int_1^\infty z^{-(1-\rho_Y(x_0))/\gamma_Y(x_0)+\delta-1}\overline{G}(z;\gamma_C(x_0),\delta_C(t_n|x_0),\rho_C(x_0))dz.$$

Since ε is arbitrary and by using calculations for the integral that are similar to those above, one finds that $T_{1,1,3} = o(\delta_Y(t_n|x_0))$. In the same way $T_{1,1,4} = o(\delta_Y(t_n|x_0))$, and

$$T_{1,1,5} = -\varepsilon_Y(t_n|x_0)\delta_Y(t_n|x_0)\frac{\gamma_T^2(x_0)\rho_Y(x_0)}{\gamma_Y(x_0)-\gamma_T(x_0)\rho_Y(x_0)} + o(\delta_Y(t_n|x_0)).$$

Analogously one can show that $T_{1,2} = o(\delta_C(t_n|x_0))$.

Collecting the terms gives then

$$\begin{aligned} T_1 &= t_n f_Y(t_n|x_0)\overline{F}_C(t_n|x_0)f_X(x_0)\gamma_T(x_0)\left[1+\delta_C(t_n|x_0)\frac{\gamma_T(x_0)\rho_C(x_0)}{\gamma_C(x_0)(\gamma_C(x_0)-\gamma_T(x_0)\rho_C(x_0))}(1+o(1))\right. \\ &\quad \left. +\delta_Y(t_n|x_0)\left(\frac{1}{\gamma_Y(x_0)}-\varepsilon_Y(t_n|x_0)\right)\frac{\gamma_T(x_0)\rho_Y(x_0)}{\gamma_Y(x_0)-\gamma_T(x_0)\rho_Y(x_0)}(1+o(1))\right]. \end{aligned}$$

Note that

$$t_n f_Y(t_n|x_0)\overline{F}_C(t_n|x_0) = \frac{\overline{F}_T(t_n|x_0)}{\gamma_Y(x_0)}\left(1-\frac{\varepsilon_Y(t_n|x_0)\delta_Y(t_n|x_0)}{1+\frac{\delta_Y(t_n|x_0)}{\gamma_Y(x_0)}}\right),$$

whence

$$\begin{aligned} T_1 &= \overline{F}_T(t_n|x_0)f_X(x_0)\frac{\gamma_T(x_0)}{\gamma_Y(x_0)}\left\{1+\delta_C(t_n|x_0)\frac{\gamma_T(x_0)\rho_C(x_0)}{\gamma_C(x_0)(\gamma_C(x_0)-\gamma_T(x_0)\rho_C(x_0))}(1+o(1))\right. \\ &\quad \left. +\,\delta_Y(t_n|x_0)\left[\left(\frac{1}{\gamma_Y(x_0)}-\varepsilon_Y(t_n|x_0)\right)\frac{\gamma_T(x_0)\rho_Y(x_0)}{\gamma_Y(x_0)-\gamma_T(x_0)\rho_Y(x_0)}-\varepsilon_Y(t_n|x_0)\right](1+o(1))\right\}. \end{aligned}$$

For T_2, we use the Hölder condition on f_X and obtain $T_2 = O(h^{\eta_{f_X}}\overline{F}_T(t_n|x_0))$.

By rearranging terms we obtain the following bound for T_3

$$|T_3| \le f_Y(t_n|x_0)\overline{F}_C(t_n|x_0)f_X(x_0)\int_{S_K}K(z)\int_{t_n}^{\infty}\frac{f_Y(y|x_0)}{f_Y(t_n|x_0)}\frac{\overline{F}_C(y|x_0)}{\overline{F}_C(t_n|x_0)}\left|\frac{\overline{F}_C(y|x_0-h_nz)}{\overline{F}_C(y|x_0)}-1\right|dydz, \quad (10)$$

and from condition $(\mathcal{H})$, for n large, and some constants M_1, M_2 and M_3,

$$\begin{aligned} \left|\frac{\overline{F}_C(y|x_0-h_nz)}{\overline{F}_C(y|x_0)}-1\right| &\le M_1\left(h_n^{\eta_{A_C}} + y^{M_2h_n^{\eta_{\gamma_C}}}h_n^{\eta_{\gamma_C}}\ln y + |\delta_C(y|x_0)|h_n^{\eta_{B_C}}\right. \\ &\quad \left. +\,|\delta_C(y|x_0)|y^{M_3h_n^{\eta_{\varepsilon_C}}}h_n^{\eta_{\varepsilon_C}}\ln y\right). \end{aligned}$$

Plugging the above inequality into (10), and computing integrals similar to those encountered above yields

$$T_3 = O\left(\overline{F}_T(t_n|x_0)(h_n^{\eta_{A_C}} + h_n^{\eta_{\gamma_C}}\ln t_n + \delta_C(t_n|x_0)h_n^{\eta_{B_C}} + \delta_C(t_n|x_0)h_n^{\eta_{\varepsilon_C}}\ln t_n)\right).$$

Using the Hölder condition on f_X one easily verifies that T_4 is of smaller order than T_3.

As for T_5, we can write

$$|T_5| \leq f_Y(t_n|x_0)\overline{F}_C(t_n|x_0)f_X(x_0)\int_{S_K} K(z)\int_{t_n}^{\infty} \frac{f_Y(y|x_0)}{f_Y(t_n|x_0)}\frac{\overline{F}_C(y|x_0)}{\overline{F}_C(t_n|x_0)}\left|\frac{f_Y(y|x_0-h_nz)}{f_Y(y|x_0)}-1\right|dydz,$$

which, combined with the inequality

$$\begin{aligned}\left|\frac{f_Y(y|x_0-h_nz)}{f_Y(y|x_0)}-1\right| \quad &\leq \quad M_1\left(h_n^{\eta_{A_Y}}+y^{M_2h_n^{\eta_{\gamma_Y}}}h_n^{\eta_{\gamma_Y}}\ln y+|\delta_Y(y|x_0)|h_n^{\eta_{B_Y}}\right.\\ &\qquad \left.+|\delta_Y(y|x_0)|y^{M_3h_n^{\eta_{\varepsilon_Y}}}h_n^{\eta_{\varepsilon_Y}}\ln y\right),\end{aligned}$$

valid for n large, where M_1, M_2 and M_3 are some constants, leads to

$$T_5 = O\left(\overline{F}_T(t_n|x_0)(h_n^{\eta_{A_Y}}+h_n^{\eta_{\gamma_Y}}\ln t_n+\delta_Y(t_n|x_0)h_n^{\eta_{B_Y}}+\delta_Y(t_n|x_0)h_n^{\eta_{\varepsilon_Y}}\ln t_n)\right).$$

After tedious calculations, but essentially involving integrals similar to the ones above, one can verify that T_6, T_7 and T_8, are of smaller order than terms that were already encountered before.

