Robust estimation of the Pickands dependence function under random right censoring

Yuri Goegebeur, Armelle Guillou, Jing Qin

Robust estimation of the Pickands dependence function under 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 robust nonparametric estimation of the Pickands dependence function under random right censoring. The estimator is obtained by applying the minimum density power divergence criterion to properly transformed bivariate observations. The asymptotic properties are investigated by making use of results for Kaplan-Meier integrals. We investigate the finite sample properties of the proposed estimator with a simulation experiment and illustrate its practical applicability on a dataset of insurance indemnity losses.
Keywords: Pickands dependence function, censoring, Kaplan-Meier integral.

1 Introduction

Multivariate extreme value statistics deals with the estimation of the tail of a multivariate distribution function based on a random sample. When studying multivariate extremes, a natural question is how to quantify extreme dependence between two or more random variables. Usually, the copula function is used as a margin-free description of the dependence structure between several random variables. Indeed, according to Sklar's theorem (Sklar, 1959), the distribution function of a pair $(Y^{(1)}, Y^{(2)})$ can be represented in terms of the two marginal distribution functions $F_{Y^{(1)}}$ and $F_{Y^{(2)}}$ of $Y^{(1)}$ and $Y^{(2)}$ respectively, and a copula function C as follows:

$$\mathbb{P}\left(Y^{(1)} \leqslant y_1, Y^{(2)} \leqslant y_2\right) = C\left(F_{Y^{(1)}}(y_1), F_{Y^{(2)}}(y_2)\right). \tag{1}$$

This function C characterizes the dependence between $Y^{(1)}$ and $Y^{(2)}$ and is called an extreme value copula if and only if it admits a representation of the form

$$C(y_1, y_2) = \exp\left(\log(y_1 y_2) A_Y\left(\frac{\log(y_2)}{\log(y_1 y_2)}\right)\right), \tag{2}$$

where A_Y: $[0,1] \to [1/2, 1]$ is the Pickands dependence function, which is convex and satisfies $\max\{t, 1-t\} \leqslant A_Y(t) \leqslant 1$, see Pickands (1981). Throughout the paper we assume that

$(Y^{(1)}, Y^{(2)})$ follows a joint distribution with an underlying extreme value copula.

Since a copula function allows to model efficiently the dependence between several random variables, it becomes more and more popular in financial or actuarial applications. To illustrate our methodology, we consider the insurance company loss and expense application by Frees and Valdez (1998). The dataset, included in the `R` package `copula`, comprises 1500 pairs containing information on general liability claims, the first component being the indemnity payment (loss) and the second one an allocated loss adjustment expense (ALAE). The latter is related to the settlement of individual claims, e.g., expenses for lawyers or claim investigation. A crucial question is the possible dependence between the two components, loss and ALAE, which has to be accounted for if we are interested in actuarial applications, such as, e.g., pricing an excess-of-loss reinsurance treaty when the reinsurer shares the claims settlement costs. For instance Micocci and Masala (2009) (see also Cebrián *et al.*, 2003) motivate the use of copula functions in that context with the aim of building a reinsurance strategy in presence of policy limits and insurer's retentions. As outlined in these contributions, a substantial mispricing can result from the usual independence assumption where the joint distribution is assumed to be the product of the marginals, and thus using copulas is the correct way to model dependence and as such to avoid the undervaluation of the reinsurance premium. However, the estimation of the joint distribution for the losses and expenses is complicated due to the presence of censoring. More specifically, for each claim there is a policy limit, and hence the losses cannot exceed this limit. In the loss-ALAE dataset, 34 observations have censored losses and these censored observations cannot be ignored since for instance the mean loss of censored claims is much higher than the corresponding mean for the uncensored claims (217 491 against 37 110, see Table 4 in Frees and Valdez, 1998). The scatterplot of the data is given in Figure 1, where the uncensored observations are in grey and the censored ones in black. Overall the scatterplot indicates a reasonably strong relationship between the two variables, but the picture is somehow obscured by censoring in the largest observations and also by potential outliers.

We have thus two issues in the dataset under consideration: the presence of censoring and potential outliers.

Concerning the first issue, we will explore in the present paper nonparametric estimation of the Pickands dependence function when there is random right censoring in the marginal distributions. More precisely, we consider the situation where $(Y^{(1)}, Y^{(2)})$ is right censored by $(C^{(1)}, C^{(2)})$, also following a bivariate distribution with an extreme value copula, but now with Pickands dependence function A_C. Thus, we observe $(\min(Y^{(1)}, C^{(1)}), \min(Y^{(2)}, C^{(2)}), \delta^{(1)}, \delta^{(2)})$, where $\delta^{(j)} := 1\!\!1_{\{Y^{(j)} \leqslant C^{(j)}\}}$, $j = 1, 2$, with $1\!\!1_E$ the indicator function on the event E, and interest is in estimating A_Y.

Random right censoring has been studied to some extent in the univariate extreme value literature, see, e.g., Beirlant *et al.* (2007), Einmahl *et al.* (2008), Gomes and Neves (2011), Worms and Worms (2014), Beirlant *et al.* (2016), among others, where focus was mainly on estimating the extreme value index and extreme quantiles. Recently, extreme value regression problems with censoring were studied by Ndao *et al.* (2016), Stupfler (2016) and Goegebeur *et al.* (2018). In

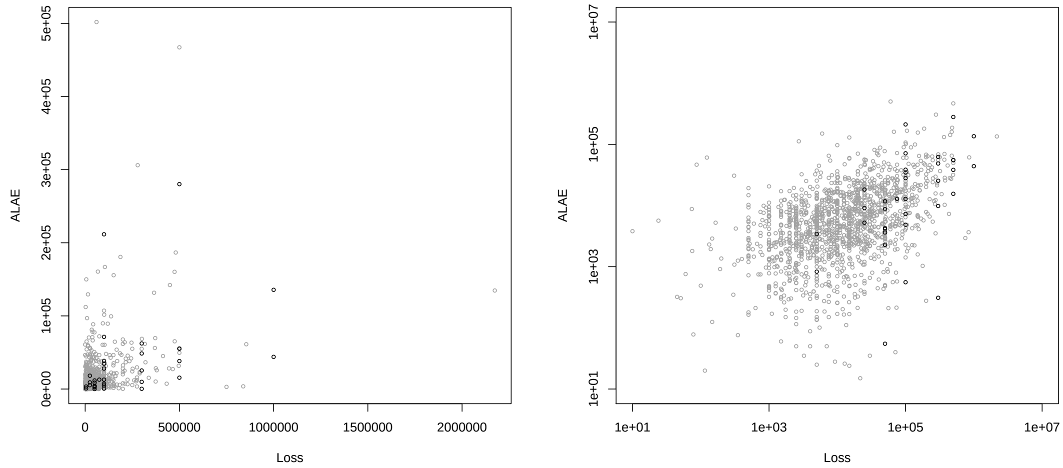

Figure 1: LOSS-ALAE data: scatterplot on the original scale (left) and on a logarithmic scale (right).

multivariate extreme value statistics a central topic is the modelling and estimation of extreme dependence between two or more random variables. Similarly to classical statistics, extreme dependence can be summarised by properly chosen dependence coefficients, like the coefficient of tail dependence, see, e.g., Ledford and Tawn (1997). Alternatively, one can use functions that give a complete characterisation of the extreme dependence like the Pickands dependence function, the stable tail dependence function and the spectral distribution function. Estimation of such dependence functions was considered in Capéraà *et al.* (1997), Fils-Villetard *et al.* (2008), Fougères *et al.* (2015) and Escobar-Bach *et al.* (2017). To the best of our knowledge censoring in the multivariate extreme value context is unexplored.

Concerning the second issue, robust methods must be proposed to prevent possible isolated outliers from completely disturbing the estimate of the joint distribution. To reach this goal, we will propose a robust estimator of the Pickands dependence function of $(Y^{(1)}, Y^{(2)})$ based on the density power divergence method introduced by Basu *et al.* (1998). In particular, the density power divergence between two density functions f and h is defined as follows

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

Here the density function h is assumed to depend on a parameter vector θ and if $Y_1, \ldots, Y_n$ is a sample of independent and identically distributed (i.i.d.) random variables according to the density function f, then the minimum density power divergence estimator (MDPDE) of θ is the point $\widehat{\theta}$ minimizing the estimated version (up to a constant independent of θ)

$$\widehat{\Delta}_\alpha(\theta) := \begin{cases} \int_{\mathbb{R}} h^{1+\alpha}(y)dy - \left(1 + \frac{1}{\alpha}\right) \int_{\mathbb{R}} h^\alpha(y) d\widehat{F}_Y(y), & \alpha > 0, \\ -\int_{\mathbb{R}} \log h(y) d\widehat{F}_Y(y), & \alpha = 0, \end{cases}$$

where $\widehat{F}_Y$ is a suitable estimator of the distribution of Y_1. In the case of no censoring, $\widehat{F}_Y$ is typically the empirical distribution function $\widehat{F}_Y^E(y) := (1/n)\sum_{i=1}^n \mathbb{1}_{\{Y_i \leqslant y\}}$, whereas in the censoring framework where Y_i is censored by an independent random variable C_i, the famous Kaplan-Meier product-limit estimator (see Kaplan and Meier, 1958) defined as

$$\widehat{F}_Y^{KM}(y) := 1 - \prod_{i=1}^n \left[1 - \frac{\delta_{[i,n]}}{n-i+1}\right]^{\mathbb{1}_{\{Z_{i,n} \leqslant y\}}}, \tag{3}$$

can be used. Here $Z_i := \min(Y_i, C_i)$, $i = 1, \ldots, n$, $Z_{i,n}$ denotes the $i-$th order statistic of $\{Z_1, \ldots, Z_n\}$ and $\delta_{[i,n]}$ is the concomitant order statistic with respect to $Z_{i,n}$, i.e., $\delta_{[i,n]} = \delta_k$ if $Z_{i,n} = Z_k$, $i = 1, \ldots, n$. The MDPD criterion depends on a parameter α which allows to make a trade-off between efficiency and robustness of the resulting estimator. Indeed, we can observe that for $\alpha = 0$ one recovers the log-likelihood function, up to the sign, which leads to an efficient but not robust estimator. By increasing α we increase the robustness of the estimator, but decrease its efficiency.

