



On a relationship between randomly and non-randomly thresholded empirical average excesses for heavy tails

Gilles Stupfler¹ 

Received: 8 August 2018 / Revised: 13 May 2019 / Accepted: 20 May 2019 /
Published online: 28 May 2019
© The Author(s) 2019

Abstract

Motivated by theoretical similarities between the classical Hill estimator of the tail index of a heavy-tailed distribution and one of its pseudo-estimator versions featuring a non-random threshold, we show a novel asymptotic representation of a class of empirical average excesses above a high random threshold, expressed in terms of order statistics, using their counterparts based on a suitable non-random threshold, which are sums of independent and identically distributed random variables. As a consequence, the analysis of the joint convergence of such empirical average excesses essentially boils down to a combination of Lyapunov's central limit theorem and the Cramér-Wold device. We illustrate how this allows to improve upon, as well as produce conceptually simpler proofs of, very recent results about the joint convergence of marginal Hill estimators for a random vector with heavy-tailed marginal distributions. These results are then applied to the proof of a convergence result for a tail index estimator when the heavy-tailed variable of interest is randomly right-truncated. New results on the joint convergence of conditional tail moment estimators of a random vector with heavy-tailed marginal distributions are also obtained.

Keywords Average excess · Conditional tail moment · Heavy-tailed distribution · Hill estimator · Joint convergence · Random right-truncation · Tail homogeneity · Tail index · Threshold

AMS 2000 Subject Classifications 62G05 · 62G20 · 62G30 · 62G32

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10687-019-00351-5>) contains supplementary material, which is available to authorized users.

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

¹ School of Mathematical Sciences, The University of Nottingham, University Park, Nottingham NG7 2RD, UK

1 Introduction and motivation

Heavy-tailed random variables appear in numerous fields of statistical applications of extreme value analysis, such as insurance and finance (see e.g. p.9 of Embrechts et al. (1997)), geoscience (see Section 1.3.5 of Beirlant et al. (2004)) and analysis of teletraffic data (see Section 8 of Resnick (2007)). A typical goal of extreme value analysis, in such contexts, is the estimation of extreme quantiles of a univariate random variable of interest, such as the daily log-return of a stock market index, the magnitude of earthquakes in a given region, or the size of data packets transferred in a computer network. A well-established procedure for extreme quantile estimation is Weissman's extrapolation method (Weissman 1978), whose essential requirement is a consistent estimator of the tail index of the underlying heavy-tailed distribution. Recent studies geared towards insurance and finance have been advocating for the use and estimation of extreme versions of alternatives to quantiles, such as Conditional Tail Moments (El Methni et al. 2014), Wang distortion risk measures (introduced in Wang (1996), and studied recently in El Methni and Stupfler (2017, 2018)), L^p -quantiles (introduced in Chen (1996), and studied recently in Daouia et al. (2018b, 2019) and extremiles (introduced and studied in Daouia et al. (2018a)). All these quantities, however, can be shown to have the heavy-tailed behaviour displayed by tail quantiles, and as such, their estimators at extreme levels can be constructed *via* straightforward adaptations of Weissman's method by relying on tail index estimators as well. Tail index estimation is therefore a central question in the statistical analysis of heavy-tailed distributions.

The most popular and well-known tail index estimator is arguably the Hill estimator (Hill 1975), which is also the maximum likelihood estimator if the underlying statistical model is purely Pareto. Assuming that X_1, \dots, X_n is a random sample of copies of a variable X having a heavy-tailed distribution with tail index γ , the Hill estimator of γ is

$$\hat{\gamma}(k) := \frac{1}{k} \sum_{i=1}^k \log \frac{X_{n-i+1,n}}{X_{n-k,n}}.$$

Here, $k = k(n)$ is such that $k \rightarrow \infty$ and $k/n \rightarrow 0$, and $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ denote the order statistics related to X_1, \dots, X_n . It is well-known that, under an appropriate second-order condition on X controlling the gap between the underlying distribution and the Pareto distribution, this estimator is \sqrt{k} -asymptotically normal; a variety of proofs of this result are available, including arguments based on Rényi's representation of order statistics (see pp.74–75 in de Haan and Ferreira (2006)), the tail empirical process (see Proposition 9.3 p.302 in Resnick (2007)), or a representation of log-spacings in terms of independent exponential random variables (see Section 4.4 in Beirlant et al. (2004)). A common aim of these approaches is to deal with the structure of the Hill estimator in terms of top order statistics, which are significantly harder to handle than the original X_1, \dots, X_n . Although this theoretical difficulty is now well-understood, it is further compounded if one wants to show the joint asymptotic normality of marginal Hill estimators for a random vector with heavy-tailed marginal distributions, since one then also needs to take into account the asymptotic dependence structure within the underlying multivariate distribution.

Such joint convergence results have only been proven very recently by Dematteo and Cl emen on (2016), Kinsvater et al. (2016) and Hoga (2018). The methods of proof therein rely on advanced theoretical methodologies, namely multivariate vague convergence in Dematteo and Cl emen on (2016) and Kinsvater et al. (2016) and multivariate empirical process theory in Hoga (2018), as well as on various *ad hoc* technical conditions. These joint asymptotic results have been found useful for testing tail homogeneity across marginals of a multivariate heavy-tailed distribution in environmental or financial contexts (see Kinsvater et al. (2016) and Hoga (2018)) and constructing improved tail index estimators by pooling (see Section 3.2 in Dematteo and Cl emen on (2016)).

The Hill estimator can be written in several different ways. When there are no ties in the sample X_1, \dots, X_n (this is for example the case when the distribution function of X is continuous), one has $k = \sum_{i=1}^n \mathbb{1}_{\{X_i > X_{n-k,n}\}}$. Noting then that

$$\widehat{\gamma}(k) = \frac{\sum_{i=1}^n [\log X_i - \log X_{n-k,n}] \mathbb{1}_{\{X_i > X_{n-k,n}\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i > X_{n-k,n}\}}}$$

suggests that a major difficulty in the theoretical analysis of the Hill estimator lies in the fact that the high threshold $X_{n-k,n}$, used to guarantee the consistency of the estimator by retaining only the high values in the sample, is random. Indeed, if we could, in the asymptotic analysis of $\widehat{\gamma}(k)$, replace the random quantity $X_{n-k,n}$, which is nothing but the empirical quantile at level $1 - k/n$, by the unknown but non-random population quantile $q(1 - k/n)$, we would find the pseudo-estimator

$$\widetilde{\gamma}(k) = \frac{\sum_{i=1}^n [\log X_i - \log q(1 - k/n)] \mathbb{1}_{\{X_i > q(1-k/n)\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i > q(1-k/n)\}}}$$

which is conceptually far easier to analyse than $\widehat{\gamma}(k)$: for independent X_1, \dots, X_n , this pseudo-estimator is a ratio of sums of independent variables constructed on the X_i , and thus can easily be handled by a combination of Lyapunov’s central limit theorem and the Cram er-Wold device. The striking and perhaps unexpected point here is that the asymptotic distributions of $\widehat{\gamma}(k)$ and $\widetilde{\gamma}(k)$ are identical; this can be seen by comparing Theorem 3.2.5 in de Haan and Ferreira (2006) and Theorem 4.3.1 in Goldie and Smith (1987). On this basis, one may therefore ask whether a relationship of the form

$$\widehat{\gamma}(k) = \widetilde{\gamma}(k) + o_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right) \tag{1}$$

can be shown. Such a relationship has not, to the best of our knowledge, been proven so far in the literature. Its validity is not obvious either, since $\widetilde{\gamma}(k)$ is obtained from $\widehat{\gamma}(k)$ by replacing $X_{n-k,n}$ with $q(1 - k/n)$, and we know that $X_{n-k,n}/q(1 - k/n) - 1$ only converges to 0 at the rate \sqrt{k} (see Theorem 2.4.1 in de Haan and Ferreira (2006)), which is the common rate of convergence of $\widetilde{\gamma}(k)$ and $\widehat{\gamma}(k)$. Let us also point out straightaway that although Eq. 1 would give an additional proof of the convergence of the Hill estimator $\widehat{\gamma}(k)$, this is not where its full value lies, since the asymptotic properties of the Hill estimator are well-known. It really becomes useful when, for instance, analysing the joint convergence of marginal Hill estimators, since, in contrast to the joint convergence of the randomly thresholded versions, the joint

convergence of the non-randomly thresholded versions is easy to obtain (under an appropriate upper tail dependence condition) by a combination of standard central limit theory and the Cramér-Wold device.

