

# On the study of extremes with dependent random right-censoring

Gilles Stupfler<sup>1</sup> 

Received: 13 October 2017 / Revised: 7 June 2018 / Accepted: 11 June 2018 /  
Published online: 26 June 2018  
© The Author(s) 2018

**Abstract** The study of extremes in missing data frameworks is a recent developing field. In particular, the randomly right-censored case has been receiving a fair amount of attention in the last decade. All studies on this topic, however, essentially work under the usual assumption that the variable of interest and the censoring variable are independent. Furthermore, a frequent characteristic of estimation procedures developed so far is their crucial reliance on particular properties of the asymptotic behaviour of the response variable  $Z$  (that is, the minimum between time-to-event and time-to-censoring) and of the probability of censoring in the right tail of  $Z$ . In this paper, we focus instead on elucidating this asymptotic behaviour in the dependent censoring case, and, more precisely, when the structure of the dependent censoring mechanism is given by an extreme value copula. We then draw a number of consequences of our results, related to the asymptotic behaviour, in this dependent context, of a number of estimators of the extreme value index of the random variable of interest that were introduced in the literature under the assumption of independent censoring, and we discuss more generally the implications of our results on the inference of the extremes of this variable.

**Keywords** Random right-censoring · Dependent censoring · Extreme value copula · Extreme value index · Tail identifiability · Tail censoring probability

**AMS Subject Classifications** 62G32 · 62N01 · 62N02

---

✉ Gilles Stupfler  
gilles.stupfler@nottingham.ac.uk

<sup>1</sup> School of Mathematical Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK

## 1 Introduction

The problem of missing data, and in particular censoring, is frequently encountered in certain fields of statistical applications. The archetypal example of censoring is arguably the study of the survival times of patients to a given chronic disease in a medical follow-up study lasting up to a fixed time  $t$ . If a patient is diagnosed with the disease at time  $s$ , then his/her survival time will be known if and only if he/she dies before time  $t$ . If this is not the case, then the only information available is that his/her survival time is not less than the censoring time  $t - s$ . In mathematical terms, the information available to the practitioner is the pair  $(Z, \delta)$ , where  $Z$  is the minimum between the survival time and censoring time, and  $\delta$  is the 0-1 variable equal to 1 if and only if the survival time is actually observed. This situation is one of the most frequent examples of random right-censoring, which shall be our framework in this paper.

Random right-censoring is also found in actuarial science: in non-life insurance, Reynkens et al. (2017) use random right-censoring to model unsettled claims, about which the insurer only knows what has been paid up to date rather than the ultimate claim amount, i.e. the sum of all payments for the claim, which is only known when the claim is fully settled. In life insurance, meanwhile, any study that monitors policyholders in a given time period contains right-censored data points, since many of the subjects still live at the end of the monitoring period. Another example is reliability data analysis: if a car company collects failure data during the warranty period, then a failure could happen not only because of a failure in the mechanics of the car, but also because of an accident or driver error. In the latter case, time-to-accident or time-to-driver-error should be treated as the censoring time, see Modarres et al. (2009). Random right-censoring should not be confused with other types of missing data mechanisms such as right-truncation, where no information is available at all when the random variable of interest is not actually observed. In a heavy-tailed context and when the right-truncation point is unknown but non-random, this problem is considered by Aban et al. (2006), while the earliest reference tackling the random right-truncation problem in the extreme value context is Gardes and Stupfler (2015).

In a random right-censoring framework, a stimulating problem is the estimation of extreme parameters of the underlying distribution of the variable of interest, a prime example of which being its extreme value index. In the aforementioned examples, this would ultimately lead to the analysis of survival times of exceptionally strong/weak patients to a given disease, extreme losses/payouts in insurance, or failure times for highly resistant/unreliable devices. This subfield of extreme value statistics has received a good amount of attention in recent years: we refer to Beirlant et al. (2007, 2010, 2016), Einmahl et al. (2008), Gomes and Neves (2011), Ndao et al. (2014, 2016), Sayah et al. (2014), Worms and Worms (2014), Brahimi et al. (2015) and Stupfler (2016). All of these papers work under the hypothesis that the variable of interest  $Y$  and the censoring variable  $T$  are independent random variables; in the case of Ndao et al. (2014, 2016) and Stupfler (2016), the condition is actually conditional independence given a suitable, in practice low-dimensional, covariate. Among others, the independence assumption allows for a very convenient expression of the cumulative distribution function of the observed variable  $Z$  and, as a result, for a simple

discussion of its extreme value properties. The asymptotic behaviour of  $Z$ , and that of  $\delta|Z$  for high  $Z$ , can then be fruitfully exploited to construct a class of simple estimators of the extreme value index of  $Y$  (see, for instance, Beirlant et al. 2007 and Einmahl et al. 2008). It should be said that since the pioneering paper of Kaplan and Meier (1958) on the product-limit estimator for the survival function, the assumption of independent censoring is arguably the standard assumption in the context of random right-censoring.

And yet, cases in which there are strong suspicions of dependence between the variable of interest and the censoring time have been reported several times over the last decades. An early reference is Lagakos (1979). In medical studies especially, a common cause of the probable violation of the independence hypothesis is a sizeable number of patient dropouts (Huang and Zhang 2008; Jackson et al. 2014). Crucially, using traditional estimators such as the Kaplan-Meier estimator when there is dependence may yield to invalid inferences, see Fisher and Kanarek (1974), Klein and Moeschberger (1987) and the introduction of Ebrahimi and Molefe (2003). Moreover, there is the additional issue of identifiability, in the sense that if the dependence structure is completely unspecified then the distribution of  $(Y, T)$  cannot be recovered from that of the pair  $(Z, \delta)$ , see Tsiatis (1975). A number of authors have suggested partial solutions to tackle the problem of dependence: some recent efforts include fitting specific types of known copulas (Li et al. 2007; Huang and Zhang 2008) or assuming weaker assumptions than independence on the pair  $(X, Y)$  (Ebrahimi et al. 2003). The integration of valuable, preferably high-dimensional covariate information may also be helpful if conditional independence given the covariate is reasonable (see Zeng 2004, Li et al. 2007 and Hsu and Taylor 2010). Let us point out that the studies by Ndao et al. (2014, 2016) and Stupfler (2016) did consider incorporating a low-dimensional covariate  $X$ , but the common idea underpinning these papers is to estimate the conditional extreme value index of  $Y$  given  $X = x$  by adapting the procedure of Einmahl et al. (2008), developed for independent censoring, using kernel-type techniques. The introduction of covariate information is therefore not motivated by a reduction of the dependence between  $Y$  and  $T$ ; in fact, these papers ignore altogether the issue of dependence and its consequences upon the inference about the extremes of the variable of interest.

Given the importance of the knowledge of the asymptotic behaviour of  $(Z, \delta)$  for  $Z$  large in the construction of extreme value estimators in the independent censoring case, it is natural to think of the consequences that dependent censoring may have on this asymptotic behaviour. As noted above, there are numerous ways to specify dependence; in this paper, we assume that the dependence structure of the pair  $(Y, T)$  is given by an extreme value copula, which is equivalent to assuming that  $(Y, T)$  has a bivariate extreme value distribution in the sense of Tawn (1988). The construction and early development of extreme value copulas date back to Galambos (1978) and Deheuvels (1984), and a recent account is provided by Gudendorf and Segers (2010). This type of copula is particularly adapted to the description of joint extreme events, i.e. of situations when both  $Y$  and  $T$  are extreme, which constitute precisely the kind of events one has to consider in order to understand the extremes of the observed variable  $Z = \min(Y, T)$ , in an effort to then get back to the extremes of  $Y$  (that would be the goal of the statistician in this context). The main results of this paper focus on,

assuming standard extreme value conditions on the distribution of  $Y$  and  $T$  together with an extreme value copula dependence model, the analysis of the extreme value properties of  $Z$  first and then of the behaviour of  $\delta$  given that  $Z$  is large, the latter variable indicating how much censoring there is in the extremes of the sample.

The basic assumption of a purely extreme value copula model may appear restrictive at first. It is indeed similar in spirit to assuming that, in the univariate case, the underlying distribution is a Generalised Extreme Value distribution (see de Haan and Ferreira 2006). This assumption shall nevertheless prove very useful in identifying several problems, such as the inconsistency of certain estimators, that may arise when there is dependence in the censoring mechanism. The idea is that any estimator (of, say, the extreme value index of  $Y$ ) which would be inconsistent in the present context cannot be expected to be consistent in general in a wider class of models (such as, for instance, the Archimax copula model of Capéràa et al. 2000), as any such wider class would contain the extreme value copula model in which the estimator is inconsistent. Let us also point out that the very popular bivariate model of Ledford and Tawn (1996, 1997) would not be appropriate here, because this model assumes that  $Y$  and  $T$  are unit Fréchet distributed. To use this model in practice, one therefore has to transform  $Y$  and  $T$  to a unit Fréchet distribution, which implies that the distributions of  $Y$  and  $T$  are known or have at least been accurately estimated beforehand. This would be an issue in the statistical analysis of extremes with censoring, since the mindset is rather that nothing is known about the distributions of  $Y$  and  $T$  and the problem is to recover the extreme value behaviour of  $Y$ .

Let us highlight the main contributions of this paper. We start by, in an extreme value copula dependent censoring model, investigating the extreme value properties of  $Z$  as well as the convergence of the proportion of censored observations in the right tail of the variable  $Z$ . After that, we shall explain how this investigation shows that a number of estimators of the extreme value index of  $Y$ , introduced in the independent censoring case, become inconsistent in the dependent censoring framework we consider, whenever  $T$  has a lighter tail than  $Y$  has. More generally, we will argue that when the dependence structure of  $(Y, T)$  is given by a non-independent extreme value copula, and if  $T$  has a lighter tail than  $Y$  has, then the identifiability of the extreme value index of  $Y$ , based on the information provided jointly by  $Z$  and  $\delta|Z$  for  $Z$  large, is unclear. This is in stark contrast with the independent censoring case, in which we know from Einmahl et al. (2008) that the problem of inferring this parameter can indeed be solved in a simple way based on the behaviour of the pair  $(Z, \delta)$  for large  $Z$  alone. This discussion will be formalised using our introduced and dedicated concept of tail identifiability, adapted to the censoring framework, and refined based on additional asymptotic considerations on the distribution of  $(Z, \delta)$  for high  $Z$ . We shall then explain why, based on the full information provided by the distribution of  $(Z, \delta)$  and if the extreme value copula describing the censoring mechanism is known, then the extreme value index of  $Y$  becomes clearly identifiable, and we shall outline a couple of possible strategies that may lead to consistent estimators of this extreme value index.

The outline of our paper is as follows. Section 2 gives further details about our assumptions and especially about our dependence framework. Section 3 gives our main results, first about the extremes of the observed variable  $Z$  in Section 3.1 and

then about the tail censoring probability in Section 3.2. Section 4 gathers statistical considerations deduced from our results about the estimation of the extreme value index of  $Y$ , relative to the inconsistency of certain estimators in Section 4.1 and then more generally to the identifiability of this parameter based on tail information in Section 4.2. Section 5 concludes by briefly discussing possible ways to design inference techniques and providing ideas for further work. Proofs of the main results are deferred to Appendix A, and auxiliary results and their proofs are relegated to Appendix B.

## 2 Framework

We assume throughout that the variable of interest  $Y$  is partially unobserved, due to the existence of a right-censoring random variable  $T$ . In other words, we only observe the pair  $(Z, \delta)$ , where  $Z = \min(Y, T)$  and  $\delta = \mathbb{I}_{\{Y \leq T\}}$ . Contrary to the standard setup, we also assume that  $Y$  and  $T$  are not independent. We describe here the dependence structure of the pair  $(Y, T)$  by the means of a copula function. The key result in order to do so is Sklar’s theorem (Sklar 1959), which says that there exists a copula  $C$  with

$$\forall y, t \in \mathbb{R}, \mathbb{P}(Y \leq y, T \leq t) = C(F_Y(y), F_T(t)),$$

in which  $F_Y$  and  $F_T$  denote the respective cumulative distribution functions of  $Y$  and  $T$ . A copula  $C$  is simply, in our case, a bivariate distribution function of a pair of standard uniform random variables. We assumed here that  $Y$  and  $T$  are not independent, so that the copula  $C$  cannot be the independent copula  $(u, v) \mapsto uv$ .

Since the ultimate goal of the statistician would be to recover the extremes of  $Y$ , we should work in a relevant extreme value framework. The condition we shall introduce, on an arbitrary distribution function  $F$ , is one of the many equivalent versions of the classical extreme value condition saying that the distribution should belong to the domain of attraction of some extreme value distribution. As we are in a randomly right-censored situation, it is convenient to write our extreme value condition on a distribution function  $F$  in terms of the survival function  $1 - F$ , which leads us to the following formulation (see Theorem 1.1.6 in de Haan and Ferreira 2006):

**Condition  $\mathcal{C}_1(\gamma)$**  There is  $\gamma \in \mathbb{R}$ , called the extreme value index of  $F$ , and a positive function  $a$  such that the distribution function  $F$  satisfies:

$$\lim_{t \uparrow \tau^*} \frac{1 - F(t + xa(t))}{1 - F(t)} = (1 + \gamma x)^{-1/\gamma}$$

for all  $x$  with  $1 + \gamma x > 0$ , where  $\tau^* = \sup\{x \in \mathbb{R}, F(x) < 1\}$  is the right endpoint of  $F$ .

