



HAL
open science

Bias-corrected estimation for conditional Pareto-type distributions with random right censoring

Yuri Goegebeur, Armelle Guillou, Jing Qin

► **To cite this version:**

Yuri Goegebeur, Armelle Guillou, Jing Qin. Bias-corrected estimation for conditional Pareto-type distributions with random right censoring. *Extremes*, 2019, 22, pp.459-498. 10.1007/s10687-019-00341-7. hal-01826112

HAL Id: hal-01826112

<https://hal.science/hal-01826112>

Submitted on 29 Jun 2018

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.

Bias-corrected estimation for conditional Pareto-type distributions with random right censoring

Yuri Goegebeur⁽¹⁾, Armelle Guillou⁽²⁾, Jing Qin⁽¹⁾

⁽¹⁾ Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark

⁽²⁾ Institut Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg et CNRS, 7 rue René Descartes, 67084 Strasbourg cedex, France

Abstract. We consider bias-reduced estimation of the extreme value index in conditional Pareto-type models with random covariates when the response variable is subject to random right censoring. The bias-correction is obtained by fitting the extended Pareto distribution locally to the relative excesses over a high threshold using the maximum likelihood method. Consistency and asymptotic normality of the estimators are established under suitable assumptions. The finite sample behaviour is illustrated with a small simulation experiment and the method is applied to AIDS survival data.

Keywords: Pareto-type model, random covariate, random right censoring, local estimation, bias-reduction.

1 Introduction

In many studies it is more interesting to focus on the tail of a distribution rather than its centre. For instance insurance companies will typically be interested in modelling the large claims generated by a portfolio, as these may pose a major threat to the solvency of the company. Accurate modelling of the upper tail of the claim size distribution is also crucial when such companies enter re-insurance contracts, in particular for the pricing of such contracts. As a second example, the design of secure coastal defense structures is naturally based on analysis of extreme storm events. Next to extremes one is in practical applications also often faced with data that are incomplete due to censoring, meaning that only a lower bound for the variable of interest is available. In this paper, we will consider the analysis of tails of heavy tailed distributions when there are random covariates and when the responses are subject to random right censoring.

Let Y be the random variable of main interest, and let C be the censoring random variable, where both have distributions that are of Pareto-type, i.e. the distribution functions are of the form

$$1 - F_{\bullet}(y) = y^{-1/\gamma} l_{\bullet}(y), \quad y > 0,$$

with \bullet denoting either Y or C , and where γ_\bullet , called the extreme value index, is a positive parameter, and l_\bullet is a slowly varying function at infinity meaning that $l_\bullet(ty)/l_\bullet(t) \rightarrow 1$ as $t \rightarrow \infty$ for all $y > 0$. The class of Pareto-type models corresponds with the max-domain of attraction of extreme value distributions with positive extreme value index and finds important applications in areas like geology, climatology, finance and actuarial science. We refer to Beirlant *et al.* (2004) for some concrete practical examples where Pareto-type models have been successfully applied. In the present paper, interest is in estimating γ_Y , but due to censoring one observes $Y \wedge C$ rather than Y , along with a censoring indicator $\mathbb{1}_{\{Y \leq C\}}$. Estimation of the extreme value index based on censored data is considered by Beirlant *et al.* (2007), Einmahl *et al.* (2008), Gomes and Neves (2011), Worms and Worms (2014), Beirlant *et al.* (2016), to mention but a few.

The above described Pareto-type model will be extended to the situation where a random covariate X is observed together with the variable of main interest. In this case, we consider the conditional distributions of Y and C given $X = x$, and assume these are of Pareto-type, meaning that γ_Y , γ_C , l_Y and l_C will depend on x . The estimation of the positive extreme value index and extreme quantiles in presence of random covariates received a lot of attention in the extreme value literature, see e.g. Wang and Tsai (2009), Daouia *et al.* (2011), Dierckx *et al.* (2014), Goegebeur *et al.* (2014b). Also extensions to the estimation of the conditional extreme value index in the general max-domain of attraction have been considered, see Daouia *et al.* (2013), Stupfler (2013) and Goegebeur *et al.* (2014a). So far there are however very few contributions for the situation where one has random covariates and random right censoring of the response variable. Ndao *et al.* (2016) introduced a nonparametric estimator for $\gamma_Y(x) > 0$ in presence of censoring, which is obtained by taking one of the local estimators of Goegebeur *et al.* (2014b), developed for the uncensored case, and dividing it by the local proportion of non-censored observations that exceed a high threshold. Similarly, in the general max-domain of attraction, Stupfler (2016) adjusted the local moment estimator, introduced in Stupfler (2013) to the censoring case, by dividing it by the local proportion of non-censored observations. Neither the Stupfler (2016) nor the Ndao *et al.* (2016) estimators explicitly address the bias-issue that is typically present when estimating tail parameters. In the current paper, we will introduce a bias-corrected estimator for the conditional extreme value index when there is random right censoring and random covariates. The estimator is obtained by fitting the extended Pareto distribution locally with the maximum likelihood method to the response observations. In this, as is classical in the context of censoring (see e.g. Lawless, 2003), the likelihood contribution of non-censored observations consists of the density of the extended Pareto distribution, while censored observations contribute the survival function of the latter. A similar idea was applied by Beirlant *et al.* (2016) in the univariate extreme value context, and in a more restrictive Pareto-type model.

The remainder of the paper is organised as follows. In the next section, we introduce a second order Pareto-type model and the local likelihood estimation method. The asymptotic properties of the obtained estimators are studied in Section 3, where we prove consistency and asymptotic normality. The finite sample behaviour is examined by a simulation experiment in Section 4, while in Section 5 we illustrate the method to AIDS survival data. The Appendix contains the proofs of the results.

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. The conditional distributions of Y and C given $X = x$, denoted $F_Y(\cdot|x)$ and $F_C(\cdot|x)$, respectively, are assumed to be of the following form.

Assumption (D) For all $x \in S_X$, the conditional survival functions of Y and C satisfy, with \bullet denoting either Y or C ,

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

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

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

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

Note that our model (D) is more general than Beirlant *et al.* (2016), where the Hall-class of Pareto-type models was considered (Hall, 1982). Also, as mentioned before, Beirlant *et al.* (2016) is in the univariate context and does not consider the situation of random covariates.

Under assumption (D), $F_Y(\cdot|x)$ and $F_C(\cdot|x)$ have density functions. Indeed, straightforward differentiation gives

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

For distribution functions satisfying (D), one can approximate the conditional distribution function of $Z := Y/t$, given that $Y > t$, or $Z := C/t$ given that $C > t$, where t denotes a high threshold value, by the extended Pareto distribution (Beirlant *et al.*, 2004, 2009), with distribution function given by

$$G(z; \gamma, \delta, \beta) = \begin{cases} 1 - [z(1 + \delta - \delta z^{-\beta})]^{-1/\gamma}, & z > 1, \\ 0, & z \leq 1, \end{cases} \quad (2)$$

and density function

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

where $\gamma > 0$, $\beta > 0$, and $\delta > \max\{-1, -1/\beta\}$. Indeed, as shown in Beirlant *et al.* (2009), one has that

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

Clearly, based on this result, one can obtain an estimator for $\gamma_Y(x)$ by fitting the extended Pareto distribution to the relative excesses over a high threshold t_n with the maximum likelihood method, adjusted for local estimation and censoring. Let $(T_i, X_i, \mathbb{1}_{\{Y_i \leq C_i\}})$, $i = 1, \dots, n$, be independent copies of the random vector $(T, X, \mathbb{1}_{\{Y \leq C\}})$, and take $x_0 \in \text{Int}(S_X)$, the interior of S_X . We estimate $(\gamma_Y(x_0), \delta_Y(t_n|x_0))$ by maximising the following local log-likelihood function

$$\begin{aligned} \ell(\gamma_Y, \delta_Y; \beta) &:= \frac{1}{n} \sum_{i=1}^n K_{h_n}(x_0 - X_i) \left\{ \ln g \left(\frac{T_i}{t_n}; \gamma_Y, \delta_Y, \beta \right) \mathbb{1}_{\{Y_i \leq C_i\}} \right. \\ &\quad \left. + \ln \overline{G} \left(\frac{T_i}{t_n}; \gamma_Y, \delta_Y, \beta \right) \mathbb{1}_{\{Y_i > C_i\}} \right\} \mathbb{1}_{\{T_i > t_n\}}, \end{aligned} \quad (3)$$

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 $h_n \rightarrow 0$ if $n \rightarrow \infty$, $\mathbb{1}_{\{A\}}$ is the indicator function on the event A , and t_n is a local non-random threshold sequence satisfying $t_n \rightarrow \infty$ if $n \rightarrow \infty$. In this paper, we only estimate $\gamma_Y(x_0)$ and $\delta_Y(t_n|x_0)$ with the local likelihood method, while the parameter β will be fixed at some value. In the framework of bias-corrected estimation it is quite common to fix the second order rate parameter at some value, see e.g. Feuerverger and Hall (1999), Gomes and Martins (2004), Dutang *et al.* (2014, 2016). As we will see later, when using the correct value for this parameter one obtains theoretically asymptotically unbiased estimators, while one loses this unbiasedness when β is mis-specified, but in the latter case, the estimators perform typically still better than uncorrected estimators in practice.

3 Asymptotic properties

In this section, we derive the asymptotic properties of solutions $(\widehat{\gamma}_{Y,n}(x_0), \widehat{\delta}_{Y,n}(t_n|x_0))$ of the likelihood equations.

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

Assumption (H) *There exist positive constants M_{f_X} , M_{A_\bullet} , M_{γ_\bullet} , M_{B_\bullet} , M_{ε_\bullet} , η_{f_X} , η_{A_\bullet} , η_{γ_\bullet} ,*

η_{B_\bullet} and $\eta_{\varepsilon_\bullet}$, where \bullet denotes either Y or C , such that for all $x, z \in S_X$:

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

For the kernel function K we assume the following:

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

For the sequel it is instructive to note that 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:

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

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

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

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

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

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

Statistics of the type $T_n^{(j)}(K, s, s'|x_0)$, $j = 1, 2$, are the basic building blocks for studying the asymptotic behaviour of the estimators for $\gamma_Y(x_0)$ and $\delta_Y(t_n|x_0)$. Indeed, the estimating equations resulting from (3) depend only on statistics of these types.

As a first step we establish the asymptotic expansions of $\mathbb{E}(T_n^{(j)}(K, s, s'|x_0))$, $j = 1, 2$. 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 $x_0 \in \text{Int}(S_X)$ with $f_X(x_0) > 0$. Then if $t_n \rightarrow \infty$ and $h_n \rightarrow 0$ as $n \rightarrow \infty$ in such a way that $h_n^\eta \ln t_n \rightarrow 0$, we have the following asymptotic expansions

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

and

$$\begin{aligned} \mathbb{E}(T_n^{(2)}(K, s, s'|x_0)) &= \bar{F}_T(t_n|x_0)f_X(x_0)\Gamma(s'+1)\frac{\gamma_T^{s'+1}(x_0)}{\gamma_Y(x_0)} \left\{ \frac{1}{(1-s\gamma_T(x_0))^{s'+1}} \right. \\ &\quad + \frac{\delta_C(t_n|x_0)}{\gamma_C(x_0)} \left[\frac{1}{(1+\beta_C(x_0)\gamma_T(x_0)-s\gamma_T(x_0))^{s'+1}} - \frac{1}{(1-s\gamma_T(x_0))^{s'+1}} \right] (1+o(1)) \\ &\quad + \frac{\delta_Y(t_n|x_0)}{\gamma_Y(x_0)} \left[\frac{1}{(1+\beta_Y(x_0)\gamma_T(x_0)-s\gamma_T(x_0))^{s'+1}} - \frac{1}{(1-s\gamma_T(x_0))^{s'+1}} \right] (1+o(1)) \\ &\quad - \frac{\delta_Y(t_n|x_0)\varepsilon_Y(t_n|x_0)}{(1+\beta_Y(x_0)\gamma_T(x_0)-s\gamma_T(x_0))^{s'+1}} (1+o(1)) \\ &\quad \left. + O(h_n^{\eta_{f_X} \wedge \eta_{A_Y} \wedge \eta_{A_C}}) + O(h_n^{\eta_{\gamma_Y} \wedge \eta_{\gamma_C}} \ln t_n) \right\}. \end{aligned}$$

From Theorem 1 we obtain the convergence in probability of our main statistics.

Corollary 1 Under the assumptions of Theorem 1 and assuming additionally that $nh_n^d \bar{F}_T(t_n|x_0) \rightarrow \infty$ we have that

$$\frac{T_n^{(1)}(K, s, s'|x_0)}{\bar{F}_T(t_n|x_0)f_X(x_0)} \xrightarrow{\mathbb{P}} \frac{\Gamma(s'+1)\gamma_T^{s'}(x_0)}{(1-s\gamma_T(x_0))^{s'+1}}, \quad (4)$$

$$\frac{T_n^{(2)}(K, s, s'|x_0)}{\bar{F}_T(t_n|x_0)f_X(x_0)} \xrightarrow{\mathbb{P}} \frac{\Gamma(s'+1)\gamma_T^{s'+1}(x_0)}{\gamma_Y(x_0)(1-s\gamma_T(x_0))^{s'+1}}. \quad (5)$$

Next we turn to establishing consistency and asymptotic normality of our local likelihood estimators. From now on we denote the true values of $\gamma_Y(x_0)$ and $\delta_Y(t_n|x_0)$ as $\gamma_Y^{(0)}(x_0)$ and $\delta_Y^{(0)}(t_n|x_0)$, respectively. Similarly, the superscript '(0)' will also be used to denote the true value of the other parameters/functions that appear in $\bar{F}_Y(y|x)$, $\bar{F}_C(y|x)$ and $\bar{F}_T(y|x)$.