The proof of Eq. 1 is the motivation for this paper. More precisely, the objective of the paper is to embed the Hill estimator in a wider class of average excesses and then provide a simple representation of those empirical average excesses above a high random threshold in terms of their pseudo-estimator versions with a non-random threshold. The class of average excesses we consider includes the Hill estimator, and can be modified in a very simple way to encompass Conditional Tail Moments (CTMs). We show in particular how the results of Dematteo and Cléménçon (2016), on the joint asymptotic normality of marginal Hill estimators for a random vector with heavy-tailed marginal distributions, can be recovered and generalised under weaker assumptions and by elementary techniques. We shall then highlight a couple of applications of this joint asymptotic normality result, including to the obtention of the asymptotic properties of the tail index estimator introduced by Gardes and Stupfler (2015) when the variable of interest is randomly right-truncated. The question of the convergence of this estimator was considered first by Gardes and Stupfler (2015) under restrictive assumptions, and then by Benchaira et al. (2015) using a delicate theoretical argument based on the weighted tail copula process and a joint tail assumption on the observed pair. Our results will make it possible to unify and extend the results of Gardes and Stupfler (2015) and Benchaira et al. (2015), without neither resorting to the former's technical conditions nor to the latter's advanced methodology and joint dependence condition. In doing so, we will also be able to give a very simple expression of the asymptotic variance of the limiting normal distribution. Finally, and motivated by the ideas of Hoga (2018), we shall apply our results to find joint convergence results on empirical CTMs that may be of independent interest, for instance in testing whether certain tail moments of two asymptotically dependent variables are equal. We will highlight how such results nicely complement earlier results proven by El Methni et al. (2014) and El Methni and Stupfler (2017).

The outline of the paper is the following. Our framework and main results are stated in Section 2. Applications of our results, to the joint convergence of Hill and CTM estimators, and to the convergence of a tail index estimator under random right-truncation are presented in Section 3. Section 4 concludes and discusses possible extensions. Proofs are deferred to a [Supplementary Material](#) document.

2 Framework and main results

We assume in this section that the data is made of independent copies X_1, \dots, X_n of a random variable X . We denote by F the distribution function of X , by $\bar{F} = 1 - F$ the related survival function, and by q the left-continuous inverse of F (that is, $q(\tau) = \inf\{t \in \mathbb{R} \mid F(t) \geq \tau\}$). We assume for ease of presentation that *the distribution function F is continuous*, so that, with probability 1, there are no ties in the sample X_1, \dots, X_n . This is not restrictive for our purposes, as our applications of the main results presented here focus on the obtention of joint convergence results for several

estimators of extreme value indicators pertaining to a multivariate random vector with heavy-tailed marginal distributions. Such an endeavour indeed typically requires the standardisation of the marginal distributions to uniform distributions, in order to obtain a meaningful description of the relevant extremal dependence structure. Our results can still be shown, at the expense of extra technical details, if F is only continuous in a neighbourhood of infinity.

Our main assumption throughout is that X has a heavy-tailed distribution. In other words, denoting by $U(t) := q(1 - t^{-1})$ the tail quantile function of X , we assume that U is a regularly varying function with index $\gamma > 0$, namely

$$\forall x > 0, \lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\gamma.$$

Since our motivating question is the obtention of the asymptotic representation (1), we shall work under a refinement of this regular variation condition that allows for the derivation of asymptotic properties of extreme value estimators. We choose here to work with the following second-order extreme value condition, which is extensively used in e.g. de Haan and Ferreira (2006):

$\mathcal{C}_2(\gamma, \rho, A)$ The function U is second-order regularly varying in a neighbourhood of $+\infty$ with index $\gamma > 0$, second-order parameter $\rho \leq 0$ and an auxiliary function A having constant sign and converging to 0 at infinity, that is,

$$\forall x > 0, \lim_{t \rightarrow \infty} \frac{1}{A(t)} \left[\frac{U(tx)}{U(t)} - x^\gamma \right] = x^\gamma \frac{x^\rho - 1}{\rho}$$

where the right-hand side should be read as $x^\gamma \log x$ when $\rho = 0$.

Let us now define the central concepts this paper will focus on. Again, let us recall our motivation, which is to link the Hill estimator $\widehat{\gamma}(k)$ to its pseudo-estimator version $\widetilde{\gamma}(k)$. One interpretation of the Hill estimator $\widehat{\gamma}(k)$ is that it is a sample version of the average log-excess $\mathbb{E}(\log X - \log U(n/k) | X > U(n/k))$ (see p.104 of Beirlant et al. (2004), and p.69 of de Haan and Ferreira (2006)); it should also be clear that $\widetilde{\gamma}(k)$ is a pseudo-estimator of this average log-excess. This motivates us to carry forward the notion of average excess with the following definition.

Definition 1 Let X be a heavy-tailed random variable and f be a continuous function on a neighbourhood of infinity such that for some t_0 , the quantity $\mathbb{E}(|f(X)| | X > t_0)$ is well-defined and finite. For any $t \geq t_0$, we define the *average f -excess of X above level t* to be

$$AE_f(t) := \mathbb{E}(f(X) - f(t) | X > t) = \frac{\mathbb{E}([f(X) - f(t)]\mathbb{1}_{\{X>t\}})}{\overline{F}(t)}$$

and the *empirical average f -excess of X above level t* to be

$$\widehat{AE}_f(t) := \frac{n^{-1} \sum_{i=1}^n [f(X_i) - f(t)]\mathbb{1}_{\{X_i>t\}}}{\widehat{F}_n(t)}, \quad \text{where } \widehat{F}_n(t) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i>t\}}.$$

We also define the *expected f -shortfall above level t* as

$$ES_f(t) := \mathbb{E}(f(X) | X > t) = AE_f(t) + f(t)$$

and the empirical expected f -shortfall above level t as

$$\widehat{ES}_f(t) := \frac{n^{-1} \sum_{i=1}^n f(X_i) \mathbb{1}_{\{X_i > t\}}}{\widehat{F}_n(t)} = \widehat{AE}_f(t) + f(t).$$

With this definition, letting $f = \log$, we find

$$\widehat{AE}_{\log}(X_{n-k,n}) = \widehat{\gamma}(k), \quad \widehat{AE}_{\log}(U(n/k)) = \widetilde{\gamma}(k)$$

$$\text{and } AE_{\log}(U(n/k)) = \mathbb{E}(\log X - \log U(n/k) \mid X > U(n/k)).$$

The other main example allowed by Definition 1 which we will consider in this paper is obtained by choosing $f = f_a : x \mapsto x^a$ for some $a > 0$. We then get, for any $t > 0$,

$$ES_{f_a}(t) = \mathbb{E}(X^a \mid X > t) =: CTM_a(t),$$

where CTM_a is the Conditional Tail Moment of order a introduced in El Methni et al. (2014). This definition makes sense for any $a < 1/\gamma$, since all conditional tail moments of X of order smaller than $1/\gamma$ are finite (a rigorous statement is Exercise 1.16 in de Haan and Ferreira (2006)). The class of CTMs has recently been used by El Methni et al. (2014) and El Methni and Stupfler (2017, 2018) for extreme risk assessment purposes in environmental, financial, and actuarial contexts. It includes in particular, for $a = 1$, the expected shortfall, whose relevance to actuarial science and interpretation has been extensively discussed recently, see e.g. Brazauskas et al. (2008), Wüthrich and Merz (2013) and Emmer et al. (2015). An empirical counterpart of $CTM_a(U(n/k)) = ES_{f_a}(U(n/k))$ is then

$$\frac{1}{k} \sum_{i=1}^k X_{n-i+1,n}^a = \widehat{ES}_{f_a}(X_{n-k,n}) = \widehat{AE}_{f_a}(X_{n-k,n}) + X_{n-k,n}^a.$$

Common features of the above two examples are that they crucially hinge on the notion of average f -excess, and most importantly that the derivative of the function f involved in their construction is a power function. This observation provides the motivation for our first main result, which provides an asymptotic relationship between $\widehat{AE}_f(X_{n-k,n})$ and $\widehat{AE}_f(U(n/k))$ when f' is regularly varying at infinity.

Theorem 1 *Suppose that X satisfies condition $\mathcal{C}_2(\gamma, \rho, A)$. Let $k = k(n) \rightarrow \infty$ be such that $k/n \rightarrow 0$ and $\sqrt{k}A(n/k) = O(1)$. Assume finally that f is continuously differentiable in a neighbourhood of infinity, ultimately increasing, and that f' is regularly varying with index $a - 1$, where $0 \leq 2a\gamma < 1$. Then we have:*

$$\frac{\widehat{AE}_f(X_{n-k,n})}{AE_f(U(n/k))} - 1 = \left(\frac{\widehat{AE}_f(U(n/k))}{AE_f(U(n/k))} - 1 \right) + a\gamma \left(\frac{\widehat{F}_n(U(n/k))}{F(U(n/k))} - 1 \right) + o_{\mathbb{P}} \left(\frac{1}{\sqrt{k}} \right).$$

In particular

$$\sqrt{k} \left(\frac{\widehat{AE}_f(X_{n-k,n})}{AE_f(U(n/k))} - 1 \right) \xrightarrow{d} \mathcal{N} \left(0, \frac{1}{1 - 2a\gamma} + a^2\gamma^2 \right).$$

Theorem 1 provides an asymptotic representation of the empirical average f -excess $\widehat{AE}_f(X_{n-k,n})$, whose expression hinges on top order statistics, in terms of sums of independent and identically distributed random variables constructed on the X_i in a simple way. Note that, due to the continuity of F , we have $\overline{F}(U(n/k)) = k/n$; writing $\overline{F}(U(n/k))$ instead of k/n , as we will do in this section, is to emphasise that the second term on the right-hand side is centred. Assumptions $k = k(n) \rightarrow \infty$, $k/n \rightarrow 0$ and $\sqrt{k}A(n/k) = O(1)$ are standard for the asymptotic analysis of extreme value estimators; the assumption $0 \leq 2a\gamma < 1$, meanwhile, ensures that $\widehat{AE}_f(U(n/k))$ is \sqrt{k} -asymptotically normal (see Lemma 4 in the [Supplementary Material](#) document), and therefore that Theorem 1 provides an expression of $\widehat{AE}_f(X_{n-k,n})$ in terms of $\widehat{AE}_f(U(n/k))$ and $\widehat{F}_n(U(n/k))$ that is meaningful for the asymptotic analysis.