Collecting the terms then establishes Theorem 1.

Proof of Theorem 2

To prove the result we make use of the Cramér-Wold device (see, e.g., van der Vaart, 1998, p.16).

Take $\xi = (\xi_1, \ldots, \xi_{J+1})^T \in \mathbb{R}^{J+1}$. Then

$$\begin{aligned}&\xi^T r_n(\mathbb{T}_n - \mathbb{E}(\mathbb{T}_n)) = \\ &\quad \sum_{i=1}^{n}\left\{\sum_{j=1}^{J}\xi_j\left(\frac{h_n^d}{n\overline{F}_T(t_n|x_0)f_X(x_0)}\right)^{1/2}K_{h_n}(x_0-X_i)\left(\frac{T_i}{t_n}\right)^{s_j}\left(\ln_+\frac{T_i}{t_n}\right)^{s_j'}1\!\!1_{\{T_i>t_n\}}\right.\\ &\quad +\xi_{J+1}\left(\frac{h_n^d}{n\overline{F}_T(t_n|x_0)f_X(x_0)}\right)^{1/2}K_{h_n}(x_0-X_i)1\!\!1_{\{Y_i\leq C_i,T_i>t_n\}}\\ &\quad -\mathbb{E}\left[\sum_{j=1}^{J}\xi_j\left(\frac{h_n^d}{n\overline{F}_T(t_n|x_0)f_X(x_0)}\right)^{1/2}K_{h_n}(x_0-X)\left(\frac{T}{t_n}\right)^{s_j}\left(\ln_+\frac{T}{t_n}\right)^{s_j'}1\!\!1_{\{T>t_n\}}\right.\\ &\quad \left.\left.+\ \xi_{J+1}\left(\frac{h_n^d}{n\overline{F}_T(t_n|x_0)f_X(x_0)}\right)^{1/2}K_{h_n}(x_0-X)1\!\!1_{\{Y\leq C,T>t_n\}}\right]\right\}\\ &\quad =: \sum_{i=1}^{n}W_i.\end{aligned}$$

We have

$$\mathbb{V}ar(W_1) = \frac{\xi^T\mathbb{C}\xi}{n},$$

where $\mathbb{C}$ has elements

$$\begin{aligned}
\mathbb{C}_{j,k} &:= \mathbb{C}ov\left(\left(\frac{h_n^d}{\overline{F}_T(t_n|x_0)f_X(x_0)}\right)^{1/2} K_{h_n}(x_0-X)\left(\frac{T}{t_n}\right)^{s_j}\left(\ln_+\frac{T}{t_n}\right)^{s'_j} 1\!\!1_{\{T>t_n\}},\right. \\
&\qquad \left.\left(\frac{h_n^d}{\overline{F}_T(t_n|x_0)f_X(x_0)}\right)^{1/2} K_{h_n}(x_0-X)\left(\frac{T}{t_n}\right)^{s_k}\left(\ln_+\frac{T}{t_n}\right)^{s'_k} 1\!\!1_{\{T>t_n\}}\right),\ j,k\in\{1,\ldots,J\}, \\
\mathbb{C}_{J+1,j} &:= \mathbb{C}ov\left(\left(\frac{h_n^d}{\overline{F}_T(t_n|x_0)f_X(x_0)}\right)^{1/2} K_{h_n}(x_0-X) 1\!\!1_{\{Y\le C, T>t_n\}},\right. \\
&\qquad \left.\left(\frac{h_n^d}{\overline{F}_T(t_n|x_0)f_X(x_0)}\right)^{1/2} K_{h_n}(x_0-X)\left(\frac{T}{t_n}\right)^{s_j}\left(\ln_+\frac{T}{t_n}\right)^{s'_j} 1\!\!1_{\{T>t_n\}}\right),\ j\in\{1,\ldots,J\}, \\
\mathbb{C}_{J+1,J+1} &:= \mathbb{V}ar\left(\left(\frac{h_n^d}{\overline{F}_T(t_n|x_0)f_X(x_0)}\right)^{1/2} K_{h_n}(x_0-X) 1\!\!1_{\{Y\le C, T>t_n\}}\right).
\end{aligned}$$

As for $\mathbb{C}_{j,k}$, with $j,k\in\{1,\ldots,J\}$, we have, by a straightforward application of Theorem 1

$$\begin{aligned}
\mathbb{C}_{j,k} &= \frac{h_n^d}{\overline{F}_T(t_n|x_0)f_X(x_0)}\left[\frac{\overline{F}_T(t_n|x_0)f_X(x_0)\|K\|_2^2\gamma_T^{s'_j+s'_k}(x_0)\Gamma(s'_j+s'_k+1)}{h_n^d(1-(s_j+s_k)\gamma_T(x_0))^{s'_j+s'_k+1}}(1+o(1))+O(\overline{F}_T^2(t_n|x_0))\right] \\
&= \Sigma_{j,k}(1+o(1)).
\end{aligned}$$

In the same way, by using Theorem 1, $\mathbb{C}_{J+1,J+1}=\Sigma_{J+1,J+1}(1+o(1))$. For $\mathbb{C}_{J+1,j}$, $j\in\{1,\ldots,J\}$, we need to evaluate an expectation of the form

$$\mathbb{E}\left[K_{h_n}(x_0-X)\left(\frac{T}{t_n}\right)^{s_j}\left(\ln_+\frac{T}{t_n}\right)^{s'_j} 1\!\!1_{\{Y\le C, T>t_n\}}\right].$$