Theorem 2 Under the assumptions of Corollary 1, with probability tending to 1 there exists a sequence of solutions $(\hat{\gamma}_{Y,n}(x_0), \hat{\delta}_{Y,n}(t_n|x_0))$ to the likelihood equations such that

$$(\hat{\gamma}_{Y,n}(x_0), \hat{\delta}_{Y,n}(t_n|x_0) - \delta_Y^{(0)}(t_n|x_0)) \xrightarrow{\mathbb{P}} (\gamma_Y^{(0)}(x_0), 0).$$

In order to obtain the asymptotic normality of the consistent estimator sequence $(\hat{\gamma}_{Y,n}(x_0), \hat{\delta}_{Y,n}(t_n|x_0))$ we need a result concerning the joint convergence in distribution of several statistics of the form

$T_n^{(1)}(K, s, s'|x_0)$ and $T_n^{(2)}(K, s, s'|x_0)$. Let

$$\mathbb{T}_n := \frac{1}{\overline{F}_T(t_n|x_0)f_X(x_0)} \begin{bmatrix} T_n^{(1)}(K, s_1, s'_1|x_0) \\ \vdots \\ T_n^{(1)}(K, s_J, s'_J|x_0) \\ T_n^{(2)}(K, s_{J+1}, s'_{J+1}|x_0) \\ \vdots \\ T_n^{(2)}(K, s_{J+L}, s'_{J+L}|x_0) \end{bmatrix},$$

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

Theorem 3 *Under the assumptions of Theorem 1 and assuming additionally that $r_n \rightarrow \infty$ we have*

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

where Σ is a symmetric $((J+L) \times (J+L))$ matrix with elements

$$\begin{aligned} \sigma_{jk} &:= \frac{\|K\|_2^2 \Gamma(s'_j + s'_k + 1) \gamma_T^{(0)(s'_j + s'_k)}(x_0)}{(1 - (s_j + s_k) \gamma_T^{(0)}(x_0))^{s'_j + s'_k + 1}}, \quad j, k \in \{1, \dots, J\}, \\ \sigma_{jk} &:= \frac{\|K\|_2^2 \Gamma(s'_j + s'_k + 1) \gamma_T^{(0)(s'_j + s'_k + 1)}(x_0)}{\gamma_Y^{(0)}(x_0) (1 - (s_j + s_k) \gamma_T^{(0)}(x_0))^{s'_j + s'_k + 1}}, \quad j, k \in \{J+1, \dots, J+L\}, \\ \sigma_{jk} &:= \frac{\|K\|_2^2 \Gamma(s'_j + s'_k + 1) \gamma_T^{(0)(s'_j + s'_k + 1)}(x_0)}{\gamma_Y^{(0)}(x_0) (1 - (s_j + s_k) \gamma_T^{(0)}(x_0))^{s'_j + s'_k + 1}}, \quad j \in \{1, \dots, J\}, k \in \{J+1, \dots, J+L\}. \end{aligned}$$

Finally, we obtain the asymptotic normality of any consistent solution sequence of the likelihood equations.

Theorem 4 *Assume the conditions of Theorem 3 and additionally that $r_n h_n^{\eta_{fX} \wedge \eta_{AY} \wedge \eta_{AC}} \rightarrow 0$, $r_n h_n^{\eta_{\gamma_Y} \wedge \eta_{\gamma_C}} \ln t_n \rightarrow 0$ and $r_n \delta_T^{(0)}(t_n|x_0) \rightarrow \lambda \in \mathbb{R}$. Then for any consistent sequence of solutions $(\widehat{\gamma}_{Y,n}(x_0), \widehat{\delta}_{Y,n}(t_n|x_0))$ to the likelihood equations one has that*

$$r_n \begin{bmatrix} \widehat{\gamma}_{Y,n}(x_0) - \gamma_Y^{(0)}(x_0) \\ \widehat{\delta}_{Y,n}(t_n|x_0) - \delta_Y^{(0)}(t_n|x_0) \end{bmatrix} \rightsquigarrow N(\lambda \zeta, \Lambda),$$

where

$$\begin{aligned} \Lambda_{11} &:= \frac{\|K\|_2^2 \gamma_Y^{(0)3}(x_0) (1 + \beta \gamma_T^{(0)}(x_0))^2}{\beta^2 \gamma_T^{(0)3}(x_0)}, \\ \Lambda_{22} &:= \frac{\|K\|_2^2 \gamma_Y^{(0)}(x_0) (1 + \beta \gamma_T^{(0)}(x_0))^2 (1 + 2\beta \gamma_T^{(0)}(x_0))}{\beta^4 \gamma_T^{(0)3}(x_0)}, \\ \Lambda_{12} &:= \frac{\|K\|_2^2 \gamma_Y^{(0)2}(x_0) (1 + \beta \gamma_T^{(0)}(x_0)) (1 + 2\beta \gamma_T^{(0)}(x_0))}{\beta^3 \gamma_T^{(0)3}(x_0)}, \end{aligned}$$

and where ζ has elements

- Case 1: $\delta_Y^{(0)}(t_n|x_0)/\delta_C^{(0)}(t_n|x_0) \rightarrow \pm\infty$

$$\begin{aligned}\zeta_1 &:= \frac{\beta_T^{(0)}(x_0)\gamma_Y^{(0)2}(x_0)(1 + \beta\gamma_T^{(0)}(x_0))(\beta_T^{(0)}(x_0) - \beta)}{\beta\gamma_T^{(0)}(x_0)(1 + \beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0))(1 + \beta\gamma_T^{(0)}(x_0) + \beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0))}, \\ \zeta_2 &:= \frac{\gamma_Y^{(0)}(x_0)(\beta + \beta_T^{(0)}(x_0) + \beta^2\gamma_T^{(0)}(x_0) + 3\beta\beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0) + \beta^2\beta_T^{(0)}(x_0)\gamma_T^{(0)2}(x_0))(\beta_T^{(0)}(x_0) - \beta)}{\beta^2\gamma_T^{(0)}(x_0)(1 + \beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0))(1 + \beta\gamma_T^{(0)}(x_0) + \beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0))},\end{aligned}$$

- Case 2: $\delta_Y^{(0)}(t_n|x_0)/\delta_C^{(0)}(t_n|x_0) \rightarrow 0$

$$\zeta_1 = \zeta_2 := 0,$$

- Case 3: $\delta_Y^{(0)}(t_n|x_0)/\delta_C^{(0)}(t_n|x_0) \rightarrow a$, where $0 < |a| < \infty$

$$\begin{aligned}\zeta_1 &:= \frac{a\beta_T^{(0)}(x_0)\gamma_Y^{(0)2}(x_0)\gamma_C^{(0)}(x_0)(1 + \beta\gamma_T^{(0)}(x_0))(\beta_T^{(0)}(x_0) - \beta)}{\beta\gamma_T^{(0)}(x_0)(1 + \beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0))(1 + \beta\gamma_T^{(0)}(x_0) + \beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0))(\gamma_Y^{(0)}(x_0) + a\gamma_C^{(0)}(x_0))}, \\ \zeta_2 &:= \frac{a\gamma_Y^{(0)}(x_0)\gamma_C^{(0)}(x_0)(\beta + \beta_T^{(0)}(x_0) + \beta^2\gamma_T^{(0)}(x_0) + 3\beta\beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0) + \beta^2\beta_T^{(0)}(x_0)\gamma_T^{(0)2}(x_0))(\beta_T^{(0)}(x_0) - \beta)}{\beta^2\gamma_T^{(0)}(x_0)(1 + \beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0))(1 + \beta\gamma_T^{(0)}(x_0) + \beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0))(\gamma_Y^{(0)}(x_0) + a\gamma_C^{(0)}(x_0))}.\end{aligned}$$

Note that in case the correct value is used for the parameter β , so $\beta = \beta_T^{(0)}(x_0)$, then the mean of the limiting distribution is zero, and we say the estimator is asymptotically unbiased. If β is mis-specified then obviously one loses this unbiasedness but the estimator still continues to perform well in practice and outperforms estimators that have no bias-correction, as is illustrated in the simulation section.

4 A simulation study

In this section we evaluate the finite sample performance of our bias-corrected estimator with a small simulation study. We compare the proposed estimator with an estimator which is not corrected for bias, obtained by constraining δ_Y to zero in (3), leading to

$$\hat{\gamma}_{Y,n}^B(x_0) := \frac{\frac{1}{n} \sum_{i=1}^n K_{h_n}(x_0 - X_i) \ln \frac{T_i}{t_n} \mathbb{1}_{\{T_i > t_n\}}}{\frac{1}{n} \sum_{i=1}^n K_{h_n}(x_0 - X_i) \mathbb{1}_{\{Y_i \leq C_i T_i > t_n\}}}.$$

This can be seen as a local Hill-type estimator, as introduced by Goegebeur *et al.* (2014b), corrected for censoring by dividing it by the local proportion of uncensored observations among the excesses over t_n . An estimator of this type was already studied by Ndao *et al.* (2016). For K , we use the bi-quadratic kernel function given by

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

For the practical implementation of the estimators we have to select the bandwidth parameter h_n and the threshold t_n , where, as is common in extreme value statistics, we use for the latter the $(k + 1)$ -th largest response observation in the ball $B(x_0, h_n)$.

The parameter h_n is selected in a data driven way by using the following cross-validation criterion

$$h_{cv} := \arg \min_{h \in \mathcal{H}} \sum_{i=1}^n \sum_{j=1}^n \left(\mathbb{1}_{\{T_i \leq T_j\}} - \widehat{F}_{n,-i}(T_j | X_i) \right)^2,$$

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

$$\widehat{F}_{n,-i}(t|x) := \frac{\sum_{j=1, j \neq i}^n K_h(x - X_j) \mathbb{1}_{\{T_j \leq t\}}}{\sum_{j=1, j \neq i}^n K_h(x - X_j)},$$

is a kernel estimator for $F_T(t|x)$. This criterion was introduced in Yao (1999), and considered in an extreme value context by Daouia *et al.* (2011, 2013) and Escobar-Bach *et al.* (2018).

We assume that the conditional distributions of Y and C given $X = x$ are Burr distributions with

$$F_{\bullet}(y|x) = 1 - \left(\frac{\eta_{\bullet}(x)}{\eta_{\bullet}(x) + y^{\tau_{\bullet}(x)}} \right)^{\lambda_{\bullet}(x)}, \quad y > 0 \text{ and } \eta_{\bullet}(x), \lambda_{\bullet}(x), \tau_{\bullet}(x) > 0.$$

We have $\gamma_{\bullet}(x) = 1/(\lambda_{\bullet}(x)\tau_{\bullet}(x))$ and $\beta_{\bullet}(x) = \tau_{\bullet}(x)$. The covariate X is assumed to be uniformly distributed on $[0, 1]$. In our simulations we set $\eta_Y(x) = \eta_C(x) = 1$ and consider the following cases:

- Case 1: $\lambda_Y(x) = 2$, $\tau_C(x) = 4$, $\lambda_C(x) = 1$, and

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

- Case 2: $\lambda_Y(x) = 1$, $\tau_C(x) = 4$, $\lambda_C(x) = 1$, and

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

- Case 3: $\lambda_Y(x) = 2/3$, $\tau_C(x) = 4$, $\lambda_C(x) = 1$, and

$$\tau_Y(x) = 3 \left[\left(\frac{1}{10} + \sin(\pi x) \right) \left(\frac{11}{10} - \frac{1}{2} \exp \left(-64 \left(x - \frac{1}{2} \right)^2 \right) \right) \right]^{-1}.$$

- Case 4: $\lambda_Y(x) = 0.5$, $\tau_C(x) = 4$, $\lambda_C(x) = 1$, and

$$\tau_Y(x) = 4 \left[\left(\frac{1}{10} + \sin(\pi x) \right) \left(\frac{11}{10} - \frac{1}{2} \exp \left(-64 \left(x - \frac{1}{2} \right)^2 \right) \right) \right]^{-1}.$$

We simulate 100 datasets of size $n = 1000$. In Figure 1 we show the mean (left) and mean squared error, MSE, (right) of $\hat{\gamma}_{Y,n}(x_0)$ for $\beta = \beta_T^{(0)}(x_0)$ (solid line), $\beta = 2$ (dashed line) and $\hat{\gamma}_{Y,n}^B(x_0)$ (dashed-dotted line) as a function of k , when $x_0 = 0.4$. In the left panels the true value of $\gamma_Y(x_0)$ is represented by a horizontal line. The four rows in the figure correspond with the four cases mentioned above. Figure 2 is constructed similarly but now with $x_0 = 0.1$. Concerning the selection of h we note the following. In a first step we compute the optimal h for each dataset using the above mentioned cross validation criterion with $\mathcal{H} = \{0.05, 0.10, \dots, 0.30\}$. This implies that the range of k varies from one dataset to the next, so means and MSE's would be based on a different number of observations when plotted as a function of k . In order to avoid this we take the median of the h_{cv} values obtained in the 100 simulations and use this for all estimations. In Figure 3 we plot the proportion of censoring (computed as $\gamma_Y(x)/(\gamma_Y(x) + \gamma_C(x))$) as a function of x . In case of $x_0 = 0.1$ and $x_0 = 0.4$ we have a proportion of censoring of 0.47 and 0.64, respectively. These can be considered as very high values, and moreover in case $x_0 = 0.1$ we are near the boundary of the covariate support, implying that at both positions the estimation is difficult. From the simulation results reported in Figures 1 and 2 we can make the following conclusions