In particular, setting $f = \log$ (and therefore $a = 0$) in Theorem 1 allows us to show that the motivating representation (1), of the Hill estimator in terms of sums of independent and identically distributed random variables, holds indeed. This is the focus of the following corollary.

Corollary 1 *Suppose that X satisfies condition $\mathcal{C}_2(\gamma, \rho, A)$. Let $k = k(n) \rightarrow \infty$ be such that $k/n \rightarrow 0$ and $\sqrt{k}A(n/k) = O(1)$. Then we have*

$$\widehat{\gamma}(k) = \widetilde{\gamma}(k) + o_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right)$$

where

$$\widehat{\gamma}(k) = \frac{1}{k} \sum_{i=1}^k \log \frac{X_{n-i+1,n}}{X_{n-k,n}} \text{ and } \widetilde{\gamma}(k) = \frac{\sum_{i=1}^n [\log X_i - \log U(n/k)] \mathbb{1}_{\{X_i > U(n/k)\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i > U(n/k)\}}}.$$

Corollary 1 can of course be used to provide yet another proof of the asymptotic normality of the Hill estimator, *via* the Lyapunov central limit theorem and the Cramér-Wold device applied to get the \sqrt{k} -asymptotic normality of $\widetilde{\gamma}(k)$. This is not, however, where the value of Corollary 1 lies, if only because its proof uses an approximation of the tail empirical process $x \mapsto \widehat{F}_n(xU(n/k))$ by a Gaussian process, which can itself be used to provide a direct proof of the asymptotic normality of the Hill estimator (see pp.162–163 of de Haan and Ferreira (2006)). A much more relevant impact of Corollary 1 lies in its potential for the analysis of the joint convergence of several Hill estimators. For instance, if X_1, \dots, X_n and Y_1, \dots, Y_n are independent copies of random variables X and Y , which are both heavy-tailed and satisfy a second-order condition, then we have under suitable assumptions on k that:

$$\begin{pmatrix} \widehat{\gamma}_X(k) \\ \widehat{\gamma}_Y(k) \end{pmatrix} = \begin{pmatrix} \widetilde{\gamma}_X(k) \\ \widetilde{\gamma}_Y(k) \end{pmatrix} + o_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right)$$

with obvious notation. The benefit of writing this is that while showing directly the joint convergence of the random pair on the left-hand side is difficult and appears to require advanced theoretical arguments (see Dematteo and Cléménçon (2016) and Hoga (2018)), the convergence of the right-hand side is much easier to obtain since it is nothing but a pair of (ratios of) sums of independent and identically distributed

random variables. We will return to this in Section 3 to show how this observation leads to conceptually simple proofs of the joint asymptotic normality of several Hill estimators.

Theorem 1 is somewhat tedious to apply if the focus is on an average shortfall, such as in the case of a CTM. The following theorem provides an analogue of Theorem 1 specifically dedicated to the analysis of $\widehat{\text{ES}}_f(X_{n-k,n})$, under a slightly stronger condition on f .

Theorem 2 *Work under the conditions of Theorem 1, under the additional assumption that $xf'(x)/f(x) \rightarrow a > 0$ as $x \rightarrow \infty$. Then:*

$$\frac{\widehat{\text{ES}}_f(X_{n-k,n})}{\text{ES}_f(U(n/k))} - 1 = \left(\frac{\widehat{\text{ES}}_f(U(n/k))}{\text{ES}_f(U(n/k))} - 1 \right) + a\gamma \left(\frac{\widehat{F}_n(U(n/k))}{F(U(n/k))} - 1 \right) + o_{\mathbb{P}} \left(\frac{1}{\sqrt{k}} \right).$$

In particular

$$\sqrt{k} \left(\frac{\widehat{\text{ES}}_f(X_{n-k,n})}{\text{ES}_f(U(n/k))} - 1 \right) \xrightarrow{d} \mathcal{N} \left(0, \frac{2a^2\gamma^2(1-a\gamma)}{1-2a\gamma} \right).$$

Setting $f = f_a$, we find back the asymptotic variance of the empirical estimator $\widehat{\text{ES}}_{f_a}(X_{n-k,n})$ of $\text{CTM}_a(U(n/k))$; see Theorem 1 of El Methni et al. (2014) in the regression case. The condition $2a\gamma < 1$, part of Theorems 1 and 2, also naturally appears in the asymptotic normality results of El Methni et al. (2014) and ensures in particular that the conditional tail variance of X^a is finite. Again, the main value of the above result does not lie in the fact that it gives the asymptotic distribution of the empirical average expected shortfall, but rather in that it allows one to establish the joint convergence of several of those estimators with no conceptual difficulty. With an eye on the latter, we state the following corollary of Theorem 2, which is easier to use in practice.

Corollary 2 *Work under the conditions of Theorem 2. Then:*

$$\frac{\widehat{\text{ES}}_f(X_{n-k,n})}{\text{ES}_f(U(n/k))} - 1 = a\gamma \left(\frac{n^{-1} \sum_{i=1}^n [f(X_i) - f(U(n/k))] \mathbb{1}_{\{X_i > U(n/k)\}}}{\mathbb{E}([f(X) - f(U(n/k))] \mathbb{1}_{\{X > U(n/k)\}})} - 1 \right) + o_{\mathbb{P}} \left(\frac{1}{\sqrt{k}} \right).$$

Let us point out that Corollary 2 gives an asymptotic representation of $\widehat{\text{ES}}_f(X_{n-k,n})$ as a single sum of independent, identically distributed and centred random variables. Its use does not even involve any linearisation (unlike Theorems 1 and 2), making it particularly simple to apply.

Our objective in the rest of this paper is to show how the main theoretical results of this section can be applied to finding a solution to two theoretical questions: the joint convergence of marginal Hill estimators, and the joint convergence of marginal Conditional Tail Moments. We shall also explore how our result on the joint convergence of Hill estimators can be fruitfully applied to solve the question of the convergence

of a specific tail index estimator, introduced by Gardes and Stupfler (2015) to tackle the case when the variable of interest is randomly right-truncated. Before that, we conclude this section on a generalisation of Theorem 1 and Corollary 1 to the case when the order statistic $X_{n-k,n}$ is replaced by an *arbitrary* \sqrt{k} -consistent estimator $\widehat{U}(n/k)$ of the quantile $U(n/k)$. In this context, $\widehat{\text{AE}}_f(\widehat{U}(n/k))$ cannot be represented by sums of independent, identically distributed and centred random variables anymore, but interestingly Corollary 1 still stands.

Theorem 3 *Work under the conditions of Theorem 1. Then we have:*

$$\frac{\widehat{\text{AE}}_f(\widehat{U}(n/k))}{\text{AE}_f(U(n/k))} - 1 = \left(\frac{\widehat{\text{AE}}_f(U(n/k))}{\text{AE}_f(U(n/k))} - 1 \right) + a\gamma \left(\frac{\widehat{F}_n(U(n/k))}{\widehat{F}_n(\widehat{U}(n/k))} - 1 \right) + o_{\mathbb{P}} \left(\frac{1}{\sqrt{k}} \right)$$

for any estimator $\widehat{U}(n/k)$ of $U(n/k)$ such that $\widehat{U}(n/k)/U(n/k) - 1 = O_{\mathbb{P}}(1/\sqrt{k})$.