Since the observed variable is  $Z = \min(Y, T)$ , it makes sense to assume that the distributions of  $Y$  and  $T$  can both be included within this extreme value framework, and our main results will then be stated under the assumption that the distributions of  $Y$  and  $T$  satisfy conditions  $\mathcal{C}_1(\gamma_Y)$  and  $\mathcal{C}_1(\gamma_T)$  respectively. Because actually

$$\mathbb{P}(Z > z) = \mathbb{P}(\min(Y, T) > z) = \mathbb{P}(Y > z, T > z)$$

it follows that the study of the extremes of  $Z$ , which is intuitively a sensible way to get information about the extremes of  $Y$ , will require a study of the situation when  $Y$  and  $T$  are jointly extreme. A very convenient assumption on the copula  $C$  in this context is then to suppose that  $C$  is an extreme value copula (see e.g. Gudendorf and Segers 2010):

$$\forall(u, v) \in (0, 1]^2 \setminus \{(1, 1)\}, C(u, v) = (uv)^{A(\log(v)/\log(uv))}$$

where  $A$  is the so-called Pickands dependence function related to  $C$ , i.e. it is a function that maps  $[0, 1]$  into  $[1/2, 1]$ , is convex and satisfies the inequalities  $\max(t, 1 - t) \leq A(t) \leq 1$  for all  $t \in [0, 1]$ . The function  $A$  characterises the copula  $C$ : in particular, the case  $A(t) = 1$  corresponds to the independent copula  $C(u, v) = uv$  (which we therefore exclude), and the case  $A(t) = \max(t, 1 - t)$  is that of the perfect dependence copula  $C(u, v) = \min(u, v)$ . In theoretical terms, these copulae arise naturally as limiting copulae of suitably normalised sequences of componentwise maxima of independent and identically distributed bivariate pairs (Joe 1997). In this sense, assuming that  $(Y, T)$  follows a bivariate extreme value distribution is analogous to, in a univariate context, assuming that the random variable of interest has a Generalised Extreme Value distribution. Working in such a context, which is a simplified version of the general case, can help to identify statistical issues, especially regarding the inconsistency of certain estimators, that would arise in a more general context as well; we will highlight such issues in Section 4.

Our first step is to, compared to the independent censoring case, quantify the influence that the dependence structure induced by  $C$  has on the distribution of the random variable  $Z$ . We do this by using a simple identity linking the survival function  $\bar{F}_Z(z) = \mathbb{P}(Z > z)$  of  $Z$  to the survival functions  $\bar{F}_Y$  and  $\bar{F}_T$  of  $Y$  and  $T$  and to the copula  $C$ . Since

$$\bar{F}_Z(z) = \mathbb{P}(Y > z, T > z) = 1 - \mathbb{P}(Y \leq z, T \leq z) - \mathbb{P}(Y \leq z, T > z) - \mathbb{P}(Y > z, T \leq z),$$

it follows, after straightforward calculations, that we can write

$$\bar{F}_Z(z) = \bar{F}_Y(z)\bar{F}_T(z) + \varphi(\bar{F}_Y(z), \bar{F}_T(z)) \text{ with } \varphi(u, v) = C(1 - u, 1 - v) - (1 - u)(1 - v). \tag{1}$$

In other words, we can write the survival function of  $Z$  as what it would be under independence of  $Y$  and  $T$ , plus the term  $\varphi(\bar{F}_Y(z), \bar{F}_T(z))$  measuring the effect that the dependence structure in  $C$  has on  $Z$ . Since the behaviour of  $\bar{F}_Y(z)\bar{F}_T(z)$  for large  $z$  (i.e. when  $z$  converges to the right endpoint of  $Z$ ) is easy to analyse, Eq. 1 suggests that, to analyse the extremes of  $Z$ , it is enough to understand the behaviour of  $\varphi(\bar{F}_Y(z), \bar{F}_T(z))$  for large  $z$ .

Let now  $\tau_Y$  and  $\tau_T$  be the right endpoints of  $Y$  and  $T$ , respectively. Since we will focus on the large values of  $Z$ , namely, near its right endpoint  $\tau = \min(\tau_Y, \tau_T)$ , we can anticipate that the relative positions of  $\tau_Y$  and  $\tau_T$  will play a major role in our context. In the case when  $\tau_T < \tau_Y$ , then the extremes of  $Y$  cannot be recovered because the distribution of  $Y$  is not identifiable past the point  $\tau_T$ . In the case  $\tau_Y < \tau_T$ , we should expect the extremes of  $Y$  to be those of  $Z$ , meaning that they can be recovered by standard techniques. The following result makes this statement precise.

**Proposition 1** *Assume that  $\tau_Y < \tau_T$  and  $C$  is an extreme value copula whose Pickands dependence function is continuously differentiable on  $[0, 1]$ . Then  $\overline{F}_Z(z)/\overline{F}_Y(z)$  has a positive and finite limit as  $z \uparrow \tau_Z = \tau_Y$ , the right endpoint of  $Z$ .*

We therefore assume in what follows that  $Y$  and  $T$  have a common right endpoint  $\tau = \tau_Y = \tau_T$ . Note that distributions with a positive extreme value index have an infinite right endpoint, while distributions with a negative extreme value index have a finite right endpoint, see Theorem 1.2.1 in de Haan and Ferreira (2006). It therefore follows from our basic assumption  $\tau_Y = \tau_T$  that the extreme value indices  $\gamma_Y$  and  $\gamma_T$  should have the same sign.

We may now summarise our hypotheses about the joint behaviour of  $Y$  and  $T$  in the following condition:

**Condition ( $\mathcal{H}$ )**  $Y$  and  $T$  have a common right endpoint  $\tau = \tau_Y = \tau_T$ , satisfy conditions  $\mathcal{C}_1(\gamma_Y)$  and  $\mathcal{C}_1(\gamma_T)$  respectively, and their joint distribution function is given by

$$\forall y, t \in \mathbb{R}, \mathbb{P}(Y \leq y, T \leq t) = C(F_Y(y), F_T(t))$$

where  $C$  is an extreme value copula whose Pickands dependence function  $A$  is twice continuously differentiable and not equal to the constant function 1.

In their pioneering paper, Einmahl et al. (2008) use the right-tail behaviour of the observed variable  $Z$ , together with an analysis of the censoring probability given that  $Z$  is large, to show that if  $Y$  and  $T$  are independent then the extreme value index of  $Y$  can be recovered exclusively from the behaviour of  $(Z, \delta)$  given that  $Z$  is large in a very simple manner under classical extreme value assumptions. A similar idea, albeit implemented differently, is used by Beirlant et al. (2007). This line of thought was then followed by Gomes and Neves (2011), Ndao et al. (2014, 2016), Brahimy et al. (2015), Beirlant et al. (2016) and Stupfler (2016) in their respective contexts. Our aim in Section 3 below is to carry out an analogue study in our dependent censoring context and see what influence the introduction of dependence in the censoring mechanism, via an extreme value copula, has on the distribution of  $(Z, \delta)$  given that  $Z$  is large. Let us finally highlight that the double differentiability assumption on  $A$  is technically convenient and allows for a unified presentation of our results in Sections 3 and 4; the results in Section 3 below can be obtained by simply assuming that  $A$  is continuously differentiable (at the expense of extra technical details).

## 3 Main results

### 3.1 The extremes of the response variable

Our first step is to analyse the extreme value behaviour of  $Z$ , in terms of the extreme value indices  $\gamma_Y$  and  $\gamma_T$  of  $Y$  and  $T$ . This is straightforward in the independent censoring case, because then the survival function of  $Z$  is the product of those of  $Y$  and  $T$ . Our aim here is to state a corresponding result in the dependence context ( $\mathcal{H}$ ); recall that this condition entails that the extreme value indices of  $Y$  and  $T$  have the same sign.

**Theorem 1** Assume that condition  $(\mathcal{H})$  holds; if  $\gamma_Y = \gamma_T$ , assume further that the ratio  $\bar{F}_T(z)/\bar{F}_Y(z)$  has a finite and positive limit as  $z \uparrow \tau$ . We have that:

- (i) If  $\gamma_Y \geq 0$  and  $\gamma_T \geq 0$ , then  $Z$  satisfies condition  $\mathcal{C}_1(\min(\gamma_Y, \gamma_T))$ .
- (ii) If  $\gamma_Y \leq 0$  and  $\gamma_T \leq 0$ , then  $Z$  satisfies condition  $\mathcal{C}_1(\max(\gamma_Y, \gamma_T))$ .

The right-tail behaviour of  $Z$  is therefore essentially that of the variable with the lightest tail in the pair  $(Y, T)$ , i.e. that of  $Y$  in the case when  $Y$  has a lighter tail than (or a similar tail to) that of  $T$ , and that of  $T$  when  $T$  has a lighter tail than  $Y$  has. Let us consider a simple illustrative example.

*Example 1* Let  $\gamma_T > 0$ . Suppose that  $Y$  and  $T$  have the Pareto distributions

$$F_Y(y) = 1 - y^{-1} \text{ and } F_T(t) = 1 - t^{-1/\gamma_T} \text{ for } y, t > 1.$$

In particular,  $Y$  (resp.  $T$ ) has extreme value index 1 (resp.  $\gamma_T$ ). Assume that the dependence structure of the pair  $(Y, T)$  is described by a Gumbel-Hougaard copula (see Gumbel 1960):

$$\forall u, v \in (0, 1], C_\theta(u, v) = \exp \left\{ - \left[ (-\log u)^\theta + (-\log v)^\theta \right]^{1/\theta} \right\}$$

where  $\theta \geq 1$  is a constant; here we choose  $\theta > 1$ , in order to ensure that  $Y$  and  $T$  are not independent. This copula is an extreme value copula, whose Pickands dependence function is

$$A(x) = (x^\theta + (1 - x)^\theta)^{1/\theta} \text{ for } x \in [0, 1].$$

Theorem 1 predicts that  $\bar{F}_Z$  should be regularly varying with index  $-1/\min(1, \gamma_T)$ . We check this by analysing the asymptotic behaviour, as  $z \rightarrow \infty$ , of

$$C_\theta(1 - \bar{F}_Y(z), 1 - \bar{F}_T(z)) = \exp \left\{ - \left[ (-\log(1 - z^{-1}))^\theta + (-\log(1 - z^{-1/\gamma_T}))^\theta \right]^{1/\theta} \right\}.$$

If  $0 < \gamma_T < 1$  then

$$\begin{aligned} \frac{C_\theta(1 - \bar{F}_Y(z), 1 - \bar{F}_T(z))}{1 - z^{-1}} &= \exp \left\{ \log(1 - z^{-1}) \left( \left[ 1 + \left( \frac{\log(1 - z^{-1/\gamma_T})}{\log(1 - z^{-1})} \right)^\theta \right]^{1/\theta} - 1 \right) \right\} \\ &= \exp \left\{ -\frac{1}{\theta} z^{\theta(1 - \gamma_T^{-1}) - 1} (1 + o(1)) \right\} \\ &= 1 + o\left(z^{-1/\gamma_T}\right) \text{ because } \theta > 1 \text{ and } \gamma_T < 1. \end{aligned}$$

If  $\gamma_T = 1$  then

$$\frac{C_\theta(1 - \bar{F}_Y(z), 1 - \bar{F}_T(z))}{1 - z^{-1}} = \exp \left\{ \log(1 - z^{-1}) (2^{1/\theta} - 1) \right\} = 1 - (2^{1/\theta} - 1)z^{-1} (1 + o(1)).$$



If  $\gamma_T > 1$  then

$$\begin{aligned} \frac{C_\theta(1 - \bar{F}_Y(z), 1 - \bar{F}_T(z))}{1 - z^{-1/\gamma_T}} &= \exp \left\{ \log(1 - z^{-1/\gamma_T}) \left( \left[ 1 + \left( \frac{\log(1 - z^{-1})}{\log(1 - z^{-1/\gamma_T})} \right)^\theta \right]^{1/\theta} - 1 \right) \right\} \\ &= \exp \left\{ -\frac{1}{\theta} z^{\theta(\gamma_T^{-1}-1)-\gamma_T^{-1}} (1 + o(1)) \right\} \\ &= 1 + o\left(z^{-1}\right) \text{ because } \theta > 1 \text{ and } \gamma_T > 1. \end{aligned}$$

Together with the identity

$$\bar{F}_Z(z) = \bar{F}_Y(z) + \bar{F}_T(z) + C(1 - \bar{F}_Y(z), 1 - \bar{F}_T(z)) - 1$$

which is a consequence of Eq. 1, this entails

$$\bar{F}_Z(z) = \begin{cases} z^{-1/\gamma_T} (1 + o(1)) & \text{if } \gamma_T < 1, \\ (2 - 2^{1/\theta})z^{-1} (1 + o(1)) & \text{if } \gamma_T = 1, \\ z^{-1} (1 + o(1)) & \text{if } \gamma_T > 1. \end{cases}$$

It follows from this computation that  $\bar{F}_Z$  is indeed regularly varying with index  $-1/\min(1, \gamma_T)$ .

Let us now highlight a couple of consequences of Theorem 1 about the tail behaviour of the observed variable  $Z$  in our setup. For ease of exposition, we assume until the end of this section that  $\gamma_Y \gamma_T > 0$ , i.e.  $Y$  and  $T$  both belong to the same max-domain of attraction, that can be either the Fréchet or Weibull domain of attraction. It is known (see Einmahl et al. 2008) that in the independent censoring case,  $Z$  then belongs to the common max-domain of attraction of  $Y$  and  $T$ , with extreme value index

$$\gamma_Z = \frac{\gamma_Y \gamma_T}{\gamma_Y + \gamma_T}; \text{ in particular, } |\gamma_Z| < \min(|\gamma_Y|, |\gamma_T|).$$

In the independent case, the absolute value of the extreme value index of the observed variable  $Z$  is therefore strictly lower than what it is in the dependent case we consider here. This means that in the dependent case, the right tail of  $Z$  is heavier than it is in the independent case. Qualitatively, this is a consequence of the positive tail dependence between  $Y$  and  $T$ , due to the dependence structure being described by an extreme value copula (see Gudendorf and Segers 2010). It is interesting to note that the expression of the dependence function itself does not affect the extreme value index of  $Z$  at all.