The remainder of the paper is organised as follows. In the next section we introduce the nonparametric MDPDE of the Pickands dependence function A_Y under random right censoring. The asymptotic properties of this estimator, consistency and finite dimensional weak convergence, are investigated in Section 3, where we use the asymptotic properties of Kaplan-Meier integrals. We illustrate the finite sample performance of the estimator with a simulation experiment in Section 4 and in Section 5 we apply the method to the dataset of insurance indemnity losses. The proofs of our results are given in Section 6.

2 Construction of the estimator

Throughout the paper, for any random variable W, we denote by F_W its distribution function. For convenience we reformulate the model as stated in (1) and (2) into standard exponential margins. Assume $F_{Y^{(j)}}$, $j = 1, 2$, are continuous. After applying the transformations $\widetilde{Y}^{(j)} = -\log F_{Y^{(j)}}(Y^{(j)})$, $j = 1, 2$, we obtain the following bivariate survival function

$$G_Y(y_1, y_2) := \mathbb{P}\left(\widetilde{Y}^{(1)} > y_1, \widetilde{Y}^{(2)} > y_2\right) = \exp\left(-(y_1 + y_2) A_Y\left(\frac{y_2}{y_1 + y_2}\right)\right),$$

for all $y_1, y_2 > 0$. A similar assumption is made for the distribution G_C of the random vector $(\widetilde{C}^{(1)}, \widetilde{C}^{(2)})$. Let $t \in [0, 1]$. Considering the univariate random variable

$$\widetilde{Y}_t := \min\left(\frac{\widetilde{Y}^{(1)}}{1-t}, \frac{\widetilde{Y}^{(2)}}{t}\right),$$

it is clear that

$$\mathbb{P}\left(\widetilde{Y}_t > z\right) = e^{-A_Y(t)z}, \ \forall z > 0.$$

Consequently, the distribution of $\widetilde{Y}_t$ is an exponential distribution with parameter $A_Y(t)$. Similarly, by defining $\widetilde{C}_t := \min\left(\frac{\widetilde{C}^{(1)}}{1-t}, \frac{\widetilde{C}^{(2)}}{t}\right)$, the random variable $\widetilde{C}_t$ follows an exponential distribution with parameter $A_C(t)$. Now, remarking that

$$\begin{aligned}
\widetilde{Z}_t &:= \min(\widetilde{Y}_t, \widetilde{C}_t) = \min\left(\frac{\min(\widetilde{Y}^{(1)}, \widetilde{C}^{(1)})}{1-t}, \frac{\min(\widetilde{Y}^{(2)}, \widetilde{C}^{(2)})}{t}\right), \\
\widetilde{\delta}_t &:= \mathbb{1}_{\{\widetilde{Y}_t \leqslant \widetilde{C}_t\}} = \delta_t^* \widetilde{\delta}^{(1)} + (1-\delta_t^*)\widetilde{\delta}^{(2)},
\end{aligned}$$

where $\delta_t^* := \mathbb{1}_{\left\{\frac{\min(\tilde{Y}^{(1)}, \tilde{C}^{(1)})}{1-t} \leqslant \frac{\min(\tilde{Y}^{(2)}, \tilde{C}^{(2)})}{t}\right\}}$ and $\widetilde{\delta}^{(j)} := \mathbb{1}_{\{\widetilde{Y}^{(j)} \leqslant \widetilde{C}^{(j)}\}}, j = 1, 2$, the pair $(\widetilde{Z}_t, \widetilde{\delta}_t)$ can actually be observed.

Let $(\widetilde{Z}_{t,i}, \widetilde{\delta}_{t,i})$, $i = 1, \ldots, n$, be independent copies of the random pair $(\widetilde{Z}_t, \widetilde{\delta}_t)$. We are now in the classical univariate censoring framework and we want to propose a nonparametric robust estimator for $A_Y(t)$ by means of the MDPD criterion, adjusted to the censoring, i.e., we minimize for $\alpha > 0$ the function

$$\widehat{\Delta}_{\alpha,t}(a) := \int_0^\infty \left(ae^{-az}\right)^{1+\alpha} dz - \left(1 + \frac{1}{\alpha}\right) \int_0^\infty \left(ae^{-az}\right)^\alpha d\widehat{F}_{\widetilde{Y}_t}^{KM}(z).$$

The MDPDE $\widehat{A}_{Y,\alpha,n}(t)$ for $A_Y(t)$ satisfies the estimating equation

$$\widehat{\Delta}_{\alpha,t}^{(1)}(\widehat{A}_{Y,\alpha,n}(t)) = 0 \qquad (4)$$

where $\widehat{\Delta}_{\alpha,t}^{(j)}(.)$ denotes the derivative of order j of $\widehat{\Delta}_{\alpha,t}(.)$. Our aim in this paper is to show the joint convergence in distribution of

$$\left[\sqrt{n}\left(\widehat{A}_{Y,\alpha,n}(t_j) - A_Y(t_j)\right), j = 1, \ldots, m\right],$$

where $\{t_1, \ldots, t_m\}$ is a grid of values in $[0, 1]$. In order to achieve this goal, a crucial step is the study of statistics of the type

$$T_n(t, \xi) := \int_0^\infty \phi_{t,\xi}(z) d\widehat{F}_{\widetilde{Y}_t}^{KM}(z), \qquad (5)$$

with $\phi_{t,\xi}(z) := z^\xi e^{-\alpha A_Y(t) z}$, $t \in [0, 1]$ and $\xi \in \mathbb{N}$, as $\widehat{\Delta}_{\alpha,t}$ and its derivatives are essentially linear combinations of such statistics, see, e.g., Section 6.1.

3 Asymptotic properties

In this section we derive the asymptotic properties of our estimator $\widehat{A}_{Y,\alpha,n}(t)$ under suitable assumptions. As a first step we need to establish the limiting behavior of (5). Kaplan-Meier integrals $\int f d\widehat{F}^{KM}$ have been studied in generality by Stute (1995), where asymptotic normality, after proper standardization, was established under suitable assumptions on the function f.

More precisely, denoting by

$$\begin{aligned}
\widetilde{H}_t^0(z) &:= \mathbb{P}\left(\widetilde{Z}_t \leqslant z, \widetilde{\delta}_t = 0\right), \\
\widetilde{H}_t^1(z) &:= \mathbb{P}\left(\widetilde{Z}_t \leqslant z, \widetilde{\delta}_t = 1\right), \\
\gamma_t^{(0)}(x) &:= \exp\left\{\int_0^x \frac{d\widetilde{H}_t^0(z)}{1 - F_{\widetilde{Z}_t}(z)}\right\}, \\
\gamma_{t,\xi}^{(1)}(x) &:= \frac{1}{1 - F_{\widetilde{Z}_t}(x)} \int_x^\infty \phi_{t,\xi}(z)\gamma_t^{(0)}(z) d\widetilde{H}_t^1(z), \\
\gamma_{t,\xi}^{(2)}(x) &:= \int_0^x \frac{1}{[1 - F_{\widetilde{Z}_t}(v)]^2} \left(\int_v^\infty \phi_{t,\xi}(z)\gamma_t^{(0)}(z) d\widetilde{H}_t^1(z)\right) d\widetilde{H}_t^0(v),
\end{aligned}$$

according to Theorem 1.1 in Stute (1995), we can obtain an i.i.d. representation of the Kaplan-Meier integral (5).

Theorem 3.1 *Let $t \in [0,1]$. Assuming $(1 + 2\alpha)A_Y(t) - A_C(t) > 0$, we have*

$$T_n(t,\xi) = \frac{1}{n}\sum_{i=1}^n \eta_{t,\xi}(\widetilde{Z}_{t,i}) + R_{n,t,\xi}$$

where

$$\eta_{t,\xi}(\widetilde{Z}_{t,i}) := \phi_{t,\xi}(\widetilde{Z}_{t,i})\gamma_t^{(0)}(\widetilde{Z}_{t,i})\widetilde{\delta}_{t,i} + \gamma_{t,\xi}^{(1)}(\widetilde{Z}_{t,i})(1 - \widetilde{\delta}_{t,i}) - \gamma_{t,\xi}^{(2)}(\widetilde{Z}_{t,i}),$$

and $R_{n,t,\xi} = o_{\mathbb{P}}(n^{-1/2})$.

From this representation, we can deduce the convergence in distribution of our key statistic T_n.

Corollary 3.1 *Under the assumption of Theorem 3.1*

$$\sqrt{n}\left(T_n(t,\xi) - \frac{\Gamma(\xi+1)}{(\alpha+1)^{\xi+1}A_Y^\xi(t)}\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2(t,\xi)),$$

where

$$\sigma^2(t,\xi) := Var(\eta_{t,\xi}(\widetilde{Z}_t)),$$

and Γ is the gamma function defined as $\Gamma(r) := \int_0^\infty t^{r-1}e^{-t}dt, \forall r > 0$.

We now derive the limiting distribution of a vector of statistics of the form (5), when properly normalized. Let $\mathbb{T}_n$ and $\mathcal{T}$ be $(m \times 1)$ vectors defined as

$$\mathbb{T}_n := (T_n(t_1, \xi_1), \ldots, T_n(t_m, \xi_m))^T,$$

and

$$\mathcal{T} := \left(\frac{\Gamma(\xi_j + 1)}{(\alpha+1)^{\xi_j+1} A_Y^{\xi_j}(t_j)}, \; j = 1, \ldots, m \right)^T,$$

for some positive integer m, where ^{T} stands for the transpose matrix. The aim of the next theorem is to provide the finite dimensional convergence result which will allow us to establish the convergence in distribution of our robust estimator of the Pickands dependence function A_Y.

Theorem 3.2 *Under the assumptions of Theorem 3.1, we have*

$$\sqrt{n}\left(\mathbb{T}_n - \mathcal{T}\right) \xrightarrow{d} \mathcal{N}_m\left(0, \Sigma\right),$$

where $\mathcal{N}_m$ denotes a $m-$dimensional normal distribution and Σ the $(m \times m)$ covariance matrix with elements $(\sigma_{j,k})_{1 \leqslant j,k \leqslant m} := (Cov(\eta_{t_j,\xi_j}(\widetilde{Z}_{t_j}), \eta_{t_k,\xi_k}(\widetilde{Z}_{t_k})))_{1 \leqslant j,k \leqslant m}$.