- In all the cases and for both positions our estimator reduces the bias considerably compared to the estimator $\hat{\gamma}_{Y,n}^B(x_0)$. In terms of MSE we are also very competitive and in general we have also a longer stable range of the MSE.
- After an extensive simulation study we observed that taking $\beta = 2$ gives overall good estimation results, whatever the cases and covariate position.
- Using the value $\beta = 2$ is competitive instead of using the unknown true value $\beta_T^{(0)}(x_0)$. To get an idea about the mis-specification, note that the true beta values are $\beta_T^{(0)}(0.4) = 1.14, 2.28, 3.41, 4$ and $\beta_T^{(0)}(0.1) = 2.22, 4, 4, 4$ in cases 1, 2, 3, 4, respectively.
- In extreme value statistics the bias of estimators is typically determined by a second order parameter $\rho_Y(x_0)$, linked to $\beta_Y(x_0)$ by $\rho_Y(x_0) = -\beta_Y(x_0)\gamma_Y(x_0)$, which is in the particular case of the Burr distribution equal to $\rho_Y(x_0) = -1/\lambda_Y(x_0)$. When $\rho_Y(x_0)$ is close to zero then estimators for tail parameters typically suffer from bias. In the figures the four cases are ordered from the more difficult ($\rho_Y(x_0) = -0.5$) to the more easy situation ($\rho_Y(x_0) = -2$). We can indeed observe that for the uncorrected estimator $\hat{\gamma}_{Y,n}^B(x_0)$, the bias is largest when $\rho_Y(x_0)$ is close to zero.
- Comparing the two positions $x_0 = 0.1$ and $x_0 = 0.4$ we can see that the estimation is more difficult near the boundary of the covariate space, but this is a well-known problem in nonparametric estimation. We observe though that the bias-corrected estimator still performs well.

5 Application to a real dataset

We illustrate in this section the proposed methodology with a dataset of patients diagnosed with AIDS in Australia before 1 July 1991. These data were provided by Dr P.J. Solomon and

the Australian National Centre in HIV Epidemiology and Clinical Research, see Venables and Ripley (2002). This dataset was formerly analysed by Einmahl *et al.* (2008) also in an extreme value context but without covariates, and more recently considered in Ndao *et al.* (2014) and Stupfler (2016) but this time in a regression context. 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, among them 2754 males, of which 1708 died. 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.

A scatterplot of the dataset is shown in Figure 4. From this figure, we can observe that most of the patients are between 20 and 65 years old at the time of diagnosis, so we only look at this interval of ages. Concerning the selection of h , we use the same procedure as in the simulation study with a grid of values in $[h_{min}, h_{max}]$ with

$$h_{min} = 0.05(x_{max} - x_{min}) \text{ and } h_{max} = 0.30(x_{max} - x_{min})$$

and six h -values evenly spaced. This grid is close to the one already used by Stupfler (2016), and with our procedure of selection it leads to the particular value $h = 11.25$.

In Figure 5 we estimate $\gamma_Y(x_0)$ at two positions $x_0 = 25$ and $x_0 = 50$ which correspond to positions close to the border of the x -values considered or close to the middle of the range of values of x , respectively. Our bias-corrected estimator $\hat{\gamma}_{Y,n}(x_0)$ is plotted in Figure 5 in full line whereas $\hat{\gamma}_{Y,n}^B(x_0)$ is in dashed line. For both estimators, as suggested in the simulation study, K is the bi-quadratic kernel and t_n is the $(k + 1)$ -th largest response observation in the ball $B(x_0, h)$. The second order rate parameter β is set at the value $\beta = 2$ as recommended also in the simulation study. It is clear from this figure that estimating $\gamma_Y(x_0)$ at $x_0 = 25$ is much more difficult than at position $x_0 = 50$. This corroborates what we have already observed in the simulation study for the position $x_0 = 0.1$. Despite, the estimation of $\gamma_Y(x_0)$ is very stable as a function of k for a wide range of k -values, on the contrary of $\hat{\gamma}_{Y,n}^B(x_0)$ which increases drastically without being stable at some stage.

Appendix

Derivatives of the log-likelihood function

Let $\ell_j(\gamma_Y, \delta_Y; \beta)$ and $\ell_{jk}(\gamma_Y, \delta_Y; \beta)$, $j, k = 1, 2$, denote the first and second order derivatives of $\ell(\gamma_Y, \delta_Y; \beta)$ with respect to γ_Y and δ_Y , respectively. Then

$$\begin{aligned}
\ell_1(\gamma_Y, \delta_Y; \beta) &= \frac{1}{n} \sum_{i=1}^n K_{h_n}(x_0 - X_i) \left[-\frac{1}{\gamma_Y} \mathbb{1}_{\{Y_i \leq C_i\}} + \frac{1}{\gamma_Y^2} \ln \frac{T_i}{t_n} + \frac{1}{\gamma_Y^2} \ln \left(1 + \delta_Y - \delta_Y \left(\frac{T_i}{t_n} \right)^{-\beta} \right) \right] \mathbb{1}_{\{T_i > t_n\}}, \\
\ell_2(\gamma_Y, \delta_Y; \beta) &= \frac{1}{n} \sum_{i=1}^n K_{h_n}(x_0 - X_i) \left[-\frac{1}{1 + \delta_Y - \delta_Y \left(\frac{T_i}{t_n} \right)^{-\beta}} \left(1 - \left(\frac{T_i}{t_n} \right)^{-\beta} \right) \mathbb{1}_{\{Y_i \leq C_i\}} \right. \\
&\quad + \frac{1}{1 + \delta_Y \left(1 - (1 - \beta) \left(\frac{T_i}{t_n} \right)^{-\beta} \right)} \left(1 - (1 - \beta) \left(\frac{T_i}{t_n} \right)^{-\beta} \right) \mathbb{1}_{\{Y_i \leq C_i\}} \\
&\quad \left. - \frac{1}{\gamma_Y} \frac{1}{1 + \delta_Y - \delta_Y \left(\frac{T_i}{t_n} \right)^{-\beta}} \left(1 - \left(\frac{T_i}{t_n} \right)^{-\beta} \right) \right] \mathbb{1}_{\{T_i > t_n\}}, \\
\ell_{11}(\gamma_Y, \delta_Y; \beta) &= \frac{1}{n} \sum_{i=1}^n K_{h_n}(x_0 - X_i) \left[\frac{1}{\gamma_Y^2} \mathbb{1}_{\{Y_i \leq C_i\}} - \frac{2}{\gamma_Y^3} \ln \frac{T_i}{t_n} - \frac{2}{\gamma_Y^3} \ln \left(1 + \delta_Y - \delta_Y \left(\frac{T_i}{t_n} \right)^{-\beta} \right) \right] \mathbb{1}_{\{T_i > t_n\}}, \\
\ell_{22}(\gamma_Y, \delta_Y; \beta) &= \frac{1}{n} \sum_{i=1}^n K_{h_n}(x_0 - X_i) \left[\frac{1}{\left(1 + \delta_Y - \delta_Y \left(\frac{T_i}{t_n} \right)^{-\beta} \right)^2} \left(1 - \left(\frac{T_i}{t_n} \right)^{-\beta} \right)^2 \mathbb{1}_{\{Y_i \leq C_i\}} \right. \\
&\quad - \frac{1}{\left(1 + \delta_Y \left(1 - (1 - \beta) \left(\frac{T_i}{t_n} \right)^{-\beta} \right) \right)^2} \left(1 - (1 - \beta) \left(\frac{T_i}{t_n} \right)^{-\beta} \right)^2 \mathbb{1}_{\{Y_i \leq C_i\}} \\
&\quad \left. + \frac{1}{\gamma_Y} \frac{1}{\left(1 + \delta_Y - \delta_Y \left(\frac{T_i}{t_n} \right)^{-\beta} \right)^2} \left(1 - \left(\frac{T_i}{t_n} \right)^{-\beta} \right)^2 \right] \mathbb{1}_{\{T_i > t_n\}}, \\
\ell_{12}(\gamma_Y, \delta_Y; \beta) &= \frac{1}{n} \sum_{i=1}^n K_{h_n}(x_0 - X_i) \frac{1}{\gamma_Y^2} \frac{1}{1 + \delta_Y - \delta_Y \left(\frac{T_i}{t_n} \right)^{-\beta}} \left(1 - \left(\frac{T_i}{t_n} \right)^{-\beta} \right) \mathbb{1}_{\{T_i > t_n\}}.
\end{aligned}$$

Proof of Theorem 1

We focus on deriving the asymptotic expansion for $\mathbb{E}(T_n^{(2)}(K, s, s'|x_0))$. 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, allowing us to obtain a more precise statement of the remainder terms using the Hölder exponents from Assumption (\mathcal{H}) .

We have

$$\begin{aligned}
\mathbb{E}(T_n^{(2)}(K, s, s'|x_0)) &= \mathbb{E} \left[K_{h_n}(x_0 - X) \left(\frac{T}{t_n} \right)^s \left(\ln_+ \frac{T}{t_n} \right)^{s'} \mathbb{1}_{\{Y \leq C, T > t_n\}} \right] \\
&= \mathbb{E} \left[K_{h_n}(x_0 - X) \mathbb{E} \left(\left(\frac{T}{t_n} \right)^s \left(\ln_+ \frac{T}{t_n} \right)^{s'} \mathbb{1}_{\{Y \leq C, T > t_n\}} \middle| X \right) \right] \\
&= \mathbb{E} \left[K_{h_n}(x_0 - X) \int_{t_n}^{\infty} \left(\frac{y}{t_n} \right)^s \left(\ln \frac{y}{t_n} \right)^{s'} f_Y(y|X) \bar{F}_C(y|X) dy \right] \\
&= \int_{\mathbb{R}^d} K_{h_n}(x_0 - u) \int_{t_n}^{\infty} \left(\frac{y}{t_n} \right)^s \left(\ln \frac{y}{t_n} \right)^{s'} f_Y(y|u) \bar{F}_C(y|u) dy f_X(u) du \\
&= \int_{S_K} K(z) \int_{t_n}^{\infty} \left(\frac{y}{t_n} \right)^s \left(\ln \frac{y}{t_n} \right)^{s'} f_Y(y|x_0 - h_n z) \bar{F}_C(y|x_0 - h_n z) dy f_X(x_0 - h_n z) dz.
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}(T_n^{(2)}(K, s, s'|x_0)) &= f_X(x_0) \int_{t_n}^{\infty} \left(\frac{y}{t_n} \right)^s \left(\ln \frac{y}{t_n} \right)^{s'} f_Y(y|x_0) \bar{F}_C(y|x_0) dy \\
&\quad + \int_{t_n}^{\infty} \left(\frac{y}{t_n} \right)^s \left(\ln \frac{y}{t_n} \right)^{s'} f_Y(y|x_0) \bar{F}_C(y|x_0) dy \int_{S_K} K(z) (f_X(x_0 - h_n z) - f_X(x_0)) dz \\
&\quad + f_X(x_0) \int_{S_K} K(z) \int_{t_n}^{\infty} \left(\frac{y}{t_n} \right)^s \left(\ln \frac{y}{t_n} \right)^{s'} f_Y(y|x_0) (\bar{F}_C(y|x_0 - h_n z) - \bar{F}_C(y|x_0)) dy dz \\
&\quad + \int_{S_K} K(z) \int_{t_n}^{\infty} \left(\frac{y}{t_n} \right)^s \left(\ln \frac{y}{t_n} \right)^{s'} f_Y(y|x_0) (\bar{F}_C(y|x_0 - h_n z) - \bar{F}_C(y|x_0)) dy (f_X(x_0 - h_n z) - f_X(x_0)) dz \\
&\quad + f_X(x_0) \int_{S_K} K(z) \int_{t_n}^{\infty} \left(\frac{y}{t_n} \right)^s \left(\ln \frac{y}{t_n} \right)^{s'} (f_Y(y|x_0 - h_n z) - f_Y(y|x_0)) \bar{F}_C(y|x_0) dy dz \\
&\quad + \int_{S_K} K(z) \int_{t_n}^{\infty} \left(\frac{y}{t_n} \right)^s \left(\ln \frac{y}{t_n} \right)^{s'} (f_Y(y|x_0 - h_n z) - f_Y(y|x_0)) \bar{F}_C(y|x_0) dy (f_X(x_0 - h_n z) - f_X(x_0)) dz \\
&\quad + f_X(x_0) \int_{S_K} K(z) \int_{t_n}^{\infty} \left(\frac{y}{t_n} \right)^s \left(\ln \frac{y}{t_n} \right)^{s'} (f_Y(y|x_0 - h_n z) - f_Y(y|x_0)) (\bar{F}_C(y|x_0 - h_n z) - \bar{F}_C(y|x_0)) dy dz \\
&\quad + \int_{S_K} K(z) \int_{t_n}^{\infty} \left(\frac{y}{t_n} \right)^s \left(\ln \frac{y}{t_n} \right)^{s'} (f_Y(y|x_0 - h_n z) - f_Y(y|x_0)) (\bar{F}_C(y|x_0 - h_n z) - \bar{F}_C(y|x_0)) dy \\
&\quad \quad \quad (f_X(x_0 - h_n z) - f_X(x_0)) dz \\
&=: T_1 + \dots + T_8.
\end{aligned}$$