Theorem 1 also has an important corollary relative to the relationship between the extremes of  $Y$  and those of the observed variable  $Z$ : whereas the extremes of  $Z$  always contain information about those of  $Y$  and those of  $T$  in the independent case, they are driven either solely by those of  $Y$  or those of  $T$  in the dependent case considered here, no matter how close to independence the dependence structure is. It should be especially emphasised that in the case when  $\gamma_Y \gamma_T > 0$  and  $|\gamma_T| < |\gamma_Y|$ , corresponding to  $T$  having a lighter tail than  $Y$  has, the extreme value index of  $Z$

is exactly that of  $T$ . The only cases when this type of behaviour is observed in the independent censoring situation are when either the right tail of  $T$  is much lighter than the right tail of  $Y$  (e.g.  $T$  is light-tailed while  $Y$  is heavy-tailed) or when  $\tau_T < \tau_Y \leq \infty$ , the latter being a case when the problem of recovering the extremes of  $Y$  has no solution since the distribution of  $Y$  is then not identifiable past the point  $\tau_T$ .

### 3.2 Tail censoring probability

The second main part of our study focuses on the information available in the censoring indicator  $\delta$  given that the observed variable  $Z$  is large. In other words, we consider the behaviour of the probability

$$P(z) := \mathbb{P}(\delta = 1 | Z > z) = \mathbb{P}(Y \leq T | Z > z) \text{ for } z \uparrow \tau.$$

When  $z$  is close to  $\tau$ , the probability  $1 - P(z)$  gives an idea of the probability of censoring in the extremes of  $Z$ . In particular, if  $P(z)$  converges to a limit  $P(\tau)$  as  $z \uparrow \tau$ , the probability  $1 - P(\tau)$  shall be called the *tail censoring probability*.

It should be noted that Einmahl et al. (2008) achieve the study of censoring in the right tail by slightly different means, as they assume that  $Y$  and  $T$  have continuous distributions and they consider

$$p(z) := \mathbb{P}(Y \leq T | Z = z) := \frac{\frac{d}{dz} [\mathbb{P}(Y \leq T, Z > z)]}{\frac{d}{dz} [\mathbb{P}(Z > z)]}.$$

They mention (without proof) that under independent censoring and suitable extreme value conditions, this function has a limit as  $z \uparrow \tau$ ; it is actually straightforward to show that this is the limit of  $P$  as well, essentially by l'Hôpital's rule. Their statistical arguments, however, use the quantity  $P(z)$  instead: in particular, they develop an estimator of the extreme value index of  $Y$  using the sample counterpart of  $P(z)$  for large  $z$ . The limit of the function  $P$  as  $z \uparrow \tau$  can be expected to play an important role in the context of extreme value analysis with right-censored data, just as the classical censoring probability  $\mathbb{P}(Y > T)$  does for the estimation of central quantities. In the classical, central setup, a positive (and less than 1) censoring probability means that the problem has a solution and that traditional estimators have to be corrected in some way in order to retain consistency. A censoring probability equal to 0 happens in totally uncensored cases, in which standard, uncorrected techniques will still be consistent. Finally, a censoring probability equal to 1 gives rise to a totally censored case, which is a situation when the estimation problem does not have a solution. In classical, central censoring problems, the condition that the censoring probability belongs to the interval  $(0, 1)$  is thus crucial for an estimation problem to be both non-trivial and workable. Our purpose here is to show that the tail censoring probability exists indeed in our dependent censoring setting, and the next section will draw conclusions from its value that are related to inference about the extreme value index of  $Y$ .

Our first step to prove that the limiting tail censoring probability exists is to obtain a convenient representation of the function  $P$ . In all what follows,  $\partial_1 C$  denotes the

partial derivative of  $C$  with respect to its first argument. It is easy to show that

$$\partial_1 C(u, v) = \frac{C(u, v)}{u} \left[ A \left( \frac{\log v}{\log u + \log v} \right) - \frac{\log v}{\log u + \log v} A' \left( \frac{\log v}{\log u + \log v} \right) \right]. \tag{2}$$

Besides, if  $Y$  has a continuous distribution with probability density function  $f_Y$  then clearly

$$\frac{\partial}{\partial y} [\mathbb{P}(Y \leq y, T > t)] = [1 - \partial_1 C(F_Y(y), F_T(t))] f_Y(y).$$

It follows that if  $T$  has a continuous distribution as well, then we can write

$$\begin{aligned} \mathbb{P}(Y \leq T, Z > z) &= \mathbb{P}(Y \leq T, Y > z) = \int_z^\tau \mathbb{P}(Y \in dy, T \geq y) \\ &= \int_z^\tau [1 - \partial_1 C(F_Y(y), F_T(y))] f_Y(y) dy, \end{aligned}$$

resulting in

$$P(z) = \frac{\mathbb{P}(Y \leq T, Z > z)}{\mathbb{P}(Z > z)} = \frac{\int_z^\tau [1 - \partial_1 C(F_Y(y), F_T(y))] f_Y(y) dy}{\bar{F}_Z(z)}. \tag{3}$$

Formula (2) and representation (3) are key to our analysis of the convergence of  $P$ , which is the focus of our second main result.

**Theorem 2** *Assume that condition (H) holds and suppose moreover that  $Y$  and  $T$  have continuous distributions. Then, as  $z \uparrow \tau$ , we have that:*

- (i)  $P(z) \rightarrow 0$  if  $|\gamma_T| < |\gamma_Y|$ ;
- (ii)  $P(z) \rightarrow 1$  if  $|\gamma_T| > |\gamma_Y|$ ;
- (iii)  $P(z) \rightarrow P_c$  if  $|\gamma_T| = |\gamma_Y|$  and  $\bar{F}_T(z)/\bar{F}_Y(z) \rightarrow c \in (0, \infty)$  as  $z \uparrow \tau$ , where:

$$\forall c > 0, P_c = \frac{1 - \frac{c}{c+1}}{1 - A\left(\frac{c}{c+1}\right)} \left\{ 1 - A\left(\frac{c}{c+1}\right) + \frac{c}{c+1} A'\left(\frac{c}{c+1}\right) \right\} \in [0, 1].$$

If moreover  $A$  is strictly convex, then  $P_c \in (0, 1)$ .

It follows from this result that, in our dependent censoring context, the tail censoring probability  $1 - P(\tau) = 1 - \lim_{z \uparrow \tau} P(z)$  exists and:

- When  $T$  has a lighter tail than  $Y$  has, then the tail censoring probability is 1;
- When  $Y$  has a lighter tail than  $T$  has, then the tail censoring probability is 0;
- The only nondegenerate case arises when  $A$  is strictly convex and  $Y$  and  $T$  have proportional tails, the tail censoring probability then being between 0 and 1 depending on the function  $A$ .

By comparison, in the independent case, provided  $Y$  and  $T$  have continuous distributions that satisfy conditions  $C_1(\gamma_Y)$  and  $C_1(\gamma_T)$  respectively, with  $\gamma_Y \gamma_T > 0$  and

equal right endpoints, the tail censoring probability is

$$1 - P(\tau) = \frac{\gamma_Y}{\gamma_Y + \gamma_T} \in (0, 1),$$

see Einmahl et al. (2008). We illustrate these conclusions further in the example below.

*Example 2* Suppose, as in Example 1, that  $Y$  and  $T$  have the Pareto distributions

$$F_Y(y) = 1 - y^{-1} \text{ and } F_T(t) = 1 - t^{-1/\gamma_T} \text{ for } y, t > 1.$$

In the independent case, the tail censoring probability would be

$$1 - P(\tau) = \frac{1}{1 + \gamma_T} \in (0, 1),$$

decreasing smoothly from 1 to 0 as the tail of  $T$  gets heavier. By contrast, in the dependent case with Pickands dependence function  $A(t) = 1 - t(1 - t)$ , the tail censoring probability is

$$1 - P(\tau) = \begin{cases} 1 & \text{if } \gamma_T < 1, \\ 1/2 & \text{if } \gamma_T = 1, \\ 0 & \text{if } \gamma_T > 1, \end{cases}$$

which, although still a decreasing function of  $\gamma_T$ , is discontinuous at 1 and piecewise constant.

Due to the nature of the limit  $P(\tau)$  in our dependent case, the above example is very much representative of the different situations that can occur, the only degrees of freedom here being the position of the discontinuity (i.e. the value of  $\gamma_Y$ , taken to be equal to 1 in the above example) and the value of  $P(\tau)$  at this discontinuity (which depends on the function  $A$ ). Again, it should be emphasised that the Pickands dependence function plays a largely insignificant role in the results, in that it does not affect the value of  $P(\tau)$  except at the discontinuity. In particular, the coefficient of upper tail dependence:

$$\lambda := \lim_{t \downarrow 0} \frac{1}{t} (2t - 1 + C(1 - t, 1 - t)) = 2(1 - A(1/2))$$

(see Gudendorf and Segers 2010), only appears when  $Y$  and  $T$  have exactly equivalent tails, and in that case

$$1 - P(\tau) = 1 - \frac{1}{\lambda} \left\{ \frac{\lambda}{2} + \frac{1}{2} A' \left( \frac{1}{2} \right) \right\} = \frac{1}{2} - \frac{A'(1/2)}{2\lambda}.$$

One reason behind this is the following: recall that the upper tail dependence coefficient is

$$\lambda = \lim_{u \uparrow 1} \mathbb{P}(U > u | V > u)$$

where  $(U, V)$  is a random pair with distribution function  $C$  (see again Gudendorf and Segers 2010). In other words, the upper tail dependence coefficient measures extremal dependence in the direction of the 45-degree line. Now  $(U, V) :=$

$(F_Y(Y), F_T(T))$  is such a random pair and

$$\begin{aligned}\bar{F}_Z(z) &= \mathbb{P}(Y > z, T > z) = \mathbb{P}(U > F_Y(z), V > F_T(z)) \\ \text{and } P(z) &= \mathbb{P}(Y \leq T | Y > z, T > z) = \mathbb{P}(Y \leq T | U > F_Y(z), V > F_T(z)).\end{aligned}$$

We should therefore only expect the upper tail dependence coefficient to appear in this problem if  $\bar{F}_Y(z)$  and  $\bar{F}_T(z)$  are equivalent for large  $z$ . More generally though, the fact that the actual expression of the function  $A$  has no influence in our results outside of the very specific case of similar tails is worthy of note.

We conclude this section by underlining what is arguably the essential difference, as far as the tail censoring probability is concerned, between the independent and our dependent censoring case. For independent censoring, when the right tails of  $Y$  and  $T$  are of the same type, the tail censoring probability is

$$1 - P(\tau) = \frac{\gamma_Y}{\gamma_Y + \gamma_T} \in (0, 1).$$

In particular, on average, a positive proportion of the highest values of  $Z$  are known to come from the variable of interest  $Y$ , meaning that the extremes of  $Y$  are indeed observed in practice and thus can be recovered by an adapted technique. By contrast, in our dependent censoring case, the tail censoring probability  $1 - P(\tau)$  is 1 as soon as  $T$  has a lighter tail than  $Y$  has; in other words, a vanishingly small proportion of the highest observations from  $Z$  will come from  $Y$  in this case. This suggests in particular that, without any further information on the tails of  $Y$  and  $T$ , the problem of inferring the extreme value index of  $Y$  from  $(Z, \delta)$  given that  $Z$  is large has no clear solution. We now seek to clarify these ideas in the next section.

## 4 Statistical consequences

In this section, we first explain what impact our main results can be expected to have on a class of estimators of the extreme value index of  $Y$  in our dependent framework. We then draw general conclusions regarding the identifiability of  $\gamma_Y$  from the information provided by the distribution of  $(Z, \delta)$  for large  $Z$ .

### 4.1 On the large-sample behaviour of a class of estimators of $\gamma_Y$

Our results imply that a dependent censoring mechanism entails consequences, in terms of consistency of estimators, that can be as serious in the study of extremes of the variable of interest as they are in the study of its central parameters. To illustrate this point, we focus on certain estimators that have been developed up to now in the literature. For ease of exposition, we work here under the conditions of Theorem 2 and we further assume that  $\gamma_Y \gamma_T > 0$ . The estimators of Beirlant et al. (2007, 2016), Einmahl et al. (2008), Gomes and Neves (2011), Ndao et al. (2014, 2016),

Brahimi et al. (2015) and Stupfler (2016) are all based on the fact that in the independent censoring case,

$$\frac{\gamma_Z}{P(\tau)} = \frac{\gamma_Y \gamma_T}{\gamma_Y + \gamma_T} \times \left\{ \frac{\gamma_T}{\gamma_Y + \gamma_T} \right\}^{-1} = \gamma_Y.$$

The common idea these authors use is then to plug a consistent estimator of  $\gamma_Z$  and a consistent estimator of  $P(\tau)$  in the left-hand side above in order to obtain a consistent estimator of  $\gamma_Y$ . One representative example of such an estimator is obtained as follows: suppose that the available data is made of independent pairs  $(Z_i, \delta_i)$ ,  $1 \leq i \leq n$ . Denote by  $Z_{1,n} \leq \dots \leq Z_{n,n}$  the order statistics of  $(Z_1, \dots, Z_n)$  and by  $\delta_{[n-i+1,n]}$  the  $\delta$ -indicator corresponding to  $Z_{n-i+1,n}$ . An empirical estimator of  $P(\tau)$  is then

$$\widehat{P}_k(\tau) := \frac{1}{k} \sum_{j=1}^k \delta_{[n-j+1,n]}$$

where  $k = k(n) \rightarrow \infty$  and  $k/n \rightarrow 0$ , see Einmahl et al. (2008). Combining this with e.g. the moment estimator  $\widehat{\gamma}_{Z,k}$  of  $\gamma_Z$  (see Dekkers et al. 1989) results in an estimator  $\widehat{\gamma}_Y = \widehat{\gamma}_{Z,k} / \widehat{P}_k(\tau)$  of  $\gamma_Y$  which is consistent under independent censoring if certain general conditions are satisfied. In the dependent case we consider, however, the equality

$$\frac{\gamma_Z}{P(\tau)} = \gamma_Y$$

is only true when the tail of  $Y$  is strictly lighter than that of  $T$ , in which case we actually have  $\gamma_Z = \gamma_Y$ ; the left-hand side is not even defined for  $|\gamma_Y| > |\gamma_T|$  since  $P(\tau)$  is then equal to 0. This leads to the inconsistency of the aforementioned estimators when  $T$  has a lighter tail than  $Y$  has, and more generally results in a considerable restriction on their applicability in our dependent censoring setup.