Note that for the result of Theorem 3.2 we need to assume $(1+2\alpha)A_Y(t_j) - A_C(t_j) > 0$ for $j = 1, \ldots, m$, which imposes a constraint on the parameter α. As a worst case scenario we could consider $A_Y(t) = \max\{t, 1-t\}$ (corresponding to complete dependence) and $A_C(t) = 1$ (corresponding to independence), and require $(1+2\alpha)A_Y(t) - A_C(t) > 0$ for all $t \in [0,1]$. By some standard calculations one can easily obtain that this will be satisfied if $\alpha > 0.5$.

By using the above results we can now prove the existence of a consistent sequence of solutions to the estimating equation (4).

Theorem 3.3 *Under the assumptions of Theorem 3.1, we have that with probability tending to one there exists a sequence $(\widehat{A}_{Y,\alpha,n}(t))_{n \geqslant 1}$ of solutions to (4), such that $\widehat{A}_{Y,\alpha,n}(t) \xrightarrow{\mathbb{P}} A_Y(t)$ as $n \to \infty$.*

We have now all the needed ingredients for proving the finite dimensional weak convergence of the MDPDE for A_Y on a grid $\{t_1, \ldots, t_m\}$ of positions in $[0,1]$. Let $\mathbb{A}$ denote a $(m \times 2m)$ matrix with elements

$$\mathbb{A}_{i,j} := \begin{cases} (1+\alpha)[A_Y(t_i)]^{\alpha-1}, & \text{if } j = 2i-1, \\ -(1+\alpha)[A_Y(t_i)]^{\alpha}, & \text{if } j = 2i, \\ 0, & \text{otherwise,} \end{cases}$$

$\mathbb{B}$ is an $(m \times m)$ diagonal matrix with entries

$$\mathbb{B}_{i,i} := \frac{(1+\alpha)^2}{[A_Y(t_i)]^{\alpha-2}(1+\alpha^2)},$$

and the matrix $\mathbb{C}$ is a $(2m \times 2m)$ matrix with elements

$$\begin{aligned}
\mathbb{C}_{2i-1,2j-1} &:= Cov(\eta_{t_i,0}(\widetilde{Z}_{t_i}), \eta_{t_j,0}(\widetilde{Z}_{t_j})), \\
\mathbb{C}_{2i-1,2j} &:= Cov(\eta_{t_i,0}(\widetilde{Z}_{t_i}), \eta_{t_j,1}(\widetilde{Z}_{t_j})), \\
\mathbb{C}_{2i,2j-1} &:= Cov(\eta_{t_i,1}(\widetilde{Z}_{t_i}), \eta_{t_j,0}(\widetilde{Z}_{t_j})), \\
\mathbb{C}_{2i,2j} &:= Cov(\eta_{t_i,1}(\widetilde{Z}_{t_i}), \eta_{t_j,1}(\widetilde{Z}_{t_j})).
\end{aligned}$$

Theorem 3.4 *Under the assumptions of Theorem 3.1, we have*

$$\sqrt{n} \begin{bmatrix} \widehat{A}_{Y,\alpha,n}(t_1) - A_Y(t_1) \\ \vdots \\ \widehat{A}_{Y,\alpha,n}(t_m) - A_Y(t_m) \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \mathbb{B}\mathbb{A}\mathbb{C}\mathbb{A}^T\mathbb{B}^T\right). \tag{6}$$

In particular, for $t \in [0,1]$, we have, as $n \to \infty$,

$$\sqrt{n}\left(\widehat{A}_{Y,\alpha,n}(t) - A_Y(t)\right) \xrightarrow{d} \mathcal{N}\left(0, \breve{\sigma}^2\right),$$

where $\breve{\sigma}^2 := N/D$, and

$$\begin{aligned}
N &:= (1+\alpha)^2 A_Y^3(t)\left[(1+4\alpha+9\alpha^2+14\alpha^3+13\alpha^4+8\alpha^5+4\alpha^6)A_Y^2(t)\right. \\
&\quad \left.-2(1+3\alpha+5\alpha^2+6\alpha^3+4\alpha^4+2\alpha^5)A_Y(t)A_C(t) + (1+2\alpha+3\alpha^2+2\alpha^3+\alpha^4)A_C^2(t)\right], \\
D &:= (1+\alpha^2)^2[(1+2\alpha)A_Y(t) - A_C(t)]^3.
\end{aligned}$$

4 Simulation experiment

In this section we illustrate the finite sample performance of the proposed estimator with a small simulation study. In first instance we consider distributions for $(Y^{(1)}, Y^{(2)})$ and $(C^{(1)}, C^{(2)})$, with an extreme value copula and unit exponential margins. The contamination is introduced by the following mixture model

$$F_\varepsilon(y_1, y_2) = (1-\varepsilon)F_\ell(y_1, y_2) + \varepsilon F_c(y_1, y_2), \tag{7}$$

where $\varepsilon \in [0,1]$ represents the fraction of contamination in the dataset, F_ℓ is the distribution function of $(T^{(1)}, T^{(2)}) := (\min(Y^{(1)}, C^{(1)}), \min(Y^{(2)}, C^{(2)}))$ and F_c is the contamination distribution function. For the main distributions of $(Y^{(1)}, Y^{(2)})$ and $(C^{(1)}, C^{(2)})$ we consider the asymmetric logistic distribution, with survival function

$$G_\bullet(y_1, y_2) = \exp\left(-(1-\psi_{1\bullet})y_1 - (1-\psi_{2\bullet})y_2 - ((\psi_{1\bullet}y_1)^{1/r_\bullet} + (\psi_{2\bullet}y_2)^{1/r_\bullet})^{r_\bullet}\right), \quad y_1, y_2 > 0,$$

where $\bullet$ denotes either Y or C, $r_\bullet \in (0,1]$ and $\psi_{1\bullet}, \psi_{2\bullet} \in [0,1]$. In this model independence is obtained for either $r_\bullet = 1$, or $\psi_{1\bullet} = 0$ or $\psi_{2\bullet} = 0$, while complete dependence is obtained for $\psi_{1\bullet} = \psi_{2\bullet} = 1$ and $r_\bullet \downarrow 0$. The logistic model is a special case of the asymmetric logistic model, and corresponds to $\psi_{1\bullet} = \psi_{2\bullet} = 1$. We consider the following settings

- Setting 1: $(r_Y, \psi_{1Y}, \psi_{2Y}) = (0.25, 1, 1)$ and $(r_C, \psi_{1C}, \psi_{2C}) = (0.75, 1, 1)$,
- Setting 2: $(r_Y, \psi_{1Y}, \psi_{2Y}) = (0.75, 1, 1)$ and $(r_C, \psi_{1C}, \psi_{2C}) = (0.25, 1, 1)$,
- Setting 3: $(r_Y, \psi_{1Y}, \psi_{2Y}) = (0.1, 0.4, 0.6)$ and $(r_C, \psi_{1C}, \psi_{2C}) = (0.05, 0.6, 0.4)$.

These settings for the main distributions are then combined with the following types of contamination:

- **First type of contamination:** the distribution function F_c is given by

$$F_c(y_1, y_2) = \frac{1}{2}\left\{1 - e^{-y_1} + 1 - e^{-y_2}\right\} \mathbb{1}_{\{y_1 \geqslant 0, y_2 \geqslant 0\}}.$$

  This means that the contamination is on the axes according to the unit exponential distribution.

- **Second type of contamination:** the distribution function F_c has completely dependent unit exponential margins.

We simulate $N = 100$ datasets of size $n = 1000$, and consider $\varepsilon = 0, 0.025$ and 0.05. We estimate $A_Y(t)$ on the grid $\{0.05, 0.10, \ldots, 0.95\}$ for t. In Figures 2 till 7 we show the boxplots of the estimates $\widehat{A}_{Y,\alpha,n}(t)$ for $t \in \{0.05, 0.10, \ldots, 0.95\}$, together with $A_Y(t)$ (blue solid line) and $A_C(t)$ (green dashed line). In each of the figures the rows correspond to the levels of contamination, while the columns correspond to $\alpha = 0.1,\ 0.5$ and 1, respectively. From these simulations we can draw the following conclusions

- In case of no contamination, we can see that increasing α improves the estimation. The robustness of the MDPD method is thus partly used to correct for the censoring of the data.
- If we increase the contamination then estimation becomes more difficult, whatever α. Note that we can handle 5% contamination still reasonably well.
- For a given percentage of contamination we observe that increasing α from 0.1 to 0.5 clearly improves estimation but increasing α further to 1 does not lead to clear further improvements.
- The contamination affects the estimators in the expected direction: axes contamination pulls the estimator up and diagonal contamination down.
- For the three settings considered, contamination on the axes seems to be more difficult than contamination on the diagonal.
- In Setting 3 with asymmetric Pickands dependence functions, we recover A_Y quite well, despite the fact that A_Y and A_C are rather close.