Concerning T_1 we have

$$T_1 = t_n f_Y(t_n|x_0) \bar{F}_C(t_n|x_0) f_X(x_0) \int_1^{\infty} z^s (\ln z)^{s'} \frac{f_Y(t_n z|x_0)}{f_Y(t_n|x_0)} \frac{\bar{F}_C(t_n z|x_0)}{\bar{F}_C(t_n|x_0)} dz.$$

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

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

This leads to the decomposition

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

From (1) we can write

$$\begin{aligned} T_{1,1} &= \int_1^{\infty} z^{s-1/\gamma_Y(x_0)-1} (\ln z)^{s'} \overline{G}(z; \gamma_C(x_0), \delta_C(t_n | x_0), \beta_C(x_0)) dz \\ &+ \frac{1}{1 + \left(\frac{1}{\gamma_Y(x_0)} - \varepsilon_Y(t_n | x_0) \right) \delta_Y(t_n | x_0)} \left[\right. \\ &\frac{\delta_Y(t_n | x_0)}{\gamma_Y(x_0)} \int_1^{\infty} z^{s-1/\gamma_Y(x_0)-1} (\ln z)^{s'} (z^{-\beta_Y(x_0)} - 1) \overline{G}(z; \gamma_C(x_0), \delta_C(t_n | x_0), \beta_C(x_0)) dz \\ &+ \frac{\delta_Y(t_n | x_0)}{\gamma_Y(x_0)} \int_1^{\infty} z^{s-1/\gamma_Y(x_0)-1} (\ln z)^{s'} \left(\frac{\delta_Y(t_n z | x_0)}{\delta_Y(t_n | x_0)} - z^{-\beta_Y(x_0)} \right) \overline{G}(z; \gamma_C(x_0), \delta_C(t_n | x_0), \beta_C(x_0)) dz \\ &- \varepsilon_Y(t_n | x_0) \delta_Y(t_n | x_0) \int_1^{\infty} z^{s-1/\gamma_Y(x_0)-1} (\ln z)^{s'} \left(\frac{\varepsilon_Y(t_n z | x_0)}{\varepsilon_Y(t_n | x_0)} - 1 \right) \frac{\delta_Y(t_n z | x_0)}{\delta_Y(t_n | x_0)} \overline{G}(z; \gamma_C(x_0), \delta_C(t_n | x_0), \beta_C(x_0)) dz \\ &\left. - \varepsilon_Y(t_n | x_0) \delta_Y(t_n | x_0) \int_1^{\infty} z^{s-1/\gamma_Y(x_0)-1} (\ln z)^{s'} \left(\frac{\delta_Y(t_n z | x_0)}{\delta_Y(t_n | x_0)} - 1 \right) \overline{G}(z; \gamma_C(x_0), \delta_C(t_n | x_0), \beta_C(x_0)) dz \right] \\ &=: T_{1,1,1} + \frac{1}{1 + \left(\frac{1}{\gamma_Y(x_0)} - \varepsilon_Y(t_n | x_0) \right) \delta_Y(t_n | x_0)} (T_{1,1,2} + \dots + T_{1,1,5}). \end{aligned}$$

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

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

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

A straightforward calculation gives then

$$\begin{aligned}
T_{1,1,1} &= \frac{\Gamma(s'+1)\gamma_T^{s'+1}(x_0)}{(1-s\gamma_T(x_0))^{s'+1}} + \delta_C(t_n|x_0) \frac{\Gamma(s'+1)\gamma_T^{s'+1}(x_0)}{\gamma_C(x_0)} \times \\
&\quad \left[\frac{1}{(1+\beta_C(x_0)\gamma_T(x_0)-s\gamma_T(x_0))^{s'+1}} - \frac{1}{(1-s\gamma_T(x_0))^{s'+1}} \right] + O(\delta_C^2(t_n|x_0)), \\
T_{1,1,2} &= \delta_Y(t_n|x_0) \frac{\Gamma(s'+1)\gamma_T^{s'+1}(x_0)}{\gamma_Y(x_0)} \left[\frac{1}{(1+\beta_Y(x_0)\gamma_T(x_0)-s\gamma_T(x_0))^{s'+1}} - \frac{1}{(1-s\gamma_T(x_0))^{s'+1}} \right] \\
&\quad + O(\delta_Y(t_n|x_0)\delta_C(t_n|x_0)).
\end{aligned}$$

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 $0 < \delta < 1/\gamma_T(x_0) + \beta_Y(x_0) - s$, arbitrary, and n sufficiently large, we have

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

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

$$\begin{aligned}
T_{1,1,5} &= -\varepsilon_Y(t_n|x_0)\delta_Y(t_n|x_0)\Gamma(s'+1)\gamma_T^{s'+1}(x_0) \left[\frac{1}{(1+\beta_Y(x_0)\gamma_T(x_0)-s\gamma_T(x_0))^{s'+1}} - \frac{1}{(1-s\gamma_T(x_0))^{s'+1}} \right] \\
&\quad + o(\delta_Y(t_n|x_0)).
\end{aligned}$$

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

Collecting the terms gives then

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

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

$$\begin{aligned}
T_1 &= \bar{F}_T(t_n|x_0)f_X(x_0)\Gamma(s'+1)\frac{\gamma_T^{s'+1}(x_0)}{\gamma_Y(x_0)}\left\{\frac{1}{(1-s\gamma_T(x_0))^{s'+1}}\right. \\
&\quad + \frac{\delta_C(t_n|x_0)}{\gamma_C(x_0)}\left[\frac{1}{(1+\beta_C(x_0)\gamma_T(x_0)-s\gamma_T(x_0))^{s'+1}}-\frac{1}{(1-s\gamma_T(x_0))^{s'+1}}\right](1+o(1)) \\
&\quad + \delta_Y(t_n|x_0)\left(\frac{1}{\gamma_Y(x_0)}-\varepsilon_Y(t_n|x_0)\right)\left[\frac{1}{(1+\beta_Y(x_0)\gamma_T(x_0)-s\gamma_T(x_0))^{s'+1}}-\frac{1}{(1-s\gamma_T(x_0))^{s'+1}}\right](1+o(1)) \\
&\quad \left.-\delta_Y(t_n|x_0)\varepsilon_Y(t_n|x_0)\frac{1}{(1-s\gamma_T(x_0))^{s'+1}}(1+o(1))\right\}.
\end{aligned}$$

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

By rearranging terms we obtain the following bound for T_3

$$\begin{aligned}
|T_3| &\leq f_Y(t_n|x_0)\bar{F}_C(t_n|x_0)f_X(x_0)\times \\
&\quad \int_{S_K} K(z)\int_{t_n}^{\infty}\left(\frac{y}{t_n}\right)^s\left(\ln\frac{y}{t_n}\right)^{s'}\frac{f_Y(y|x_0)}{f_Y(t_n|x_0)}\frac{\bar{F}_C(y|x_0)}{\bar{F}_C(t_n|x_0)}\left|\frac{\bar{F}_C(y|x_0-h_nz)}{\bar{F}_C(y|x_0)}-1\right|dydz, \quad (6)
\end{aligned}$$

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

$$\begin{aligned}
\left|\frac{\bar{F}_C(y|x_0-h_nz)}{\bar{F}_C(y|x_0)}-1\right| &\leq M_1\left(h_n^{\eta_{AC}}+y^{M_2h_n^{\eta_{\gamma C}}}h_n^{\eta_{\gamma C}}\ln y+|\delta_C(y|x_0)|h_n^{\eta_{BC}}\right. \\
&\quad \left.+|\delta_C(y|x_0)|y^{M_3h_n^{\eta_{\varepsilon C}}}h_n^{\eta_{\varepsilon C}}\ln y\right).
\end{aligned}$$

Plugging the above inequality into (6), and computing integrals similar to those encountered above yields

$$T_3 = O\left(\bar{F}_T(t_n|x_0)(h_n^{\eta_{AC}}+h_n^{\eta_{\gamma C}}\ln t_n+\delta_C(t_n|x_0)h_n^{\eta_{BC}}+\delta_C(t_n|x_0)h_n^{\eta_{\varepsilon C}}\ln t_n)\right).$$

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

$$\begin{aligned}
|T_5| &\leq f_Y(t_n|x_0)\bar{F}_C(t_n|x_0)f_X(x_0)\times \\
&\quad \int_{S_K} K(z)\int_{t_n}^{\infty}\left(\frac{y}{t_n}\right)^s\left(\ln\frac{y}{t_n}\right)^{s'}\frac{f_Y(y|x_0)}{f_Y(t_n|x_0)}\frac{\bar{F}_C(y|x_0)}{\bar{F}_C(t_n|x_0)}\left|\frac{f_Y(y|x_0-h_nz)}{f_Y(y|x_0)}-1\right|dydz,
\end{aligned}$$

which, combined with the inequality

$$\begin{aligned}
\left|\frac{f_Y(y|x_0-h_nz)}{f_Y(y|x_0)}-1\right| &\leq M_1\left(h_n^{\eta_{AY}}+y^{M_2h_n^{\eta_{\gamma Y}}}h_n^{\eta_{\gamma Y}}\ln y+|\delta_Y(y|x_0)|h_n^{\eta_{BY}}\right. \\
&\quad \left.+|\delta_Y(y|x_0)|y^{M_3h_n^{\eta_{\varepsilon Y}}}h_n^{\eta_{\varepsilon Y}}\ln y\right),
\end{aligned}$$

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

$$T_5 = O\left(\overline{F}_T(t_n|x_0)(h_n^{\eta_{AY}} + h_n^{\eta_{\gamma Y}} \ln t_n + \delta_Y(t_n|x_0)h_n^{\eta_{BY}} + \delta_Y(t_n|x_0)h_n^{\eta_{\varepsilon Y}} \ln t_n)\right).$$

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 Corollary 1

We only give the details for $T_n^{(1)}(K, s, s'|x_0)/(\overline{F}_T(t_n|x_0)f_X(x_0))$. The convergence in probability of $T_n^{(2)}(K, s, s'|x_0)/(\overline{F}_T(t_n|x_0)f_X(x_0))$ can be established analogously.

From Theorem 1 we have immediately

$$\mathbb{E}\left(\frac{T_n^{(1)}(K, s, s'|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)}\right) \rightarrow \frac{\Gamma(s'+1)\gamma_T^{s'}(x_0)}{(1-s\gamma_T(x_0))^{s'+1}}.$$

By independence

$$\mathbb{V}ar\left(\frac{T_n^{(1)}(K, s, s'|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)}\right) = \frac{\mathbb{V}ar\left(K_{h_n}(x_0 - X)\left(\frac{T}{t_n}\right)^s\left(\ln_+\frac{T}{t_n}\right)^{s'}\mathbb{1}_{\{T>t_n\}}\right)}{n\overline{F}_T^2(t_n|x_0)f_X^2(x_0)}.$$

Using (\mathcal{K}) and Theorem 1 we obtain

$$\begin{aligned} & \mathbb{V}ar\left(K_{h_n}(x_0 - X)\left(\frac{T}{t_n}\right)^s\left(\ln_+\frac{T}{t_n}\right)^{s'}\mathbb{1}_{\{T>t_n\}}\right) \\ &= \frac{\|K\|_2^2}{h_n^d}\mathbb{E}\left(\frac{1}{\|K\|_2^2 h_n^d}K^2\left(\frac{x_0 - X}{h_n}\right)\left(\frac{T}{t_n}\right)^{2s}\left(\ln_+\frac{T}{t_n}\right)^{2s'}\mathbb{1}_{\{T>t_n\}}\right) + O(\overline{F}_T^2(t_n|x_0)) \\ &= \overline{F}_T(t_n|x_0)f_X(x_0)\frac{\|K\|_2^2\Gamma(2s'+1)\gamma_T^{2s'}(x_0)}{h_n^d(1-2s\gamma_T(x_0))^{2s'+1}}(1+o(1)), \end{aligned}$$

and thus

$$\mathbb{V}ar\left(\frac{T_n^{(1)}(K, s, s'|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)}\right) = \frac{1}{nh_n^d\overline{F}_T(t_n|x_0)f_X(x_0)}\frac{\|K\|_2^2\Gamma(2s'+1)\gamma_T^{2s'}(x_0)}{(1-2s\gamma_T(x_0))^{2s'+1}}(1+o(1)),$$

which tends to zero under the assumption $nh_n^d\overline{F}_T(t_n|x_0) \rightarrow \infty$. This establishes (4).