By arguments similar to those used in deriving the expansion for $\mathbb{E}(T_n^{(2)}(K|x_0))$ we obtain

$$\mathbb{E}\left[K_{h_n}(x_0-X)\left(\frac{T}{t_n}\right)^{s_j}\left(\ln_+\frac{T}{t_n}\right)^{s'_j} 1\!\!1_{\{Y\le C, T>t_n\}}\right] = \overline{F}_T(t_n|x_0)f_X(x_0)\frac{\gamma_T^{s'_j+1}(x_0)\Gamma(s'_j+1)}{\gamma_Y(x_0)(1-s_j\gamma_T(x_0))^{s'_j+1}}(1+o(1)). \tag{11}$$

In order to establish the weak convergence to a Gaussian random variable we need to verify the Lyapounov condition (see, e.g., Billingsley, 1995, p. 362), which simplifies in our setting to showing that $\lim_{n\to\infty} n\mathbb{E}(|W_1|^3)=0$. To this aim, note that W_1 is of the form $V-\mathbb{E}(V)$, leading to the inequality

$$\mathbb{E}(|W_1|^3)\le \mathbb{E}(|V^3|)+3\mathbb{E}(V^2)\mathbb{E}(|V|)+4(\mathbb{E}(|V|))^3.$$

Again using the result from Theorem 1 and (11), we obtain the following orders

$$
\begin{aligned}
\mathbb{E}(|V|^3) &= O\left(\frac{1}{n^{3/2}\sqrt{h_n^d \overline{F}_T(t_n|x_0)}}\right), \\
\mathbb{E}(V^2)\mathbb{E}(|V|) &= O\left(\frac{\sqrt{h_n^d \overline{F}_T(t_n|x_0)}}{n^{3/2}}\right), \\
(E(|V|))^3 &= O\left(\left(\frac{h_n^d \overline{F}_T(t_n|x_0)}{n}\right)^{3/2}\right),
\end{aligned}
$$

so that $n\mathbb{E}(|W_1|^3) \to 0$ under our assumption $r_n \to \infty$.

Proof of Theorem 3

Let

$$
\begin{aligned}
\widetilde{\mathbb{S}}_n^{(j)} &:= \left\{ r_n \left[\frac{T_n^{(1)}(K,s,j|x_0)}{\overline{F}(t_n|x_0)f_X(x_0)} - \mathbb{E}\left(\frac{T_n^{(1)}(K,s,j|x_0)}{\overline{F}(t_n|x_0)f_X(x_0)}\right)\right]; s \in [S,0] \right\}, \qquad j \in \{0,1,2,3\}, \\
\widetilde{\mathbb{S}}_n^{(4)} &:= r_n \left[\frac{T_n^{(2)}(K|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)} - \mathbb{E}\left(\frac{T_n^{(2)}(K|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)}\right)\right].
\end{aligned}
$$

The weak convergence of the individual processes $\widetilde{\mathbb{S}}_n^{(j)}$, $j \in \{0,1,2,3\}$, to tight, zero centered, Gaussian processes, with a covariance structure as given in the statement of the theorem, can be established following the arguments in the proof of Theorem 1 in Dierckx *et al.* (2014), while the weak convergence of $\widetilde{\mathbb{S}}_n^{(4)}$ to a zero centered Gaussian random variable with variance as in the statement of the theorem follows from Theorem 2. Then joint tightness will follow from the individual tightness. The joint tightness combined with the finite dimensional convergence from Theorem 2 leads then to the joint weak convergence of $(\widetilde{\mathbb{S}}_n^{(0)}, \widetilde{\mathbb{S}}_n^{(1)}, \widetilde{\mathbb{S}}_n^{(2)}, \widetilde{\mathbb{S}}_n^{(3)}, \widetilde{\mathbb{S}}_n^{(4)})$. It remains to verify the expressions for the expected values of $\mathbb{S}^{(j)}$, $j \in \{0,1,\dots,4\}$. To this aim we study

$$
r_n \left[\mathbb{E}\left(\frac{T_n^{(1)}(K,s,j|x_0)}{\overline{F}(t_n|x_0)f_X(x_0)}\right) - \frac{j!\gamma_T^j(x_0)}{(1-s\gamma_T(x_0))^{j+1}} \right],
$$

and

$$
r_n \left[\mathbb{E}\left(\frac{T_n^{(2)}(K|x_0)}{\overline{F}(t_n|x_0)f_X(x_0)}\right) - \frac{\gamma_T(x_0)}{\gamma_Y(x_0)} \right].
$$

A straightforward application of Theorem 1, and taking the link between $\delta_T(t_n|x_0)$, $\delta_Y(t_n|x_0)$ and $\delta_C(t_n|x_0)$ into account, one easily obtains the expressions for the expected values of the limiting processes $\mathbb{S}^{(0)}$, $\mathbb{S}^{(1)}$, $\mathbb{S}^{(2)}$ and $\mathbb{S}^{(3)}$, and the random variable $\mathbb{S}^{(4)}$.

Proof of Theorem 4

The consistency of $\widehat{\gamma}_{T,n}(x_0|\tilde{\rho})$ and $\widehat{\delta}_{T,n}(x_0|\tilde{\rho})$ for $\gamma_T^{(0)}(x_0)$ and 0, respectively, follows from Proposition 1 in Dierckx *et al.* (2014).

Concerning $\widehat{p}_n(x_0)$ we have from Theorem 1

$$\begin{aligned}
\mathbb{E}\left(\frac{T_n^{(1)}(K,0,0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)}\right) &= 1+o(1), \\
\mathbb{E}\left(\frac{T_n^{(2)}(K|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)}\right) &= \frac{\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)}+o(1),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{V}ar\left(\frac{T_n^{(1)}(K,0,0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)}\right) &= O\left(\frac{1}{nh_n^d\overline{F}_T(t_n|x_0)}\right), \\
\mathbb{V}ar\left(\frac{T_n^{(2)}(K|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)}\right) &= O\left(\frac{1}{nh_n^d\overline{F}_T(t_n|x_0)}\right).
\end{aligned}$$

Thus

$$\frac{T_n^{(1)}(K,0,0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)} \xrightarrow{P} 1 \qquad \text{and} \qquad \frac{T_n^{(2)}(K|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)} \xrightarrow{P} \frac{\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)},$$

and hence, by the continuous mapping theorem, $\widehat{p}_n(x_0) \xrightarrow{P} \gamma_T^{(0)}(x_0)/\gamma_Y^{(0)}(x_0)$. Another application of the continuous mapping theorem gives then $\widehat{\gamma}_{Y,n}(x_0|\tilde{\rho}) \xrightarrow{P} \gamma_Y^{(0)}(x_0)$.