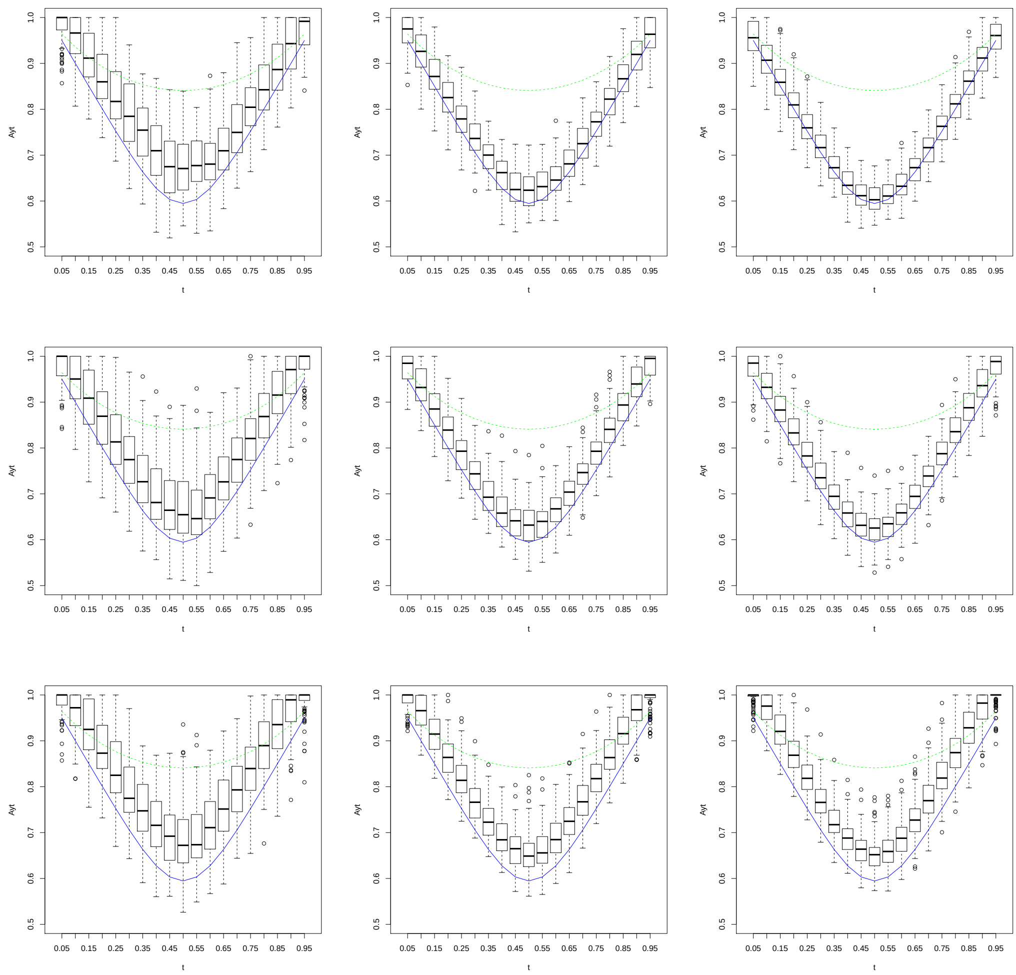

Figure 2: Setting 1, contamination on the axes. First row: no contamination, second row: 2.5% contamination, third row: 5% contamination. First column: $\alpha = 0.1$, second column: $\alpha = 0.5$, third column: $\alpha = 1$.

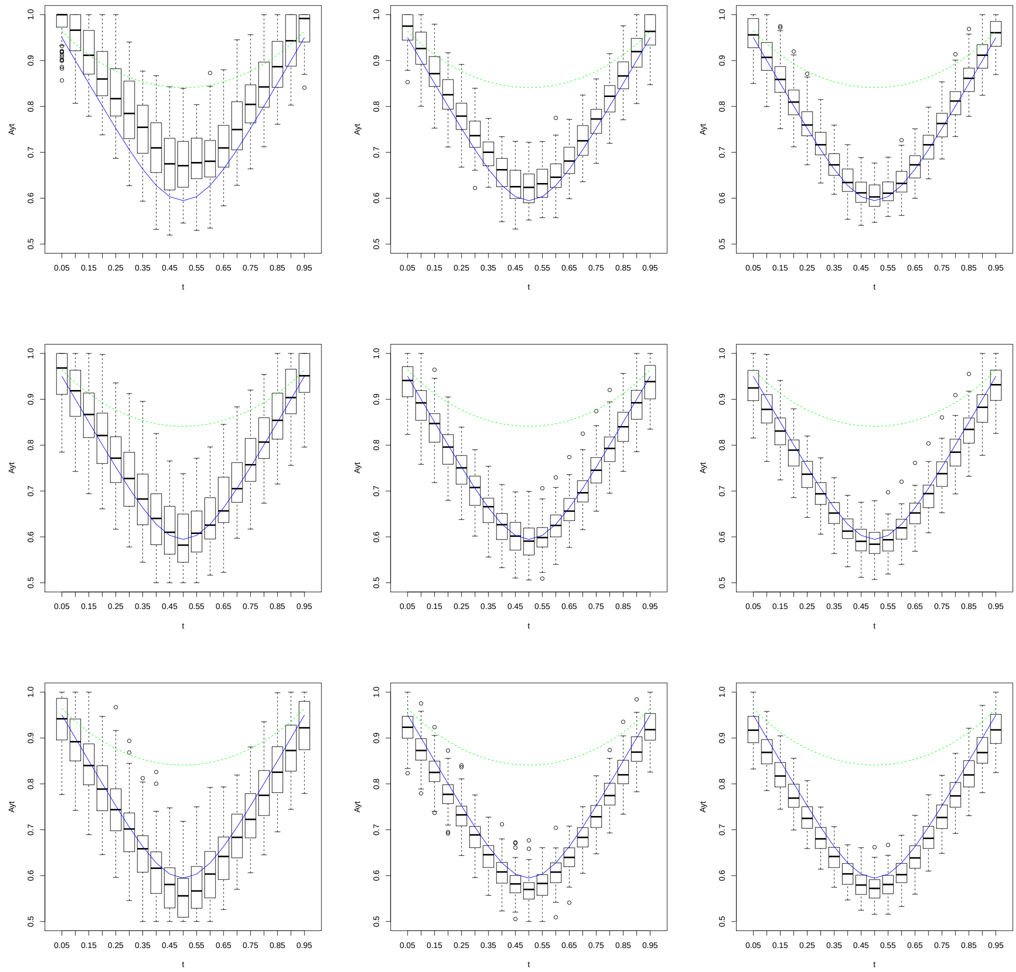

Figure 3: Setting 1, contamination on the diagonal. First row: no contamination, second row: 2.5% contamination, third row: 5% contamination. First column: $\alpha = 0.1$, second column: $\alpha = 0.5$, third column: $\alpha = 1$.

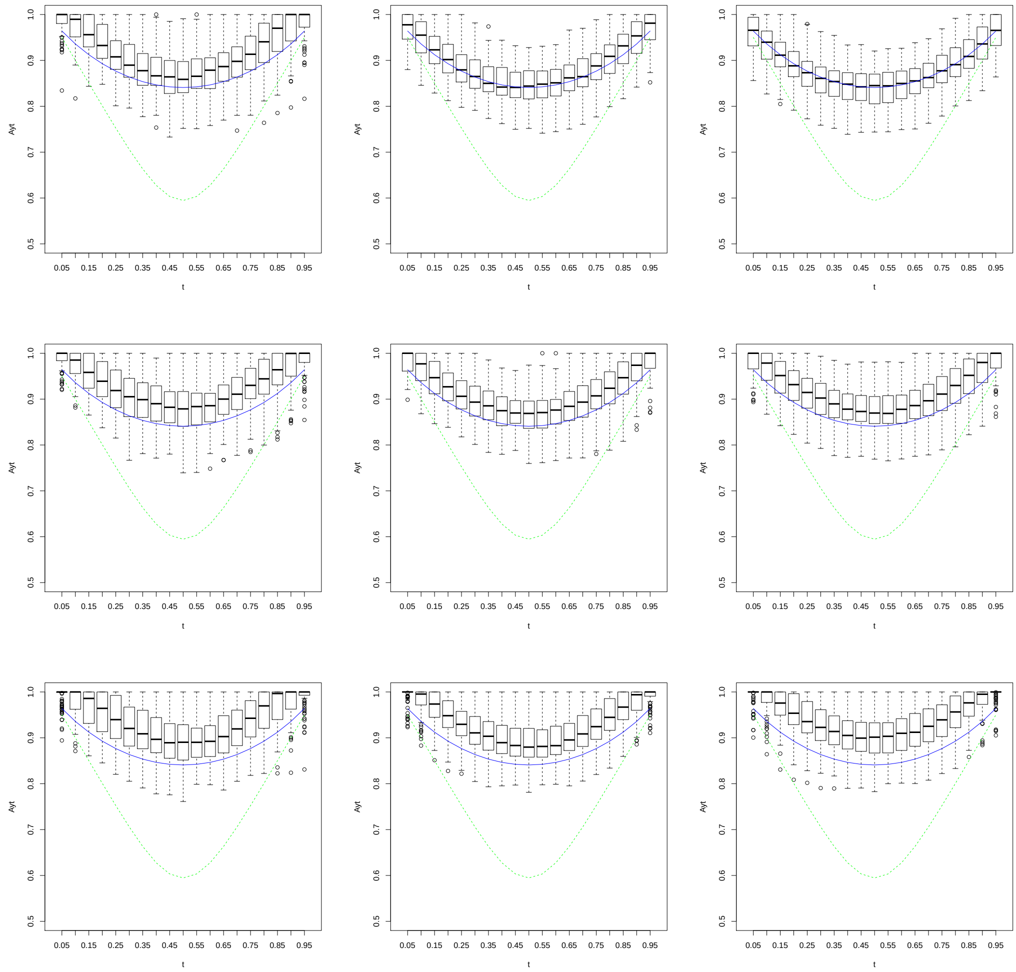

Figure 4: Setting 2, contamination on the axes. First row: no contamination, second row: 2.5% contamination, third row: 5% contamination. First column: $\alpha = 0.1$, second column: $\alpha = 0.5$, third column: $\alpha = 1$.