Corollary 3 *Suppose that X satisfies condition $\mathcal{C}_2(\boldsymbol{\gamma}, \boldsymbol{\rho}, \mathbf{A})$. Let $k = k(n) \rightarrow \infty$ be such that $k/n \rightarrow 0$ and $\sqrt{k}\mathbf{A}(n/k) = O(1)$. Assume finally that $\widehat{U}(n/k)$ is an estimator of $U(n/k)$ such that $\widehat{U}(n/k)/U(n/k) - 1 = O_{\mathbb{P}}(1/\sqrt{k})$. Then we have*

$$\bar{\gamma}(k) = \tilde{\gamma}(k) + o_{\mathbb{P}} \left(\frac{1}{\sqrt{k}} \right)$$

where

$$\bar{\gamma}(k) = \frac{\sum_{i=1}^n [\log X_i - \log \widehat{U}(n/k)] \mathbb{1}_{\{X_i > \widehat{U}(n/k)\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i > \widehat{U}(n/k)\}}}$$

and

$$\tilde{\gamma}(k) = \frac{\sum_{i=1}^n [\log X_i - \log U(n/k)] \mathbb{1}_{\{X_i > U(n/k)\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i > U(n/k)\}}}.$$

We observe that Theorem 3 and Corollary 3 are indeed generalisations of Theorem 1 and Corollary 1: when $\widehat{U}(n/k) = X_{n-k,n}$, one has $\widehat{F}_n(\widehat{U}(n/k)) = \widehat{F}_n(X_{n-k,n}) = k/n = \overline{F}(U(n/k))$ by continuity of F and $\bar{\gamma}(k) = \widehat{\gamma}(k)$.

3 Applications

3.1 Joint convergence of marginal Hill estimators

Let $(\mathbf{X}_i)_{1 \leq i \leq n}$, with $\mathbf{X}_i = (X_i^{(1)}, X_i^{(2)}, \dots, X_i^{(d)})$, be a sample of independent copies of a random vector $\mathbf{X} = (X^{(1)}, X^{(2)}, \dots, X^{(d)})$ such that each component $X^{(j)}$ has a continuous distribution function F_j and satisfies condition $\mathcal{C}_2(\boldsymbol{\gamma}_j, \boldsymbol{\rho}_j, \mathbf{A}_j)$. Our first goal is to establish a joint convergence result for the Hill estimators of the γ_j built on this sample, that is:

$$\widehat{\gamma}_j(k_j) = \frac{1}{k_j} \sum_{i=1}^{k_j} \log \left(X_{n-i+1,n}^{(j)} \right) - \log \left(X_{n-k_j,n}^{(j)} \right)$$

where $k_j = k_j(n) \rightarrow \infty$, with $k_j/n \rightarrow 0$. This theoretical question is addressed in Dematteo and Cl emen on (2016) and further discussed in Kinsvater et al. (2016) under the assumption $\gamma_j = \gamma$ for all $j \in \{1, \dots, d\}$.

Since each $\widehat{\gamma}_j(k_j)$ is built on top order statistics from the corresponding $X^{(j)}$, our objective of analysing the joint convergence of the $\widehat{\gamma}_j(k_j)$ calls for some sort of extremal dependence assumption between the $X^{(j)}$. We shall work under the following condition on the pairwise upper tail dependence between any two components $X^{(j)}$ and $X^{(l)}$:

J(R) For any (j, l) with $1 \leq j < l \leq d$, there is a function $R_{j,l}$ on $[0, \infty]^2 \setminus \{(\infty, \infty)\}$ such that we have the convergence

$$\lim_{t \rightarrow \infty} t \mathbb{P} \left(\overline{F}_j(X^{(j)}) \leq \frac{x_j}{t}, \overline{F}_l(X^{(l)}) \leq \frac{x_l}{t} \right) = R_{j,l}(x_j, x_l)$$

for any $(x_j, x_l) \in [0, \infty]^2 \setminus \{(\infty, \infty)\}$.

This condition appears, among others, in Cai et al. (2015) as well as, in a slightly different form, in Hoga (2018). It is a convenient way to describe the asymptotic dependence structure of the bivariate random vector $(X^{(j)}, X^{(l)})$, while being weaker than a pairwise bivariate regular variation assumption (in the sense of e.g. Resnick (1987)) and *a fortiori* weaker than a multivariate regular variation assumption on the random vector \mathbf{X} such as the one of Dematteo and Cl emen on (2016). The function $R_{j,l}$ is sometimes called the tail copula of $(X^{(j)}, X^{(l)})$ (see Schmidt and Stadtm uller (2006)). We mention that, as noted by Cai et al. (2015), the function $R_{j,l}$ characterises the stable tail dependence function of the pair $(X^{(j)}, X^{(l)})$ as defined by Drees and Huang (1998).

Second-order regular variation of each marginal distribution and the above pairwise upper tail dependence assumption turn out to be sufficient to obtain the joint convergence of the $\widehat{\gamma}_j(k_j)$, as the next result shows.

Theorem 4 Assume that, for each $1 \leq j \leq d$, $X^{(j)}$ satisfies condition $\mathbf{C}_2(\gamma_j, \rho_j, \mathbf{A}_j)$, and suppose that condition **J(R)** holds. Let, for $1 \leq j \leq d$, $k_j = k_j(n) \rightarrow \infty$ be such that:

- $k_j/n \rightarrow 0$ and $\sqrt{k_j} A_j(n/k_j) \rightarrow \lambda_j \in \mathbb{R}$;
- $k_1/k_j \rightarrow c_j \in (0, \infty)$ (with then $c_1 = 1$).

Set $\mathbf{c} = (c_1, c_2, \dots, c_d)$. Then we have:

$$\sqrt{k_1} (\widehat{\gamma}_j(k_j) - \gamma_j)_{1 \leq j \leq d} \xrightarrow{d} \mathcal{N}_d(\mathbf{b}(\mathbf{c}), \mathbf{V}(\mathbf{c}))$$

with

$$\mathbf{b}_j(\mathbf{c}) = \sqrt{c_j} \frac{\lambda_j}{1 - \rho_j} \text{ and } \mathbf{V}_{j,l}(\mathbf{c}) = \begin{cases} c_j \gamma_j^2 & \text{if } j = l, \\ \gamma_j \gamma_l R_{j,l}(c_l, c_j) & \text{if } j < l. \end{cases}$$

Let us briefly highlight that the structure of the asymptotic bias component $\mathbf{b}(\mathbf{c})$ is quite strongly constrained by the values of the second-order parameters ρ_j : for all its components to be non-zero (or equivalently, for all the constants λ_j to be non-zero), all the second-order parameters ρ_j should be equal. This is due to the fact that the k_j

are proportional and each function $|A_j|$ is regularly varying with index ρ_j . For the same reason, if $\rho^* = \max_{1 \leq j \leq d} \rho_j$, then $\mathbf{b}_j(\mathbf{c}) = 0$ whenever j is such that $\rho_j < \rho^*$.

Our Theorem 4 contains Theorem 3.3 and Corollary 3.4 in Dematteo and Cl  men  on (2016), which are stated under more restrictive conditions, including equality of all tail indices γ_j , a multivariate regular variation assumption on \mathbf{X} , von Mises conditions on each marginal distribution of \mathbf{X} , and a uniform analogue of condition $\mathcal{J}(\mathbf{R})$ [note also that Corollary 3.4 in Dematteo and Cl  men  on (2016) should, with their notation, read $\Gamma_{i,j} = c_i c_j \nu_{i,j} \left(c_i^{1/\alpha}, c_j^{1/\alpha} \right)$ in the case $i < j$; as stated, their covariance matrix may fail to be positive semi-definite, for instance for $d = 2$ and small c_2 . Compare with Proposition 1 in Kinsvater et al. (2016)]. Theorem 4 also contains Proposition 1 in Jiang et al. (2017), which is limited to the case $d = 2$ and $X^{(2)} = -X^{(1)}$. Another result related to Theorem 4 is Proposition 3 in Hoga (2018), although the present result and that of Hoga (2018) are more difficult to compare since the latter is stated within the particular context of time series analysis. Our proof of Theorem 4 is also conceptually less involved than that of Dematteo and Cl  men  on (2016), which rests upon a multivariate functional central limit theorem (see Theorem 7.1 and Corollary 7.2 therein). Finally, and without taking the time series framework into account, the result of Hoga (2018) is based on delicate arguments involving a *multivariate*, joint Skorokhod construction of Gaussian approximations for marginal tail empirical processes. Our proof, meanwhile, rests on the standard Lyapunov central limit theorem and Corollary 1, whose proof is based only on a *univariate* Gaussian approximation of the tail empirical process. It should nonetheless be made clear once again that Proposition 3 in Hoga (2018) holds in a framework of β -mixing time series, and as such is not restricted to independent and identically distributed observations, unlike our Theorem 4. Section 4 discusses a possible way of extending Theorem 4 to the time series context as well as the implications this would have on the other results presented within this paper.

Results such as Theorem 4 may be applied to define a test of tail homogeneity, that is, equality of tail indices across marginals. If there is no evidence to reject this assumption, one may then define improved estimators of the common value of the tail index by pooling together the marginal Hill estimators. These ideas are considered in Dematteo and Cl  men  on (2016) and Kinsvater et al. (2016). To illustrate how our results can be applied to such problems, we state a corollary of Theorem 4 in the case $d = 2$ of a bivariate distribution with heavy-tailed marginals.