Proof of Theorem 5

Let $\widetilde{\Delta}_\alpha(\gamma,\delta;\tilde{\rho}) := \widehat{\Delta}_\alpha(\gamma,\delta;\tilde{\rho})/(\overline{F}_T(t_n|x_0)f_X(x_0))$, and let $\widetilde{\Delta}_{\alpha,u}(\gamma,\delta;\tilde{\rho})$, $u=1,2$, denote the derivatives with respect to γ and δ, respectively, apart from the common scale factor $1+\alpha$. Similarly, $\widetilde{\Delta}_{\alpha,u,v}(\gamma,\delta;\tilde{\rho})$ and $\widetilde{\Delta}_{\alpha,u,v,w}(\gamma,\delta;\tilde{\rho})$, $u,v,w=1,2$, will denote second and third order derivatives (again apart from the common scaling by $1+\alpha$).

We apply a Taylor series expansion of the estimating equations (8) and (9) around $(\gamma_T^{(0)}(x_0),0)$, and extend these with $T_n^{(1)}(K,0,0|x_0)$ and $T_n^{(2)}(K|x_0)$ to obtain

$$r_n \begin{bmatrix} -\widetilde{\Delta}_{\alpha,1}(\gamma_T^{(0)}(x_0),0;\tilde{\rho}) \\ -\widetilde{\Delta}_{\alpha,2}(\gamma_T^{(0)}(x_0),0;\tilde{\rho}) \\ \frac{T_n^{(1)}(K,0,0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)} - 1 \\ \frac{T_n^{(2)}(K|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)} - \frac{\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)} \end{bmatrix} =$$

$$\begin{bmatrix} \overline{\Delta}_{\alpha,1,1}(\gamma_T^{(0)}(x_0),0;\tilde{\rho}) & \overline{\Delta}_{\alpha,1,2}(\gamma_T^{(0)}(x_0),0;\tilde{\rho}) & 0 & 0 \\ \overline{\Delta}_{\alpha,1,2}(\gamma_T^{(0)}(x_0),0;\tilde{\rho}) & \overline{\Delta}_{\alpha,2,2}(\gamma_T^{(0)}(x_0),0;\tilde{\rho}) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_n(\widehat{\gamma}_{T,n}(x_0|\tilde{\rho}) - \gamma_T^{(0)}(x_0)) \\ r_n\widehat{\delta}_{T,n}(x_0|\tilde{\rho}) \\ r_n\left(\frac{T_n^{(1)}(K,0,0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)} - 1\right) \\ r_n\left(\frac{T_n^{(2)}(K|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)} - \frac{\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)}\right) \end{bmatrix}$$

with

$$\begin{aligned} \overline{\Delta}_{\alpha,1,1}(\gamma_T^{(0)}(x_0),0;\tilde{\rho}) &:= \widetilde{\Delta}_{\alpha,1,1}(\gamma_T^{(0)}(x_0),0;\tilde{\rho}) + \frac{1}{2}\Big[\widetilde{\Delta}_{\alpha,1,1,1}(\breve{\gamma}_{T,n}(x_0|\tilde{\rho}),\breve{\delta}_{T,n}(x_0|\tilde{\rho});\tilde{\rho})(\widehat{\gamma}_{T,n}(x_0|\tilde{\rho}) - \gamma_T^{(0)}(x_0)) \\ &\quad +\widetilde{\Delta}_{\alpha,1,1,2}(\breve{\gamma}_{T,n}(x_0|\tilde{\rho}),\breve{\delta}_{T,n}(x_0|\tilde{\rho});\tilde{\rho})\widehat{\delta}_{T,n}(x_0|\tilde{\rho})\Big], \\ \overline{\Delta}_{\alpha,1,2}(\gamma_T^{(0)}(x_0),0;\tilde{\rho}) &:= \widetilde{\Delta}_{\alpha,1,2}(\gamma_T^{(0)}(x_0),0;\tilde{\rho}) + \frac{1}{2}\Big[\widetilde{\Delta}_{\alpha,1,2,2}(\breve{\gamma}_{T,n}(x_0|\tilde{\rho}),\breve{\delta}_{T,n}(x_0|\tilde{\rho});\tilde{\rho})\widehat{\delta}_{T,n}(x_0|\tilde{\rho}) \\ &\quad +\widetilde{\Delta}_{\alpha,1,1,2}(\breve{\gamma}_{T,n}(x_0|\tilde{\rho}),\breve{\delta}_{T,n}(x_0|\tilde{\rho});\tilde{\rho})(\widehat{\gamma}_{T,n}(x_0|\tilde{\rho}) - \gamma_T^{(0)}(x_0))\Big], \\ \overline{\Delta}_{\alpha,2,2}(\gamma_T^{(0)}(x_0),0;\tilde{\rho}) &:= \widetilde{\Delta}_{\alpha,2,2}(\gamma_T^{(0)}(x_0),0;\tilde{\rho}) + \frac{1}{2}\Big[\widetilde{\Delta}_{\alpha,2,2,2}(\breve{\gamma}_{T,n}(x_0|\tilde{\rho}),\breve{\delta}_{T,n}(x_0|\tilde{\rho});\tilde{\rho})\widehat{\delta}_{T,n}(x_0|\tilde{\rho}) \\ &\quad +\widetilde{\Delta}_{\alpha,1,2,2}(\breve{\gamma}_{T,n}(x_0|\tilde{\rho}),\breve{\delta}_{T,n}(x_0|\tilde{\rho});\tilde{\rho})(\widehat{\gamma}_{T,n}(x_0|\tilde{\rho}) - \gamma_T^{(0)}(x_0))\Big], \end{aligned}$$

and where $(\breve{\gamma}_{T,n}(x_0|\tilde{\rho}),\breve{\delta}_{T,n}(x_0|\tilde{\rho}))$ is a point on the line segment connecting $(\gamma_T^{(0)}(x_0),0)$ and $(\widehat{\gamma}_{T,n}(x_0|\tilde{\rho}),\widehat{\delta}_{T,n}(x_0|\tilde{\rho}))$.