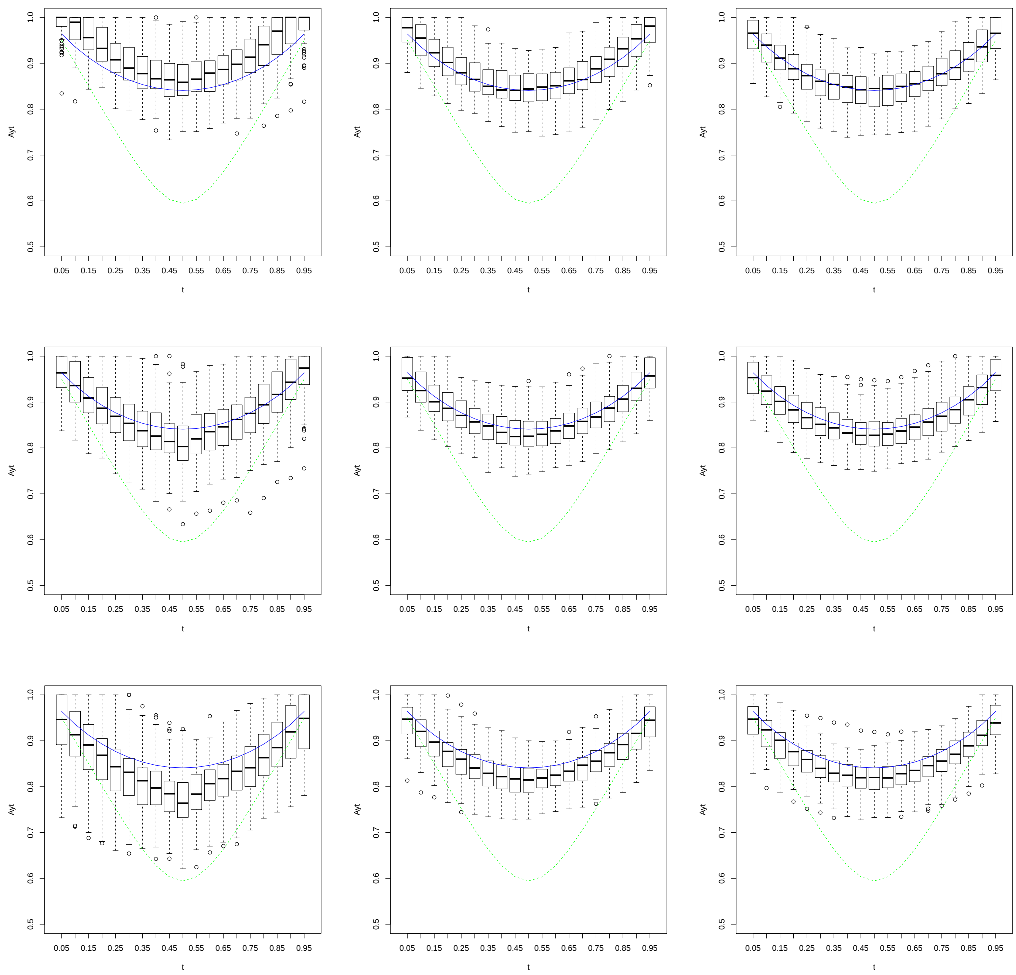

Figure 5: Setting 2, contamination on the diagonal. First row: no contamination, second row: 2.5% contamination, third row: 5% contamination. First column: $\alpha = 0.1$, second column: $\alpha = 0.5$, third column: $\alpha = 1$.

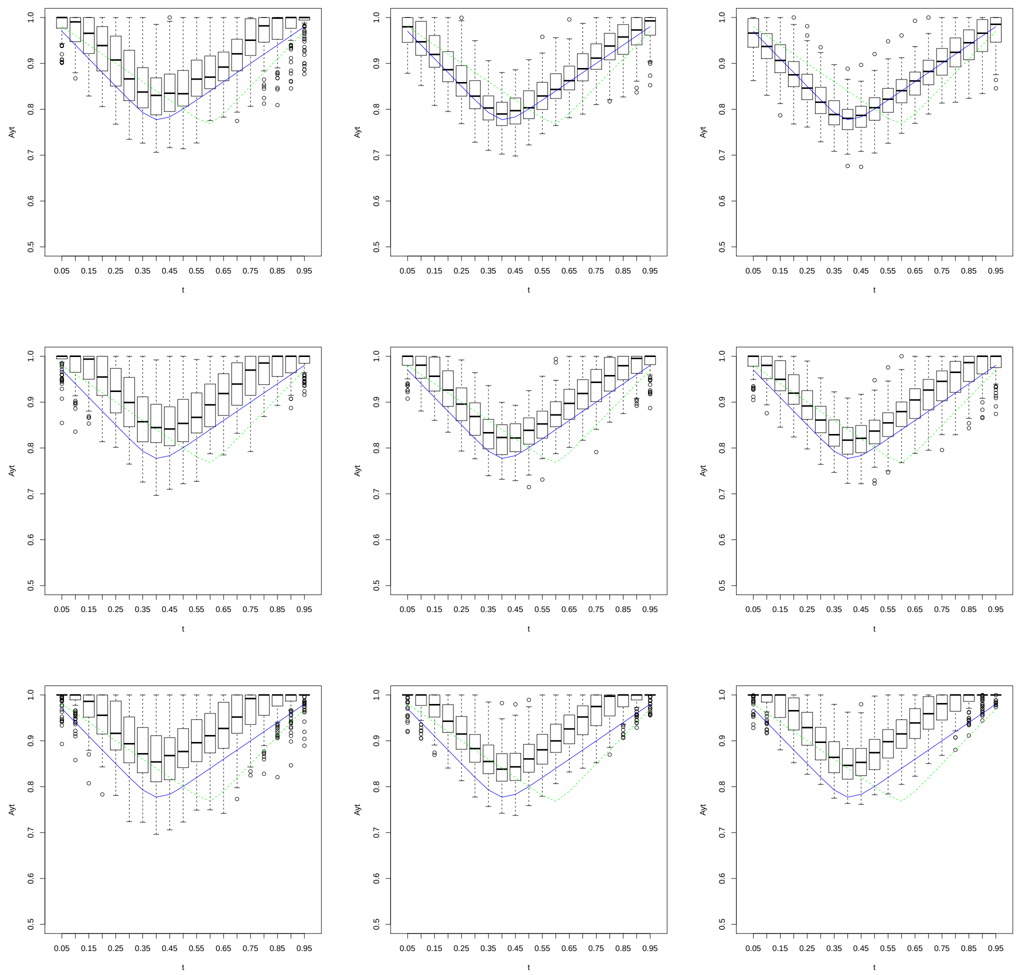

Figure 6: Setting 3, contamination on the axes. First row: no contamination, second row: 2.5% contamination, third row: 5% contamination. First column: $\alpha = 0.1$, second column: $\alpha = 0.5$, third column: $\alpha = 1$.

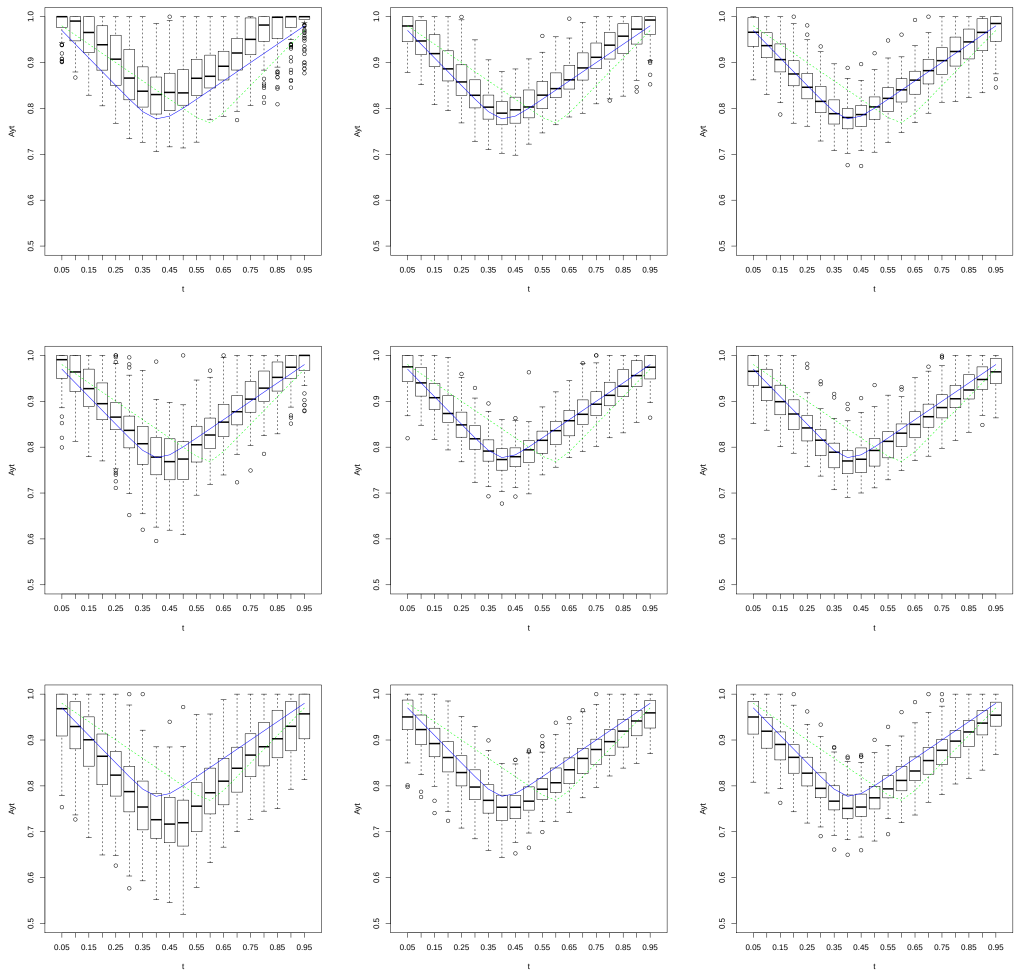

Figure 7: Setting 3, contamination on the diagonal. First row: no contamination, second row: 2.5% contamination, third row: 5% contamination. First column: $\alpha = 0.1$, second column: $\alpha = 0.5$, third column: $\alpha = 1$.

Next, we illustrate the situation where the marginal distributions are not unit exponential. We consider the case of a logistic Pickands dependence function for $(Y^{(1)}, Y^{(2)})$ and $(C^{(1)}, C^{(2)})$ with $r_Y = 0.25$ and $r_C = 0.75$, respectively. The marginal distributions of $Y^{(1)}$ and $Y^{(2)}$ are Exp(2), and those of $C^{(1)}$ and $C^{(2)}$ are Fréchet(2) shifted by 0.75, i.e. $F_{C^{(j)}}(c) = \exp(-(c - 0.75)^{-1/2})$, $c > 0.75$, $j = 1, 2$. This gives about 5% censoring. As before, we combine this setting with the two types of contamination, whereafter the observations are transformed to approximate unit exponential by $-\log \widehat{F}^{KM}_{Y^{(j)}}$, $j = 1, 2$. The results are shown in Figures 8 and 9, which have a layout that is the same as before. In case of no contamination, we have that overall we can capture the shape of the Pickands dependence function but the estimate is biased downwards. Using $\alpha = 0.5$ gives slightly better results than $\alpha = 0.1$, especially in the centre of the range for t, and increasing α to one does not lead to further improvements. When adding contamination the estimates behave again as expected, in particular they become pulled up under axes contamination and pulled down under diagonal contamination. Using $\alpha = 0.5$ gives some protection against contamination, in the sense that the results are close to those obtained under the uncontaminated case, and gives slightly less biased results than $\alpha = 1$.

5 Data example

In this section we illustrate the nonparametric MDPDE for A_Y on the dataset of insurance company indemnity claims introduced in Section 1.

