University of Reading

Reduced-bias estimation of the residual dependence index with unspecified marginals

Article

Published Version

Israelsson, J. ORCID: https://orcid.org/0000-0001-6181-1702, Black, E. ORCID: https://orcid.org/0000-0003-1344-6186, Neves, C. ORCID: https://orcid.org/0000-0003-1201-5720 and Walshaw, D. (2026) Reduced-bias estimation of the residual dependence index with unspecified marginals. Electronic journal of statistics, 20 (2). pp. 2853-2891. ISSN 1935-7524 doi: 10.1214/26-EJS2548 Available at https://centaur.reading.ac.uk/131553/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1214/26-EJS2548

Publisher: Institute of Mathematical Statistics

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

University of Reading

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Electronic Journal of Statistics
Vol. 20 (2026) 2853–2891
ISSN: 1935-7524
https://doi.org/10.1214/26-EJS2548

Reduced-bias estimation of the residual dependence index with unspecified marginals

Jennifer Israelsson[1], **Emily Black**[2], **Cláudia Neves**[3] **and David Walshaw**[4]

[1]*Institute for Global Health, University College London, United Kingdom,*
e-mail: j.israelsson@ucl.ac.uk

[2]*Department of Meteorology, University of Reading, United Kingdom,*
e-mail: e.c.l.black@reading.ac.uk

[3]*Department of Mathematics, King's College London, United Kingdom,*
e-mail: claudia.neves@kcl.ac.uk

[4]*School of Mathematics & Statistics, University of Newcastle, United Kingdom,*
e-mail: david.walshaw@newcastle.ac.uk

Abstract: This paper addresses important weaknesses in current methodology for the estimation of multivariate extreme event distributions. The estimation of the residual dependence index $\eta \in (0, 1]$ is notoriously problematic. This motivates us to introduce a flexible class of gradient estimators for this parameter, designed to ameliorate the usual problems of threshold selection while furnishing a unifying approach to the usual marginal transforms. The efficacy of this semi-parametric gradient estimation of η originates from a hitherto neglected exponentially decaying term in the hidden regular variation characterisation. Numerical simulations aimed at assessing the performance for finite samples over a range of familiar copulas indicate an improved performance relatively to the existing estimators, notably the Hill's. Paired with a reduced-bias procedure that we also construct, the gradient estimation offers much desired statistical guarantees as it further pinpoints the optimal number of intermediate data points to enter the estimation (in terms of minimising asymptotic mean squared error) in an automatic and demonstrable way. Our leading application illustrates how asymptotic independence can be discerned from monsoon-related rainfall occurrences at different locations in Ghana. The considerations involved in extending this framework to inference on the extreme value index drawing on max-domains of attraction are briefly discussed.

MSC2020 subject classifications: Primary 60F17, 62G32; secondary 62A99, 62P12.
Keywords and phrases: Asymptotic independence, Brownian motion, empirical processes, extreme value theory, monsoon rainfall, regular variation, tail dependence coefficient.

Received October 2025.

1. Introduction

In recent years, statistical methodology focusing on multivariate extremes has garnered ever greater interest brought by the many challenges faced when attempting to model compound to nearly simultaneous extreme events. For example, according to [34], asymptotic independence must be of conjecture when modelling extreme rainfall, especially given the convective nature of extreme tropical precipitation. Although this example focuses on extremal behaviour, having a good grasp of the strength of dependence at play amongst the whole of the data

arXiv: 2112.02899

is crucially important in applied science at large. Quantifying dependence is commonplace in statistics. Wherever one sits on the sliding scale of data complexity and dimensionality – ranging from the bivariate setting to the infinite-dimensional case relating stochastic processes and to multivariate stochastic processes – accounting for an estimator's standard error is often required as much as it is unavoidable. And more so, accounting for uncertainty associated with model approximation is crucial to obtaining feasible confidence intervals that can ward off overly optimistic inference, especially if residual dependence is overlooked [see e.g. 16, 36].