After tedious, but straightforward, derivations one obtains

$$\begin{aligned} &\widetilde{\Delta}_{\alpha,1}(\gamma_T^{(0)}(x_0),0;\tilde{\rho}) \\ &= (\gamma_T^{(0)}(x_0))^{-\alpha-2}\left[-\frac{\alpha\gamma_T^{(0)}(x_0)(1+\gamma_T^{(0)}(x_0))}{[1+\alpha(1+\gamma_T^{(0)}(x_0))]^2}\frac{T_n^{(1)}(K,0,0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)}\right. \\ &\quad \left. +\gamma_T^{(0)}(x_0)\frac{T_n^{(1)}(K,-\alpha(1+\gamma_T^{(0)}(x_0))/\gamma_T^{(0)}(x_0),0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)} - \frac{T_n^{(1)}(K,-\alpha(1+\gamma_T^{(0)}(x_0))/\gamma_T^{(0)}(x_0),1|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)}\right], \end{aligned}$$

$$
\begin{aligned}
&\widetilde{\Delta}_{\alpha,2}(\gamma_T^{(0)}(x_0),0;\tilde{\rho})\\
&= (\gamma_T^{(0)}(x_0))^{-\alpha-1}\Bigg[-\frac{\alpha\tilde{\rho}(1+\gamma_T^{(0)}(x_0))}{[1+\alpha(1+\gamma_T^{(0)}(x_0))][1-\tilde{\rho}+\alpha(1+\gamma_T^{(0)}(x_0))]}\frac{T_n^{(1)}(K,0,0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)}\\
&\quad+\frac{T_n^{(1)}(K,-\alpha(1+\gamma_T^{(0)}(x_0))/\gamma_T^{(0)}(x_0),0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)} -(1-\tilde{\rho})\frac{T_n^{(1)}(K,-(\alpha(1+\gamma_T^{(0)}(x_0))-\tilde{\rho})/\gamma_T^{(0)}(x_0),0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)}\Bigg],
\end{aligned}
$$

and

$$
\begin{aligned}
&\widetilde{\Delta}_{\alpha,1,1}(\gamma_T^{(0)}(x_0),0;\tilde{\rho})\\
&= (\gamma_T^{(0)}(x_0))^{-\alpha-2}\Bigg[\Bigg(\frac{\alpha+2}{1+\alpha(1+\gamma_T^{(0)}(x_0))}-\frac{2\alpha+4}{[1+\alpha(1+\gamma_T^{(0)}(x_0))]^2}\\
&\qquad+\frac{2\alpha+2}{[1+\alpha(1+\gamma_T^{(0)}(x_0))]^3}\Bigg)\frac{T_n^{(1)}(K,0,0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)}\\
&\quad-(\alpha+1)\frac{T_n^{(1)}(K,-\alpha(1+\gamma_T^{(0)}(x_0))/\gamma_T^{(0)}(x_0),0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)} +\frac{2\alpha+2}{\gamma_T^{(0)}(x)}\frac{T_n^{(1)}(K,-\alpha(1+\gamma_T^{(0)}(x_0))/\gamma_T^{(0)}(x_0),1|x_0)}{\overline{F}(t_n|x_0)f_X(x_0)}\\
&\quad-\frac{\alpha}{(\gamma_T^{(0)}(x_0))^2}\frac{T_n^{(1)}(K,-\alpha(1+\gamma_T^{(0)}(x_0))/\gamma_T^{(0)}(x_0),2|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)}\Bigg],
\end{aligned}
$$

$$
\begin{aligned}
&\widetilde{\Delta}_{\alpha,1,2}(\gamma_T^{(0)}(x_0),0;\tilde{\rho})\\
&= (\gamma_T^{(0)}(x_0))^{-\alpha-2}\Bigg[\Bigg(\frac{1+\alpha(2+\alpha)(1+\gamma_T^{(0)}(x_0))}{[1+\alpha(1+\gamma_T^{(0)}(x_0))]^2}\\
&\quad-\frac{(1-\tilde{\rho})^2-\alpha[\tilde{\rho}(1-\tilde{\rho})-2(1+\gamma_T^{(0)}(x_0))(1-\tilde{\rho})]+\alpha^2(1+\gamma_T^{(0)}(x_0))(1-\tilde{\rho})}{[1-\tilde{\rho}+\alpha(1+\gamma_T^{(0)}(x_0))]^2}\Bigg)\\
&\quad\times\frac{T_n^{(1)}(K,0,0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)}-(1+\alpha)\frac{T_n^{(1)}(K,-\alpha(1+\gamma_T^{(0)}(x_0))/\gamma_T^{(0)}(x_0),0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)}\\
&\quad+(\alpha+1)(1-\tilde{\rho})\frac{T_n^{(1)}(K,-(\alpha(1+\gamma_T^{(0)}(x_0))-\tilde{\rho})/\gamma_T^{(0)}(x_0),0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)}\\
&\quad+\frac{\alpha}{\gamma_T^{(0)}(x_0)}\frac{T_n^{(1)}(K,-\alpha(1+\gamma_T^{(0)}(x_0))/\gamma_T^{(0)}(x_0),1|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)}\\
&\quad-\frac{(\alpha-\tilde{\rho})(1-\tilde{\rho})}{\gamma_T^{(0)}(x_0)}\frac{T_n^{(1)}(K,-(\alpha(1+\gamma_T^{(0)}(x_0))-\tilde{\rho})/\gamma_T^{(0)}(x_0),1|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)}\Bigg],
\end{aligned}
$$