In Frees and Valdez (1998) these data were analysed by fitting parametric copula models to the data using the maximum likelihood method. Based on their findings, the Gumbel-Hougaard copula (corresponding to the logistic model considered here) provided the best fit to the data, which gives an indication that the distribution underlying the data has an extreme value copula. Also in Cebrián *et al.* (2003), a similar loss-ALAE dataset was modelled with extreme value copulas.

The nonparametric MDPDE of A_Y is shown in Figure 10 for $\alpha = 0.1$ (solid line), $\alpha = 0.5$ (dotted line) and $\alpha = 1$ (dashed-dotted line), along with the parametric maximum likelihood estimate obtained by Frees and Valdez (1998) (dashed line). Our nonparametric MDPDE with $\alpha = 0.1$ is close to the parametric estimate from Frees and Valdez (1998), and this can be seen as a further confirmation of the fit of the parametric Gumbel-Hougaard copula proposed in the latter. Note that both these estimates are not robust with respect to outliers. To overcome this we also applied the MDPD method with $\alpha = 0.5$ and $\alpha = 1$, and the obtained estimates differ slightly from the MDPDE with $\alpha = 0.1$ and the Frees and Valdez (1998) estimate. This might indicate the presence of a few outliers in the loss-ALAE data, as was also suggested by the scatterplots given in Figure 1, though these do not seem to have a disturbing effect on the estimation, which is reasonable given a total sample size of $n = 1500$.

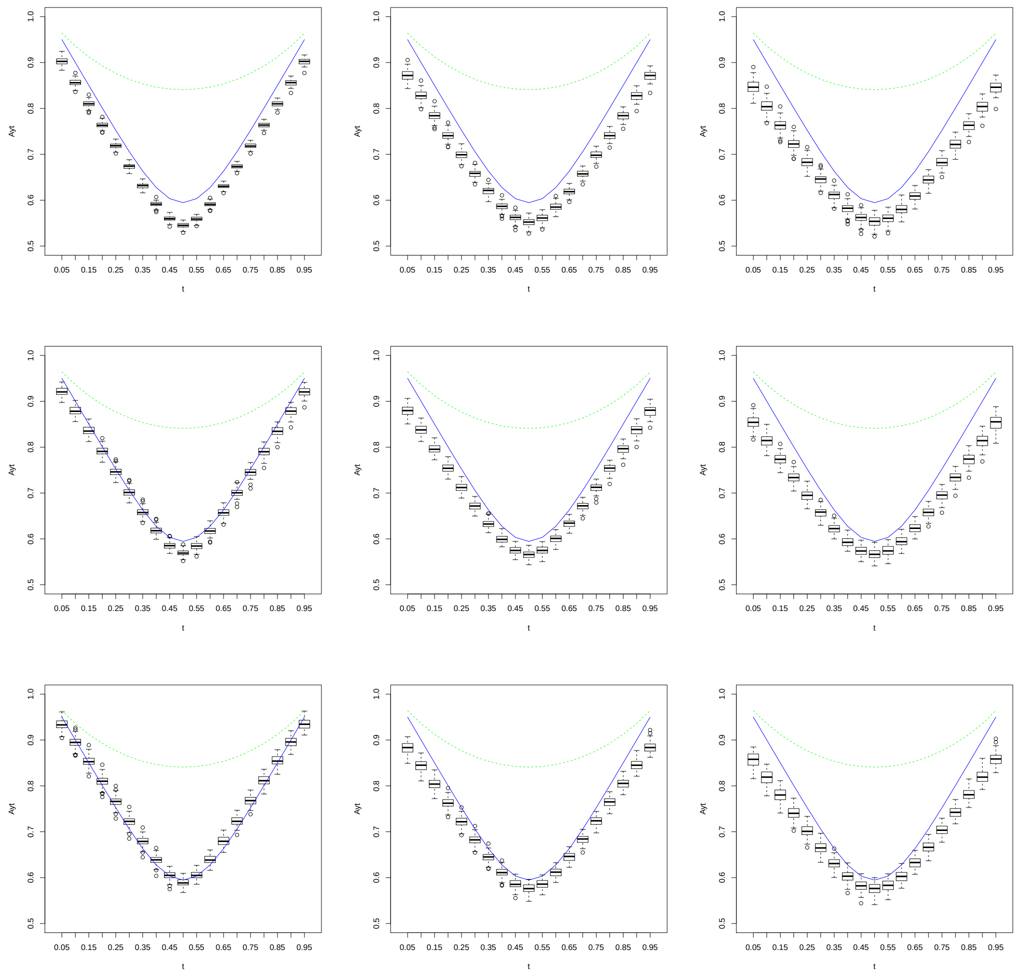

Figure 8: Setting 1, contamination on the axes, margins transformed to unit exponential using $-\log \widehat{F}_Y^{KM}$. First row: no contamination, second row: 2.5% contamination, third row: 5% contamination. First column: $\alpha = 0.1$, second column: $\alpha = 0.5$, third column: $\alpha = 1$.

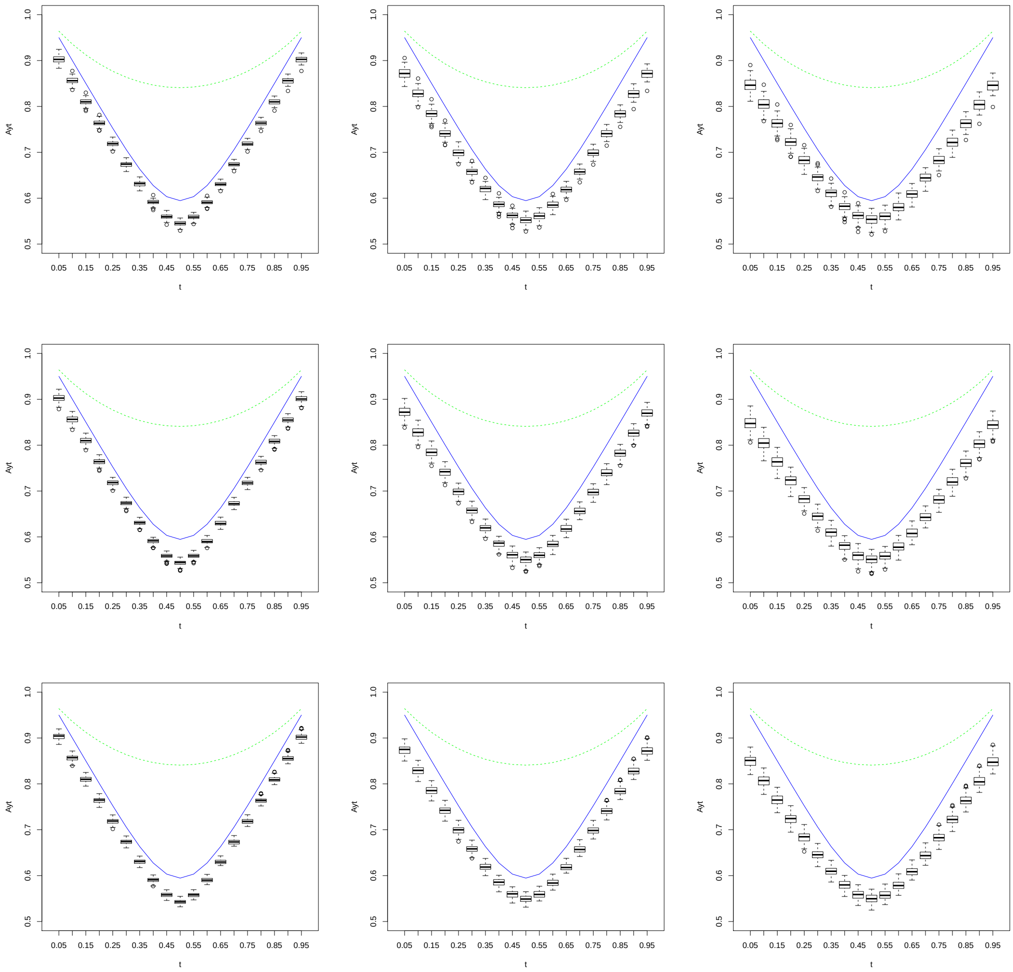

Figure 9: Setting 1, contamination on the diagonal, margins transformed to unit exponential using $-\log \widehat{F}_Y^{KM}$. First row: no contamination, second row: 2.5% contamination, third row: 5% contamination. First column: $\alpha = 0.1$, second column: $\alpha = 0.5$, third column: $\alpha = 1$.

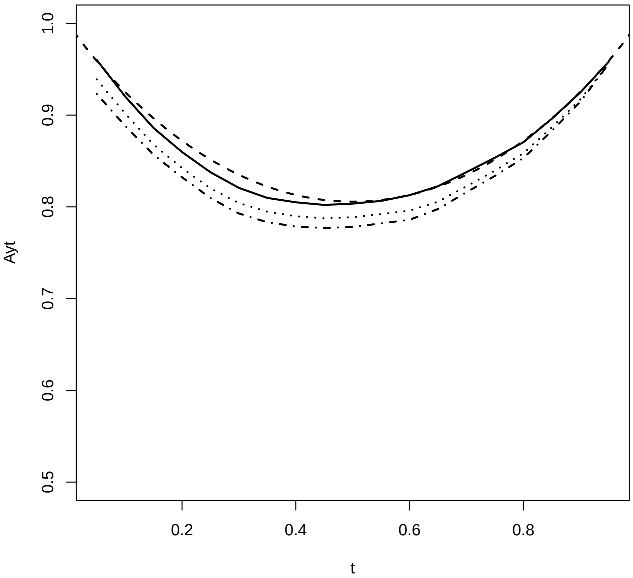

Figure 10: LOSS-ALAE data: nonparametric estimate $\widehat{A}_{Y,\alpha,n}$ for $\alpha = 0.1$ (solid line), $\alpha = 0.5$ (dotted line) and $\alpha = 1$ (dashed-dotted line). The parametric Gumbel-Hougaard estimate of Frees and Valdez (1998) is given by the dashed line.