The importance of quantifying the strength of any remaining shred of association between two or more random variables within the asymptotic independence realm has long been recognised. But it was not until the compelling case made by [37] that the concept of residual dependence really became central to extreme value statistics [see 26, 35]. The example put forward in [37] – that the componentwise maxima of a bivariate normal distribution with correlation coefficient satisfying $|\rho| < 1$ are asymptotically independent – spurred on many important developments in the statistical modelling of extreme events. It provided a compelling argument for devoting efforts to discerning between exact independence, asymptotic independence and dependence [13, 9], and it has seen greater resurgence in the development of modern statistical methods with particular focus on smooth transitions from dependence to asymptotic independence [23].

The overarching goal of the present paper is to capture penultimate dependence among extreme values, within which context we tackle the problem of reduced-bias estimation for the so-called residual dependence index. This index, also termed the coefficient of tail dependence, was introduced by [26] and further investigated by [32], [20] and [14].

To fix the idea, let (X_i, Y_i), $i = 1, ..., n$, be independent copies of the random vector (X, Y) with joint distribution function F, whose marginal distributions we denote by F_1 and F_2, i.e. $F_1(x) := F(x, \infty)$ and $F_2(y) := F(\infty, y)$, assumed continuous. The interest is in evaluating the probability of the two components being simultaneously large, more formally $P(X > u_1, Y > u_2)$, for u_1, u_2 usually much larger than the respective sample realisations for the componentwise maxima $M_{X,n} := \max(X_1, \ldots, X_n)$, $M_{Y,n} := \max(Y_1, \ldots, Y_n)$. From the applied illustrative perspective, this paper directs its focus to random phenomena that exhibit positively quadrant residual dependence [see e.g. 40], meaning that the probability that two random variables are simultaneously large is greater than that if they were independent. In environmental science, this can apply for example to the joint modelling of strong wind-speeds and extreme wave heights; or to two large events of the same physical process occurring at nearby locations, for example heavy rainfall at two neighbouring cities. In the financial context, different variables (e.g. different financial instruments) may face common shocks that make them to correlate positively, often leading to overly optimistic inference when dependence is ignored, as highlighted in [18]. Because we are solely interested in larger values, extreme value theory for threshold exceedances is the appropriate theory to rely upon.

1.1. Asymptotic independence setting

Unlike univariate extreme value theory, multivariate extreme value distributions cannot be specified through a finite-dimensional parameter family of distributions. Instead, the key features of multivariate extremes are mirrored in the dependence structure of componentwise maxima, which must be dissociated from the limiting extreme behaviour of its marginal

distribution functions before a proper characterisation of extreme domains of attractions can be determined.

We assume that the independent and identically distributed pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ have common joint distribution function F belonging to the domain of attraction of a bivariate extreme value distribution G. Formally, this entails that there exist constants $a_n, c_n > 0$ and $b_n, d_n \in \mathbb{R}$ such that, for all continuity points (x, y) of G the following limit relation for the linearly normalised component-wise partial maxima, $M_{X,n}$ and $M_{Y,n}$, is satisfied:

$$\lim_{n\to\infty} P\Big\{\frac{M_{X,n} - b_n}{a_n} \le x, \frac{M_{Y,n} - d_n}{c_n} \le y\Big\} = \lim_{n\to\infty} F^n(a_n + b_n x,\ c_n + d_n y) = G(x, y). \quad (1)$$

We shall condense the above in the notation $F \in \mathcal{D}(G)$ throughout. The domain of attraction condition (1) implies convergence of the marginal distributions to the corresponding $G(x, \infty) =: G_1(x)$ and $G(\infty, y) =: G_2(y)$, whereby it is possible to redefine the constants a_n, c_n, b_n, d_n in such a way that the generalised extreme value (GEV) marginal distributions $G_1(x) = \exp\{-(1 + \gamma_1 x)^{-1/\gamma_1}\}$ and $G_2(y) = \exp\{-(1 + \gamma_2 y)^{-1/\gamma_2}\}$ are attained on each component, with ascribed extreme value indices $\gamma_1, \gamma_2 \in \mathbb{R}$. This characterisation grants a semi-parametric approach to statistical inference for extreme events.