Corollary 4 *Suppose that X and Y have continuous distribution functions F_X and F_Y , which satisfy conditions $\mathcal{C}_2(\gamma_X, \rho_X, \mathbf{A}_X)$ and $\mathcal{C}_2(\gamma_Y, \rho_Y, \mathbf{A}_Y)$, respectively. Assume that there is a function R on $[0, \infty]^2 \setminus \{(\infty, \infty)\}$ such that we have the convergence*

$$\lim_{t \rightarrow \infty} t \mathbb{P} \left(\overline{F}_X(X) \leq \frac{x}{t}, \overline{F}_Y(Y) \leq \frac{y}{t} \right) = R(x, y) \text{ for any } (x, y) \in [0, \infty]^2 \setminus \{(\infty, \infty)\}.$$

Let $k_X = k_X(n)$, $k_Y = k_Y(n)$ be such that $k_X, k_Y \rightarrow \infty$ and:

- $\sqrt{k_X} A_X(n/k_X) \rightarrow 0$ and $\sqrt{k_Y} A_Y(n/k_Y) \rightarrow 0$;
- $k_X/k_Y \rightarrow c \in (0, \infty)$.

Let also

$$R_i^X := \sum_{j=1}^n \mathbb{1}_{\{X_j \leq X_i\}} \text{ and } R_i^Y := \sum_{j=1}^n \mathbb{1}_{\{Y_j \leq Y_i\}}$$

denote the ranks of observations X_i and Y_i and $\widehat{R}_k(x, y)$ be the nonparametric estimator of $R(x, y)$ defined by

$$\widehat{R}_k(x, y) := \frac{1}{k} \sum_{i=1}^n \mathbb{1}_{\{R_i^X \geq n-kx, R_i^Y \geq n-ky\}}$$

(see e.g. Drees and Huang (1998)). Then the following hold:

- (i) Unless $c = 1$ and (X, Y) are asymptotically perfectly dependent (in the sense that $R(x, y) = \min(x, y)$), we have:

$$k_X \frac{[\widehat{\gamma}_X(k_X) - \widehat{\gamma}_Y(k_Y)]^2}{\widehat{\gamma}_X^2(k_X) + c\widehat{\gamma}_Y^2(k_Y) - 2\widehat{\gamma}_X(k_X)\widehat{\gamma}_Y(k_Y)\widehat{R}_{k_X}(c, 1)} \xrightarrow{d} \begin{cases} \chi_1^2 & \text{if } \gamma_X = \gamma_Y, \\ +\infty & \text{if } \gamma_X \neq \gamma_Y. \end{cases}$$

- (ii) If moreover $\gamma_X = \gamma_Y = \gamma$, then the estimator

$$\widehat{\gamma}(k_X, k_Y) := \frac{c - \widehat{R}_{k_X}(c, 1)}{1 + c - 2\widehat{R}_{k_X}(c, 1)} \widehat{\gamma}_X(k_X) + \frac{1 - \widehat{R}_{k_X}(c, 1)}{1 + c - 2\widehat{R}_{k_X}(c, 1)} \widehat{\gamma}_Y(k_Y)$$

satisfies

$$\sqrt{k_X} (\widehat{\gamma}(k_X, k_Y) - \gamma) \xrightarrow{d} \mathcal{N}\left(0, \gamma^2 \frac{c - R^2(c, 1)}{1 + c - 2R(c, 1)}\right)$$

and has the minimal possible asymptotic variance among the class of convex combinations of $\widehat{\gamma}_X(k_X)$ and $\widehat{\gamma}_Y(k_Y)$.

Corollary 4(i) complements Proposition 2 in Kinsvater et al. (2016) in the case $d = 2$: the latter is stated under significantly stronger assumptions, although it includes the possibility of different sample sizes for the X and Y samples, which may be relevant in specific applied setups. Corollary 4(ii) provides, under the tail homogeneity condition $\gamma_X = \gamma_Y = \gamma$, a simple expression of the convex combination of $\widehat{\gamma}_X(k_X)$ and $\widehat{\gamma}_Y(k_Y)$ that is optimal for the estimation of γ in terms of asymptotic variance. It therefore constitutes, in the case $d = 2$, an explicit version of the BEAR estimator of Dematteo and Cl  men  on (2016, pp.159–160). With our notation, this estimator can be written, for a general value of d , as

$$\widehat{\gamma}_{\text{BEAR}}(k_1, \dots, k_d) := \sum_{j=1}^d \widehat{\mu}_j \widehat{\gamma}_j(k_j), \text{ with } (\widehat{\mu}_1, \dots, \widehat{\mu}_d) = \arg \min_{\substack{\mathbf{u} \in [0, 1]^d \\ u_1 + \dots + u_d = 1}} \mathbf{u}^\top \widehat{\mathbf{V}} \mathbf{u}$$

and $\widehat{\mathbf{V}}$ is a consistent estimator of the asymptotic covariance matrix \mathbf{V} defined in Theorem 4. Let us highlight that the estimator $\widehat{\gamma}(k_X, k_Y)$ is analysed here under weaker conditions than those of Dematteo and Cl  men  on (2016) and, due to the particular

choice $d = 2$, without having to resort to a numerical optimisation routine for the calculation of the estimator. It is worth noting that since

$$\frac{c - R^2(c, 1)}{1 + c - 2R(c, 1)} - \min(c, 1) = -\frac{[\min(c, 1) - R(c, 1)]^2}{1 + c - 2R(c, 1)} \leq 0,$$

the estimator analysed in Corollary 4(ii) indeed has an asymptotic variance which is lower than each of the asymptotic variances of $\widehat{\gamma}_X(k_X)$ and $\widehat{\gamma}_Y(k_Y)$, and we can quantify this improvement. More precisely, since the function

$$r \mapsto \frac{[\min(c, 1) - r]^2}{1 + c - 2r}$$

is decreasing on $[0, \min(c, 1)]$, the improvement in asymptotic variance brought by the use of the pooled estimator $\widehat{\gamma}(k_X, k_Y)$ gets stronger as the asymptotic dependence structure of (X, Y) gets closer to asymptotic independence. In the case of asymptotic independence, we have

$$\sqrt{k_X} (\widehat{\gamma}(k_X, k_Y) - \gamma) \xrightarrow{d} \mathcal{N}\left(0, \gamma^2 \frac{c}{1 + c}\right).$$

This is rather intuitive: since $\widehat{\gamma}(k_X, k_Y)$ is then essentially the weighted average of two independent quantities, made respectively of k_X independent log-excesses from X (with weighting $c/(1 + c)$) and $k_Y \approx k_X/c$ independent log-excesses from Y (with weighting $1/(1 + c)$), each with individual variance γ^2 by Corollary 1, we should expect the total asymptotic variance to be

$$\gamma^2 \left[\frac{c^2}{(1 + c)^2} + c \frac{1}{(1 + c)^2} \right] = \gamma^2 \frac{c}{1 + c},$$

just as we found through our rigorous asymptotic analysis.

3.2 Convergence of a tail index estimator for right-truncated samples

We now illustrate how our results can be used to complete the analysis of the convergence of a tail index estimator when the available data is subject to random right-truncation. The context we consider is the following: let $(Y_1, T_1), \dots, (Y_n, T_n)$ be n independent copies of a random pair (Y, T) , where Y and T are independent, nonnegative, and have continuous marginal distribution functions F_Y and F_T . Assume also that Y and T have heavy-tailed distributions with tail indices γ_Y and γ_T . In the random right-truncation problem considered here, it is assumed that the pair (Y_i, T_i) is observed if and only if $Y_i \leq T_i$; otherwise, no information on this pair is available at all. The objective is to estimate γ_Y .

This problem has only been considered very recently, starting with Gardes and Stupfler (2015), who were ultimately interested in the estimation of extreme quantiles of Y . Several studies have since then proposed alternative techniques for tail index estimation in this context; we refer to Benchaira et al. (2016a, b), Worms and Worms (2016) and Haouas et al. (2017). The random right-truncation context should not be mistaken for random right-censoring, where the available information is made of the pairs $(\min(Y_i, T_i), \mathbb{1}_{\{Y_i \leq T_i\}})$, $1 \leq i \leq n$. The latter context has received a substantial

amount of attention over the last decade: we refer to Beirlant et al. (2007, 2010, 2016), Einmahl et al. (2008), Gomes and Neves (2011), Ndao et al. (2014), Sayah et al. (2014), Worms and Worms (2014), Brahimy et al. (2015), Ndao et al. (2016), Stupfler (2016), Dierckx et al. (2018) and Stupfler (2019).