Let us also point out that the alternative estimators of Beirlant et al. (2010), Sayah et al. (2014) and Worms and Worms (2014) do not directly use the ratio of  $\gamma_Z$  and  $P(\tau)$  in order to design a consistent estimator. They are however based on a Kaplan-Meier estimate of  $F_Y$  in its right tail, so that they would also be seriously affected by a violation of the independent censoring assumption, as dependence can cause the Kaplan-Meier estimator to become inconsistent (see Fisher and Kanarek 1974, Klein and Moeschberger 1987 and Ebrahimi and Molefe 2003). An adaptation of these methods therefore requires in particular the construction of an adapted estimator of  $F_Y$ , which necessitates more information than that brought by the pair  $(Z, \delta)$  for  $Z$  large: we shall return to this in Section 5.

### 4.2 Tail identifiability

In order to draw some consequences of our results on the problem of inferring the extreme value index of  $Y$  from the behaviour of  $(Z, \delta)$  conditional on  $Z$  being large, we start by introducing a dedicated concept of tail identifiability. Recall that a statistical model  $\{\mathbb{P}_\theta, \theta \in \Theta\}$ , where  $\mathbb{P}_\theta$  denotes a distribution described by a parameter

$\theta$  belonging to a parameter space  $\Theta$ , is called identifiable if the mapping  $\theta \mapsto \mathbb{P}_\theta$  specifying the model is one-to-one, i.e.

$$\mathbb{P}_{\theta_1} = \mathbb{P}_{\theta_2} \Rightarrow \theta_1 = \theta_2.$$

In the statistical literature, the phrases “identifiable statistical model” and “identifiable parameter” are very often used interchangeably: a parameter (of a model) is then said to be identifiable if two different values of the parameter yield two different models. Here, our interest lies primarily in the extreme value index  $\gamma_Y$  of  $Y$ . A complication is that  $\gamma_Y$  does not completely specify the distribution of  $Y$  in general; to put it differently, two distributions can have the same extreme value index and still not be identical (such as a Pareto distribution and a nontrivial mixture of this same Pareto distribution with a uniform distribution). However, the extreme value index precisely describes the type of tail a distribution has, by the following simple observation: if  $U$  and  $V$  satisfy conditions  $\mathcal{C}_1(\gamma_U)$  and  $\mathcal{C}_1(\gamma_V)$  respectively, and  $U$  and  $V$  have an equal right endpoint  $\tau$ , then

$$\lim_{x \uparrow \tau} \frac{\mathbb{P}(U > x)}{\mathbb{P}(V > x)} = 1 \Rightarrow \gamma_U = \gamma_V.$$

This assertion is essentially a consequence of Theorem 1.2.6 in de Haan and Ferreira (2006).

This simple remark is the basis for our concept of tail identifiability in the random censoring framework, which we define below. Note that this definition bears no relationship to the concept of tail identifiability of a family of copulas introduced in Ding (2010, 2012).

**Definition 1** Let  $\mathcal{K}_\tau$  be the class of joint distributions of  $(Y, T)$  such that  $\tau_Y = \tau_T = \tau$  and the distributions of  $Y$  and  $T$  satisfy conditions  $\mathcal{C}_1(\gamma_Y)$  and  $\mathcal{C}_1(\gamma_T)$  respectively. Let  $\mathcal{M}_\tau \subset \mathcal{K}_\tau$ . We say that  $\gamma_Y$  is tail identifiable in the random censoring model  $\mathcal{M}_\tau$  if, whenever the distributions of  $(Y_1, T_1)$  and  $(Y_2, T_2)$  both belong to  $\mathcal{M}_\tau$  then, denoting by  $Z_i = \min(Y_i, T_i)$  and  $\delta_i = \mathbb{I}_{\{Y_i \leq T_i\}}$ , we have

$$\left[ \lim_{z \uparrow \tau} \frac{\mathbb{P}(Z_1 > z)}{\mathbb{P}(Z_2 > z)} = 1 \text{ and } \lim_{z \uparrow \tau} [\mathbb{P}(\delta_1 = 1 | Z_1 > z) - \mathbb{P}(\delta_2 = 1 | Z_2 > z)] = 0 \right] \Rightarrow \gamma_{Y_1} = \gamma_{Y_2}.$$

The above definition thus essentially says that  $\gamma_Y$  is tail identifiable if its value can be inferred from the first meaningful asymptotics in the distribution of  $(Z, \delta)$  given that  $Z$  is large; in practice,  $\gamma_Y$  will be tail identifiable if its value can be computed from the extreme value index of  $Z$  and the tail censoring probability. Since the extreme value condition  $\mathcal{C}_1(\gamma)$  itself only gives information about the first-order asymptotics of a survival function near its right endpoint, Definition 1 appears to be a reasonable definition for our purpose. Finally, we note that, as we should expect,  $\gamma_Y$  is tail identifiable in the submodel of  $\mathcal{K}_\tau$  restricted to independent censoring and  $\gamma_Y \gamma_T > 0$ : this is because in that particular case,  $\gamma_Y = \gamma_Z / P(\tau)$ .

We shall now summarise Theorems 1 and 2 with a corollary in terms of tail identifiability of  $\gamma_Y$  in our random censoring model.

**Corollary 1** *For a fixed  $\tau$ , let  $\mathcal{M}_\tau$  be the class of joint distributions of  $(Y, T)$  satisfying condition  $(\mathcal{H})$ , and such  $Y$  and  $T$  have continuous distributions. Let further*

- $\mathcal{M}_\tau^<$  denote the submodel of  $\mathcal{M}_\tau$  in which  $|\gamma_Y| < |\gamma_T|$ ;
- $\mathcal{M}_\tau^>$  denote the submodel of  $\mathcal{M}_\tau$  in which  $|\gamma_Y| > |\gamma_T|$ ;
- $\mathcal{M}_\tau^=$  denote the submodel of  $\mathcal{M}_\tau$  in which  $\gamma_Y = \gamma_T$  and  $\overline{F}_T(z)/\overline{F}_Y(z)$  has a finite positive limit as  $z \uparrow \tau$ .

Then:

- (i)  $\gamma_Y$  is tail identifiable in the random censoring models  $\mathcal{M}_\tau^<$  and  $\mathcal{M}_\tau^=$ .
- (ii)  $\gamma_Y$  is not tail identifiable in the random censoring model  $\mathcal{M}_\tau^>$ .
- (iii)  $\gamma_Y$  is not tail identifiable in the general random censoring model  $\mathcal{M}_\tau$ .

Corollary 1 has essentially three consequences. The first one is that in general, if the dependence structure is described by an extreme value copula distinct from the independence copula, then the parameter  $\gamma_Y$  is not tail identifiable. In other words, the information contained in the first-order asymptotics of the distribution of the pair  $(Z, \delta)$ , for large  $Z$ , is in general not sufficient to identify the value of  $\gamma_Y$ . The second one is that even if it is known that  $Y$  has a heavier tail than  $T$  has, the parameter  $\gamma_Y$  is still not tail identifiable. The third and final consequence is that, if by contrast it is known that  $Y$  does not have a heavier tail than that of  $T$ , then  $Y$  is indeed tail identifiable; this is a corollary of Theorem 1, stating that in this case  $\gamma_Y$  is actually the extreme value index of  $Z$ .

Statistically speaking, it follows from Corollary 1 that, when there is dependent censoring induced by an extreme value copula and contrary to the independent censoring case, there is no “obvious” way to estimate  $\gamma_Y$  consistently based on the highest values of the observed variable in a sample together with their censoring indicators, if no further information is provided. This result does not, however, consider what can be obtained by looking at the asymptotics of the functions  $z \mapsto \overline{F}_Z(z) = \mathbb{P}(Z > z)$  and  $z \mapsto P(z) = \mathbb{P}(\delta = 1|Z > z)$  in greater detail. For instance, recalling that, from Eq. 1:

$$\overline{F}_Z(z) = \overline{F}_Y(z) + \overline{F}_T(z) + C(1 - \overline{F}_Y(z), 1 - \overline{F}_T(z)) - 1,$$

it can be thought that, although the influence of  $\gamma_Y$  (or equivalently  $\overline{F}_Y(z)$ ) is not necessarily visible in the first-order asymptotics of the survival function of  $Z$ , it may well affect its second-order asymptotics. Similarly, tail identifiability is only linked to the limit of the function  $z \mapsto P(z)$  at the endpoint of  $Z$ , but it does not consider the rate of convergence of this function to its limit. In the problematic, non-tail-identifiable case  $|\gamma_Y| > |\gamma_T|$ , when  $\gamma_Z = \gamma_T$  and  $P(z) \rightarrow 0$ , the equality

$$\mathbb{P}(Y > z|Y \leq T) = \frac{\overline{F}_Z(z)P(z)}{\mathbb{P}(Y \leq T)}$$

suggests that the rate of convergence of  $P(z)$  to 0 should contain information about  $\gamma_Y$ . Our final result, which is a refinement of Theorems 1 and 2, sheds some light on this by giving asymptotic expansions of  $z \mapsto \overline{F}_Z(z)$  and  $z \mapsto P(z)$  in the case  $|\gamma_Y| > |\gamma_T|$ .



**Proposition 2** Assume that condition  $(\mathcal{H})$  holds and suppose moreover that  $Y$  and  $T$  have continuous distributions. If  $|\gamma_Y| > |\gamma_T|$  then, as  $z \uparrow \tau$ , we have

$$\begin{aligned} \bar{F}_Z(z) &= -A'(0)\bar{F}_T(z) \left( 1 - \frac{1 + A'(0)}{A'(0)}\bar{F}_Y(z) + \frac{A''(0)}{2A'(0)} \frac{\bar{F}_T(z)}{\bar{F}_Y(z)} + o(\bar{F}_Y(z)) \right) \\ &\quad + o\left(\frac{\bar{F}_T(z)}{\bar{F}_Y(z)}\right) \\ \text{and } P(z) &= - \frac{(1 + A'(0) + o(1)) \int_z^\tau \bar{F}_T(t) f_Y(t) dt + \frac{1}{2}(A''(0) + o(1)) \int_z^\tau (\bar{F}_T(t)/\bar{F}_Y(t))^2 f_Y(t) dt}{A'(0)\bar{F}_T(z)}. \end{aligned}$$

It follows from Proposition 2 that the extreme value behaviour of  $Y$  does indeed theoretically have an influence on the higher-order asymptotic properties of  $z \mapsto \bar{F}_Z(z)$  and  $z \mapsto P(z)$ . However, this information is still not sufficient in practice to identify  $\gamma_Y$ , as the next example shows.

*Example 3* Consider the following two models:

- Model 1:  $Y_1$  has a Pareto distribution with parameter  $\gamma_{Y_1} = 3$ , and  $T_1 = T$  is an equally weighted mixture of two Pareto distributions with respective parameters 1 and  $6/7$ ,
- Model 2:  $Y_2$  has a Pareto distribution with parameter  $\gamma_{Y_2} = 3/2$ , and  $T_2 = T$  is an equally weighted mixture of two Pareto distributions with respective parameters 1 and  $6/7$ ,

with a common dependence structure given by the Pickands dependence function

$$A(t) = 1 - \frac{1}{2}t(1 - t); \text{ in particular } A'(0) = -\frac{1}{2} \text{ and } A''(0) = 1.$$

Then for any  $i \in \{1, 2\}$ ,

$$\forall t > 1, \bar{F}_{T_i}(t) = \frac{1}{2}t^{-1}(1 + t^{-1/6}).$$

In particular

$$\frac{\bar{F}_{T_1}(z)}{\bar{F}_{Y_1}(z)} = \frac{z^{-2/3}}{2}(1 + o(1)) \text{ and } \frac{\bar{F}_{T_2}(z)}{\bar{F}_{Y_2}(z)} = \frac{z^{-1/3}}{2}(1 + o(1)) \text{ as } z \rightarrow \tau = \infty,$$

and it follows from Proposition 2 that, for any  $i \in \{1, 2\}$ , we have

$$\bar{F}_{Z_i}(z) = \frac{1}{4z} \left( 1 + z^{-1/6} + o\left(z^{-1/6}\right) \right)$$

and

$$P_i(z) := \mathbb{P}(\delta_i = 1 | Z_i > z) = \frac{z^{-1/3}}{4} (1 + o(1)).$$

The functions  $z \mapsto \bar{F}_{Z_i}(z)$  (resp.  $z \mapsto P_i(z)$ ) therefore have the same second-order expansion (resp. asymptotic equivalent) in both models, although  $\gamma_{Y_1} = 3 \neq 3/2 = \gamma_{Y_2}$ .