Next, let G_0 define the class of simple max-stable distributions, i.e., the class of distribution functions arising in the limit (1) given by

$$G_0(x, y) := G\Big(\frac{x^{\gamma_1} - 1}{\gamma_1}, \frac{x^{\gamma_2} - 1}{\gamma_2}\Big),$$

signifying that both marginal distributions are fixed as unit Fréchet: $G_0(x, \infty) = G_0(\infty, x) = \exp(-1/x)$, for $x > 0$. With the marginal quantile functions $V_i := \big(1/-\log F_i\big)^{\leftarrow}$, $i = 1, 2$, where $^{\leftarrow}$ indicates the generalised inverse function, we say that $F \in \mathcal{D}(G)$ if and only if

$$\lim_{t\to\infty} t g(tx, ty) = -\log G_0(x, y), \quad (2)$$

where $g(x, y) := -\log F(V_1(x - 1/2), V_2(y - 1/2))$, and F_1 and F_2 is in the domain of attraction of G_1 and G_2 with respective extreme value indices $\gamma_1, \gamma_2 \in \mathbb{R}$.

The extreme value condition (2) implies that $1 - F\big(V_1(t - 1/2), V_2(t - 1/2)\big)$ is regularly varying at infinity with index -1 (see details in the Appendix). A positive function h is of regular variation with index α, i.e. $h \in RV_\alpha$, if $\lim_{t\to\infty} h(tx)/h(t) = x^\alpha$, for all $x > 0$. Relation (2) also emphasises the key role of curve $s := \{V_1(t - 1/2), V_2(t - 1/2)\}$ to capturing extremal dependence that manifests at higher levels of F. In the Appendix, s is proved asymptotically equivalent to the curve $\{(b(t), d(t)\}$ borne out of (1). Other marginal standardisation transforms amenable to different definitions of s, supporting equivalent extreme value characterisations for tractable tail dependence measures, are of course possible. A good catalogue can be found in [4, Sections 8.2-8.3]. For example, Laplace marginal transforms are known to simplify the regression structure compared to other transformations [38].

Departing from and yet complementing the results in [30, 13] [see also 19, Sections 6.2 and 7.5], we proceed with the domain of attraction condition that follows directly from (2) via Taylor's expansion of $-\log F$ about $F \to 1$, i.e.,

$$\lim_{t\to\infty} t P\big\{X > V_1(t/x - 1/2), Y > V_2(t/y - 1/2)\big\} = x + y - L(x, y) = R(x, y). \quad (3)$$

The function L is the so-called stable dependence function and the R-function, also thrown out by the limit (3), is commonly referred to as the tail copula [see 19, p.225]. Although

assuming dependence is paramount to multivariate distributions that deal with more than one variable at the time, such an assumption can be questionable or even false at extreme values. Therefore, a clear understanding of the residual dependence regime held above the curve s by concomitant component-wise large values of (X, Y) is desirable.

The earliest and still central measure for the only possible positively quadrant dependence in bivariate extreme value distributions is the dependence coefficient λ introduced by Sibuya [37]; it results from (3) evaluated at $x = y = 1$:

$$\lambda := \lim_{t\to\infty} tP\{X > V_1(t - 1/2), Y > V_2(t - 1/2)\} = 2 - L(1, 1) = R(1, 1). \tag{4}$$

The boundary cases are $\lambda = 1$, implying full positive dependence of X and Y, whereas $\lambda = 0$ is ascribed to asymptotic independence between X and Y. Momentous works by [26, 27] [see also 19, Chapter 7] address the asymptotically independent case for extremes that arises from (1) if $\lim_{n\to\infty} P\{X > a_n x + b_n, Y > c_n y + d_n\}/(1 - F(b_n, d_n)) = 0$. Equivalently, the case of $\lambda = 0$ induces the following sub-model of (3) (and hence of (1)) that offers the characterisation pursued in this paper: X and Y are asymptotically independent if, for all $x, y > 0$, with a positive function α,