$$
\begin{aligned}
&\widetilde{\Delta}_{\alpha,2,2}(\gamma_T^{(0)}(x_0),0;\tilde{\rho})\\
= \quad & (\gamma_T^{(0)}(x_0))^{-\alpha-2}\left[\left(\frac{1+\alpha+\gamma_T^{(0)}(x_0)}{1+\alpha(1+\gamma_T^{(0)}(x_0))}-\frac{2(1-\tilde{\rho})(1+\gamma_T^{(0)}(x_0)+\alpha)}{1-\tilde{\rho}+\alpha(1+\gamma_T^{(0)}(x_0))}\right.\right.\\
&\left.+\frac{(1+\gamma_T^{(0)}(x_0))(1-2\tilde{\rho})+\alpha(1-\tilde{\rho})^2}{1-2\tilde{\rho}+\alpha(1+\gamma_T^{(0)}(x_0))}\right)\frac{T_n^{(1)}(K,0,0|x_0)}{\overline{F}(t_n|x_0)f_X(x_0)}\\
&-(\alpha+\gamma_T^{(0)}(x_0))\frac{T_n^{(1)}(K,-\alpha(1+\gamma_T^{(0)}(x_0))/\gamma_T^{(0)}(x_0),0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)}\\
&+2(1-\tilde{\rho})(\alpha+\gamma_T^{(0)}(x_0))\frac{T_n^{(1)}(K,-(\alpha(1+\gamma_T^{(0)}(x_0))-\tilde{\rho})/\gamma_T^{(0)}(x_0),0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)}\\
&\left.-[(1+\gamma_T^{(0)}(x_0))(1-2\tilde{\rho})+(\alpha-1)(1-\tilde{\rho})^2]\frac{T_n^{(1)}(K,-(\alpha(1+\gamma_T^{(0)}(x_0))-2\tilde{\rho})/\gamma_T^{(0)}(x_0),0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)}\right].
\end{aligned}
$$

For brevity, the expressions for the third order derivatives are omitted from the paper.

Now let

$$
\mathbb{U}_n(\tilde{\rho}) := \frac{1}{\overline{F}_T(t_n|x_0)f_X(x_0)}\begin{bmatrix} T_n^{(1)}(K,0,0|x_0) \\ T_n^{(1)}(K,-\alpha(1+\gamma_T^{(0)}(x_0))/\gamma_T^{(0)}(x),0|x_0) \\ T_n^{(1)}(K,-(\alpha(1+\gamma_T^{(0)}(x_0))-\tilde{\rho})/\gamma_T^{(0)}(x),0|x_0) \\ T_n^{(1)}(K,-\alpha(1+\gamma_T^{(0)}(x_0))/\gamma_T^{(0)}(x_0),1|x_0) \\ T_n^{(2)}(K|x_0) \end{bmatrix},
$$

$$
\overline{\mathbb{U}}(\tilde{\rho}) := \begin{bmatrix} 1 \\ \frac{1}{1+\alpha(1+\gamma_T^{(0)}(x_0))} \\ \frac{1}{1-\tilde{\rho}+\alpha(1+\gamma_T^{(0)}(x_0))} \\ \frac{\gamma_T^{(0)}(x_0)}{[1+\alpha(1+\gamma_T^{(0)}(x_0))]^2} \\ \frac{\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)} \end{bmatrix}.
$$

Then by Theorem 3

$$
\mathbb{W}_n(\tilde{\rho}) := r_n(\mathbb{U}_n(\tilde{\rho})-\overline{\mathbb{U}}(\tilde{\rho})) \rightsquigarrow N(\lambda\mu(\tilde{\rho}),\Sigma(\tilde{\rho})),
$$

where $\mu(\tilde{\rho})$ a (5×1) vector with elements

$$\begin{aligned}
\mu_1(\tilde{\rho}) &= 0, \\
\mu_2(\tilde{\rho}) &= -\frac{\alpha \rho_T^{(0)}(x_0)(1+\gamma_T^{(0)}(x_0))}{\gamma_T^{(0)}(x_0)[1+\alpha(1+\gamma_T^{(0)}(x_0))][1-\rho_T^{(0)}(x_0)+\alpha(1+\gamma_T^{(0)}(x_0))]}, \\
\mu_3(\tilde{\rho}) &= -\frac{[\alpha(1+\gamma_T^{(0)}(x_0))-\tilde{\rho}]\rho_T^{(0)}(x_0)}{\gamma_T^{(0)}(x_0)[1-\tilde{\rho}+\alpha(1+\gamma_T^{(0)}(x_0))][1-\rho_T^{(0)}(x_0)-\tilde{\rho}+\alpha(1+\gamma_T^{(0)}(x_0))]}, \\
\mu_4(\tilde{\rho}) &= \frac{\rho_T^{(0)}(x_0)(1-\rho_T^{(0)}(x_0))-\alpha^2\rho_T^{(0)}(x_0)(1+\gamma_T^{(0)}(x_0))^2}{[1+\alpha(1+\gamma_T^{(0)}(x_0))]^2[1-\rho_T^{(0)}(x_0)+\alpha(1+\gamma_T^{(0)}(x_o))]^2}, \\
\mu_5(\tilde{\rho}) &= \mu,
\end{aligned}$$

where μ is given by (7), apart from the factor λ, and $\Sigma(\tilde{\rho})$ a (5×5) symmetric matrix with elements

$$\begin{aligned}
\Sigma_{11}(\tilde{\rho}) &:= \|K\|_2^2, \\
\Sigma_{21}(\tilde{\rho}) &:= \frac{\|K\|_2^2}{1+\alpha(1+\gamma_T^{(0)}(x_0))}, \\
\Sigma_{22}(\tilde{\rho}) &:= \frac{\|K\|_2^2}{1+2\alpha(1+\gamma_T^{(0)}(x_0))}, \\
\Sigma_{31}(\tilde{\rho}) &:= \frac{\|K\|_2^2}{1-\tilde{\rho}+\alpha(1+\gamma_T^{(0)}(x_0))}, \\
\Sigma_{32}(\tilde{\rho}) &:= \frac{\|K\|_2^2}{1-\tilde{\rho}+2\alpha(1+\gamma_T^{(0)}(x_0))}, \\
\Sigma_{33}(\tilde{\rho}) &:= \frac{\|K\|_2^2}{1-2\tilde{\rho}+2\alpha(1+\gamma_T^{(0)}(x_0))}, \\
\Sigma_{41}(\tilde{\rho}) &:= \frac{\gamma_T^{(0)}(x_0)\|K\|_2^2}{[1+\alpha(1+\gamma_T^{(0)}(x_0))]^2}, \\
\Sigma_{42}(\tilde{\rho}) &:= \frac{\gamma_T^{(0)}(x_0)\|K\|_2^2}{[1+2\alpha(1+\gamma_T^{(0)}(x_0))]^2}, \\
\Sigma_{43}(\tilde{\rho}) &:= \frac{\gamma_T^{(0)}(x_0)\|K\|_2^2}{[1-\tilde{\rho}+2\alpha(1+\gamma_T^{(0)}(x_0))]^2}, \\
\Sigma_{44}(\tilde{\rho}) &:= \frac{2(\gamma_T^{(0)}(x_0))^2\|K\|_2^2}{[1+2\alpha(1+\gamma_T^{(0)}(x_0))]^3},
\end{aligned}$$