Our focus in this section is to revisit the asymptotic properties of the estimator of Gardes and Stupfler (2015) using the tools we have developed in Sections 2 and 3.1. Let us introduce some notation beforehand: let N be the total (random) number of observed pairs (Y_i, T_i) such that $Y_i \leq T_i$. Such pairs shall be denoted as (Y_i^*, T_i^*) , $1 \leq i \leq N$. It is straightforward to show the following:

- N has a binomial distribution with parameters n and $p := \mathbb{P}(Y \leq T)$, where we assume throughout this section that $p > 0$;
- Given N , the pairs (Y_i^*, T_i^*) , $1 \leq i \leq N$ are independent copies of a random pair (Y^*, T^*) having joint distribution function $H^*(y, t) = \mathbb{P}(Y \leq y, T \leq t | Y \leq T)$, and corresponding marginal survival functions

$$\bar{F}_Y^*(y) := \frac{1}{p} \int_y^\infty \bar{F}_T(z) dF_Y(z) \quad \text{and} \quad \bar{F}_T^*(t) := \frac{1}{p} \int_t^\infty F_Y(z) dF_T(z).$$

It is also reasonably easy to show that in this context, \bar{F}_Y^* and \bar{F}_T^* are heavy-tailed, with tail indices $\gamma_Y^* := \gamma_Y \gamma_T / (\gamma_Y + \gamma_T)$ and γ_T (see Lemma 3 in Gardes and Stupfler (2015), as well as the Introduction of Benchaira et al. (2015)). Rewriting this equality as $\gamma_Y = \gamma_Y^* \gamma_T / (\gamma_T - \gamma_Y^*)$ motivates the following estimator of γ_Y :

$$\widehat{\gamma}_Y(k_N, k'_N) := \frac{\widehat{\gamma}_Y^*(k_N) \widehat{\gamma}_T(k'_N)}{\widehat{\gamma}_T(k'_N) - \widehat{\gamma}_Y^*(k_N)}$$

$$\text{with } \widehat{\gamma}_Y^*(k_N) := \frac{1}{k_N} \sum_{i=1}^{k_N} \log \frac{Y_{N-i+1, N}^*}{Y_{N-k_N, N}^*} \quad \text{and} \quad \widehat{\gamma}_T(k'_N) := \frac{1}{k'_N} \sum_{i=1}^{k'_N} \log \frac{T_{N-i+1, N}^*}{T_{N-k'_N, N}^*}.$$

Here (k_m) and (k'_m) are two nonrandom sequences of integers and $(k_N, k'_N) := (k_m, k'_m)$ given $N = m$. Gardes and Stupfler (2015) examine the asymptotic distribution of this estimator, under the condition that $k_m/k'_m \rightarrow 0$ or $k'_m/k_m \rightarrow 0$ as $m \rightarrow \infty$. This technical restriction was imposed because the analysis of the dependence between the two Hill estimators $\widehat{\gamma}_Y^*(k_N)$ and $\widehat{\gamma}_T(k'_N)$ is difficult; assuming that either $k_m/k'_m \rightarrow 0$ or $k'_m/k_m \rightarrow 0$ ensures that one of the estimators converges at a slower rate than the other, and therefore imposes its asymptotic distribution to $\widehat{\gamma}_Y(k_N, k'_N)$. Benchaira et al. (2015) subsequently studied the asymptotic distribution of this estimator when $k_m = k'_m$ and under a condition on the asymptotic dependence between Y^* and T^* . The main result of this section, which we state now, examines the general case $k_m/k'_m \rightarrow c \in [0, \infty]$.

Theorem 5 *Assume that Y^* and T^* respectively satisfy conditions $\mathcal{C}_2(\gamma_Y^*, \rho_Y^*, A_Y^*)$ and $\mathcal{C}_2(\gamma_T, \rho_T^*, A_T^*)$. Let (k_m) and (k'_m) be two sequences of integers tending to infinity such that*

- $\max(k_m, k'_m)/m \rightarrow 0, \sqrt{k_m} A_Y^*(m/k_m) \rightarrow \lambda_1 \in \mathbb{R}$ and $\sqrt{k'_m} A_T^*(m/k'_m) \rightarrow \lambda_2 \in \mathbb{R}$;
- $k_m/k'_m \rightarrow c \in [0, \infty]$.

Then the following hold, as $n \rightarrow \infty$:

(i) If $c = 0$,

$$\sqrt{k_N}(\widehat{\gamma}_Y(k_N, k'_N) - \gamma_Y) \xrightarrow{d} \mathcal{N}\left(\frac{\gamma_Y^2}{\gamma_T^2} \frac{\lambda_1}{1 - \rho_Y^*} \left[1 + \frac{\gamma_T}{\gamma_Y}\right]^2, \frac{\gamma_Y^4}{\gamma_T^2} \left[1 + \frac{\gamma_T}{\gamma_Y}\right]^2\right).$$

(ii) If $c = \infty$,

$$\sqrt{k'_N}(\widehat{\gamma}_Y(k_N, k'_N) - \gamma_Y) \xrightarrow{d} \mathcal{N}\left(-\frac{\gamma_Y^2}{\gamma_T^2} \frac{\lambda_2}{1 - \rho_T^*}, \frac{\gamma_Y^4}{\gamma_T^2}\right).$$

(iii) If $0 < c < \infty$,

$$\begin{aligned} & \sqrt{k_N}(\widehat{\gamma}_Y(k_N, k'_N) - \gamma_Y) \\ & \xrightarrow{d} \mathcal{N}\left(\frac{\gamma_Y^2}{\gamma_T^2} \left\{ \frac{\lambda_1}{1 - \rho_Y^*} \left[1 + \frac{\gamma_T}{\gamma_Y}\right]^2 - \sqrt{c} \frac{\lambda_2}{1 - \rho_T^*} \right\}, \frac{\gamma_Y^4}{\gamma_T^2} \left\{ c + \left[1 + \frac{\gamma_T}{\gamma_Y}\right]^2 \right\}\right). \end{aligned}$$

With convergences (i) and (ii), we essentially find back Theorem 3 in Gardes and Stupfler (2015), which was stated under a slightly different set of second-order extreme value conditions. The value of our result, however, mainly resides in the general convergence result (iii), whose proof, unlike that of Theorem 2.1 in Benchaira et al. (2015), does not hinge on a Gaussian construction for the weighted tail copula process. Most importantly for practical setups, and contrary to Theorem 2.1 in Benchaira et al. (2015), Theorem 5(iii) does not rely on any assumption on the form of asymptotic dependence between Y^* and T^* . The reason for this is that the combination of the independence assumption between Y and T with the heavy-tailed framework is actually sufficient to ensure that Y^* and T^* are asymptotically independent, in the sense that

$$\lim_{t \rightarrow \infty} t \mathbb{P}\left(\overline{F}_Y^*(Y^*) \leq \frac{x_1}{t}, \overline{F}_T^*(T^*) \leq \frac{x_2}{t}\right) = 0$$

for any $(x_1, x_2) \in [0, \infty)^2$ (see the proof of Theorem 5). With this in mind, our result offers a simplified expression of the asymptotic variance of $\widehat{\gamma}_Y(k_N, k'_N)$, compared to the one provided in Benchaira et al. (2015). We also point out that, taking this asymptotic independence result into account, the asymptotic distribution we find in the case $c = 1$ essentially coincides with that of Benchaira et al. (2015), although it should be pointed out that the bias term μ therein should read like a difference rather than a sum (this is revealed by inspecting Equation (3.10) therein) and their variance σ^2 should be divided by 2 (otherwise the Gaussian representation stated early in their Theorem 2.1 would contradict their asymptotic normality result). Finally, our result includes the possibility of taking proportional sequences k_N and k'_N , which is useful since one may obtain better finite-sample performance by selecting a value k'_N different from k_N if the marginal distributions of Y^* and T^* have very different second-order parameters ρ_Y^* and ρ_T^* (a related point about the estimation of a common tail index based on several samples of data is made in the Introduction of Dematteo and Cl  men  on (2016).

3.3 Joint convergence of marginal conditional tail moments

Another consequence of our theoretical results can be formulated in terms of the joint convergence of high marginal Conditional Tail Moments (CTMs). As in Section 3.1, let $(\mathbf{X}_i)_{1 \leq i \leq n}$, with $\mathbf{X}_i = (X_i^{(1)}, X_i^{(2)}, \dots, X_i^{(d)})$, be a sample of independent copies of a random vector $\mathbf{X} = (X^{(1)}, X^{(2)}, \dots, X^{(d)})$ such that each component $X^{(j)}$ has a continuous distribution function F_j and satisfies condition $\mathcal{C}_2(\gamma_j, \rho_j, A_j)$. We let also, for $1 \leq j \leq d, k_j = k_j(n) \rightarrow \infty$ be such that $k_j/n \rightarrow 0$. The CTM of order a_j of the variable $X^{(j)}$ above its high quantile $U_j(n/k_j)$ is then

$$\text{CTM}_{a_j}^{(j)}(U_j(n/k_j)) = \mathbb{E}([X^{(j)}]^{a_j} \mid X^{(j)} > U_j(n/k_j)).$$

With the language of Definition 1, this quantity is exactly the expected f_{a_j} -shortfall of $X^{(j)}$ above the quantile $U_j(n/k_j)$, where $f_{a_j}(x) = x^{a_j}$ for $x > 0$. Following the ideas of Section 2, its empirical counterpart is

$$\widehat{\text{CTM}}_{a_j}^{(j)}(U_j(n/k_j)) = \frac{1}{k_j} \sum_{i=1}^{k_j} [X_{n-i+1,n}^{(j)}]^{a_j}.$$

This estimator is studied in El Methni et al. (2014) and El Methni and Stupfler (2017). The asymptotic results therein provide information on the joint convergence of estimators of several CTMs of a single variable X . Our next result below adopts the point of view of joint convergence of these estimators *across marginals*.