The idea behind the construction of the above example is that, even if  $\gamma_Y$  should theoretically have an influence on the higher-order asymptotic properties of  $z \mapsto \overline{F}_Z(z)$ , this influence can be completely masked by the asymptotic behaviour of the survival function of  $T$ . For instance, if  $\gamma_Y > \gamma_T > 0$ , and

$$\begin{aligned} \overline{F}_Y(y) &= y^{-1/\gamma_Y} (a_Y + b_Y y^{\rho_Y/\gamma_Y} + o(y^{\rho_Y/\gamma_Y})) \text{ as } y \rightarrow \infty \\ \text{and } \overline{F}_T(t) &= t^{-1/\gamma_T} (a_T + b_T t^{\rho_T/\gamma_T} + o(t^{\rho_T/\gamma_T})) \text{ as } t \rightarrow \infty, \end{aligned}$$

where  $a_Y, a_T > 0, b_Y, b_T \neq 0$  and  $\rho_Y, \rho_T < 0$ , then by Proposition 2

$$\overline{F}_Z(z) = -A'(0)z^{-1/\gamma_T} (a_T + b_T z^{\rho_T/\gamma_T} + o(z^{\rho_T/\gamma_T})) \text{ if } |\rho_T| < \min\left(\frac{\gamma_T}{\gamma_Y}, 1 - \frac{\gamma_T}{\gamma_Y}\right).$$

In other words, the extreme value index  $\gamma_Y$  may not appear at all in the second-order asymptotics of  $\overline{F}_Z$  if the second-order parameter  $\rho_T$  of  $T$  is sufficiently close to 0. This discussion carries over to higher-order asymptotics; if the third-order parameter of  $T$  (in the sense of Goegebeur and de Wet 2012) is small as well, then  $\overline{F}_Z$  and  $\overline{F}_T$  will (up to the third asymptotic order) have the same asymptotic behaviour and in particular,  $\gamma_Y$  will not feature in the asymptotic properties of  $\overline{F}_Z$  up to that order.

The same kind of phenomenon happens when considering the function  $P$ . Calculations similar to those that lead to Proposition 2 show that, when  $A$  is three times differentiable and  $\gamma_Y > \gamma_T > 0$ , then one can push further the asymptotic expansion of  $P$  given in Proposition 2 to get

$$\begin{aligned} P(z) &= \frac{1}{\overline{F}_Z(z)} \left( \int_z^\infty \left[ (1 + A'(0)) + \left( \frac{A''(0)}{2} + o(1) \right) \frac{\overline{F}_T(t)}{\overline{F}_Y(t)} \right] \overline{F}_T(t) f_Y(t) dt \right. \\ &\quad \left. + \int_z^\infty \left[ \frac{A''(0)}{2} + \left( \frac{A^{(3)}(0)}{3} - A''(0) + o(1) \right) \frac{\overline{F}_T(t)}{\overline{F}_Y(t)} \right] \left( \frac{\overline{F}_T(t)}{\overline{F}_Y(t)} \right)^2 f_Y(t) dt \right). \end{aligned}$$

The function  $P$  is then generally regularly varying with index  $-\min(\gamma_Y^{-1}, \gamma_T^{-1} - \gamma_Y^{-1})$ . Even if  $\gamma_T$  is identified, the knowledge of  $\min(\gamma_Y^{-1}, \gamma_T^{-1} - \gamma_Y^{-1})$  is not sufficient in general to know  $\gamma_Y$ . Furthermore, the second-order (and higher-order) terms in  $\overline{F}_Y$  and  $\overline{F}_T$  can once again hide the influence of  $\gamma_Y$  on the asymptotic behaviour of the function  $P$  if they converge to 0 sufficiently slowly. An example, at the second order, is obtained by altering Example 3 as follows: consider the models

- Model 1:  $Y_1$  has a survival function such that

$$\overline{F}_{Y_1}(y) = y^{-1/3} \left( 1 + \frac{1}{5}y^{-1/12} + o(y^{-1/12}) \right) \text{ as } y \rightarrow \infty$$

(one such example is  $\overline{F}_{Y_1}(y) = y^{-1/3}(1 + y^{-1/12})^{1/5}$  for  $y$  large enough) and  $T_1 = T$  is an equally weighted mixture of two Pareto distributions with respective parameters 1 and 6/7,

- Model 2:  $Y_2$  has a survival function such that

$$\overline{F}_{Y_2}(y) = y^{-2/3} \left( 1 - \frac{2}{7}y^{-1/12} + o(y^{-1/12}) \right) \text{ as } y \rightarrow \infty$$

(one such example is  $\bar{F}_{Y_2}(y) = y^{-2/3}(1 - y^{-1/12})^{2/7}$  for  $y$  large enough) and  $T_2 = T$  is an equally weighted mixture of two Pareto distributions with respective parameters 1 and  $6/7$ ,

with a common dependence structure given by the Pickands dependence function  $A(t) = 1 - \frac{1}{2}t(1 - t)$ . It is then readily checked that, in both models,

$$\bar{F}_{Z_i}(z) = \frac{1}{4z} \left( 1 + z^{-1/6} + o\left(z^{-1/6}\right) \right)$$

and

$$P_i(z) := \mathbb{P}(\delta_i = 1 | Z_i > z) = \frac{z^{-1/3}}{4} \left( 1 + \frac{4}{17}z^{-1/12} + o(z^{-1/12}) \right).$$

The functions  $z \mapsto \bar{F}_{Z_i}(z)$  and  $z \mapsto P_i(z)$  therefore have the same second-order expansion in both models, although the extreme value indices of  $Y_1$  and  $Y_2$  are different. As we argued above, our arguments can be extended to construct further pairs of models in which higher-order (such as third-order) expansions of  $\bar{F}_Z$  and  $P$  coincide, although extreme value indices of  $Y$  are different. It is therefore not clear that the information provided jointly by  $Z$  and  $\delta|Z$  for  $Z$  large is generally sufficient to recover the extreme value index  $\gamma_Y$ .

### 5 Discussion and ideas for further work

Our arguments in Section 4.2 show that, when the censoring mechanism is described by a non-independent extreme value copula, then the information contained in the extremes of  $Z$ , and in the distribution of  $\delta$  given that  $Z$  is high, does not appear to allow one to identify the extreme value index of  $Y$ . Let us stress that our results do not show that the extreme value index of  $Y$  cannot be recovered; they suggest however that, if one wants to estimate the extreme value index of  $Y$ , then one should use more information than what is provided by the distribution of  $(Z, \delta)$  for  $Z$  large. This stands in contrast with the independent censoring situation, in which the extreme value index of  $Y$  is generally the ratio of the extreme value index  $\gamma_Z$  and of the tail proportion of uncensored observations, both quantities being solely determined by the information provided by  $Z$  and  $\delta|Z$  for  $Z$  large.

One possible way of incorporating more information into the model is to integrate covariate information that makes the assumption of conditional independence between censoring time and variable of interest plausible. While Ndao et al. (2014, 2016) and Stupfler (2016) consider a conditional extreme value model, they do not consider the question of dependence in the censoring mechanism. The focus of Ndao et al. (2014) and Stupfler (2016) is rather the analysis of extreme survival times to AIDS given age at diagnosis. It is unlikely that such a covariate would be enough to eliminate all source of dependence in the censoring mechanism if it is believed to be present; more relevant information would include prognostic covariates, which are typically high-dimensional (see Hsu and Taylor 2010). It can also happen that the covariate information which could be used by the investigator to alleviate the dependence is simply not available from the data at hand.

Here, staying within the context of an extreme value copula dependence structure, we show how assuming that the copula  $C$  is fully known could lead to estimators of  $\gamma_Y$ . One possible idea for that is to come back to the interpretation of  $\gamma_Y$  as a shape parameter for the tail of the survival function  $\bar{F}_Y$ . For instance, if  $\gamma_Y > 0$ , then  $\gamma_Y$  can be obtained as a limiting average log-excess:

$$\gamma_Y = \lim_{u \rightarrow \infty} \mathbb{E}(\log Y - \log u \mid Y > u).$$

In other words (see equation (3.2.1) in de Haan and Ferreira 2006):

$$\gamma_Y = \lim_{u \rightarrow \infty} \frac{\int_u^\infty (\log y - \log u) dF_Y(y)}{1 - F_Y(u)}.$$

In general, we have the relationship

$$\gamma_Y = \lim_{u \rightarrow \tau_Y} \left( I_Y^{(1)}(u) + 1 - \frac{1}{2} \left[ 1 - \frac{[I_Y^{(1)}(u)]^2}{I_Y^{(2)}(u)} \right]^{-1} \right) \text{ with}$$

$$I_Y^{(k)}(u) = \frac{\int_u^{\tau_Y} (\log y - \log u)^k dF_Y(y)}{1 - F_Y(u)}. \tag{4}$$

This is the convergence behind the well-known moment estimator of Dekkers et al. (1989). One consequence of this relationship is that we have an explicit, relatively simple limit relationship linking  $\gamma_Y$  to  $F_Y$ . Besides, in the context of model  $(\mathcal{H})$ , the partial derivatives of  $C$  exist and are continuous, being given by

$$\partial_1 C(u, v) = \frac{C(u, v)}{u} \left[ A \left( \frac{\log v}{\log u + \log v} \right) - \frac{\log v}{\log u + \log v} A' \left( \frac{\log v}{\log u + \log v} \right) \right]$$

and

$$\partial_2 C(u, v) = \frac{C(u, v)}{v} \left[ A \left( \frac{\log v}{\log u + \log v} \right) + \frac{\log u}{\log u + \log v} A' \left( \frac{\log v}{\log u + \log v} \right) \right].$$

These nice expressions and regularity properties guarantee that, if  $C$  (or equivalently the Pickands dependence function  $A$ ) is known, then  $F_Y$  is identifiable (see Zheng and Klein 1995 and Carrière 1995) and can be consistently estimated by solving iteratively a system of nonlinear differential equations (Carrière 1995), or by the so-called Copula-Graphic estimator of Zheng and Klein (1995); see also Lo and Wilke (2010). The identifiability of  $F_Y$  then guarantees that of  $\gamma_Y$ , by Formula (4), and this formula also suggests a potential plug-in strategy for the construction of an estimator of  $\gamma_Y$  if  $F_Y$  has been accurately estimated beforehand, although the particular structure of the aforementioned estimators of  $F_Y$  may make this task computationally challenging. Let us also note that, while the Copula-Graphic estimator of Zheng and Klein (1995) has in general only an implicit form, it admits a simple closed form if the copula  $C$  is moreover Archimedean, namely

$$C(u, v) = \phi^{-1}(\phi(u) + \phi(v))$$

where  $\phi : (0, 1] \rightarrow [0, \infty]$  is a decreasing convex function satisfying  $\phi(x) \rightarrow \infty$  as  $x \rightarrow 0$  and  $\phi(1) = 0$ . This result is due to Rivest and Wells (2001; see Formula (4) therein). It happens that the Gumbel-Hougaard copula, introduced in Example 1, is

the only non-independent extreme value copula that is also Archimedean (see Genest and Rivest 1989). In the independent case, the Copula-Graphic estimator is actually the classical Kaplan-Meier estimator (see again Rivest and Wells 2001).

The idea we have just developed was essentially that, since it is not clear that  $\gamma_Y$  can in general be identified from the information provided by  $Z$  and  $\delta|Z$  for large  $Z$  only, a good idea is to look at what further information the central part of the distribution of  $Z$  (and the associated conditional distribution  $\delta|Z$ ) can bring. If the copula  $C$  is moreover known, then  $F_Y$  is identifiable from  $(Z, \delta)$  and  $\gamma_Y$  can then at least conceptually be recovered as a (tail) shape parameter for the distribution of  $Y$  via, for instance, Formula (4). The assumption of a known copula is not as costly as it may seem; in general, the overarching aim of extreme value analysis is to estimate extreme quantiles of  $Y$ , or more generally get some understanding of the distribution of  $Y$  in its right tail which is more comprehensive than the knowledge of its extreme value index. Extreme quantile estimation, in particular, is nothing but the estimation of the generalised inverse of  $F_Y$  at specific levels, and this inference therefore translates into inference about  $F_Y$  on part of its range. It is then appropriate, in this context, to make an assumption adapted to the estimation of  $F_Y$ . The approaches we have just outlined could, however, result in potentially implicit and/or computationally challenging estimators of  $\gamma_Y$ , whose asymptotic properties are not yet known and outside the scope of this paper.