$$\lim_{t\to\infty} \frac{tP\{X > V_1(t/x - 1/2), Y > V_2(t/y - 1/2)\}}{t\alpha(t)} =: c(x, y), \tag{5}$$

exists and is positive. Without loss of generality, it is possible to take $c(1, 1) = 1$, whereby the function α, is found regularly varying at infinity with index $-1/\eta$ for some *residual dependence index* $\eta \in (0, 1]$. We define α as

$$\alpha(t) := P\{X > V_1(t-1/2), Y > V_2(t-1/2)\} = P\Big\{\frac{1}{-\log F_1(X)}+\frac{1}{2} > t, \frac{1}{-\log F_2(Y)}+\frac{1}{2} > t\Big\}. \tag{6}$$

Hence, the regular variation of α makes it possible to specify the rate of decay to zero of the linearly normalised joint exceedances' probability featuring in (3), under which characterisation $L(x, y) = x + y$. This concept of asymptotic independence, addressed in detail as part of the next section, is key to this paper.

In what follows, we are going to exploit the fact that the tail distribution of the corrected Fréchet marginals of (6), henceforth denoted by Z_i, $i = 1, 2$, delivers a best possible approximation to the standard Pareto-tail for large enough t:

$$\overline{F}_{Z_i}(t) = 1 - \exp\{-(t - 1/2)^{-1}\} = t^{-1}\Big(1 - \frac{t^{-2}}{12} + o(t^{-2})\Big), \quad i = 1, 2. \tag{7}$$

Noticeably with a zero (non-existing) second-order term in the above Taylor's expansion, the stage is set for unifying marginal transforms. Since $\alpha(t) = P\{Z_1 > t, Z_2 > t\} \le P\{Z_1 > t\} \sim t^{-1}$ and this approximation is sharp, it implies that $\limsup t\alpha(t) < 1$; owing to (3) with $R(x, y) \equiv 0$ (sustaining the asymptotic independence case of $\lambda = 0$), necessarily $\eta \le 1$. While these conform to analogous results by [13], they also bring to the fore how the type of slow variation dictated by $\mathcal{L}(t) = t^{1/\eta}\alpha(t)$ influences the property of asymptotic independence. Compounded with an index η near 1, a possibly non-constant $\mathcal{L}(t) = o(1)$ can single-handedly take hold of the overall convergence towards $\lambda = 0$, thus becoming the primary determinant of the asymptotic independence attached to $\eta = 1$. It is important to hold on to this idea because of the non-negligible differences that will provably manifest through different marginal transforms in the estimation of the *residual dependence index* η.

In a nutshell, $\eta \in (0, 1)$ implies but is not equivalent to asymptotic independence. If for $\eta \in (0, 1)$ the extremes of X and Y are asymptotically independent, a residual dependence index $\eta = 1$ instead does not imply asymptotic dependence [cf. 19, Section 7.6]. The equivalence can nonetheless be guaranteed at the expense of imposing an extra second order condition which would essentially state that $t^{1/\eta} P\{X > V_1(t/x - 1/2), Y > V_2(t/y - 1/2)\}$, featuring in the primary condition (5), is in a certain sense proportional to $t^{1/\eta}$ [cf. 30, Eq.2.1]. We proceed in the wider scope of the non-equivalence by assuming that $\lambda = 0$. However, we will purposefully leverage on a slower (i.e., worse) convergence rate than that entitled by the widely accepted second order conditions in [13, 19] so as to seize estimation advantages.