6 Appendix: Proofs of the results

6.1 Derivatives of the MDPD objective function

The arguments used to establish the consistency and asymptotic normality of the MDPDE depend on the MDPD objective function and its derivates. For convenience, in this section we give $\widehat{\Delta}^{(j)}_{\alpha,t}(a)$, for $j = 1, 2$ and 3. Straightforward computations for $\alpha > 0$, give

$$\begin{aligned}
\widehat{\Delta}^{(1)}_{\alpha,t}(a) &= \frac{\alpha}{1+\alpha} a^{\alpha-1} - (1+\alpha)a^{\alpha-1} \int_0^\infty e^{-a\alpha z}(1-az)\, d\widehat{F}^{KM}_{\widetilde{Y}_t}(z), \\
\widehat{\Delta}^{(2)}_{\alpha,t}(a) &= \frac{\alpha(\alpha-1)}{1+\alpha} a^{\alpha-2} - (1+\alpha)a^{\alpha-2} \int_0^\infty e^{-a\alpha z}\left(\alpha - 1 - 2\alpha a z + \alpha a^2 z^2\right) d\widehat{F}^{KM}_{\widetilde{Y}_t}(z), \\
\widehat{\Delta}^{(3)}_{\alpha,t}(a) &= \frac{\alpha(\alpha-1)(\alpha-2)}{1+\alpha} a^{\alpha-3} \\
&\quad -(1+\alpha)a^{\alpha-3} \int_0^\infty e^{-a\alpha z}\left[(\alpha-1)(\alpha-2) + 3\alpha(1-\alpha)az + 3\alpha^2 a^2 z^2 - \alpha^2 a^3 z^3\right] d\widehat{F}^{KM}_{\widetilde{Y}_t}(z).
\end{aligned}$$

6.2 Proof of Theorem 3.1

Direct computations yield, for $z > 0$,

$$\begin{aligned}
\widetilde{H}^0_t(z) &= \frac{A_C(t)}{A_C(t) + A_Y(t)} \left\{1 - e^{-[A_C(t)+A_Y(t)]z}\right\}, \\
\widetilde{H}^1_t(z) &= \frac{A_Y(t)}{A_C(t) + A_Y(t)} \left\{1 - e^{-[A_C(t)+A_Y(t)]z}\right\}, \\
\gamma^{(0)}_t(z) &= \exp\{A_C(t)z\}, \\
\gamma^{(1)}_{t,\xi}(z) &= A_Y(t) e^{[A_C(t)+A_Y(t)]z} \int_z^\infty \omega^\xi\, e^{-(1+\alpha)A_Y(t)\omega}\, d\omega, \\
\gamma^{(2)}_{t,\xi}(z) &= \frac{A_Y(t)A_C(t)}{A_Y(t)+A_C(t)} \int_0^\infty \omega^\xi\, e^{-(1+\alpha)A_Y(t)\omega} \left\{e^{[A_C(t)+A_Y(t)]\min(z,\omega)} - 1\right\} d\omega.
\end{aligned}$$

Then, essentially we need to verify the conditions (1.5) and (1.6) in Stute (1995). In our context, as soon as $(1+2\alpha)A_Y(t) - A_C(t) > 0$, we have

$$\begin{aligned}
\int_0^\infty [\phi_{t,\xi}(z)\gamma^{(0)}_t(z)]^2 d\widetilde{H}^1_t(z) &< \infty, \\
\text{and } \int_0^\infty |\phi_{t,\xi}(z)| \sqrt{C(z)}\, dF_{\widetilde{Y}_t}(z) &< \infty,
\end{aligned}$$

where

$$C(z) := \int_0^z \frac{dF_{\widetilde{C}_t}(y)}{[1 - F_{\widetilde{Z}_t}(y)][1 - F_{\widetilde{C}_t}(y)]}.$$

Using the fact that

$$\int_0^\infty \phi_{t,\xi}(\omega)\, dF_{\widetilde{Y}_t}(\omega) = \frac{\Gamma(\xi+1)}{(\alpha+1)^{\xi+1} A^\xi_Y(t)},$$

a direct application of Theorem 1.1 in Stute (1995) achieves the proof of Theorem 3.1.

6.3 Proof of Corollary 3.1

Similarly as Corollary 1.2 in Stute (1995), Corollary 3.1 is a direct consequence of the result in our Theorem 3.1, combined with the classical central limit theorem and Slutsky's theorem.

6.4 Proof of Theorem 3.2

To prove this theorem, we will make use of the Cramér-Wold device (see, e.g., Severini, 2005, p. 337), according to which it is sufficient to show that

$$\beta^T \sqrt{n}\,(\mathbb{T}_n - \mathcal{T}) \xrightarrow{d} \mathcal{N}(0, \beta^T \Sigma \beta), \tag{8}$$

for all $\beta \in \mathbb{R}^m$. A straightforward rearrangement of the terms leads to

$$\begin{aligned} \beta^T \sqrt{n}\,(\mathbb{T}_n - \mathcal{T}) &= \sqrt{n}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\beta_j\left(\eta_{t_j,\xi_j}(\widetilde{Z}_{t_j,i}) - \frac{\Gamma(\xi_j+1)}{(\alpha+1)^{\xi_j+1}A_Y^{\xi_j}(t_j)}\right) + \sqrt{n}\sum_{j=1}^{m}\beta_j R_{n,t_j,\xi_j} \\ &=: T_{1,n} + T_{2,n}. \end{aligned}$$

According to Theorem 3.1, $T_{2,n} = o_{\mathbb{P}}(1)$ as $n \to \infty$. Now, concerning $T_{1,n}$, we have $T_{1,n} = \sum_{i=1}^{n} V_i/\sqrt{n}$, where $V_1, \ldots, V_n$ are independent and identically distributed centered random variables with variance given by

$$Var(V_i) = \sum_{j=1}^{m}\sum_{k=1}^{m}\beta_j\beta_k Cov\left(\eta_{t_j,\xi_j}(\widetilde{Z}_{t_j,i}), \eta_{t_k,\xi_k}(\widetilde{Z}_{t_k,i})\right), \text{ for } 1 \leqslant i \leqslant n.$$

This implies that $Var(T_{1,n}) = \beta^T\Sigma\beta$ where $\Sigma = (Cov(\eta_{t_j,\xi_j}(\widetilde{Z}_{t_j}), \eta_{t_k,\xi_k}(\widetilde{Z}_{t_k})))_{1\leqslant j,k\leqslant m}$. By invoking the central limit theorem, $T_{1,n} \xrightarrow{d} \mathcal{N}(0, \beta^T\Sigma\beta)$. Now, by Slutsky's theorem we also have (8), and Theorem 3.2 follows.

6.5 Proof of Theorem 3.3

To prove the existence and consistency of $\widehat{A}_{Y,\alpha,n}(t)$ we adapt the proof of Theorem 5.1 in Chapter 6 of Lehmann and Casella (1998), where existence and consistency of solutions of the likelihood equations is established, to the MDPDE framework. First we show that for any r sufficiently small

$$\mathbb{P}_{A_Y(t)}(\widehat{\Delta}_{\alpha,t}(A_Y(t)) < \widehat{\Delta}_{\alpha,t}(a) \text{ for } a = A_Y(t) \pm r) \to 1. \tag{9}$$

Note that r should be such that $A_Y(t) \pm r$ belongs to the parameter space, here $[1/2 - \delta, 1 + \delta]$, where $\delta > 0$.

By Taylor's theorem

$$\begin{aligned}\widehat{\Delta}_{\alpha,t}(a) - \widehat{\Delta}_{\alpha,t}(A_Y(t)) &= \widehat{\Delta}^{(1)}_{\alpha,t}(A_Y(t))(a - A_Y(t)) + \frac{1}{2}\widehat{\Delta}^{(2)}_{\alpha,t}(A_Y(t))(a - A_Y(t))^2 \\ &\quad + \frac{1}{6}\widehat{\Delta}^{(3)}_{\alpha,t}(\breve{A}_Y(t))(a - A_Y(t))^3 \\ &=: S_1 + S_2 + S_3,\end{aligned}$$

where $\breve{A}_Y(t)$ is a value between a and $A_Y(t)$.

Using Section 6.1 and the result of Corollary 3.1, we have $\widehat{\Delta}^{(1)}_{\alpha,t}(A_Y(t)) \xrightarrow{\mathbb{P}} 0$, as $n \to \infty$, and hence $|S_1| < r^3$ with probability tending to 1. As for S_2, we have

$$\widehat{\Delta}^{(2)}_{\alpha,t}(A_Y(t)) \xrightarrow{\mathbb{P}} \frac{1+\alpha^2}{(1+\alpha)^2}[A_Y(t)]^{\alpha-2},$$

and thus there exists a $c > 0$ such that $S_2 > cr^2$ with probability tending to one. For S_3 we use the fact that

$$\sup_{a \in [1/2-\delta, 1+\delta]} |\widehat{\Delta}^{(3)}_{\alpha,t}(a)| \leqslant M,$$

where $M \xrightarrow{\mathbb{P}} d < \infty$. Thus $|S_3| < br^3$ with probability tending to one, where $b := d/3$. Combining the above we find that with probability tending to 1,

$$S_1 + S_2 + S_3 > cr^2 - (1+b)r^3,$$

for $a = A_Y(t) \pm r$. Clearly, since the right-hand side of the above inequality is positive if $r < c/(1+b)$, (9) follows.

To complete the proof of the existence and consistency we adjust the line of argumentation of Theorem 3.7 in Chapter 6 of Lehmann and Casella (1998). For $r > 0$, small enough such that $A_Y(t) \pm r \in [1/2-\delta, 1+\delta]$ we let

$$S_n(r) := \{\widehat{\Delta}_{\alpha,t}(A_Y(t)) < \widehat{\Delta}_{\alpha,t}(a) \text{ for } a = A_Y(t) \pm r\}.$$

From the above we have that $\mathbb{P}_{A_Y(t)}(S_n(r)) \to 1$ for any such r, and hence there exists a sequence $r_n \downarrow 0$ such that $\mathbb{P}_{A_Y(t)}(S_n(r_n)) \to 1$ as $n \to \infty$. By the differentiability of $\widehat{\Delta}_{\alpha,t}$ we have that $v \in S_n(r_n)$ implies that there exists a value $\widehat{A}_{Y,n} \in [A_Y(t) - r_n, A_Y(t) + r_n]$ for which $\widehat{\Delta}_{\alpha,t}$ attains a local minimum, and thus $\widehat{\Delta}^{(1)}_{\alpha,t}(\widehat{A}_{Y,n}) = 0$. Now let $\widehat{A}^*_{Y,n} := \widehat{A}_{Y,n}$ for $v \in S_n(r_n)$ and arbitrary otherwise. Clearly

$$\mathbb{P}_{A_Y(t)}(\widehat{\Delta}_{\alpha,t}(\widehat{A}^*_{Y,n}) = 0) \geqslant \mathbb{P}_{A_Y(t)}(S_n(r_n)) \to 1,$$

as $n \to \infty$. Thus with probability tending to 1 there exists a sequence of solutions to the estimating equation (4). Then, for any fixed $r > 0$ and n sufficiently large

$$\mathbb{P}_{A_Y(t)}(|\widehat{A}^*_{Y,n} - A_Y(t)|) < r) \geqslant \mathbb{P}_{A_Y(t)}(|\widehat{A}^*_{Y,n} - A_Y(t)|) < r_n) \geqslant \mathbb{P}_{A_Y(t)}(S_n(r_n)) \to 1,$$

which establishes the consistency of the sequence $(\widehat{A}^*_{Y,n})_{n \geqslant 1}$.