Proof of Theorem 2

To prove consistency we adapt the argument of the proof of Theorem 5.1 in Chapter 6 of Lehmann and Casella (1998) to our context. Let $\tilde{\ell}(\gamma_Y, \delta_Y; \beta) := \ell(\gamma_Y, \delta_Y; \beta) / (\overline{F}_T(t_n|x_0) f_X(x_0))$, and similarly for $\tilde{\ell}_j$ and $\tilde{\ell}_{jk}$. We denote by ℓ_{jkl} , $j, k, l = 1, 2$, the third order derivatives of ℓ , and by $\tilde{\ell}_{jkl}$ their rescaled version. As a first step we show that for any $\nu > 0$ sufficiently small

$$\mathbb{P}_{(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0))}(\tilde{\ell}(\gamma_Y, \delta_Y; \beta) < \tilde{\ell}(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta)) \text{ for all } (\gamma_Y, \delta_Y) \text{ on the surface of } Q_{\nu, n} \rightarrow 1$$

as $n \rightarrow \infty$, where $Q_{\nu, n}$ denotes the circle centered at $(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0))$ with radius ν , and $Q_{\nu, n}$ is assumed to belong to the parameter space. By Taylor's theorem

$$\begin{aligned} & \tilde{\ell}(\gamma_Y, \delta_Y; \beta) - \tilde{\ell}(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) \\ &= \tilde{\ell}_1(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta)(\gamma_Y - \gamma_Y^{(0)}(x_0)) + \tilde{\ell}_2(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta)(\delta_Y - \delta_Y^{(0)}(t_n|x_0)) \\ & \quad + \frac{1}{2}[\tilde{\ell}_{11}(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta)(\gamma_Y - \gamma_Y^{(0)}(x_0))^2 + \tilde{\ell}_{22}(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta)(\delta_Y - \delta_Y^{(0)}(t_n|x_0))^2 \\ & \quad + 2\tilde{\ell}_{12}(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta)(\gamma_Y - \gamma_Y^{(0)}(x_0))(\delta_Y - \delta_Y^{(0)}(t_n|x_0))] \\ & \quad + \frac{1}{6}[\tilde{\ell}_{111}(\tilde{\gamma}_Y, \tilde{\delta}_Y; \beta)(\gamma_Y - \gamma_Y^{(0)}(x_0))^3 + \tilde{\ell}_{222}(\tilde{\gamma}_Y, \tilde{\delta}_Y; \beta)(\delta_Y - \delta_Y^{(0)}(t_n|x_0))^3 \\ & \quad + 3\tilde{\ell}_{112}(\tilde{\gamma}_Y, \tilde{\delta}_Y; \beta)(\gamma_Y - \gamma_Y^{(0)}(x_0))^2(\delta_Y - \delta_Y^{(0)}(t_n|x_0)) \\ & \quad + 3\tilde{\ell}_{122}(\tilde{\gamma}_Y, \tilde{\delta}_Y; \beta)(\gamma_Y - \gamma_Y^{(0)}(x_0))(\delta_Y - \delta_Y^{(0)}(t_n|x_0))^2] \\ &=: S_1 + S_2 + S_3, \end{aligned}$$

where $(\tilde{\gamma}_Y, \tilde{\delta}_Y)$ is a point on the line segment connecting (γ_Y, δ_Y) and $(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0))$.

After linearisation, $\tilde{\ell}_1$ and $\tilde{\ell}_2$ can be written in terms of $T_n^{(1)}(K, s, s'|x_0)$ and $T_n^{(2)}(K, s, s'|x_0)$ as follows

$$\begin{aligned} \tilde{\ell}_1(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) &= -\frac{1}{\gamma_Y^{(0)}(x_0)} \frac{T_n^{(2)}(K, 0, 0|x_0)}{\overline{F}_T(t_n|x_0) f_X(x_0)} + \frac{1}{\gamma_Y^{(0)2}(x_0)} \frac{T_n^{(1)}(K, 0, 1|x_0)}{\overline{F}_T(t_n|x_0) f_X(x_0)} \\ & \quad + \delta_Y^{(0)}(t_n|x_0) \frac{1}{\gamma_Y^{(0)2}(x_0)} \left[\frac{T_n^{(1)}(K, 0, 0|x_0)}{\overline{F}_T(t_n|x_0) f_X(x_0)} - \frac{T_n^{(1)}(K, -\beta, 0|x_0)}{\overline{F}_T(t_n|x_0) f_X(x_0)} \right] \\ & \quad + O_{\mathbb{P}}(\delta_Y^{(0)2}(t_n|x_0)), \end{aligned} \tag{7}$$

$$\begin{aligned}
\tilde{\ell}_2(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) &= \beta \frac{T_n^{(2)}(K, -\beta, 0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)} - \frac{1}{\gamma_Y^{(0)}(x_0)} \frac{T_n^{(1)}(K, 0, 0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)} + \frac{1}{\gamma_Y^{(0)}(x_0)} \frac{T_n^{(1)}(K, -\beta, 0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)} \\
&+ \delta_Y^{(0)}(t_n|x_0) \left[(2\beta - \beta^2) \frac{T_n^{(2)}(K, -2\beta, 0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)} - 2\beta \frac{T_n^{(2)}(K, -\beta, 0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)} \right. \\
&+ \frac{1}{\gamma_Y^{(0)}(x_0)} \frac{T_n^{(1)}(K, 0, 0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)} + \frac{1}{\gamma_Y^{(0)}(x_0)} \frac{T_n^{(1)}(K, -2\beta, 0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)} \\
&\left. - \frac{2}{\gamma_Y^{(0)}(x_0)} \frac{T_n^{(1)}(K, -\beta, 0|x_0)}{\overline{F}_T(t_n|x_0)f_X(x_0)} \right] + O_{\mathbb{P}}(\delta_Y^{(0)2}(t_n|x_0)). \tag{8}
\end{aligned}$$

Using Corollary 1 we have

$$\tilde{\ell}_1(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) \xrightarrow{\mathbb{P}} 0 \text{ and } \tilde{\ell}_2(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) \xrightarrow{\mathbb{P}} 0,$$

so, for any $\nu > 0$ we have that $|\tilde{\ell}_j(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta)| < \nu^2$, $j = 1, 2$, with probability tending to 1, and hence on $Q_{\nu, n}$, $|S_1| < 2\nu^3$ with probability tending to 1.

Using again linearisation and Corollary 1 one has

$$\begin{aligned}
\tilde{\ell}_{11}(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) &\xrightarrow{\mathbb{P}} \dot{\ell}_{11} := -\frac{\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)3}(x_0)}, \\
\tilde{\ell}_{22}(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) &\xrightarrow{\mathbb{P}} \dot{\ell}_{22} := -\frac{\beta^2 \gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)(1 + 2\beta \gamma_T^{(0)}(x_0))}, \\
\tilde{\ell}_{12}(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) &\xrightarrow{\mathbb{P}} \dot{\ell}_{12} := \frac{\beta \gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)2}(x_0)(1 + \beta \gamma_T^{(0)}(x_0))}.
\end{aligned}$$

Now write

$$\begin{aligned}
2S_2 &= \dot{\ell}_{11}(\gamma_Y - \gamma_Y^{(0)}(x_0))^2 + \dot{\ell}_{22}(\delta_Y - \delta_Y^{(0)}(t_n|x_0))^2 + 2\dot{\ell}_{12}(\gamma_Y - \gamma_Y^{(0)}(x_0))(\delta_Y - \delta_Y^{(0)}(t_n|x_0)) \\
&+ (\tilde{\ell}_{11}(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) - \dot{\ell}_{11})(\gamma_Y - \gamma_Y^{(0)}(x_0))^2 \\
&+ (\tilde{\ell}_{22}(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) - \dot{\ell}_{22})(\delta_Y - \delta_Y^{(0)}(t_n|x_0))^2 \\
&+ 2(\tilde{\ell}_{12}(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) - \dot{\ell}_{12})(\gamma_Y - \gamma_Y^{(0)}(x_0))(\delta_Y - \delta_Y^{(0)}(t_n|x_0)).
\end{aligned}$$

It can easily be verified that the first three terms are a deterministic negative definite quadratic form in $\gamma_Y - \gamma_Y^{(0)}(x_0)$ and $\delta_Y - \delta_Y^{(0)}(t_n|x_0)$. By the spectral decomposition this quadratic form can be rewritten as $\lambda_1 \xi_1^2 + \lambda_2 \xi_2^2$, where $\lambda_2 \leq \lambda_1 < 0$ are the eigenvalues and ξ_1 and ξ_2 are the orthogonal transformations of $\gamma_Y - \gamma_Y^{(0)}(x_0)$ and $\delta_Y - \delta_Y^{(0)}(t_n|x_0)$. Note that in the new coordinate system $Q_{\nu, n}$ becomes $\xi_1^2 + \xi_2^2 = \nu^2$. Thus $\lambda_1 \xi_1^2 + \lambda_2 \xi_2^2 \leq \lambda_1 \nu^2$. For the random terms of S_2 we use the above convergences in probability and conclude that the random part is in absolute value less than $4\nu^3$ with probability tending to 1. Combined, we have that there exists $c > 0$

and $\nu_0 > 0$ such that for $\nu < \nu_0$ one has $S_2 < -c\nu^2$ with probability tending to 1.

For the third order derivatives one has that $|\tilde{\ell}_{jkl}(\gamma_Y, \delta_Y; \beta)| \leq M_{jkl}(\mathbf{V})$ for $(\gamma_Y, \delta_Y) \in Q_{\nu,n}$, where $\mathbf{V} := \{(T_i, X_i, \mathbb{1}_{\{Y_i \leq C_i\}}), i = 1, \dots, n\}$, and $M_{jkl} \xrightarrow{\mathbb{P}} m_{jkl}$, bounded, $j, k, l = 1, 2$. These derivations are straightforward and for brevity omitted from the paper. We have then that, with probability tending to one, $|\tilde{\ell}_{jkl}(\tilde{\gamma}_Y, \tilde{\delta}_Y; \beta)| < 2m_{jkl}$ and thus $|S_3| < e\nu^3$ on $Q_{\nu,n}$, where $e = 1/3 \sum_{j=1}^2 \sum_{k=1}^2 \sum_{l=1}^2 m_{jkl}$.

Combining the above, with probability tending to 1,

$$\max_{Q_{\nu,n}}(S_1 + S_2 + S_3) < -c\nu^2 + (2 + e)\nu^3,$$

which is negative if $\nu < c/(2 + e)$.

To establish existence and consistency we adjust the line of argumentation in the proof of Theorem 3.7 in Chapter 6 of Lehmann and Casella (1998). For $\nu > 0$, small enough such that $Q_{\nu,n}$ is a subset of the parameter space, consider

$$S_n(\nu) := \{\mathbf{v} : \tilde{\ell}(\gamma_Y, \delta_Y; \beta) < \tilde{\ell}(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) \text{ for all } (\gamma_Y, \delta_Y) \text{ on the surface of } Q_{\nu,n}\}.$$

From the above we have that $\mathbb{P}_{(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0))}(S_n(\nu)) \rightarrow 1$ for any such ν , and hence there exists a sequence $\nu_n \downarrow 0$ such that $\mathbb{P}_{(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0))}(S_n(\nu_n)) \rightarrow 1$ as $n \rightarrow \infty$. By the differentiability of $\tilde{\ell}(\gamma_Y, \delta_Y; \beta)$ we have that $\mathbf{v} \in S_n(\nu_n)$ implies that there exists a point $(\hat{\gamma}_{Y,n}(\nu_n), \hat{\delta}_{Y,n}(\nu_n)) \in Q_{\nu,n}$ for which $\tilde{\ell}(\gamma_Y, \delta_Y; \beta)$ attains a local maximum, and thus $\tilde{\ell}_j(\hat{\gamma}_{Y,n}(\nu_n), \hat{\delta}_{Y,n}(\nu_n); \beta) = 0$, $j = 1, 2$. Now let $(\hat{\gamma}_{Y,n}^*(x_0), \hat{\delta}_{Y,n}^*(t_n|x_0)) := (\hat{\gamma}_{Y,n}(\nu_n), \hat{\delta}_{Y,n}(\nu_n))$ for $\mathbf{v} \in S_n(\nu_n)$ and arbitrary otherwise. Clearly

$$\begin{aligned} \mathbb{P}_{(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0))}(\tilde{\ell}_1(\hat{\gamma}_{Y,n}^*(x_0), \hat{\delta}_{Y,n}^*(t_n|x_0)); \beta) = 0, \tilde{\ell}_2(\hat{\gamma}_{Y,n}^*(x_0), \hat{\delta}_{Y,n}^*(t_n|x_0)); \beta) = 0) \\ \geq \mathbb{P}_{(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0))}(S_n(\nu_n)) \rightarrow 1, \end{aligned}$$

as $n \rightarrow \infty$. Thus with probability tending to 1 there exists a sequence of solutions to the likelihood equations. Also, for any fixed $\nu > 0$ and n sufficiently large, and denoting by d the usual Euclidean distance,

$$\begin{aligned} \mathbb{P}_{(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0))}(d((\hat{\gamma}_{Y,n}^*(x_0), \hat{\delta}_{Y,n}^*(t_n|x_0)), (\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0))) < \nu) \\ \geq \mathbb{P}_{(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0))}(d((\hat{\gamma}_{Y,n}^*(x_0), \hat{\delta}_{Y,n}^*(t_n|x_0)), (\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0))) < \nu_n) \\ \geq \mathbb{P}_{(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0))}(S_n(\nu_n)) \rightarrow 1, \end{aligned}$$

which establishes the consistency of the sequence $(\hat{\gamma}_{Y,n}^*(x_0), \hat{\delta}_{Y,n}^*(t_n|x_0))$.