1.2. Our contribution

There are numerous joint distribution functions F possessing the asymptotic independence property, in particular such that (4) holds with $\lambda = 0$. More often than not, a tail dependence regime arises seemingly in complete disconnection from the dependence structure prevalent at moderate levels, of which a salient example is the bivariate normal distribution with correlation coefficient $|\rho| < 1$ [see e.g. 26, 30, 33]. Despite the wealth of literature published to date on the estimation of the residual dependence index η, this paper adds substantively to the current body of knowledge mainly in two ways:

(i) we introduce a class of reduced-bias gradient estimators for the residual dependence index $\eta \in (0, 1]$ that staves off the problem of (random) threshold selection. The optimal number of tail-related bivariate data points entering the estimation is automatically obtained at the region where the unfurling gradient eventually subsides to a single beam of estimates' trajectories tightly wrapped around the true value η. This provides a practical solution to the problem of threshold selection, and one with a theoretical foundation: both consistency and asymptotic normality of the gradient estimators are obtained under a new set of regularity conditions at the confluence of those typically used in the characterisation of the max-domains and those specialised in peaks-over-threshold (pot) exceedances; thus in a reconciliatory stance rather than having to adjudicate between the two incumbent data transforms from the literature, both Pareto and Fréchet [see 3, 7, 6, the latter for max and pot approaches in comparison].

(ii) We devise an all-encompassing Hall-Welsh-type class of models [cf. 21] of which the canonical distribution is the unit Fréchet equipped with a shift by $1/2$. It turns out that such corrected Fréchet distribution is optimal for instancing the Pareto-type limit characterised by η. The new Hall-Welsh-type representation is key to the novel strand of development that runs through this paper. While bringing together classical sets of conditions for regularly varying tail quantile functions (either suited to larger block-maxima or peaks-over-threshold), at which intersection reduced-bias estimation is then devised, our results can attend to broader applications where the estimation of the index of regular variation is required with enhanced efficacy. The proposed Hall-Welsh characterisation will likely forge a wider framework for achieving significant bias reduction, especially when handling familiar semi-parametric estimators of η [cf. 5, 13], most prominently the Hill estimator [22].

Finally, our theoretical results are mirrored in the numerical experiments implemented as part of a comprehensive simulation study. In particular, the proposed Hall-Welsh-type class that underpins the reduced-bias estimation of $\eta \in (0, 1]$ is demonstrated successful in

averting common pitfalls such as that highlighted in [17, item (i), Section 4.3], that "The Hill estimator is generally biased, though the bias seems to be a more severe problem for unit Fréchet marginal distributions than for unit Pareto marginal distributions". Our findings are consistent with [10, p.152], where it is found that the unit Fréchet is the one marginal's transform conducive to admissible estimates for the extremal dependence coefficient.

1.3. Structure of the paper

The remainder of this paper is organised as follows. In Section 2 we introduce the class of gradient estimators for $\eta \in (0, 1]$ at the core of this paper, with accompanying large-sample properties of consistency and asymptotic normality. In Section 3, the latter distributional expansion culminates in a bespoke bias correction procedure with view to the asymptotically optimal estimation of the residual dependence index η. All the theoretical results developed to this endeavour, starting from a basic invariance distribution-free principle in terms of empirical processes, are collated in Section 4. The asymptotic normality of the proposed reduced-bias estimators is proved on the basis of a bespoke tail empirical process whose asymptotic representation is of Donsker type [25]. Upon inversion, the resulting uniform approximation for the quantile empirical process obtained in this paper conforms to analogous results established in [13, 17] and discussed in [4, 19]. Reduced-bias estimators from Section 3 are devised in such a way as to exploit the convergence rate in the new Hall-Welsh class of models alluded to in Section 1.2 for ameliorating the inevitable discrepancies between the actual residual dependence held in the data and the level of association reflected by the limit in the characterisation (5). Proofs of the key theorems from Sections 2 and 3, as well as those of their ancillary results in Section 4, are deferred to Section 7. A number of these proofs is non-trivial, especially in the wider extreme values context.

Section 5 is devoted to the application of the proposed gradient estimators to monsoon rainfall data collected from a highly dense network of gauging stations in Ghana, furthering prior results in [24]. Section 6 contains simulation results aimed at evaluating finite-sample performance of representative estimators of the new class of gradient estimators with emphasis on their reduced-bias variants, where the automatic threshold selection is in effect. The numerical experiments in Section 6 have also been devised in such as way as to inform the decision making as part of Section 5's tropical rainfall study.