$$
\begin{aligned}
\Sigma_{51}(\tilde\rho) &:= \frac{\gamma_T^{(0)}(x_0)\|K\|_2^2}{\gamma_Y^{(0)}(x_0)}, \\
\Sigma_{52}(\tilde\rho) &:= \frac{\gamma_T^{(0)}(x_0)\|K\|_2^2}{\gamma_Y^{(0)}(x_0)[1+\alpha(1+\gamma_T^{(0)}(x_0))]}, \\
\Sigma_{53}(\tilde\rho) &:= \frac{\gamma_T^{(0)}(x_0)\|K\|_2^2}{\gamma_Y^{(0)}(x_0)[1-\tilde\rho+\alpha(1+\gamma_T^{(0)}(x_0))]}, \\
\Sigma_{54}(\tilde\rho) &:= \frac{(\gamma_T^{(0)}(x_0))^2\|K\|_2^2}{\gamma_Y^{(0)}(x_0)[1+\alpha(1+\gamma_T^{(0)}(x_0))]^2}, \\
\Sigma_{55}(\tilde\rho) &:= \frac{\gamma_T^{(0)}(x_0)\|K\|_2^2}{\gamma_Y^{(0)}(x_0)}.
\end{aligned}
$$

We have

$$
r_n \begin{bmatrix} -\widetilde\Delta_{\alpha,1}(\gamma_T^{(0)}(x_0),0;\tilde\rho) \\ -\widetilde\Delta_{\alpha,2}(\gamma_T^{(0)}(x_0),0;\tilde\rho) \\ \frac{T_n^{(1)}(K,0,0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)} - 1 \\ \frac{T_n^{(2)}(K|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)} - \frac{\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)} \end{bmatrix} = B(\tilde\rho)\mathbb{W}_n(\tilde\rho) \rightsquigarrow N(\lambda B(\tilde\rho)\mu(\tilde\rho), B(\tilde\rho)\Sigma(\tilde\rho)B(\tilde\rho)^T),
$$

where

$$
B(\tilde\rho) := \begin{bmatrix} b_{11}(\tilde\rho) & -(\gamma_T^{(0)}(x_0))^{-\alpha-1} & 0 & (\gamma_T^{(0)}(x_0))^{-\alpha-2} & 0 \\ b_{21}(\tilde\rho) & -(\gamma_T^{(0)}(x_0))^{-\alpha-1} & (\gamma_T^{(0)}(x_0))^{-\alpha-1}(1-\tilde\rho) & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},
$$

and

$$
\begin{aligned}
b_{11}(\tilde\rho) &:= \frac{\alpha(1+\gamma_T^{(0)}(x_0))}{(\gamma_T(x_0))^{\alpha+1}[1+\alpha(1+\gamma_T^{(0)}(x_0))]^2}, \\
b_{21}(\tilde\rho) &:= \frac{\alpha\tilde\rho(1+\gamma_T^{(0)}(x_0))}{(\gamma_T(x_0))^{\alpha+1}[1+\alpha(1+\gamma_T^{(0)}(x_0))][1-\tilde\rho+\alpha(1+\gamma_T^{(0)}(x_0))]}.
\end{aligned}
$$

As for $\overline{\Delta}_{\alpha,u,v}(\gamma_T^{(0)}(x_0),0;\tilde\rho)$, $u,v=1,2$, we have by Theorems 1, 3 and 4, and because $|\widetilde\Delta_{\alpha,u,v,w}(\gamma,\delta;\tilde\rho)| \le M_{u,v,w}$, in some open neighborhood of $(\gamma_T^{(0)}(x_0),0)$ with $M_{u,v,w} = O_{\mathbb{P}}(1)$, $u,v,w=1,2$, that

$\overline{\Delta}_{\alpha,u,v}(\gamma_T^{(0)}(x_0), 0; \tilde{\rho}) \xrightarrow{P} \Delta_{\alpha,u,v}(\gamma_T^{(0)}(x_0), 0; \tilde{\rho})$, where

$$\Delta_{\alpha,1,1}(\gamma_T^{(0)}(x_0), 0; \tilde{\rho}) := (\gamma_T^{(0)}(x_0))^{-\alpha-2} \frac{1 + \alpha^2(1 + \gamma_T^{(0)}(x_0))^2}{[1 + \alpha(1 + \gamma_T^{(0)}(x_0))]^3},$$

$$\Delta_{\alpha,1,2}(\gamma_T^{(0)}(x_0), 0; \tilde{\rho}) := (\gamma_T^{(0)}(x_0))^{-\alpha-2} \frac{\tilde{\rho}(1 - \tilde{\rho})[1 + \alpha(1 + \gamma_T^{(0)}(x_0)) + \alpha^2(1 + \gamma_T^{(0)}(x_0))^2] + \alpha^3 \tilde{\rho}(1 + \gamma_T^{(0)}(x_0))^3}{[1 + \alpha(1 + \gamma_T^{(0)}(x_0))]^2 [1 - \tilde{\rho} + \alpha(1 + \gamma_T^{(0)}(x_0))]^2},$$

$$\Delta_{\alpha,2,2}(\gamma_T^{(0)}(x_0), 0; \tilde{\rho}) := (\gamma_T^{(0)}(x_0))^{-\alpha-2} \frac{(1 - \tilde{\rho})\tilde{\rho}^2 + \alpha \tilde{\rho}^2 (1 + \gamma_T^{(0)}(x_0))[\alpha(1 + \gamma_T^{(0)}(x_0)) - \tilde{\rho}]}{[1 + \alpha(1 + \gamma_T^{(0)}(x_0))][1 - \tilde{\rho} + \alpha(1 + \gamma_T^{(0)}(x_0))][1 - 2\tilde{\rho} + \alpha(1 + \gamma_T^{(0)}(x_0))]}.$$