**Acknowledgements** The author acknowledges an anonymous Associate Editor and two anonymous reviewers for their helpful comments that led to an improved version of this paper.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## Appendix A: Proofs of the main results

*Proof of Proposition 1* Use Eq. 1 to get

$$\frac{\bar{F}_Z(z)}{\bar{F}_Y(z)} = \bar{F}_T(z) + \frac{\varphi(\bar{F}_Y(z), \bar{F}_T(z))}{\bar{F}_Y(z)}.$$

Set  $v_0 = \lim_{z \uparrow \tau_Y} \bar{F}_T(z) > 0$  and apply Lemma 2 to get

$$\lim_{z \uparrow \tau_Y} \frac{\varphi(\bar{F}_Y(z), \bar{F}_T(z))}{\bar{F}_Y(z)} = (1 - v_0)A'(1).$$

This is a positive constant, by Lemma 1. Thus

$$\lim_{z \uparrow \tau_Y} \frac{\bar{F}_Z(z)}{\bar{F}_Y(z)} = v_0 + (1 - v_0)A'(1) > 0$$

which completes the proof. □

*Proof of Theorem 1* Apply Lemma 3 to get:

- (i) If  $\gamma_Y > \gamma_T \geq 0$ , then  $\tau = \infty$  and  $\overline{F}_T(z)/\overline{F}_Y(z) \rightarrow 0$  as  $z \rightarrow \infty$ .
- (ii) If  $\gamma_T > \gamma_Y \geq 0$ , then  $\tau = \infty$  and  $\overline{F}_T(z)/\overline{F}_Y(z) = [\overline{F}_Y(z)/\overline{F}_T(z)]^{-1} \rightarrow \infty$  as  $z \rightarrow \infty$ .
- (iii) If  $0 \geq \gamma_Y > \gamma_T$ , then  $\tau < \infty$  and  $\overline{F}_T(z)/\overline{F}_Y(z) = [\overline{F}_Y(z)/\overline{F}_T(z)]^{-1} \rightarrow \infty$  as  $z \uparrow \tau$ .
- (iv) If  $0 \geq \gamma_T > \gamma_Y$ , then  $\tau < \infty$  and  $\overline{F}_T(z)/\overline{F}_Y(z) \rightarrow 0$  as  $z \uparrow \tau$ .

In each of these four cases, the result is then a direct corollary of Lemma 5(i) and (ii). In the case  $\gamma_Y = \gamma_T$ , the result is a consequence of Lemma 5(iii). □

*Proof of Theorem 2* Recall that

$$P(z) = \frac{\int_z^\tau [1 - \partial_1 C(F_Y(t), F_T(t))] f_Y(t) dt}{\overline{F}_Z(z)}.$$

To prove (i), remark that if  $\gamma_Y \gamma_T \geq 0$  and  $|\gamma_Y| > |\gamma_T|$ , then  $\overline{F}_T(z)/\overline{F}_Y(z) \rightarrow 0$  as  $z \uparrow \tau$  by Lemma 3. In that case, Lemma 5(i) entails together with Lemma 6(i) that

$$P(z) = \frac{\int_z^\tau \left[ (1 + A'(0) + o(1)) \overline{F}_T(t) + \frac{1}{2} (A''(0) + o(1)) (\overline{F}_T(t)/\overline{F}_Y(t))^2 \right] f_Y(t) dt}{-A'(0) \overline{F}_T(z) (1 + o(1))}. \tag{5}$$

The first part of the integral in the numerator is controlled by noting that

$$0 \leq \frac{\int_z^\tau \overline{F}_T(t) f_Y(t) dt}{\overline{F}_T(z)} \leq \int_z^\tau f_Y(t) dt = \overline{F}_Y(z) \rightarrow 0 \text{ as } z \rightarrow \tau.$$

To control the second one, we first write

$$0 \leq \frac{1}{\overline{F}_T(z)} \int_z^\tau \left( \frac{\overline{F}_T(t)}{\overline{F}_Y(t)} \right)^2 f_Y(t) dt \leq \int_z^\tau \left\{ \frac{\overline{F}_T(t)}{\overline{F}_Y(t)} \right\} \left\{ \frac{f_Y(t)}{\overline{F}_Y(t)} \right\} dt.$$

When  $\gamma_Y > 0$ , and therefore  $\tau = \infty$ , we apply Lemma 3 to obtain that there is  $\varepsilon > 0$  such that

$$\begin{aligned} 0 \leq \frac{1}{\overline{F}_T(z)} \int_z^\infty \left( \frac{\overline{F}_T(t)}{\overline{F}_Y(t)} \right)^2 f_Y(t) dt &\leq \int_z^\infty \left\{ \frac{f_Y(t)}{\overline{F}_Y(t)} \right\} t^{-\varepsilon} dt \\ &= \int_z^\infty \left\{ \frac{d}{dt} [-\log(\overline{F}_Y(t)) t^{-\varepsilon}] - \varepsilon \log(\overline{F}_Y(t)) t^{-1-\varepsilon} \right\} dt. \end{aligned}$$

Now, by Theorem 1.2.1 in de Haan and Ferreira (2006), the function  $\overline{F}_Y$  is regularly varying with index  $-1/\gamma_Y$ . In other words, we can write this function as

$$\overline{F}_Y(t) = t^{-1/\gamma_Y} L_Y(t)$$

where  $L_Y$  is a slowly varying function at infinity. Since  $\log(L_Y(t))/\log t \rightarrow 0$  as  $t \rightarrow \infty$  (see Proposition 1.3.6 p.16 in Bingham et al. 1987) we get  $\log(\overline{F}_Y(t))t^{-\varepsilon/2} \rightarrow 0$  as  $t \rightarrow \infty$ . In particular:

$$|\log(\overline{F}_Y(t))t^{-\varepsilon}| \rightarrow 0 \text{ and } \left| \log(\overline{F}_Y(t))t^{-1-\varepsilon} \right| = o(t^{-1-\varepsilon/2}).$$

Therefore

$$0 \leq \frac{1}{\overline{F}_T(z)} \int_z^\infty \left( \frac{\overline{F}_T(t)}{\overline{F}_Y(t)} \right)^2 f_Y(t) dt \leq \log(\overline{F}_Y(z))z^{-\varepsilon} - \varepsilon \int_z^\infty \log(\overline{F}_Y(t))t^{-1-\varepsilon} dt \rightarrow 0$$

as  $z \rightarrow \infty$

which concludes the proof in this case. When  $\gamma_Y < 0$  and  $\tau < \infty$ , we apply Lemma 3 again to obtain that there is  $\varepsilon > 0$  such that

$$0 \leq \frac{1}{\overline{F}_T(z)} \int_z^\tau \left( \frac{\overline{F}_T(t)}{\overline{F}_Y(t)} \right)^2 f_Y(t) dt \leq \int_z^\tau \left\{ \frac{f_Y(t)}{\overline{F}_Y(t)} \right\} (\tau - t)^\varepsilon dt.$$

Use the change of variables  $T = (\tau - t)^{-1}$  to obtain

$$0 \leq \frac{1}{\overline{F}_T(z)} \int_z^\tau \left( \frac{\overline{F}_T(t)}{\overline{F}_Y(t)} \right)^2 f_Y(t) dt \leq \int_{(\tau-z)^{-1}}^\infty \left\{ \frac{f_Y(\tau - T^{-1})}{\overline{F}_Y(\tau - T^{-1})} \frac{d}{dT} (\tau - T^{-1}) \right\} T^{-\varepsilon} dT.$$

Since by Theorem 1.2.1 in de Haan and Ferreira (2006) the function  $T \mapsto \overline{F}_Y(\tau - T^{-1})$  is regularly varying with index  $1/\gamma_Y < 0$ , we can exactly mimic the proof of the case  $\gamma_Y > 0$  to obtain

$$\frac{1}{\overline{F}_T(z)} \int_z^\tau \left( \frac{\overline{F}_T(t)}{\overline{F}_Y(t)} \right)^2 f_Y(t) dt \rightarrow 0 \text{ as } z \uparrow \tau.$$

The proof of (i) is then complete.

To prove (ii), we note that in this case  $\overline{F}_T(z)/\overline{F}_Y(z) = [\overline{F}_Y(z)/\overline{F}_T(z)]^{-1} \rightarrow \infty$  as  $z \uparrow \tau$  by Lemma 3. Apply then Lemma 5(ii) together with Lemma 6(ii) to get

$$P(z) = \frac{\int_z^\tau (A'(1) + o(1)) f_Y(t) dt}{A'(1)\overline{F}_Y(z)(1 + o(1))}.$$

Since  $A'(1) > 0$  by Lemma 1(ii), the numerator is clearly equivalent to  $A'(1)\overline{F}_Y(z)$ , which entails that  $P(z) \rightarrow 1$  as  $z \rightarrow \tau$  and concludes the proof of (ii).

Finally, to show (iii), apply Lemma 5(iii) together with Lemma 6(iii) to obtain

$$P(z) = \frac{\int_z^\tau \left( 1 - \left[ A\left(\frac{c}{c+1}\right) - \frac{c}{c+1} A'\left(\frac{c}{c+1}\right) \right] + o(1) \right) f_Y(t) dt}{(c+1) \left[ 1 - A\left(\frac{c}{c+1}\right) \right] \overline{F}_Y(z)(1 + o(1))}.$$

If the constant  $1 - \left[ A\left(\frac{c}{c+1}\right) - \frac{c}{c+1}A'\left(\frac{c}{c+1}\right) \right]$  is zero then

$$P(z) = \frac{o(\bar{F}_Y(z))}{(c+1) \left[ 1 - A\left(\frac{c}{c+1}\right) \right] \bar{F}_Y(z)(1 + o(1))} = o(1) \rightarrow 0 = P_c \text{ as } z \uparrow \tau$$

as required. Otherwise we clearly have

$$P(z) \rightarrow \frac{1 - \left[ A\left(\frac{c}{c+1}\right) - \frac{c}{c+1}A'\left(\frac{c}{c+1}\right) \right]}{(c+1) \left[ 1 - A\left(\frac{c}{c+1}\right) \right]} = P_c \text{ as } z \uparrow \tau,$$

and this completes the proof. □

*Proof of Proposition 2* Recall that  $\bar{F}_T(z)/\bar{F}_Y(z) \rightarrow 0$  as  $z \uparrow \tau$  by Lemma 3. Then, use Lemma 7 together with the equality

$$\bar{F}_Z(z) = \bar{F}_Y(z)\bar{F}_T(z) + \varphi(\bar{F}_Y(z), \bar{F}_T(z))$$

to obtain the statement about  $\bar{F}_Z(z)$ . The result on  $P(z)$  is a reformulation of Eq. 5 in the proof of Theorem 2. □

### Appendix B: Auxiliary results and proofs

In this Appendix,  $A$  denotes a Pickands dependence function,  $C$  is the corresponding extreme value copula

$$C(u, v) = (uv)^{A(\log(v)/\log(uv))}$$

which we assume not to be equal to the independence copula, and  $\varphi$  is the function defined by:

$$\varphi(u, v) = C(1 - u, 1 - v) - (1 - u)(1 - v).$$

The first lemma gathers a few results about Pickands dependence functions.

**Lemma 1** *Let  $A$  be a Pickands dependence function, and assume that the related copula  $C$  is not the independent copula. Then:*

- (i) *It holds that  $A(t) < 1$  for all  $t \in (0, 1)$ . If moreover  $A$  is continuously differentiable on  $[0, 1]$ , then:*
- (ii) *We have  $A'(0) < 0$  and  $A'(1) > 0$ .*
- (iii) *We have*

$$\forall x \in [0, 1], 0 \leq A(x) - xA'(x) \leq 1.$$

*These inequalities are all strict on  $(0, 1)$  should  $A$  be strictly convex.*

*Proof of Lemma 1* To show (i), assume that there is  $t_0 \in (0, 1)$  such that  $A(t_0) = 1$ . Recall the increasing slopes inequality:

$$\forall x, y, z \in [0, 1], x < y < z \Rightarrow \frac{A(y) - A(x)}{y - x} \leq \frac{A(z) - A(x)}{z - x} \leq \frac{A(z) - A(y)}{z - y}.$$



Apply this first with  $x = t \in (0, t_0)$ ,  $y = t_0$  and  $z = 1$  to get

$$\frac{1 - A(t)}{1 - t} \leq 0$$

so that  $A(t) \geq 1$  and therefore  $A(t) = 1$  for all  $t \in (0, t_0)$ . Apply now the above set of inequalities with  $x = 0$ ,  $y = t_0$  and  $z = t \in (t_0, 1)$  to obtain

$$0 \leq \frac{A(t) - 1}{t}$$

so that again  $A(t) = 1$  for all  $t \in (t_0, 1)$ . Consequently,  $A$  is the constant function 1, which is impossible since  $C$  is not the independent copula.

We turn to the proof of (ii). Since  $A$  is continuously differentiable on  $[0, 1]$  and convex, its derivative  $A'$  is nondecreasing. Consequently, if  $A'(0)$  were nonnegative, then so would be  $A'(x)$  for any  $x \in (0, 1]$ , and thus  $A$  would be nondecreasing on  $[0, 1]$ . Since  $A(0) = A(1) = 1$ , this would entail that  $A$  is the constant function 1, which is a contradiction. Similarly,  $A'(1)$  cannot be nonpositive since if it were, then  $A$  would be nonincreasing on  $[0, 1]$ , which is a contradiction in virtue of the equality  $A(0) = A(1) = 1$ .

We now show (iii). Notice that since  $A(x) \geq x$ , we have

$$A(x) - xA'(x) \geq x(1 - A'(x)).$$

Assume that there is  $x_0 \in (0, 1)$  such that  $A'(x_0) > 1$ . Since  $A$  is convex, the function  $A'$  is nondecreasing and therefore  $A'(x) > 1$  on  $[x_0, 1]$ . Integrating yields

$$1 - A(x_0) = A(1) - A(x_0) = \int_{x_0}^1 A'(x)dx > 1 - x_0$$

which entails  $A(x_0) < x_0$  and a contradiction. As a consequence,  $A'(x) \leq 1$  for all  $x \in (0, 1)$ , and this should also be true on the closed interval  $[0, 1]$  by continuity of  $A'$ . Finally,  $A(x) - xA'(x) \geq x(1 - A'(x)) \geq 0$  for all  $x \in [0, 1]$ . Use now the increasing slopes inequality to write

$$\forall x, t \in (0, 1), x < t \Rightarrow \frac{A(x) - 1}{x} = \frac{A(x) - A(0)}{x - 0} \leq \frac{A(t) - A(x)}{t - x}.$$

Let  $t \downarrow x$  to get

$$\forall x \in (0, 1), \frac{A(x) - 1}{x} \leq A'(x)$$

which entails  $A(x) - xA'(x) \leq 1$  for all  $x \in (0, 1)$ . Using the continuity of  $A$  and  $A'$ , this inequality is of course also true for  $x \in \{0, 1\}$ , which concludes the proof of the first set of desired inequalities if  $A$  is convex.