Proof of Theorem 3

The result of the theorem will be established by using the Cramér-Wold device (see e.g. van der Vaart, 1998, p.16). Take $\theta := (\theta_1, \dots, \theta_{J+L})^T \in \mathbb{R}^{J+L}$. Then

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

Note that $\text{Var}(W_1) = \theta^T \mathbb{A} \theta / n$, where \mathbb{A} has elements

$$\begin{aligned}
\mathbb{A}_{jk} &:= \text{Cov} \left(\left(\frac{h_n^d}{\bar{F}_T(t_n|x_0)f_X(x_0)} \right)^{1/2} K_{h_n}(x_0 - X) \left(\frac{T}{t_n} \right)^{s_j} \left(\ln_+ \frac{T}{t_n} \right)^{s'_j} \mathbb{1}_{\{T > t_n\}}, \right. \\
&\quad \left. \left(\frac{h_n^d}{\bar{F}_T(t_n|x_0)f_X(x_0)} \right)^{1/2} K_{h_n}(x_0 - X) \left(\frac{T}{t_n} \right)^{s_k} \left(\ln_+ \frac{T}{t_n} \right)^{s'_k} \mathbb{1}_{\{T > t_n\}} \right), \quad j, k \in \{1, \dots, J\}, \\
\mathbb{A}_{jk} &:= \text{Cov} \left(\left(\frac{h_n^d}{\bar{F}_T(t_n|x_0)f_X(x_0)} \right)^{1/2} K_{h_n}(x_0 - X) \left(\frac{T}{t_n} \right)^{s_j} \left(\ln_+ \frac{T}{t_n} \right)^{s'_j} \mathbb{1}_{\{Y \leq C, T > t_n\}}, \right. \\
&\quad \left. \left(\frac{h_n^d}{\bar{F}_T(t_n|x_0)f_X(x_0)} \right)^{1/2} K_{h_n}(x_0 - X) \left(\frac{T}{t_n} \right)^{s_k} \left(\ln_+ \frac{T}{t_n} \right)^{s'_k} \mathbb{1}_{\{Y \leq C, T > t_n\}} \right), \\
&\quad \quad \quad j, k \in \{J+1, \dots, J+L\}, \\
\mathbb{A}_{jk} &:= \text{Cov} \left(\left(\frac{h_n^d}{\bar{F}_T(t_n|x_0)f_X(x_0)} \right)^{1/2} K_{h_n}(x_0 - X) \left(\frac{T}{t_n} \right)^{s_j} \left(\ln_+ \frac{T}{t_n} \right)^{s'_j} \mathbb{1}_{\{T > t_n\}}, \right. \\
&\quad \left. \left(\frac{h_n^d}{\bar{F}_T(t_n|x_0)f_X(x_0)} \right)^{1/2} K_{h_n}(x_0 - X) \left(\frac{T}{t_n} \right)^{s_k} \left(\ln_+ \frac{T}{t_n} \right)^{s'_k} \mathbb{1}_{\{Y \leq C, T > t_n\}} \right), \\
&\quad \quad \quad j \in \{1, \dots, J\}, k \in \{J+1, \dots, J+L\}.
\end{aligned}$$

Using Theorem 1 we obtain that $\mathbb{A}_{jk} = \Sigma_{jk}(1+o(1))$, $j, k \in \{1, \dots, J+L\}$, and thus $\text{Var}(\theta^T r_n(\mathbb{T}_n - \mathbb{E}(\mathbb{T}_n))) = \theta^T \Sigma \theta (1+o(1))$.

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 $\lim_{n \rightarrow \infty} n\mathbb{E}(|W_1|^3) = 0$. To this aim, note that W_1 is of the form $V - \mathbb{E}(V)$, leading to the inequality

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

Again using the result from Theorem 1, we obtain the following orders

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

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

Proof of Theorem 4

We apply a Taylor series expansion of $\tilde{\ell}_j(\hat{\gamma}_{Y,n}(x_0), \hat{\delta}_{Y,n}(t_n|x_0); \beta)$, $j = 1, 2$, around $(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0))$ and obtain, after rearranging terms

$$\begin{aligned} -r_n \begin{bmatrix} \tilde{\ell}_1(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) \\ \tilde{\ell}_2(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) \end{bmatrix} &= \\ \begin{bmatrix} \check{\ell}_{11}(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) & \check{\ell}_{12}(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) \\ \check{\ell}_{12}(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) & \check{\ell}_{22}(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) \end{bmatrix} \begin{bmatrix} r_n(\hat{\gamma}_{Y,n}(x_0) - \gamma_Y^{(0)}(x_0)) \\ r_n(\hat{\delta}_{Y,n}(t_n|x_0) - \delta_Y^{(0)}(t_n|x_0)) \end{bmatrix} & \quad (9) \end{aligned}$$

where

$$\begin{aligned} \check{\ell}_{11}(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) &= \tilde{\ell}_{11}(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) + \frac{1}{2}[\tilde{\ell}_{111}(\tilde{\gamma}_{Y,n}, \tilde{\delta}_{Y,n}; \beta)(\hat{\gamma}_{Y,n}(x_0) - \gamma_Y^{(0)}(x_0)) \\ &\quad + \tilde{\ell}_{112}(\tilde{\gamma}_{Y,n}, \tilde{\delta}_{Y,n}; \beta)(\hat{\delta}_{Y,n}(t_n|x_0) - \delta_Y^{(0)}(t_n|x_0))], \\ \check{\ell}_{22}(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) &= \tilde{\ell}_{22}(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) + \frac{1}{2}[\tilde{\ell}_{122}(\tilde{\gamma}_{Y,n}, \tilde{\delta}_{Y,n}; \beta)(\hat{\gamma}_{Y,n}(x_0) - \gamma_Y^{(0)}(x_0)) \\ &\quad + \tilde{\ell}_{222}(\tilde{\gamma}_{Y,n}, \tilde{\delta}_{Y,n}; \beta)(\hat{\delta}_{Y,n}(t_n|x_0) - \delta_Y^{(0)}(t_n|x_0))], \\ \check{\ell}_{12}(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) &= \tilde{\ell}_{12}(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) + \frac{1}{2}[\tilde{\ell}_{112}(\tilde{\gamma}_{Y,n}, \tilde{\delta}_{Y,n}; \beta)(\hat{\gamma}_{Y,n}(x_0) - \gamma_Y^{(0)}(x_0)) \\ &\quad + \tilde{\ell}_{122}(\tilde{\gamma}_{Y,n}, \tilde{\delta}_{Y,n}; \beta)(\hat{\delta}_{Y,n}(t_n|x_0) - \delta_Y^{(0)}(t_n|x_0))], \end{aligned}$$

and $(\tilde{\gamma}_{Y,n}, \tilde{\delta}_{Y,n})$ is a random value on the line segment between $(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0))$ and $(\hat{\gamma}_{Y,n}(x_0), \hat{\delta}_{Y,n}(t_n|x_0))$.

We start by analysing the left hand side of (9). As a first step we rewrite the expansion for $\mathbb{E}(T_n^{(2)}(K, s, s'|x_0))$ in terms of $\delta_T^{(0)}(t_n|x_0)$.

- Case 1: $\delta_Y^{(0)}(t_n|x_0)/\delta_C^{(0)}(t_n|x_0) \rightarrow \pm\infty$

$$\begin{aligned} & \mathbb{E}(T_n^{(2)}(K, s, s'|x_0)) \\ &= \bar{F}_T(t_n|x_0)f_X(x_0)\Gamma(s'+1)\frac{\gamma_T^{(0)(s'+1)}(x_0)}{\gamma_Y^{(0)}(x_0)}\left\{\frac{1}{(1-s\gamma_T^{(0)}(x_0))^{s'+1}}\right. \\ & \quad +\delta_T^{(0)}(t_n|x_0)\frac{\gamma_Y^{(0)}(x_0)}{\gamma_T^{(0)}(x_0)}\left[\frac{1}{\gamma_Y^{(0)}(x_0)}\left(\frac{1}{(1+\beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0)-s\gamma_T^{(0)}(x_0))^{s'+1}}-\frac{1}{(1-s\gamma_T^{(0)}(x_0))^{s'+1}}\right)\right. \\ & \quad \left.\left.+\frac{\beta_T^{(0)}(x_0)}{(1+\beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0)-s\gamma_T^{(0)}(x_0))^{s'+1}}\right](1+o(1))+O(h_n^{\eta_{f_X}\wedge\eta_{A_Y}\wedge\eta_{A_C}})+O(h_n^{\eta_{\gamma_Y}\wedge\eta_{\gamma_C}}\ln t_n)\right\}. \end{aligned}$$

- Case 2: $\delta_Y^{(0)}(t_n|x_0)/\delta_C^{(0)}(t_n|x_0) \rightarrow 0$

$$\begin{aligned} & \mathbb{E}(T_n^{(2)}(K, s, s'|x_0)) \\ &= \bar{F}_T(t_n|x_0)f_X(x_0)\Gamma(s'+1)\frac{\gamma_T^{(0)(s'+1)}(x_0)}{\gamma_Y^{(0)}(x_0)}\left\{\frac{1}{(1-s\gamma_T^{(0)}(x_0))^{s'+1}}\right. \\ & \quad +\frac{\delta_T^{(0)}(t_n|x_0)}{\gamma_T^{(0)}(x_0)}\left[\frac{1}{(1+\beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0)-s\gamma_T^{(0)}(x_0))^{s'+1}}-\frac{1}{(1-s\gamma_T^{(0)}(x_0))^{s'+1}}\right](1+o(1)) \\ & \quad \left.+O(h_n^{\eta_{f_X}\wedge\eta_{A_Y}\wedge\eta_{A_C}})+O(h_n^{\eta_{\gamma_Y}\wedge\eta_{\gamma_C}}\ln t_n)\right\}. \end{aligned}$$

- Case 3: $\delta_Y^{(0)}(t_n|x_0)/\delta_C^{(0)}(t_n|x_0) \rightarrow a$, where $0 < |a| < \infty$

$$\begin{aligned} & \mathbb{E}(T_n^{(2)}(K, s, s'|x_0)) \\ &= \bar{F}_T(t_n|x_0)f_X(x_0)\Gamma(s'+1)\frac{\gamma_T^{(0)(s'+1)}(x_0)}{\gamma_Y^{(0)}(x_0)}\left\{\frac{1}{(1-s\gamma_T^{(0)}(x_0))^{s'+1}}\right. \\ & \quad +\delta_T^{(0)}(t_n|x_0)\frac{\gamma_Y^{(0)}(x_0)\gamma_C^{(0)}(x_0)a}{\gamma_T^{(0)}(x_0)(\gamma_Y^{(0)}(x_0)+a\gamma_C^{(0)}(x_0))}\left[\left(\frac{1}{\gamma_Y^{(0)}(x_0)}+\frac{1}{\gamma_C^{(0)}(x_0)a}\right)\times\right. \\ & \quad \left.\left(\frac{1}{(1+\beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0)-s\gamma_T^{(0)}(x_0))^{s'+1}}-\frac{1}{(1-s\gamma_T^{(0)}(x_0))^{s'+1}}\right)\right. \\ & \quad \left.\left.+\frac{\beta_T^{(0)}(x_0)}{(1+\beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0)-s\gamma_T^{(0)}(x_0))^{s'+1}}\right](1+o(1))+O(h_n^{\eta_{f_X}\wedge\eta_{A_Y}\wedge\eta_{A_C}})+O(h_n^{\eta_{\gamma_Y}\wedge\eta_{\gamma_C}}\ln t_n)\right\}. \end{aligned}$$

Now let

$$\mathbb{U}_n := \frac{1}{\overline{F}_T(t_n|x_0)f_X(x_0)} \begin{bmatrix} T_n^{(1)}(K, 0, 0|x_0) \\ T_n^{(1)}(K, -\beta, 0|x_0) \\ T_n^{(1)}(K, 0, 1|x_0) \\ T_n^{(2)}(K, 0, 0|x_0) \\ T_n^{(2)}(K, -\beta, 0|x_0) \end{bmatrix}, \quad \mu := \begin{bmatrix} 1 \\ \frac{1}{1+\beta\gamma_T^{(0)}(x_0)} \\ \gamma_T^{(0)}(x_0) \\ \frac{\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)} \\ \frac{\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)} \\ \frac{\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)(1+\beta\gamma_T^{(0)}(x_0))} \end{bmatrix}.$$