Let

$$\Delta(\tilde{\rho}) := \begin{bmatrix} \Delta_{\alpha,1,1}(\gamma_T^{(0)}(x_0), 0; \tilde{\rho}) & \Delta_{\alpha,1,2}(\gamma_T^{(0)}(x_0), 0; \tilde{\rho}) \\ \Delta_{\alpha,1,2}(\gamma_T^{(0)}(x_0), 0; \tilde{\rho}) & \Delta_{\alpha,2,2}(\gamma_T^{(0)}(x_0), 0; \tilde{\rho}) \end{bmatrix}$$

and

$$\breve{\Delta}(\tilde{\rho}) := \begin{bmatrix} \Delta(\tilde{\rho}) & 0 \\ 0 & I_2 \end{bmatrix},$$

where I_2 is the (2×2) identity matrix. It can be verified that $\Delta(\tilde{\rho})$ is positive definite and thus invertible. Then, by Lemma 5.2 in Chapter 6 of Lehmann and Casella (1998), we have

$$r_n \begin{bmatrix} \widehat{\gamma}_{T,n}(x_0|\tilde{\rho}) - \gamma_T^{(0)}(x_0) \\ \widehat{\delta}_{T,n}(x_0|\tilde{\rho}) \\ \frac{T_n^{(1)}(K,0,0|x_0)}{\overline{F}_T(t_n|x_0) f_X(x_0)} - 1 \\ \frac{T_n^{(2)}(K|x_0)}{\overline{F}_T(t_n|x_0) f_X(x_0)} - \frac{\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)} \end{bmatrix} \rightsquigarrow N(\lambda \breve{\Delta}^{-1}(\tilde{\rho}) B(\tilde{\rho}) \mu(\tilde{\rho}), \breve{\Delta}^{-1}(\tilde{\rho}) B(\tilde{\rho}) \Sigma(\tilde{\rho}) B(\tilde{\rho})^T \breve{\Delta}^{-1}(\tilde{\rho})).$$

Finally, a straightforward application of the delta method gives

$$r_n(\widehat{\gamma}_{Y,n}(x_0|\tilde{\rho}) - \gamma_Y^{(0)}(x_0)) \rightsquigarrow N(\lambda L^T \breve{\Delta}^{-1}(\tilde{\rho}) B(\tilde{\rho}) \mu(\tilde{\rho}), L^T \breve{\Delta}^{-1}(\tilde{\rho}) B(\tilde{\rho}) \Sigma(\tilde{\rho}) B(\tilde{\rho})^T \breve{\Delta}^{-1}(\tilde{\rho}) L),$$

with $L^T := [\gamma_Y^{(0)}(x_0)/\gamma_T^{(0)}(x_0), 0, \gamma_Y^{(0)}(x_0), -(\gamma_Y^{(0)}(x_0))^2/\gamma_T^{(0)}(x_0)]$.

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., Bardoutsos, A., de Wet, T., Gijbels, I., 2016. Bias reduced tail estimation for censored Pareto type distributions. Statistics and Probability Letters, 109, 78–88.

[3] Beirlant, J., Guillou, A., Dierckx, G., Fils-Villetard, A., 2007. Estimation of the extreme value index and extreme quantiles under random censoring. Extremes, 10, 151–174.

[4] 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.

[5] Billingsley, P., 1995. Probability and measure. Wiley, New York.

[6] Byrd, R.H., Lu, P., Nocedal, J., Zhu, C., 1995. A limited memory algorithm for bound constrained optimization. SIAM Journal on Scientific Computing, 16, 1190–1208.

[7] Daouia, A., Gardes, L., Girard, S., 2013. On kernel smoothing for extremal quantile regression. Bernoulli, 19, 2557–2589.

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

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

[10] Dierckx, G., Goegebeur, Y., Guillou, A., 2013. An asymptotically unbiased minimum density power divergence estimator for the Pareto-tail index. Journal of Multivariate Analysis, 121, 70–86.

[11] Dierckx, G., Goegebeur, Y., Guillou, A., 2014. Local robust and asymptotically unbiased estimation of conditional Pareto-type tails. TEST, 23, 330–355.

[12] Drees, H. (1998). On smooth statistical tail functionals. Scandinavian Journal of Statistics 25, 187–210.

[13] Einmahl, J.H.J., Fils-Villetard, A., Guillou, A., 2008. Statistics of extremes under random censoring. Bernoulli, 14, 207–227.

[14] Escobar-Bach, M., Goegebeur, Y., Guillou, A., 2018. Local robust estimation of the Pickands dependence function. Accepted for publication in *Annals of Statistics*.

[15] Goegebeur, Y., Guillou, A., Osmann, M., 2014a. A local moment type estimator for the extreme value index in regression with random covariates. Canadian Journal of Statistics, 42, 487–507.

[16] Goegebeur, Y., Guillou, A., Qin, J., 2018. Bias-corrected estimation for conditional Pareto-type distributions with random right censoring. Submitted.

[17] Goegebeur, Y., Guillou, A., Schorgen, A., 2014b. Nonparametric regression estimation of conditional tails - the random covariate case. Statistics, 48, 732–755.

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

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

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

[21] Lehmann, E.L., Casella, G., 1998. Theory of Point Estimation. Springer.

[22] Ndao, P., Diop, A., Dupuy, J.-F., 2014. Nonparametric estimation of the conditional tail index and extreme quantiles under random censoring. Computational Statistics and Data Analysis, 79, 63–79.

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

[24] Stupfler, G., 2013. A moment estimator for the conditional extreme-value index. Electronic Journal of Statistics, 7, 2298–2353.

[25] Stupfler, G., 2016. Estimating the conditional extreme-value index under random right-censoring. Journal of Multivariate Analysis, 144, 1–24.

[26] van der Vaart, A.W., 1998. Asymptotic statistics. Cambridge University Press, Cambridge.

[27] Venables, W.N., Ripley, B.D., 2002. Modern applied statistics with S. 4th ed., New York, Springer.

[28] Wang, H., Li, D., He, X., 2012. Estimation of high conditional quantiles for heavy-tailed distributions. Journal of the American Statistical Association, 107, 1453–1464.

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

[30] Worms, J., Worms, R., 2014. New estimators of the extreme value index under random right censoring, for heavy-tailed distributions. Extremes, 17, 337–358.