If  $A$  is moreover strictly convex, assume that there is  $x_0 \in (0, 1)$  such that  $A'(x_0) \geq 1$ . Then  $A'(x) \geq 1$  on  $[x_0, 1]$  and integrating yields

$$\forall t \in [x_0, 1] \cap [1/2, 1], 1 - A(t) = A(1) - A(t) = \int_t^1 A'(x)dx \geq 1 - t$$

which entails  $A(t) \leq t$  and therefore  $A(t) = t$  on an interval with nonempty interior, which is a contradiction by the strict convexity of  $A$ . This entails  $A(x) - xA'(x) \geq$

$x(1 - A'(x)) > 0$  for all  $x \in (0, 1)$ . Finally, to prove that  $A(x) - xA'(x) < 1$  for all  $x \in (0, 1)$ , we apply again the increasing slopes inequality: for all  $x, t \in (0, 1)$ ,

$$x < t \Rightarrow \frac{A(x) - 1}{x} = \frac{A(x) - A(0)}{x - 0} < \frac{A(x) - A(x/2)}{x/2} < \frac{A(t) - A(x/2)}{t - (x/2)} < \frac{A(t) - A(x)}{t - x}$$

where all inequality signs are strict because of the strict convexity of  $A$ . Let  $t \downarrow x$  to get

$$\forall x \in (0, 1), \frac{A(x) - 1}{x} < \frac{A(x) - A(x/2)}{x/2} \leq A'(x)$$

which entails  $A(x) - xA'(x) < 1$  for all  $x \in (0, 1)$  as required. □

The second lemma contains an equivalent of  $\varphi(u, v)$  as  $u \rightarrow 0$  and  $v \rightarrow v_0 > 0$ .

**Lemma 2** *Assume that  $A$  is continuously differentiable on  $[0, 1]$ . Then, if  $u \rightarrow 0$  and  $v \rightarrow v_0 > 0$ , we have that:*

$$\varphi(u, v) = (1 - v_0)A'(1)u(1 + o(1)).$$

*Proof of Lemma 2* Note that, since  $u \rightarrow 0$  and  $v \rightarrow v_0 > 0$ ,

$$\begin{aligned} \varphi(u, v) &= (1 - u)(1 - v) \left\{ [(1 - u)(1 - v)]^{A(\log(1-v)/\log([1-u][1-v]))-1} - 1 \right\} \\ &= (1 - v_0 + o(1)) \left\{ [(1 - u)(1 - v)]^{A(\log(1-v)/\log([1-u][1-v]))-1} - 1 \right\}. \end{aligned}$$

Besides, we get, as  $u \rightarrow 0$  and  $v \rightarrow v_0 > 0$ :

$$\frac{\log(1 - v)}{\log([1 - u][1 - v])} \rightarrow 1 \text{ and therefore } A\left(\frac{\log(1 - v)}{\log([1 - u][1 - v])}\right) \rightarrow A(1) = 1.$$

In particular:

$$\begin{aligned} A\left(\frac{\log(1 - v)}{\log([1 - u][1 - v])}\right)^{-1} &= \left[ \frac{\log(1 - v)}{\log([1 - u][1 - v])} - 1 \right] A'(1) + o\left(\left| \frac{\log(1 - v)}{\log([1 - u][1 - v])} - 1 \right|\right) \\ &= \left[ \frac{-\log(1 - u)}{\log([1 - u][1 - v])} \right] A'(1) + o\left(\frac{\log(1 - u)}{\log([1 - u][1 - v])}\right). \end{aligned}$$

Since as  $u \rightarrow 0$  and  $v \rightarrow v_0 > 0$  we have:

$$\frac{1}{u} \times \frac{-\log(1 - u)}{\log([1 - u][1 - v])} \rightarrow \frac{1}{\log(1 - v_0)} < 0,$$

we get

$$\log([1 - u][1 - v]) \left[ A\left(\frac{\log(1 - v)}{\log([1 - u][1 - v])}\right) - 1 \right] = uA'(1) + o(u).$$

Because this quantity converges to 0 and

$$\varphi(u, v) = (1 - v_0 + o(1)) \left\{ \exp\left(\log([1 - u][1 - v]) \left[ A\left(\frac{\log(1 - v)}{\log([1 - u][1 - v])}\right) - 1 \right]\right) - 1 \right\}$$

the conclusion readily follows by using a Taylor expansion of the exponential function in a neighbourhood of 0. □

Lemma 3 is a very useful result about the asymptotic interactions between the survival functions of  $Y$  and  $T$ . We state it for the sake of clarity: its proof essentially consists in applying Theorem 1.2.6 in de Haan and Ferreira (2006) repeatedly.

**Lemma 3** *Assume that condition  $(\mathcal{H})$  holds and that the distributions of  $Y$  and  $T$  satisfy conditions  $\mathcal{C}_1(\gamma_Y)$  and  $\mathcal{C}_1(\gamma_T)$  respectively. If in addition  $\gamma_Y\gamma_T \geq 0$  and  $|\gamma_Y| > |\gamma_T|$ , then when  $\tau = \infty$  we have*

$$\exists \varepsilon > 0, \frac{\overline{F}_T(z)}{\overline{F}_Y(z)} = O(z^{-\varepsilon}) \text{ as } z \rightarrow \infty.$$

When otherwise  $\tau < \infty$ , then

$$\exists \varepsilon > 0, \frac{\overline{F}_T(z)}{\overline{F}_Y(z)} = O((\tau - z)^\varepsilon) \text{ as } z \uparrow \tau.$$

The next result contains an equivalent of  $\varphi(u, v)$  as  $u, v \rightarrow 0$ .

**Lemma 4** *Assume that  $A$  is continuously differentiable on  $[0, 1]$ . Then, if  $u, v \rightarrow 0$ , we have that:*

- $\varphi(u, v) = -A'(0)v(1 + o(1))$  if  $v/u \rightarrow 0$ ;
- $\varphi(u, v) = A'(1)u(1 + o(1))$  if  $v/u \rightarrow \infty$ ;
- $\varphi(u, v) = (c + 1) \left[ 1 - A\left(\frac{c}{c + 1}\right) \right] u(1 + o(1))$  if  $v/u \rightarrow c \in (0, \infty)$ .

*Proof of Lemma 4* Start by writing, as  $u, v \rightarrow 0$ ,

$$\begin{aligned} \varphi(u, v) &= (1 - u)(1 - v) \left\{ [(1 - u)(1 - v)]^{A(\log(1-v)/\log([1-u][1-v]))-1} - 1 \right\} \\ &= \left\{ [(1 - u)(1 - v)]^{A(\log(1-v)/\log([1-u][1-v]))-1} - 1 \right\} (1 + o(1)). \end{aligned}$$

Since  $u, v \rightarrow 0$ , we get

$$A\left(\frac{\log(1-v)}{\log([1-u][1-v])}\right) = A\left(\frac{v(1 + o(1))}{u(1 + o(1)) + v(1 + o(1))}\right) \rightarrow \begin{cases} A(0) = 1 & \text{if } v/u \rightarrow 0, \\ A(1) = 1 & \text{if } v/u \rightarrow \infty, \\ A(c/(c + 1)) & \text{if } v/u \rightarrow c \in (0, \infty). \end{cases}$$

In particular:

$$A\left(\frac{\log(1-v)}{\log([1-u][1-v])}\right) - 1 = \begin{cases} \frac{v}{u} A'(0)(1 + o(1)) & \text{if } v/u \rightarrow 0, \\ -\frac{u}{v} A'(1)(1 + o(1)) & \text{if } v/u \rightarrow \infty, \\ A(c/(c + 1)) - 1 + o(1) & \text{if } v/u \rightarrow c \in (0, \infty). \end{cases}$$

Meanwhile

$$\log([1 - u][1 - v]) = \begin{cases} -u(1 + o(1)) & \text{if } v/u \rightarrow 0, \\ -v(1 + o(1)) & \text{if } v/u \rightarrow \infty, \\ -(c + 1)u(1 + o(1)) & \text{if } v/u \rightarrow c \in (0, \infty) \end{cases}$$

so that

$$\begin{aligned} & \log([1 - u][1 - v]) \left[ A \left( \frac{\log(1 - v)}{\log([1 - u][1 - v])} \right) - 1 \right] \\ &= \begin{cases} -vA'(0)(1 + o(1)) & \text{if } v/u \rightarrow 0, \\ uA'(1)(1 + o(1)) & \text{if } v/u \rightarrow \infty, \\ (c + 1)[1 - A(c/(c + 1))]u(1 + o(1)) & \text{if } v/u \rightarrow c \in (0, \infty). \end{cases} \end{aligned}$$

Since this quantity converges to 0 and

$$\varphi(u, v) = (1 + o(1)) \left\{ \exp \left( \log([1 - u][1 - v]) \left[ A \left( \frac{\log(1 - v)}{\log([1 - u][1 - v])} \right) - 1 \right] \right) - 1 \right\}$$

the conclusion readily follows by using a Taylor expansion of the exponential function in a neighbourhood of 0. □

Lemma 5, which provides an asymptotic equivalent of  $\bar{F}_Z$ , is a direct corollary of Lemma 4. Note that it gives indeed a true asymptotic equivalent of  $\bar{F}_Z$ , since  $A'(0) < 0$ ,  $A'(1) > 0$  and  $A(t) < 1$  for any  $t \in (0, 1)$ , see Lemma 1(i) and (ii).

**Lemma 5** *Assume that condition (H) holds and that A is continuously differentiable on [0, 1]. Then, as  $z \uparrow \tau$ , we have that:*

- (i)  $\bar{F}_Z(z) = -A'(0)\bar{F}_T(z)(1 + o(1))$  if  $\bar{F}_T(z)/\bar{F}_Y(z) \rightarrow 0$  as  $z \uparrow \tau$ ;
- (ii)  $\bar{F}_Z(z) = A'(1)\bar{F}_Y(z)(1 + o(1))$  if  $\bar{F}_T(z)/\bar{F}_Y(z) \rightarrow \infty$  as  $z \uparrow \tau$ ;
- (iii)  $\bar{F}_Z(z) = (c + 1) \left[ 1 - A \left( \frac{c}{c + 1} \right) \right] \bar{F}_Y(z)(1 + o(1))$  if  $\bar{F}_T(z)/\bar{F}_Y(z) \rightarrow c \in (0, \infty)$ .

*Proof of Lemma 5.* Just use Eq. 1 together with Lemma 4. □

The next lemma contains an equivalent of the quantity  $1 - \partial_1 C(1 - u, 1 - v)$  as  $u, v \rightarrow 0$ , and is fundamental to prove Theorem 2.

**Lemma 6** *Assume that A is twice continuously differentiable on [0, 1]. Then, if  $u, v \rightarrow 0$ , we have that:*

- (i)  $1 - \partial_1 C(1 - u, 1 - v) = (1 + A'(0))v + \frac{1}{2}A''(0)(v/u)^2 + o(v) + o((v/u)^2)$  if  $v/u \rightarrow 0$ ;
- (ii)  $1 - \partial_1 C(1 - u, 1 - v) \rightarrow A'(1)$  if  $v/u \rightarrow \infty$ ;
- (iii)  $1 - \partial_1 C(1 - u, 1 - v) \rightarrow 1 - \left[ A \left( \frac{c}{c + 1} \right) - \frac{c}{c + 1} A' \left( \frac{c}{c + 1} \right) \right]$  if  $v/u \rightarrow c \in (0, \infty)$ .