2. Residual dependence index: Pareto meets Fréchet

When assessing asymptotic independence, a few distributions have been used to perform marginal standardisations that allow to leave aside the modelling of true marginal distributions and proceed with the intended focus on a certain measure for extremal dependence. However, it is rarely the case that such estimation procedures take into account the uncertainty brought by transforming the data into the prescribed common distribution for the marginals. The tail dependence coefficient η, here momentarily denominated in the nomenclature of [26] for ease of reference, originated from the finding that the probability of joint exceedance of a common threshold $t > 0$ by random pair $(E_1, E_2) := (-1/\log F_1(X), -1/\log F_2(Y))$ having unit Fréchet marginals can be expressed as $P\{E_1 > t,\, E_2 > t\} = \mathcal{L}(t)\big(P\{E_1 > t\}\big)^{1/\eta} \sim t^{-1/\eta}$, as $t \to \infty$ [cf. 41, p.2094]. Obviously, the sharper the approximation of the joint exceedances'

probability to the canonical $t^{-1/\eta}$-regularly varying limit, the more accurate the estimation of the residual dependence coefficient as draws on smaller approximation bias. The purpose of this section is to present some meaningful derivations for two popular marginal transforms, both standard Pareto and unit Fréchet, that will serve to elicit conditions under which the two are unified. These conditions, rooted in the theory of regular variation, are aimed at harnessing the convergence rate in (5), which will be pivotal to the bias reduction tackled in this paper.

We assume that $F \in \mathcal{D}(G)$ satisfies the limiting relation (3). Since our aim is to assess the degree of residual dependence at an asymptotic level in the realm of asymptotic independence, we follow on from the sub-domain (5) and cast focus on the ensuing fact that α is defined as in (6), while noting that the limiting function c is homogeneous of order $1/\eta$. In order to proceed, we set $x = y$ hence readily obtaining the condition of regular variation:

$$\lim_{t\to\infty} \frac{P\Big\{\big(\frac{1}{-\log F_1(X)} + \frac{1}{2}\big) \wedge \big(\frac{1}{-\log F_2(Y)} + \frac{1}{2}\big) > \frac{t}{x}\Big\}}{P\Big\{\big(\frac{1}{-\log F_1(X)} + \frac{1}{2}\big) \wedge \big(\frac{1}{-\log F_2(Y)} + \frac{1}{2}\big) > t\Big\}} = c(x,x) = x^{1/\eta} c(1,1) = x^{1/\eta}\,, \qquad (8)$$

for all $x > 0$. Hence, the index $-1/\eta$ can be regarded as the tail index of the minimum between two components with common unit Fréchet marginal distributions equipped with a location-correction by $1/2$. The fact that the minimum of independent Pareto random variables is again a Pareto random variable, exhibiting an exact power-law coherent with the limit (8), makes it the obvious canonical form for marginal's standardisation. In particular, the case of the maximal depth ascribed to exact independence produces $tP\{1/(1-F_1(X)) \wedge 1/(1-F_2(Y)) > t\} = t^{-1}$, thus giving out $\eta = 1/2$ from an exact equality for every t rather than attaining it by way of a limit for sufficiently large t. Having this case as point of departure, values of η greater than $1/2$ would indicate a positive association between tail-related values of X and Y. However, an asymptotic regime based on other marginal transforms than Pareto thus rendering alternative tail dependence asymptotic relations, might nonetheless prove its merits. This is indeed the situation unraveled in the next example. Practical implications of these considerations on rank-transforms for tail dependence estimation are studied in [10].