6.6 Proof of Theorem 3.4

First we prove the result for a specific, single $t \in [0,1]$, since this is more explicit. Our starting point is the estimating equation (4). By applying a Taylor series expansion around the true value $A_Y(t)$, we get

$$0 = \widehat{\Delta}^{(1)}_{\alpha,t}(A_Y(t)) + \left(\widehat{A}_{Y,\alpha,n}(t) - A_Y(t)\right)\widehat{\Delta}^{(2)}_{\alpha,t}(A_Y(t)) + \frac{1}{2}\left(\widehat{A}_{Y,\alpha,n}(t) - A_Y(t)\right)^2 \widehat{\Delta}^{(3)}_{\alpha,t}(\widetilde{A}_Y(t))$$

where $\widetilde{A}_Y(t)$ is a random value between $A_Y(t)$ and $\widehat{A}_{Y,\alpha,n}(t)$. A straightforward rearrangement of the above display gives

$$\sqrt{n}\left(\widehat{A}_{Y,\alpha,n}(t) - A_Y(t)\right) = \frac{-\sqrt{n}\,\widehat{\Delta}^{(1)}_{\alpha,t}(A_Y(t))}{\widehat{\Delta}^{(2)}_{\alpha,t}(A_Y(t)) + \frac{1}{2}\widehat{\Delta}^{(3)}_{\alpha,t}(\widetilde{A}_Y(t))\left(\widehat{A}_{Y,\alpha,n}(t) - A_Y(t)\right)}. \tag{10}$$

Remark that, according to Section 6.1, the numerator of the right-hand side of (10) can be rewritten as

$$\begin{aligned} -\sqrt{n}\widehat{\Delta}^{(1)}_{\alpha,t}(A_Y(t)) &= (1+\alpha)[A_Y(t)]^{\alpha-1}\left\{\sqrt{n}\left(T_n(t,0) - \frac{1}{1+\alpha}\right)\right. \\ &\quad \left. -A_Y(t)\sqrt{n}\left(T_n(t,1) - \frac{1}{(1+\alpha)^2 A_Y(t)}\right)\right\}. \end{aligned}$$

After some tedious computations we obtain

$$\begin{aligned} Var(\eta_{t,0}(\widetilde{Z}_t)) &= \frac{\alpha^2 A_Y(t)}{(1+\alpha)^2[(1+2\alpha)A_Y(t) - A_C(t)]}, \\ Var(\eta_{t,1}(\widetilde{Z}_t)) &= \frac{(1+2\alpha+2\alpha^4)A_Y^2(t) - 2(1+\alpha-\alpha^2)A_Y(t)A_C(t) + A_C^2(t)}{(1+\alpha)^4 A_Y(t)[(1+2\alpha)A_Y(t) - A_C(t)]^3}, \\ Cov(\eta_{t,0}(\widetilde{Z}_t), \eta_{t,1}(\widetilde{Z}_t)) &= -\frac{\alpha(1+\alpha-\alpha^2)A_Y(t) - \alpha A_C(t)}{(1+\alpha)^3[(1+2\alpha)A_Y(t) - A_C(t)]^2}. \end{aligned}$$

Then, by Theorem 3.2

$$-\sqrt{n}\widehat{\Delta}^{(1)}_{\alpha}(A_Y(t)) \xrightarrow{d} \mathcal{N}(0, \bar{\sigma}^2),$$

where

$$\begin{aligned} \bar{\sigma}^2 \;:=\; & \frac{[A_Y(t)]^{2\alpha-1}}{(1+\alpha)^2[(1+2\alpha)A_Y(t) - A_C(t)]^3}\left\{(1+4\alpha+9\alpha^2+14\alpha^3+13\alpha^4+8\alpha^5+4\alpha^6)[A_Y(t)]^2\right. \\ & \left. -2(1+3\alpha+5\alpha^2+6\alpha^3+4\alpha^4+2\alpha^5)A_Y(t)A_C(t) + (1+2\alpha+3\alpha^2+2\alpha^3+\alpha^4)[A_C(t)]^2\right\}. \end{aligned}$$

Now, concerning the denominator of the right-hand side of (10), according to Section 6.1 and Corollary 3.1, $\widehat{\Delta}^{(3)}_{\alpha,t}(a)$ is bounded in probability and

$$\widehat{\Delta}^{(2)}_{\alpha,t}(A_Y(t)) \xrightarrow{\mathbb{P}} \frac{1+\alpha^2}{(1+\alpha)^2}[A_Y(t)]^{\alpha-2}.$$

This implies that

$$\widehat{\Delta}^{(2)}_{\alpha,t}(A_Y(t)) + \frac{1}{2}\widehat{\Delta}^{(3)}_{\alpha,t}(\widetilde{A}_Y(t))\left(\widehat{A}_{Y,\alpha,n}(t) - A_Y(t)\right) \xrightarrow{\mathbb{P}} \frac{1+\alpha^2}{(1+\alpha)^2}[A_Y(t)]^{\alpha-2}.$$

Again by Slutsky's theorem, we have that Theorem 3.4 follows for a single $t \in [0,1]$.

Now we focus on the general finite dimensional convergence. Let

$$\mathbb{X}_n := \begin{bmatrix} \sqrt{n}\left(T_n(t_1,0) - \frac{1}{1+\alpha}\right) \\ \sqrt{n}\left(T_n(t_1,1) - \frac{1}{(1+\alpha)^2 A_Y(t_1)}\right) \\ \vdots \\ \sqrt{n}\left(T_n(t_m,0) - \frac{1}{1+\alpha}\right) \\ \sqrt{n}\left(T_n(t_m,1) - \frac{1}{(1+\alpha)^2 A_Y(t_m)}\right) \end{bmatrix}.$$

Then

$$\mathbb{D}_n := -\begin{bmatrix} \sqrt{n}\widehat{\Delta}^{(1)}_{\alpha,t_1}(A_Y(t_1)) \\ \vdots \\ \sqrt{n}\widehat{\Delta}^{(1)}_{\alpha,t_m}(A_Y(t_m)) \end{bmatrix} = \mathbb{A}\mathbb{X}_n,$$

and we have $\mathbb{D}_n \xrightarrow{d} \mathcal{N}_m(0, \mathbb{A}\mathbb{C}\mathbb{A}^T)$. Now let $\mathbb{B}_n$ be a $(m \times m)$ diagonal matrix with entries

$$\mathbb{B}_{n,i,i} := \frac{1}{\widehat{\Delta}^{(2)}_{\alpha,t_i}(A_Y(t_i)) + \frac{1}{2}\widehat{\Delta}^{(3)}_{\alpha,t_i}(\widetilde{A}_Y(t_i))\left(\widehat{A}_{Y,\alpha,n}(t_i) - A_Y(t_i)\right)}.$$

Then

$$\sqrt{n}\begin{bmatrix} \widehat{A}_{Y,\alpha,n}(t_1) - A_Y(t_1) \\ \vdots \\ \widehat{A}_{Y,\alpha,n}(t_m) - A_Y(t_m) \end{bmatrix} = \mathbb{B}_n\mathbb{D}_n,$$

and the result follows.

References

[1] Basu, A., Harris, I. R., Hjort, N. L. and Jones, M. C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika*, **85**, 549–559.

[2] Beirlant, J., Bardoutsos, A., de Wet, T. and 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. and Fils-Villetard, A. (2007). Estimation of the extreme value index and extreme quantiles under random censoring. *Extremes*, **10**, 151–174.

[4] Capéraà, P., Fougères, A. L. and Genest, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, **84**, 567–577.

[5] Cebrián, A. C., Denuit, M. and Lambert, P. (2003). Analysis of bivariate tail dependence using extreme value copulas: an application to the SOA medical large claims database. *Belgian Actuarial Bulletin*, **3**, 33–41.

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

[7] Escobar-Bach, M., Goegebeur, Y., Guillou, A. and You, A. (2017). Bias-corrected and robust estimation of the bivariate stable tail dependence function. *TEST*, **26**, 284–307.

[8] Fils-Villetard, A., Guillou, A. and Segers, J. (2008). Projection estimators of Pickands dependence functions. *Canadian Journal of Statistics*, **36**, 369–382.

[9] Fougères, A. L., de Haan, L. and Mercadier, C. (2015). Bias correction in multivariate extremes. *Annals of Statistics*, **43**, 903–934.

[10] Frees, E. and Valdez, E. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, **2**, 1–25.

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

[12] Gomes, M. I. and Neves, M. M. (2011). Estimation of the extreme value index for randomly censored data. *Biometrical Letters*, **48**, 1–22.

[13] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**, 457–481.

[14] Ledford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society Series B*, **59**, 475–499.

[15] Lehmann, E. L. and Casella, G. (1998). *Theory of point estimation*, Springer.

[16] Micocci, M. and Masala, G. (2009). Loss-ALAE modelling through a copula dependence structure. *Investment Management and Financial Innovations*, **6**, 67–80.

[17] Ndao, P., Diop, A. and 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.

[18] Pickands, J. (1981). Multivariate extreme value distributions. *Bulletin of the International Statistical Institut*, **49**, 859–878.

[19] Severini, T. A. (2005). *Elements of distribution theory*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge.

[20] Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, **8**, 229–231.

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

[22] Stute, W. (1995). The central limit theorem under random censorship. *The Annals of Statistics*, **23**, 422–439.

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