Theorem 6 *Work under the conditions of Theorem 4, with $\sqrt{k_j}A_j(n/k_j) \rightarrow \lambda_j \in \mathbb{R}$ replaced by the weaker assumption $\sqrt{k_j}A_j(n/k_j) = O(1)$. Let also $a_1, a_2, \dots, a_d > 0$ be such that $2a_j\gamma_j < 1$ for any j . Then*

$$\sqrt{k_1} \left(\frac{\widehat{\text{CTM}}_{a_j}^{(j)}(U_j(n/k_j))}{\text{CTM}_{a_j}^{(j)}(U_j(n/k_j))} - 1 \right)_{1 \leq j \leq d} \xrightarrow{d} \mathcal{N}_d(0, \Sigma(\mathbf{a}, \mathbf{c}))$$

with

$$\Sigma_{j,l}(\mathbf{a}, \mathbf{c}) = \begin{cases} c_j \frac{2a_j^2\gamma_j^2(1 - a_j\gamma_j)}{1 - 2a_j\gamma_j} & \text{if } j = l, \\ a_j\gamma_j a_l\gamma_l(1 - a_j\gamma_j)(1 - a_l\gamma_l) \int_0^1 \int_0^1 \frac{R_{j,l}(c_l u, c_j v)}{u^{a_j\gamma_j+1} v^{a_l\gamma_l+1}} du dv & \text{if } j < l. \end{cases}$$

A corollary of Theorem 6 is the following result on the joint convergence of several CTM estimators

$$\widehat{\text{CTM}}_{a_j}(U(n/k)) = \frac{1}{k} \sum_{i=1}^k X_{n-i+1,n}^{a_j}$$

of a *single* heavy-tailed random variable X . This corollary immediately follows from the fact that the random pair (X, X) is asymptotically perfectly dependent, in the sense that

$$\lim_{t \rightarrow \infty} t \mathbb{P} \left(\overline{F}(X) \leq \frac{x_1}{t}, \overline{F}(X) \leq \frac{x_2}{t} \right) = \min(x_1, x_2)$$

for any $(x_1, x_2) \in [0, \infty]^2 \setminus \{(\infty, \infty)\}$.

Corollary 5 *Suppose that X satisfies condition $\mathcal{C}_2(\boldsymbol{\gamma}, \boldsymbol{\rho}, \mathbf{A})$. Let $k = k(n) \rightarrow \infty$ be such that $k/n \rightarrow 0$ and $\sqrt{k}A(n/k) = O(1)$. Let also $a_1, a_2, \dots, a_d > 0$ be such that $\gamma < 1/2a_j$ for any j . Then*

$$\sqrt{k} \left(\frac{\widehat{\text{CTM}}_{a_j}(U(n/k))}{\text{CTM}_{a_j}(U(n/k))} - 1 \right)_{1 \leq j \leq d} \xrightarrow{d} \mathcal{N}_d(0, \mathbf{M}(\mathbf{a}))$$

with

$$\mathbf{M}_{j,l}(\mathbf{a}) = \frac{a_j a_l \gamma^2 (2 - [a_j + a_l] \gamma)}{1 - (a_j + a_l) \gamma}.$$

This result complements Theorem 2 in El Methni and Stupfler (2017), for distortion functions all equal to the identity function. A related result, in the presence of a finite-dimensional covariate, is Theorem 1 in El Methni et al. (2014).

While Theorem 6 is informative regarding the structure of the dependence between CTM estimators at high levels, its scope may however be limited from the practical point of view, as the focus in applied setups is usually on risk measures at out-of-sample levels. To put it differently, one would generally want to estimate a CTM of the form

$$\text{CTM}_{a_j}^{(j)}(U_j(1/p_n)) = \mathbb{E}([X^{(j)}]^{a_j} | X^{(j)} > U_j(1/p_n))$$

where $p_n \rightarrow 0$ is such that $np_n = O(1)$, a typical choice then being $p_n = 1/n$ (see e.g. recently Cai et al. (2015) and Gong et al. (2015)). In this context, following El Methni et al. (2014) and El Methni and Stupfler (2017), one may replace the estimator $\widehat{\text{CTM}}_{a_j}^{(j)}$ by an extrapolated Weissman-type version:

$$\overline{\text{CTM}}_{a_j}^{(j)}(U_j(1/p_n)) := \left(\frac{k_j}{np_n} \right)^{a_j \widehat{\gamma}_j(k_j)} \widehat{\text{CTM}}_{a_j}^{(j)}(U_j(n/k_j)).$$

Here $\widehat{\gamma}_j(k_j)$ is the j th marginal Hill estimator introduced in Section 3.1. A combination of Theorems 4 and 6 makes it possible to obtain the following joint asymptotic normality result for these extrapolated estimators across marginals.

Corollary 6 *Work under the conditions of Theorem 4. Let $a_1, a_2, \dots, a_d > 0$ be such that $2a_j \gamma_j < 1$ for any j . Assume also that $\rho_j < 0$ for any j , and that $np_n \rightarrow C < \infty$ and $\sqrt{k_1}/\log(k_1/[np_n]) \rightarrow \infty$. Then*

$$\sqrt{k_1} \left(\frac{1}{a_j \log(k_j/[np_n])} \left[\frac{\overline{\text{CTM}}_{a_j}^{(j)}(U_j(1/p_n))}{\text{CTM}_{a_j}^{(j)}(U_j(1/p_n))} - 1 \right] \right)_{1 \leq j \leq d} \xrightarrow{d} \mathcal{N}_d(\mathbf{b}(\mathbf{c}), \mathbf{V}(\mathbf{c}))$$

with the notation of Theorem 4.

In the case $a_1 = a_2 = \dots = a_d$, such a result may then be used to test the equality of CTMs across marginals. For instance, if $d = 2$ and $a_1 = a_2 = 1$, one may test whether the two variables $X^{(1)}$ and $X^{(2)}$ have the same Expected Shortfall at high levels. The possibility of assessing the equality of financial tail risk within a multivariate context using such results is explored in Sections 3 and 4 in Hoga (2018); a related result is Proposition 3 in Hoga (2018), albeit for a different type of estimator of the Expected Shortfall, and in a time series context.

4 Conclusion and discussion

Writing empirical average f -excesses and empirical expected f -shortfalls in terms of their pseudo-estimator counterparts with non-random threshold, as in Theorems 1 and 2, makes it possible to obtain joint convergence theorems for certain extreme value estimators under weak assumptions, and with conceptually simple proofs. The results of this paper, however, are limited to independent and identically distributed random variables. As Hoga (2018) argues, assessing tail homogeneity of the marginals of a random vector is of particular interest in certain financial applications, where it is important to develop asymptotic theory for dependent but stationary sequences. Since the cornerstone of the proofs of Theorems 1 and 2 is a weighted Gaussian approximation of the univariate tail empirical process, it is reasonable to expect that analogues of these two results can also be shown in this kind of framework: for instance, Drees (2003) proves such an approximation in the case of β -mixing sequences under certain regularity conditions and upper bounds on the size of clusters of exceedances. This would therefore result in analogues of Theorems 1 and 2 and ultimately of Theorems 4 and 6. We note though that the expression of the asymptotic covariance matrices in such analogues of Theorems 4 and 6 would certainly be more complicated than in the present paper, which, among others, makes the derivation of a testing procedure of tail homogeneity a difficult task. Hoga (2018) avoids this issue by resorting to a self-normalised test statistic in the spirit of Shao (2010), although the limiting distribution does not have a simple expression. An interesting open question is to, based on an analogue of Theorem 4 in a dependent and stationary setting, construct an asymptotically chi-squared test statistic of tail homogeneity and contrast its finite-sample performance with that of the test statistic of Hoga (2018).

It would also be potentially very fruitful to extend the present results to the regression case, in order to allow a general theoretical analysis of conditional, randomly thresholded tail index estimators based on kernel smoothing. There has been a significant body of work in the area of conditional extreme value analysis over the last decade, among which we can refer to Daouia et al. (2011, 2013), Goegebeur et al. (2014, 2015). However, the estimators introduced in the first two of these papers only take a fixed number of order statistics into account, and the theoretical results of the latter two only apply to a non-randomly thresholded conditional version of the Hill estimator. By contrast, the estimators of Stupfler (2013) and Gardes and Stupfler (2014) are built on non-random thresholds and their asymptotic normality is studied, but they do not include the possibility of kernel smoothing. The explanation for

this is that kernel smoothing adds dependent and non-stationary weightings to the log-spacings appearing in the randomly thresholded estimator, making its theoretical analysis a very difficult mathematical task. Developing an extension of the present theoretical results to the regression case would allow to bridge the gap between the aforementioned studies, and thus provide stronger theoretical ground for conditional extreme value analysis based on kernel smoothing.