Example 1. Let Φ_Σ denote the bivariate standard normal distribution function with correlation coefficient $\theta \in (-1,1)$, both marginal expectations equal to zero and unit variances. With Φ standing for standard normal distribution function, the Gaussian copula generated by $C_\Sigma(u,v) = \Phi_\Sigma\big(\Phi^{-1}(u), \Phi^{-1}(v)\big)$, $0 < u, v < 1$, satisfies the standard Pareto analogue of (5) with $c(x,y) = (xy)^{1/(1+\theta)}$ and $\eta = (1+\theta)/2$. This is verified as follows: firstly, we define the marginal tail quantile functions $U_i(t) := \big(1/(1-F_i)\big)^{\leftarrow}(t)$. Borrowing the developments in [26, 33], we write

$$\begin{aligned} \overline{q}(t) &:= P\big\{X > U_1(t), Y > U_2(t)\big\} = P\Big\{\frac{1}{1-F_1(X)} \wedge \frac{1}{1-F_2(Y)} > t\Big\} \\ &= k(\theta)\,(\log t)^{-\theta/(1+\theta)} t^{-2/(1+\theta)} \Big(1 + \frac{\theta}{1+\theta}\frac{\log(\log t)}{2\log t}\Big), \end{aligned}$$

with $k(\theta) := (1+\theta)^{3/2}/(1-\theta)^{1/2}(4\pi)^{-\theta/1+\theta}$. Clearly, we have $t\overline{q}(t) = \mathcal{L}_{\overline{q}}(t) t^{-\frac{1-\theta}{1+\theta}} = \mathcal{L}_{\overline{q}}(t) t^{1-1/\eta}$ dictating asymptotic independence if $\theta < 1$ ($\lambda = 0$, as $t \to \infty$), with the residual dependence index readily identified as $\eta = (1+\theta)/2$. We also have that $\mathcal{L}_{\overline{q}}(t) = o(1)$ for every $\theta > 0$. Analogously, but now in terms of $\alpha(t)$ featuring in (8), we find that:

$$\alpha(t) = P\Big\{X > U_1\big(\frac{1}{1-\exp\{-(t-1/2)^{-1}\}}\big), Y > U_2\big(\frac{1}{1-\exp\{-(t-1/2)^{-1}\}}\Big\}$$

$$
\begin{aligned}
&= t^{-\frac{2}{1+\theta}} k(\theta) \left(\log t - \frac{t^{-2}}{12} + o(t^{-2})\right)^{-\frac{\theta}{1+\theta}} \\
&\qquad \times \left(1 + \frac{t^{-2}}{6\theta} + o(t^{-2})\right)^{-\frac{\theta}{1+\theta}} \left(1 + \frac{\theta}{1+\theta} \frac{\log(\log t)}{2\log t}(1 + o(1))\right) \\
&= k(\theta)\, t^{-\frac{2}{1+\theta}} \left(\log t + \frac{t^{-2}}{6\theta} + o(t^{-2})\right)^{-\frac{\theta}{1+\theta}} \left(1 + \frac{\theta}{1+\theta} \frac{\log(\log t)}{2\log t}(1 + o(1))\right).
\end{aligned}
$$

Hence, $(t^{\eta}\alpha(t))/(t^{\eta}\overline{q}(t)) - 1 = -1/\big(6(1+\theta)\big)t^{-2}(\log t)^{-1} + o(t^{-2}/\log t)$, as $t \to \infty$, with $\eta = (1+\theta)/2$ as before. The rate of convergence equal to -2 surrendered by the approximating function of regular variation can be fairly associated to a very fast convergence [see e.g. 8].

Example 1 underlines that a principled compromise between standard Pareto and unit Fréchet can reveal a fertile middle ground for achieving substantive improvements in the estimation of η. Crucially, the asymptotic equivalence in distribution inherited from (7) will enable seamless transitions between sub-models of asymptotic dependence framed around the basic functions α and $\overline{q}$ as in Example 1 for the bivariate gaussian case. This presently identified and yet untapped link provides the basis for the new class of gradient estimators for η that can dispense with overly prescriptive marginal rank-transforms.

2.1. Gradient estimation of the residual dependence index

If there is one aspect the condition (8) strikes very clearly, it is that the problem of estimating the residual dependence coefficient η is inexorably attached to the problem of estimating the tail index within the context of heavy-tailed, univariate extremes. Purely on analytical grounds, the advantage of Pareto over Fréchet marginals in capturing the tail behaviour of the joint distribution function F is self-explanatory: the possibility of replacing limiting statements, that conceal a varying degree of approximation to an extreme value distribution, with exact ones lends legitimate appeal to the archetypal standard Pareto transform. What expansion (7) now carves out is that the unit Fréchet transform plus $1/2$ can be used in place of Pareto in a wholly optimal way. The goal in this paper is to develop this idea further and crystallise it into practice with a fully-fledged semi-parametric framework that builds on condition (8) for estimating the residual dependence index η.