From Theorem 3 we have

$$r_n(\mathbb{U}_n - \mu) \rightsquigarrow N(\lambda\tilde{\mu}, \Sigma)$$

where Σ is a (5×5) symmetric matrix with elements

$$\begin{aligned} \sigma_{11} &:= \|K\|_2^2 & \sigma_{12} &:= \|K\|_2^2 \frac{1}{1 + \beta\gamma_T^{(0)}(x_0)} \\ \sigma_{22} &:= \|K\|_2^2 \frac{1}{1 + 2\beta\gamma_T^{(0)}(x_0)} & \sigma_{13} &:= \|K\|_2^2 \gamma_T^{(0)}(x_0) \\ \sigma_{33} &:= \|K\|_2^2 2\gamma_T^{(0)2}(x_0) & \sigma_{14} &:= \|K\|_2^2 \frac{\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)} \\ \sigma_{44} &:= \|K\|_2^2 \frac{\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)} & \sigma_{15} &:= \|K\|_2^2 \frac{\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)(1 + \beta\gamma_T^{(0)}(x_0))} \\ \sigma_{55} &:= \|K\|_2^2 \frac{\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)(1 + 2\beta\gamma_T^{(0)}(x_0))} & \sigma_{34} &:= \|K\|_2^2 \frac{\gamma_T^{(0)2}(x_0)}{\gamma_Y^{(0)}(x_0)} \\ \sigma_{23} &:= \|K\|_2^2 \frac{\gamma_T^{(0)}(x_0)}{(1 + \beta\gamma_T^{(0)}(x_0))^2} & \sigma_{35} &:= \|K\|_2^2 \frac{\gamma_T^{(0)2}(x_0)}{\gamma_Y^{(0)}(x_0)(1 + \beta\gamma_T^{(0)}(x_0))^2} \\ \sigma_{24} &:= \|K\|_2^2 \frac{\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)(1 + \beta\gamma_T^{(0)}(x_0))} & \sigma_{45} &:= \|K\|_2^2 \frac{\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)(1 + \beta\gamma_T^{(0)}(x_0))} \\ \sigma_{25} &:= \|K\|_2^2 \frac{\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)(1 + 2\beta\gamma_T^{(0)}(x_0))} \end{aligned}$$

and $\tilde{\mu}$ is a vector with elements

$$\begin{aligned} \tilde{\mu}_1 &:= 0, \\ \tilde{\mu}_2 &:= \frac{\beta\beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0)}{(1 + \beta\gamma_T^{(0)}(x_0))(1 + \beta\gamma_T^{(0)}(x_0) + \beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0))}, \\ \tilde{\mu}_3 &:= -\frac{\beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0)}{1 + \beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0)}, \end{aligned}$$

and

- Case 1: $\delta_Y^{(0)}(t_n|x_0)/\delta_C^{(0)}(t_n|x_0) \rightarrow \pm\infty$

$$\tilde{\mu}_4 := -\frac{\beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)(1 + \beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0))} + \frac{\beta_T^{(0)}(x_0)}{1 + \beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0)},$$

$$\tilde{\mu}_5 := -\frac{\beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)(1 + \beta_T^{(0)}(x_0))(1 + \beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0))} + \frac{\beta_T^{(0)}(x_0)}{1 + \beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0)},$$

- Case 2: $\delta_Y^{(0)}(t_n|x_0)/\delta_C^{(0)}(t_n|x_0) \rightarrow 0$

$$\tilde{\mu}_4 := -\frac{\beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)(1 + \beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0))},$$

$$\tilde{\mu}_5 := -\frac{\beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)(1 + \beta_T^{(0)}(x_0))(1 + \beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0))},$$

- Case 3: $\delta_Y^{(0)}(t_n|x_0)/\delta_C^{(0)}(t_n|x_0) \rightarrow a$, where $0 < |a| < \infty$

$$\tilde{\mu}_4 := -\frac{\beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)(1 + \beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0))} + \frac{a\beta_T^{(0)}(x_0)\gamma_C^{(0)}(x_0)}{(\gamma_Y^{(0)}(x_0) + a\gamma_C^{(0)}(x_0))(1 + \beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0))},$$

$$\tilde{\mu}_5 := -\frac{\beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)(1 + \beta_T^{(0)}(x_0))(1 + \beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0))} + \frac{a\beta_T^{(0)}(x_0)\gamma_C^{(0)}(x_0)}{(\gamma_Y^{(0)}(x_0) + a\gamma_C^{(0)}(x_0))(1 + \beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0))}.$$

Based on (7) and (8) we can write

$$r_n \tilde{\ell}_1(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) = \xi_1 r_n (\mathbb{U}_n - \mu) + r_n \delta_Y^{(0)}(t_n|x_0) \frac{\beta \gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)2}(x_0)(1 + \beta \gamma_T^{(0)}(x_0))} (1 + o_{\mathbb{P}}(1)),$$

$$r_n \tilde{\ell}_2(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) = \xi_2 r_n (\mathbb{U}_n - \mu) - r_n \delta_Y^{(0)}(t_n|x_0) \frac{\beta^2 \gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)(1 + 2\beta \gamma_T^{(0)}(x_0))} (1 + o_{\mathbb{P}}(1)),$$

where

$$\xi_1 := \left[0, 0, \frac{1}{\gamma_Y^{(0)2}(x_0)}, -\frac{1}{\gamma_Y^{(0)}(x_0)}, 0 \right],$$

$$\xi_2 := \left[-\frac{1}{\gamma_Y^{(0)}(x_0)}, \frac{1}{\gamma_Y^{(0)}(x_0)}, 0, 0, \beta \right],$$

and thus

$$r_n \begin{bmatrix} \tilde{\ell}_1(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) \\ \tilde{\ell}_2(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) \end{bmatrix} \rightsquigarrow N(\lambda\kappa, \Delta),$$

where Δ is a symmetric (2×2) matrix with elements

$$\begin{aligned} \Delta_{11} &:= \|K\|_2^2 \frac{\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)3}(x_0)}, \\ \Delta_{22} &:= \|K\|_2^2 \frac{\beta^2 \gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)(1 + 2\beta\gamma_T^{(0)}(x_0))}, \\ \Delta_{12} &:= -\|K\|_2^2 \frac{\beta\gamma_T^{(0)}(x_0)}{\gamma_Y^{(0)2}(x_0)(1 + \beta\gamma_T^{(0)}(x_0))}, \end{aligned}$$

and κ has elements

- Case 1: $\delta_Y^{(0)}(t_n|x_0)/\delta_C^{(0)}(t_n|x_0) \rightarrow \pm\infty$

$$\begin{aligned} \kappa_1 &:= -\frac{1}{\gamma_Y^{(0)}(x_0)} \left[\frac{\beta_T^{(0)}(x_0)}{1 + \beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0)} - \frac{\beta}{1 + \beta\gamma_T^{(0)}(x_0)} \right], \\ \kappa_2 &:= \frac{\beta\beta_T^{(0)}(x_0)}{1 + \beta\gamma_T^{(0)}(x_0) + \beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0)} - \frac{\beta^2}{1 + 2\beta\gamma_T^{(0)}(x_0)}, \end{aligned}$$

- Case 2: $\delta_Y^{(0)}(t_n|x_0)/\delta_C^{(0)}(t_n|x_0) \rightarrow 0$

$$\kappa_1 = \kappa_2 = 0,$$

- Case 3: $\delta_Y^{(0)}(t_n|x_0)/\delta_C^{(0)}(t_n|x_0) \rightarrow a$, where $0 < |a| < \infty$

$$\begin{aligned} \kappa_1 &:= -\frac{a\gamma_C^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0)(\gamma_Y^{(0)}(x_0) + a\gamma_C^{(0)}(x_0))} \left[\frac{\beta_T^{(0)}(x_0)}{1 + \beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0)} - \frac{\beta}{1 + \beta\gamma_T^{(0)}(x_0)} \right], \\ \kappa_2 &:= \frac{a\beta\gamma_C^{(0)}(x_0)}{\gamma_Y^{(0)}(x_0) + a\gamma_C^{(0)}(x_0)} \left[\frac{\beta_T^{(0)}(x_0)}{1 + \beta\gamma_T^{(0)}(x_0) + \beta_T^{(0)}(x_0)\gamma_T^{(0)}(x_0)} - \frac{\beta}{1 + 2\beta\gamma_T^{(0)}(x_0)} \right]. \end{aligned}$$

Now we focus on the right hand side of (9). Concerning the terms $\tilde{\ell}_{jk}(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta)$, $j, k = 1, 2$, we have by the consistency of $(\hat{\gamma}_{Y,n}(x_0), \hat{\delta}_{Y,n}(t_n|x_0))$ and because $|\tilde{\ell}_{jkl}(\gamma_Y, \delta_Y; \beta)| \leq M_{jkl}(\mathbf{V})$, in some open neighborhood of $(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0))$, with $M_{jkl}(\mathbf{V}) = O_{\mathbb{P}}(1)$, $j, k, l = 1, 2$, that $\tilde{\ell}_{jk}(\gamma_Y^{(0)}(x_0), \delta_Y^{(0)}(t_n|x_0); \beta) \xrightarrow{\mathbb{P}} \dot{\ell}_{jk}$, $j, k = 1, 2$. Let

$$\mathbb{C} := \begin{bmatrix} \dot{\ell}_{11} & \dot{\ell}_{12} \\ \dot{\ell}_{12} & \dot{\ell}_{22} \end{bmatrix}.$$

From the proof of the consistency, we know that \mathbb{C} is a negative definite matrix, and thus invertible. Then, according to Lemma 5.2 in Chapter 6 of Lehmann and Casella (1998), for the solution of the system of equations (9), we have the following convergence

$$r_n \begin{bmatrix} \widehat{\gamma}_{Y,n}(x_0) - \gamma_Y^{(0)}(x_0) \\ \widehat{\delta}_{Y,n}(t_n|x_0) - \delta_Y^{(0)}(t_n|x_0) \end{bmatrix} \rightsquigarrow N(-\lambda \mathbb{C}^{-1} \kappa, \mathbb{C}^{-1} \Delta \mathbb{C}^{-1}).$$

Working out the matrix products appearing in the mean and variance of the limiting distribution yields the result of the theorem.

Acknowledgement

This work was supported by a research grant (VKR023480) from VILLUM FONDEN.

References

- [1] 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.
- [2] Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J.L., 2004. *Statistics of Extremes – Theory and Applications*. Wiley.
- [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] Daouia, A., Gardes, L., Girard, S., 2013. On kernel smoothing for extremal quantile regression. *Bernoulli*, 19, 2557–2589.
- [7] Daouia, A., Gardes, L., Girard, S., Lekina, A., 2011. Kernel estimators of extreme level curves. *Test*, 20, 311–333.
- [8] de Haan, L., Ferreira, A., 2006. *Extreme Value Theory: An Introduction*. Springer.
- [9] Dierckx, G., Goegebeur, Y., Guillou, A., 2014. Local robust and asymptotically unbiased estimation of conditional Pareto-type tails. *TEST*, 23, 330–355.
- [10] Drees, H., 1998. On smooth statistical tail functionals. *Scandinavian Journal of Statistics*, 25, 187–210.
- [11] Dutang, C., Goegebeur, Y., Guillou, A., 2014. Robust and bias-corrected estimation of the coefficient of tail dependence. *Insurance: Mathematics and Economics*, 57, 46–57.

- [12] Dutang, C., Goegebeur, Y., Guillou, A., 2016. Robust and bias-corrected estimation of the probability of extreme failure sets. *Sankhya A*, 78, 52–86.
- [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 estimation of the conditional stable tail dependence function. To appear in *Scandinavian Journal of Statistics*.
- [15] Feuerverger, A., Hall, P., 1999. Estimating a tail exponent by modelling departure from a Pareto distribution. *Annals of Statistics*, 27, 760–781.
- [16] 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.
- [17] Goegebeur, Y., Guillou, A., Schorgen, A., 2014b. Nonparametric regression estimation of conditional tails - the random covariate case. *Statistics*, 48, 732–755.
- [18] Gomes, M.I., Martins, M.J., 2004. Bias-reduction and explicit semi-parametric estimation of the tail index. *Journal of Statistical Planning and Inference*, 124, 361–378.
- [19] Gomes, M.I., Neves, M.M., 2011. Estimation of the extreme value index for randomly censored data. *Biometrical Letters*, 48, 1–22.
- [20] Hall, P., 1982. On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society Series B*, 44, 37–42.
- [21] Lawless, J.F., 2003. *Statistical models and methods for lifetime data*. Wiley.
- [22] Lehmann, E.L., Casella, G., 1998. *Theory of Point Estimation*. Springer.
- [23] Ndao, P., Diop, A., Dupuy, J.F., 2014. Non parametric estimation of the conditional tail index and extreme quantiles under random censoring. *Computational Statistics and Data Analysis*, 79, 63-79.
- [24] Ndao, P., Diop, A., Dupuy, J.F., 2016. Nonparametric estimation of the conditional extreme value index with random covariates and censoring. *Journal of Statistical Planning and Inference*, 168, 20–37.
- [25] Stupfler, G., 2013. A moment estimator for the conditional extreme-value index. *Electronic Journal of Statistics*, 7, 2298–2353.
- [26] Stupfler, G., 2016. Estimating the conditional extreme-value index under random right-censoring. *Journal of Multivariate Analysis*, 144, 1–24.
- [27] van der Vaart, A.W., 1998. *Asymptotic statistics*. Cambridge University Press, Cambridge.
- [28] Venables, W.N., Ripley, B.D., 2002. *Modern Applied Statistics with S*, fourth edition, Springer.

- [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.
- [31] Yao, Q., 1999. Conditional predictive regions for stochastic processes. Technical report, Institute of Mathematics and Statistics, University of Kent at Canterbury.