Supplementary Material A Supplementary Material document contains all necessary proofs of the results of this paper.

Acknowledgments The author acknowledges an anonymous Associate Editor and two anonymous reviewers for their helpful comments that led to a substantially improved version of this paper, and in particular to the writing of Theorem 3 and Corollary 3. The author also gratefully acknowledges Laurens de Haan for a couple of very instructive conversations about tail empirical processes which led him to write 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.

References

- Beirlant, J., Bardoutsos, A., de Wet, T., Gijbels, I.: Bias reduced tail estimation for censored Pareto type distributions. *Statist. Probab. Lett.* **109**, 78–88 (2016)
- Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J.: *Statistics of extremes: Theory and applications*. Chichester, Wiley (2004)
- 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. *Communications in Statistics – Theory and Methods* **39**, 1158–1179 (2010)
- Benchaira, S., Meraghni, D., Necir, A.: On the asymptotic normality of the extreme value index for right-truncated data. *Statist. Probab. Lett.* **107**, 378–384 (2015)
- Benchaira, S., Meraghni, D., Necir, A.: Tail product-limit process for truncated data with application to extreme value index estimation. *Extremes* **19**, 219–251 (2016a)
- Benchaira, S., Meraghni, D., Necir, A.: Kernel estimation of the tail index of a right-truncated Pareto-type distribution. *Statist. Probab. Lett.* **119**, 186–193 (2016b)
- Brahimi, B., Meraghni, D., Necir, A.: Gaussian approximation to the extreme value index estimator of a heavy-tailed distribution under random censoring. *Math. Methods Statist.* **24**, 266–279 (2015)
- Brazauskas, V., Jones, B.L., Puri, M.L., Zitikis, R.: Estimating conditional tail expectation with actuarial applications in view. *Journal of Statistical Planning and Inference* **138**, 3590–3604 (2008)
- Cai, J.-J., Einmahl, J.H.J., de Haan, L., Zhou, C.: Estimation of the marginal expected shortfall: the mean when a related variable is extreme. *J. R. Stat. Soc. Ser. B* **77**, 417–442 (2015)
- Chen, Z.: Conditional L_p -quantiles and their application to testing of symmetry in non-parametric regression. *Statist. Probab. Lett.* **29**, 107–115 (1996)
- Daouia, A., Gardes, L., Girard, S., Lekina, A.: Kernel estimators of extreme level curves. *Test* **20**, 311–333 (2011)
- Daouia, A., Gardes, L., Girard, S.: On kernel smoothing for extremal quantile regression. *Bernoulli* **19**, 2557–2589 (2013)
- Daouia, A., Gijbels, I., Stupfler, G.: Extremiles: A new perspective on asymmetric least squares, to appear in *Journal of the American Statistical Association* (2018a)
- Daouia, A., Girard, S., Stupfler, G.: Estimation of tail risk based on extreme expectiles. *J. R. Stat. Soc. Ser. B* **80**, 263–292 (2018b)

- Daouia, A., Girard, S., Stupfler, G.: Extreme M-quantiles as risk measures: From L^1 to L^p optimization. *Bernoulli* **25**, 264–309 (2019)
- de Haan, L., Ferreira, A.: *Extreme value theory: an introduction*. Springer, New York (2006)
- Dematteo, A., Cléménçon, S.: On tail index estimation based on multivariate data. *Journal of Nonparametric Statistics* **28**, 152–176 (2016)
- Dierckx, G., Goegebeur, Y., Guillou, A.: Local robust estimation of Pareto-type tails with random right censoring, preprint hal-01829618 (2018)
- Drees, H., Huang, X.: Best attainable rates of convergence for estimators of the stable tail dependence function. *J. Multivar. Anal.* **64**, 25–47 (1998)
- Drees, H.: Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli* **9**, 617–657 (2003)
- Einhahl, J.H.J., Fils-Villetard, A., Guillou, A.: Statistics of extremes under random censoring. *Bernoulli* **14**, 207–227 (2008)
- El Methni, J., Gardes, L., Girard, S.: Nonparametric estimation of extreme risks from conditional heavy-tailed distributions. *Scand. J. Stat.* **41**, 988–1012 (2014)
- El Methni, J., Stupfler, G.: Extreme versions of Wang risk measures and their estimation for heavy-tailed distributions. *Stat. Sin.* **27**, 907–930 (2017)
- El Methni, J., Stupfler, G.: Improved estimators of extreme Wang distortion risk measures for very heavy-tailed distributions. *Econ. Stat.* **6**, 129–148 (2018)
- Embrechts, P., Klüppelberg, C., Mikosch, T.: *Modelling extremal events for insurance and finance*. Springer, Berlin-Heidelberg (1997)
- Emmer, S., Kratz, M., Tasche, D.: What is the best risk measure in practice? A comparison of standard measures. *J. Risk* **18**, 31–60 (2015)
- Gardes, L., Stupfler, G.: Estimation of the conditional tail index using a smoothed local Hill estimator. *Extremes* **17**, 45–75 (2014)
- Gardes, L., Stupfler, G.: Estimating extreme quantiles under random truncation. *TEST* **24**, 207–227 (2015)
- Goegebeur, Y., Guillou, A., Schorgen, A.: Nonparametric regression estimation of conditional tails: the random covariate case. *Statistics* **48**, 732–755 (2014)
- Goegebeur, Y., Guillou, A., Stupfler, G.: Uniform asymptotic properties of a nonparametric regression estimator of conditional tails. *Annales de l’Institut Henri Poincaré (B): Probability and Statistics* **51**, 1190–1213 (2015)
- Goldie, C.M., Smith, R.L.: Slow variation with remainder: theory and applications. *Q. J. Math.* **38**, 45–71 (1987)
- Gomes, M.I., Neves, M.M.: Estimation of the extreme value index for randomly censored data. *Biometrical Lett.* **48**, 1–22 (2011)
- Gong, J., Li, Y., Peng, L., Yao, Q.: Estimation of extreme quantiles for functions of dependent random variables. *J. R. Stat. Soc. Ser. B* **77**, 1001–1024 (2015)
- Haouas, N., Necir, A., Meraghni, D., Brahimi, B.: A Lynden-Bell integral estimator for the tail index of right-truncated data with a random threshold. *Afr. Stat.* **12**, 1159–1170 (2017)
- Hill, B.M.: A simple general approach to inference about the tail of a distribution. *Ann. Stat.* **3**, 1163–1174 (1975)
- Hoga, Y.: Detecting tail risk differences in multivariate time series. *J. Time Ser. Anal.* **39**, 665–689 (2018)
- Jiang, Y., Sun, H., Zhang, Y., Long, H.: Non-parametric tests for the tail equivalence via empirical likelihood. *Communications in Statistics – Theory and Methods* **46**, 10640–10656 (2017)
- Kinsvater, P., Fried, R., Lilienthal, J.: Regional extreme value index estimation and a test of tail homogeneity. *Environmetrics* **27**, 103–115 (2016)
- 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. *Journal of Statistical Planning and Inference* **168**, 20–37 (2016)
- Resnick, S.: *Extreme values, regular variation and point processes*. Springer, New York (1987)
- Resnick, S.: *Heavy-tail phenomena: Probabilistic and statistical modeling*. Springer, New York (2007)
- Sayah, A., Yahia, D., Brahimi, B.: On robust tail index estimation under random censorship. *Afr. Stat.* **9**, 671–683 (2014)
- Schmidt, R., Stadtmüller, U.: Non-parametric estimation of tail dependence. *Scand. J. Stat.* **33**, 307–335 (2006)

- Shao, X.: A self-normalized approach to confidence interval construction in time series. *J. R. Stat. Soc. Ser. B* **72**, 343–366 (2010)
- Stupfler, G.: A moment estimator for the conditional extreme-value index. *Electronic Journal of Statistics* **7**, 2298–2343 (2013)
- Stupfler, G.: Estimating the conditional extreme-value index under random right-censoring. *J. Multivar. Anal.* **144**, 1–24 (2016)
- Stupfler, G.: On the study of extremes with dependent random right-censoring. *Extremes* **22**, 97–129 (2019)
- Wang, S.S.: Premium calculation by transforming the layer premium density. *ASTIN Bulletin* **26**, 71–92 (1996)
- Weissman, I.: Estimation of parameters and large quantiles based on the k largest observations. *J. Am. Stat. Assoc.* **73**, 812–815 (1978)
- Worms, J., Worms, R.: New estimators of the extreme value index under random right censoring, for heavy-tailed distributions. *Extremes* **17**, 337–358 (2014)
- Worms, J., Worms, R.: A Lynden-Bell integral estimator for extremes of randomly truncated data. *Statist. Probab. Lett.* **109**, 106–117 (2016)
- Wüthrich, M., Merz, M.: *Financial modeling, actuarial valuation and solvency in insurance*. Springer, Berlin-Heidelberg (2013)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.