*Proof of Lemma 6* Recall the identity

$$\partial_1 C(1 - u, 1 - v) = \frac{C(1 - u, 1 - v)}{1 - u} B\left(\frac{\log(1 - v)}{\log(1 - u) + \log(1 - v)}\right)$$

where  $B(x) = A(x) - xA'(x)$ . We first prove (i). It is straightforward to show that

$$B(x) = 1 - \frac{x^2}{2} A''(0) + o(x^2).$$

Since in this case we assume  $v/u \rightarrow 0$ , we get

$$\frac{\log(1 - v)}{\log(1 - u) + \log(1 - v)} = \frac{v}{u}(1 + o(1)).$$

Therefore

$$\partial_1 C(1 - u, 1 - v) = \frac{C(1 - u, 1 - v)}{1 - u} \left[ 1 - \frac{A''(0)}{2} \left(\frac{v}{u}\right)^2 + o\left(\left[\frac{v}{u}\right]^2\right) \right]. \tag{6}$$

It remains to compute an asymptotic expansion of  $C(1 - u, 1 - v)/(1 - u)$ . To this end, we rewrite this term as

$$\frac{C(1 - u, 1 - v)}{1 - u} = \exp\left\{A\left(\frac{\log(1 - v)}{\log(1 - u) + \log(1 - v)}\right) [\log(1 - u) + \log(1 - v)] - \log(1 - u)\right\}.$$

Note now that

$$\begin{aligned} \frac{\log(1 - v)}{\log(1 - u) + \log(1 - v)} &= \frac{\log(1 - v)}{\log(1 - u)} \left(1 + \frac{\log(1 - v)}{\log(1 - u)}\right)^{-1} \\ &= \frac{\log(1 - v)}{\log(1 - u)} \left(1 - \frac{\log(1 - v)}{\log(1 - u)} + o\left(\frac{v}{u}\right)\right). \end{aligned}$$

Plugging this into the Taylor expansion

$$A(x) = A(0) + xA'(0) + \frac{x^2}{2} A''(0) + o(x^2) = 1 + xA'(0) + \frac{x^2}{2} A''(0) + o(x^2)$$

and rearranging yields

$$\begin{aligned} A\left(\frac{\log(1 - v)}{\log(1 - u) + \log(1 - v)}\right) &= 1 + \frac{\log(1 - v)}{\log(1 - u)} A'(0) + \left(\frac{\log(1 - v)}{\log(1 - u)}\right)^2 \left(\frac{A''(0)}{2} - A'(0)\right) \\ &\quad + o\left(\left[\frac{v}{u}\right]^2\right). \end{aligned}$$

Now

$$\log(1 - u) + \log(1 - v) = \log(1 - u) \left[1 + \frac{\log(1 - v)}{\log(1 - u)}\right]$$

so that

$$\begin{aligned}
 & A \left( \frac{\log(1-v)}{\log(1-u) + \log(1-v)} \right) [\log(1-u) + \log(1-v)] \\
 &= \log(1-u) \left( 1 + \frac{\log(1-v)}{\log(1-u)} (1+A'(0)) + \left( \frac{\log(1-v)}{\log(1-u)} \right)^2 \frac{A''(0)}{2} + o \left( \left[ \frac{v}{u} \right]^2 \right) \right)
 \end{aligned} \tag{7}$$

and therefore

$$\begin{aligned}
 & A \left( \frac{\log(1-v)}{\log(1-u) + \log(1-v)} \right) [\log(1-u) + \log(1-v)] - \log(1-u) \\
 &= (1 + A'(0)) \log(1-v) + \frac{[\log(1-v)]^2 A''(0)}{\log(1-u)} \frac{1}{2} + o \left( \frac{v^2}{u} \right).
 \end{aligned}$$

We combine this with a Taylor expansion of the exponential function in a neighbourhood of 0 to get

$$\begin{aligned}
 \frac{C(1-u, 1-v)}{1-u} &= 1 + (1 + A'(0)) \log(1-v) + \frac{[\log(1-v)]^2 A''(0)}{\log(1-u)} \frac{1}{2} + o \left( \frac{v^2}{u} \right) \\
 &= 1 - (1 + A'(0))v + o(v).
 \end{aligned} \tag{8}$$

Combining Eqs. 6 and 8 and rearranging terms concludes the proof of (i). □

We now turn to the proof of (ii) and (iii): as  $u$  and  $v \rightarrow 0$ , we have

$$\partial_1 C(1-u, 1-v) = B \left( \frac{\log(1-v)}{\log(1-u) + \log(1-v)} \right) (1 + o(1)).$$

Besides,

$$B \left( \frac{\log(1-v)}{\log(1-u) + \log(1-v)} \right) = B \left( \frac{v(1 + o(1))}{u(1 + o(1)) + v(1 + o(1))} \right).$$

The argument of  $B$  here converges to 1 if  $v/u \rightarrow \infty$ , and to  $c/(c+1)$  if  $v/u \rightarrow c \in (0, \infty)$ . By continuity of  $B$ , it follows that  $1 - \partial_1 C(1-u, 1-v) \rightarrow 1 - B(1) = A'(1)$  if  $v/u \rightarrow \infty$ , and

$$1 - \partial_1 C(1-u, 1-v) \rightarrow 1 - B \left( \frac{c}{c+1} \right) = 1 - \left[ A \left( \frac{c}{c+1} \right) - \frac{c}{c+1} A' \left( \frac{c}{c+1} \right) \right]$$

if  $v/u \rightarrow c \in (0, \infty)$ . This completes the proof. □

The final lemma contains, together with Lemma 6, the core result to show Proposition 2.

**Lemma 7** *Assume that  $A$  is twice continuously differentiable on  $[0, 1]$ . Then, if  $u, v \rightarrow 0$  with  $v/u \rightarrow 0$ , we have that*

$$\varphi(u, v) = -A'(0)v + v \left( A'(0)u - \frac{A''(0)}{2} \frac{v}{u} + o(u) + o \left( \frac{v}{u} \right) \right).$$

*Proof of Lemma 7* Recall from the proof of Lemma 4 that

$$\varphi(u, v) = (1 - u)(1 - v) \left\{ [(1 - u)(1 - v)]^{A(\log(1-v)/\log((1-u)[1-v]))-1} - 1 \right\}.$$

Now, by Eq. 7 in the proof of Lemma 6,

$$\begin{aligned} & \left[ A \left( \frac{\log(1 - v)}{\log(1 - u) + \log(1 - v)} \right) - 1 \right] [\log(1 - u) + \log(1 - v)] \\ &= \log(1 - v)A'(0) + \frac{\log^2(1 - v)}{\log(1 - u)} \frac{A''(0)}{2} + o\left(\frac{v^2}{u}\right) \\ &= -vA'(0) - \frac{v^2}{u} \frac{A''(0)}{2} + o\left(\frac{v^2}{u}\right). \end{aligned}$$

Consequently, using a Taylor expansion,

$$\begin{aligned} & [(1 - u)(1 - v)]^{A(\log(1-v)/\log((1-u)[1-v]))-1} - 1 \\ &= \exp \left( \left[ A \left( \frac{\log(1 - v)}{\log(1 - u) + \log(1 - v)} \right) - 1 \right] [\log(1 - u) + \log(1 - v)] \right) - 1 \\ &= -vA'(0) - \frac{v^2}{u} \frac{A''(0)}{2} + o\left(\frac{v^2}{u}\right). \end{aligned}$$

Multiplying this expansion by  $(1 - u)(1 - v)$  and expanding again concludes the proof. □

## References

Aban, I.B., Meerschaert, M.M., Panorska, A.K.: Parameter estimation for the truncated Pareto distribution. *J. Am. Stat. Assoc.* **101**, 270–277 (2006)

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

Beirlant, J., Guillou, A., Toulemonde, G.: Peaks-Over-Threshold modeling under random censoring. *Commun. Stat. – Theory Methods* **39**, 1158–1179 (2010)

Beirlant, J., Bardoutsos, A., de Wet, T., Gijbels, I.: Bias reduced tail estimation for censored Pareto type distributions. *Stat. Probab. Lett.* **109**, 78–88 (2016)

Bingham, N.H., Goldie, C.M., Teugels, J.L.: *Regular Variation*. Cambridge University Press, Cambridge (1987)

Brahimi, B., Meraghni, D., Necir, A.: Gaussian approximation to the extreme value index estimator of a heavy-tailed distribution under random censoring. *Math. Methods Stat.* **24**, 266–279 (2015)

Capéràa, P., Fougères, A.-L., Genest, C.: Bivariate distributions with given extreme value attractor. *J. Multivar. Anal.* **72**, 30–49 (2000)

Carrière, J.F.: Removing cancer when it is correlated with other causes of death. *Biom. J.* **37**, 339–350 (1995)

de Haan, L., Ferreira, A.: *Extreme Value Theory: an Introduction*. Springer, New York (2006)

Deheuvels, P.: Probabilistic aspects of multivariate extremes. In: Tiago de Oliveira, J. (ed.) *Statistical extremes and applications*, pp. 117–130 (1984)

Dekkers, A.L.M., Einmahl, J.H.J., de Haan, L.: A moment estimator for the index of an extreme-value distribution. *Ann. Stat.* **17**, 1833–1855 (1989)

Ding, A.A.: Identifiability conditions for covariate effects model on survival times under informative censoring. *Stat. Probab. Lett.* **80**, 911–915 (2010)

- Ding, A.A.: Copula identifiability conditions for dependent truncated data model. *Lifetime Data Anal.* **18**, 397–407 (2012)
- Ebrahimi, N., Molefe, D.: Survival function estimation when lifetime and censoring time are dependent. *J. Multivar. Anal.* **87**, 101–132 (2003)
- Ebrahimi, N., Molefe, D., Ying, Z.: Identifiability and censored data. *Biometrika* **90**, 724–727 (2003)
- Einmahl, J.H.J., Fils-Villetard, A., Guillou, A.: Statistics of extremes under random censoring. *Bernoulli* **14**, 207–227 (2008)
- Fisher, L., Kanarek, P.: Presenting censored survival data when censoring and survival times may not be independent. In: Proschan, F., Serfling, R. (eds.) *Reliability and Biometry. Statistical Analysis of Life Length*, pp. 303–326. Society for Industrial and Applied Mathematics (1974)
- Galambos, J.: *The Asymptotic Theory of Extreme Order Statistics*, Wiley Series in Probability and Mathematical Statistics. Wiley, New York-Chichester-Brisbane (1978)
- Gardes, L., Stupfler, G.: Estimating extreme quantiles under random truncation. *TEST* **24**, 207–227 (2015)
- Genest, C., Rivest, L.-P.: A characterization of Gumbel's family of extreme value distributions. *Stat. Probab. Lett.* **8**, 207–211 (1989)
- Goegebeur, Y., de Wet, T.: Estimation of the third-order parameter in extreme value statistics. *TEST* **21**, 330–354 (2012)
- Gomes, M.I., Neves, M.M.: Estimation of the extreme value index for randomly censored data. *Biom. Lett.* **48**, 1–22 (2011)
- Gudendorf, G., Segers, J.: Extreme-value copulas. In: Jaworski, P., Durante, F., Härdle, W.K., Rychlik, T. (eds.) *Copula Theory and Its Applications. Lecture Notes in Statistics*, vol. 198, pp. 127–145 (2010)
- Gumbel, E.J.: Bivariate exponential distributions. *J. Am. Stat. Assoc.* **55**, 698–707 (1960)
- Hsu, C.-H., Taylor, J.M.G.: A robust weighted Kaplan-Meier approach for data with dependent censoring using linear combinations of prognostic covariates. *Stat. Med.* **29**, 2215–2223 (2010)
- Huang, X., Zhang, N.: Regression survival analysis with an assumed copula for dependent censoring: a sensitivity analysis approach. *Biometrics* **64**, 1090–1099 (2008)
- Jackson, D., White, I.R., Seaman, S., Evans, H., Baisley, K., Carpenter, J.: Relaxing the independent censoring assumption in the Cox proportional hazards model using multiple imputation. *Stat. Med.* **33**, 4681–4694 (2014)
- Joe, H.: *Multivariate Models and Dependence Concepts*. Chapman and Hall/CRC, Boca Raton (1997)
- Kaplan, E.L., Meier, P.: Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **53**, 457–481 (1958)
- Klein, J.P., Moeschberger, M.L.: Independent or dependent competing risks: does it make a difference. *Commun. Stat. Simul. Comput.* **16**, 507–533 (1987)
- Lagakos, S.W.: General right censoring and its impact on the analysis of survival data. *Biometrics* **35**, 139–156 (1979)
- Ledford, A.W., Tawn, J.A.: Statistics for near independence in multivariate extreme values. *Biometrika* **83**, 169–187 (1996)
- Ledford, A.W., Tawn, J.A.: Modelling dependence within joint tail regions. *J. R. Stat. Soc. Ser. B* **59**, 475–499 (1997)
- Li, Y., Tiwari, R.C., Guha, S.: Mixture cure survival models with dependent censoring. *J. R. Stat. Soc. Ser. B* **69**, 285–306 (2007)
- Lo, S.M.S., Wilke, R.A.: A copula model for dependent competing risks. *J. R. Stat. Soc. Ser. C* **59**, 359–376 (2010)
- Modarres, M., Kaminskiy, M.P., Krivtsov, V.: *Reliability Engineering and Risk Analysis: a Practical Guide*, 2nd edn. CRC Press, Boca Raton (2009)
- Ndao, P., Diop, A., Dupuy, J.-F.: Nonparametric estimation of the conditional tail index and extreme quantiles under random censoring. *Comput. Stat. Data Anal.* **79**, 63–79 (2014)
- Ndao, P., Diop, A., Dupuy, J.-F.: Nonparametric estimation of the conditional extreme-value index with random covariates and censoring. *J. Stat. Plan. Infer.* **168**, 20–37 (2016)
- Reynkens, T., Verbelen, R., Beirlant, J., Antonio, K.: Modelling censored losses using splicing: a global fit strategy with mixed Erlang and extreme value distributions. *Insur.: Math. Econ.* **77**, 65–77 (2017)
- Rivest, L.-P., Wells, M.T.: A martingale approach to the Copula-Graphic estimator for the survival function under dependent censoring. *J. Multivar. Anal.* **79**, 138–155 (2001)
- Sayah, A., Yahia, D., Brahim, B.: On robust tail index estimation under random censorship. *Afr. Stat.* **9**, 671–683 (2014)



- Sklar, A.: Fonctions de répartition à  $n$  dimensions et leurs marges, vol. 8. Publications de l'Institut de Statistique de l'Université de Paris, pp. 229–231 (1959)
- Stupfler, G.: Estimating the conditional extreme-value index under random right-censoring. *J. Multivar. Anal.* **144**, 1–24 (2016)
- Tawn, J.A.: Bivariate extreme value theory: Models and estimation. *Biometrika* **75**, 397–415 (1988)
- Tsiatis, A.: A non-identifiability aspect of the problem of competing risks. *Proc. Natl. Acad. Sci.* **72**, 20–22 (1975)
- Worms, J., Worms, R.: New estimators of the extreme value index under random right censoring, for heavy-tailed distributions. *Extremes* **17**, 337–358 (2014)
- Zeng, D.: Estimating marginal survival function by adjusting for dependent censoring using many covariates. *Ann. Stat.* **32**, 1533–1555 (2004)
- Zheng, M., Klein, J.P.: Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika* **82**, 127–138 (1995)