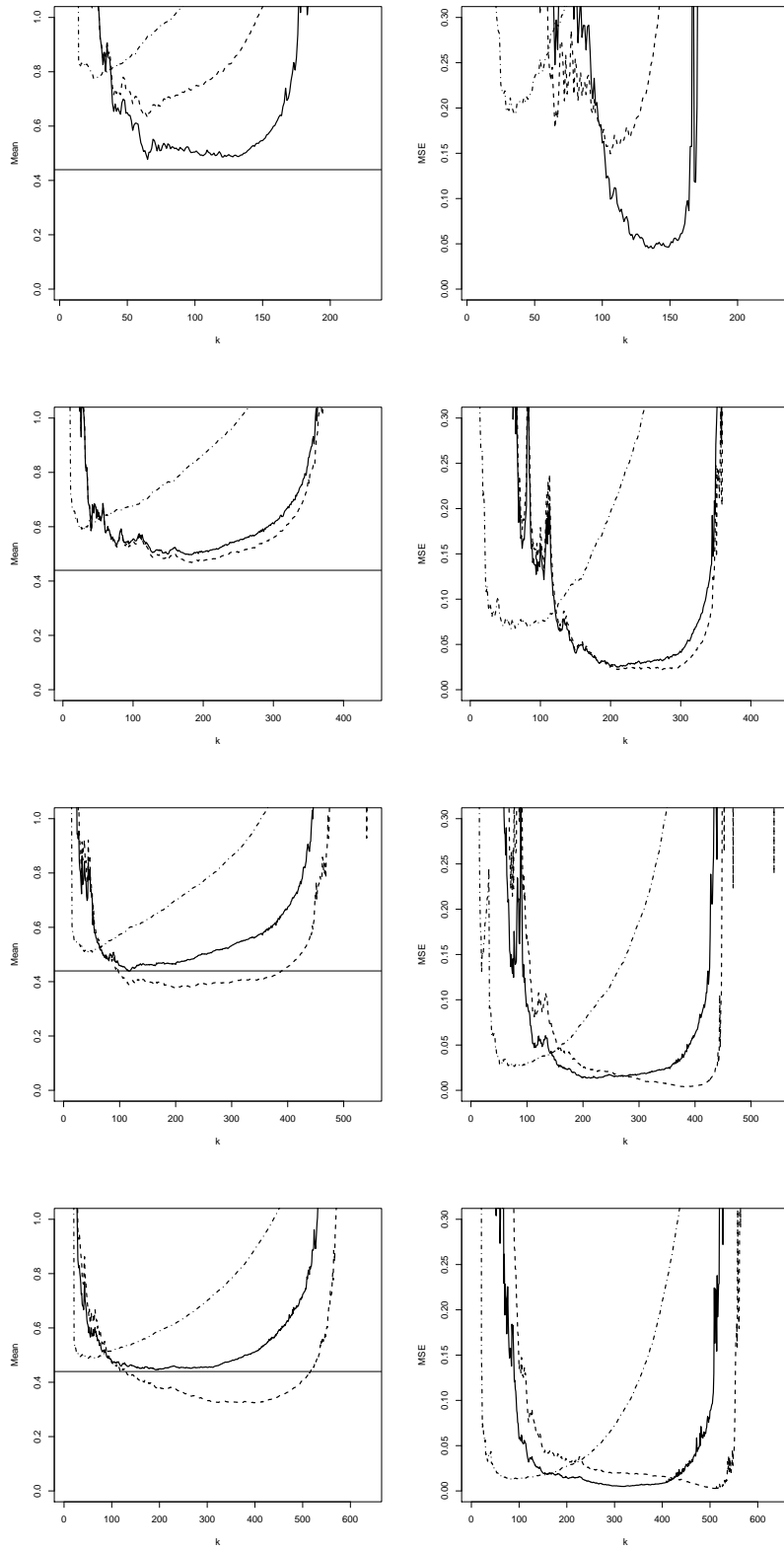


Figure 1: Estimation of $\gamma_Y(0.4)$, case 1 (row 1), case 2 (row 2), case 3 (row 3), case 4 (row 4). Mean (left) and MSE (right) as a function of k for $\hat{\gamma}_{Y,n}^B(0.4)$ (dashed-dotted line), and $\hat{\gamma}_{Y,n}(0.4)$ with $\beta = \beta_T^{(0)}(0.4)$ (solid line) and $\beta = 2$ (dashed line).

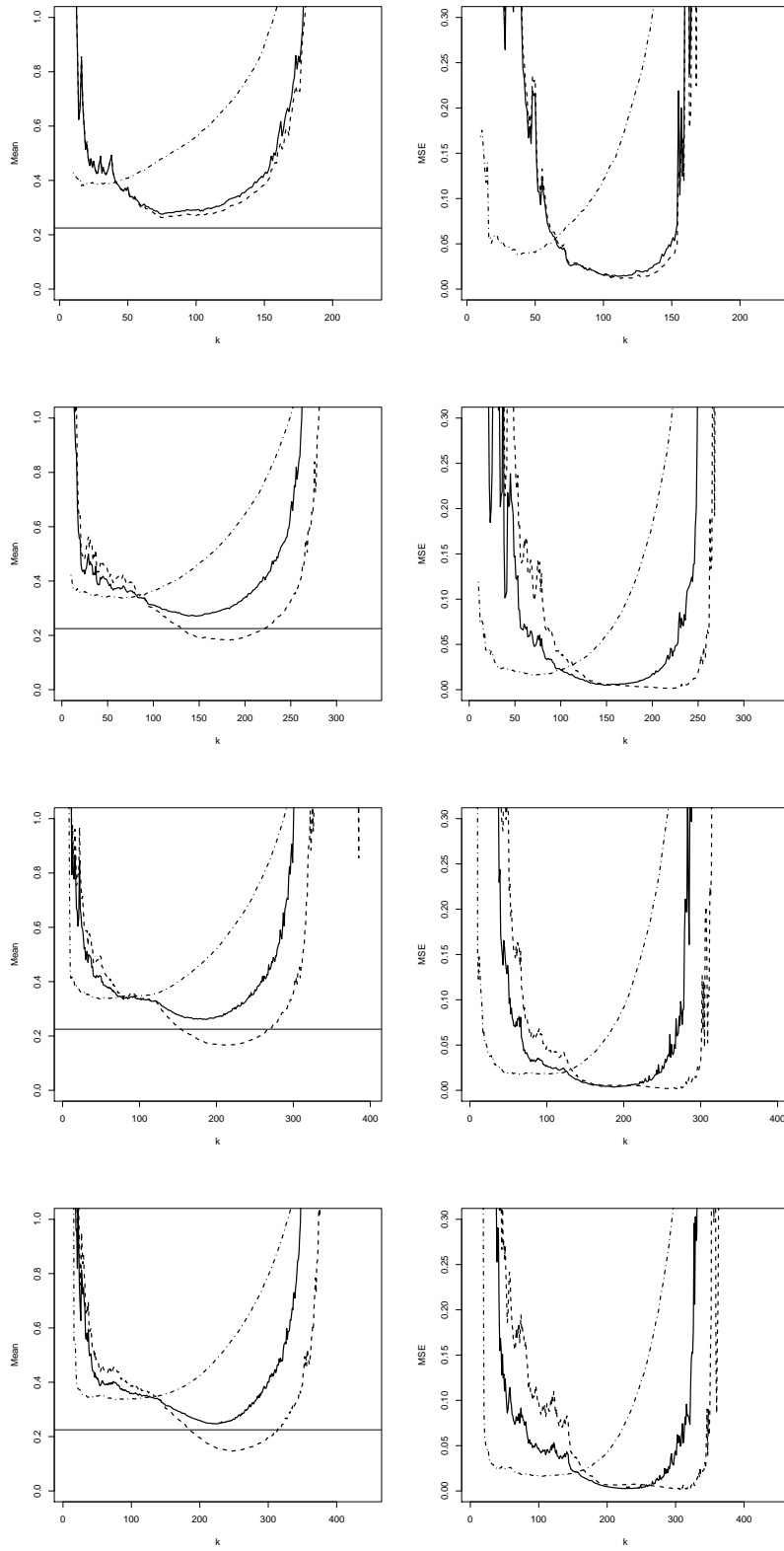


Figure 2: Estimation of $\gamma_Y(0.1)$, case 1 (row 1), case 2 (row 2), case 3 (row 3), case 4 (row 4). Mean (left) and MSE (right) as a function of k for $\hat{\gamma}_{Y,n}^B(0.1)$ (dashed-dotted line), and $\hat{\gamma}_{Y,n}(0.1)$ with $\beta = \beta_T^{(0)}(0.1)$ (solid line) and $\beta = 2$ (dashed line).

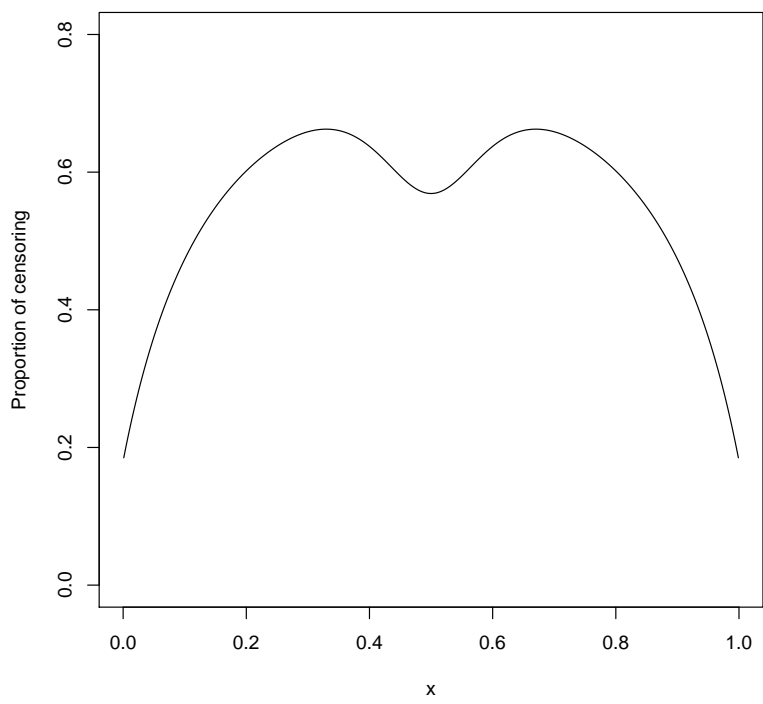


Figure 3: Proportion of censoring as a function of x .

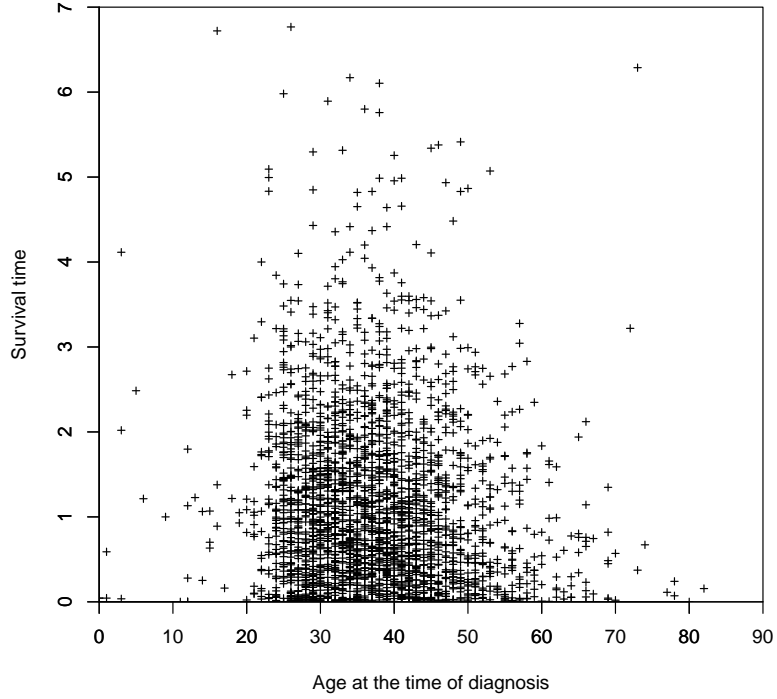


Figure 4: Scatterplot of the AIDS survival dataset.

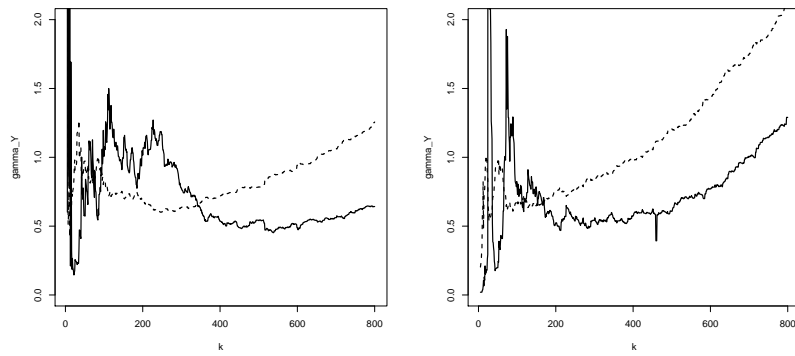


Figure 5: AIDS survival dataset: Estimation of $\gamma_Y(x_0)$ by $\hat{\gamma}_{Y,n}(x_0)$ with $\beta = 2$ (full line) and $\hat{\gamma}_{Y,n}^B(x_0)$ (dashed line). On the left $x_0 = 25$, and on the right $x_0 = 50$.