We start off by recalling that, through the basic relation (8), the tail distribution function $\overline{F}_Z := 1 - F_Z$ of the random variable

$$
Z := \frac{1}{-\log F_1(X) \vee -\log F_2(Y)} + \frac{1}{2} \tag{9}
$$

is deemed regularly varying with index $-1/\eta$, with $\eta \in (0, 1]$. In a similar vein to [19, p. 265], which endorses the Hill estimator as especially purpose-fit when it comes to the problem of estimating the residual dependence index η from standard Pareto-originating pseudo-observables, we proceed to defining the sequence $\{Z_i^{(n)}\}_{i=1}^n$ of random variables

$$
Z_i^{(n)} := \left\{\left(-\log \frac{R(X_i)}{n+1}\right) \vee \left(-\log \frac{R(Y_i)}{n+1}\right)\right\}^{-1} + \frac{1}{2}, \tag{10}
$$

where $R(X_i) := \sum_{j=1}^n \mathbb{1}\{X_j \le X_i\}$ defines the rank of X_i among $(X_1, X_2, \ldots, X_n)$ and $R(Y_i)$ is the corresponding rank of Y_i among $(Y_1, Y_2, \ldots, Y_n)$. Let $F_Z^{(n)}$ denote the empirical distribution

function of the random variables $Z_i^{(n)}$. By construction, there is weak dependence across these identically distributed random variables that seek to emulate those taken from the population Z if the marginal distribution functions F_i were known.

The class of gradient estimators for $\eta \in (0, 1]$ introduced in this paper is given by the general functional

$$\hat{\eta}_{a,b} = \frac{1}{b}\left(\left\{\frac{n}{m}\int_{Z_{n,n-m}}^{\infty}\left(\frac{x}{Z_{n,n-m}}\right)^a dF_Z^{(n)}(x)\right\}^{b/a} - 1\right), \tag{11}$$

for some $a, b \in [-\infty, \infty]$, $a \neq 0$, of which the Hill estimator is a member by letting $a \to 0$ and $a/b \to -1$. In order to make this class trackable, we consider the order statistics $Z_{n,1} \leq \ldots \leq Z_{n,n}$ associated with the sample (10) and let $p \in \mathbb{R}$, $q > 0$ be conjugate constants in the sense that $1/p + 1/q = 1$ (or $1/p + 1/q \to 1$). Hence the q-gradient estimator drawing on the corrected unit Fréchet marginals:

$$\hat{\eta}_q^{(S)} := \frac{1 - \left\{\frac{1}{m}\sum_{i=0}^{m-1}\left(\frac{Z_{n,n-i}}{Z_{n,n-m}}\right)^{1-\frac{1}{q}}\right\}^{-1}}{1 - 1/q}. \tag{12}$$

The conjugate constants assigned to $(a, b) = (1/p, 1/q - 1)$ in (11) do not signify a parameter that would be fine-tuned to the estimator's optimal performance. Instead, q and p are allowed to vary within a certain gradient of values that will work in unison and eventually coalesce around the proper stable regime in the upper tail quadrant harbouring extremal residual dependence between joint exceedances in terms of the initial F (see (5) and related text).

2.2. Second order regular variation

More than building directly on the regular variation property established for the tail distribution function $\overline{F}_Z$ for the gradient estimation of η, we need to quantify the inherent approximation error to (8), which must be kept within acceptable bounds. To this end, we consider the following tentative second order strengthening of (5): with functions $\alpha > 0$ and β,

$$\lim_{t\to\infty} \frac{\frac{P\left\{\frac{1}{1-F_1(X)} > t/x, \frac{1}{1-F_2(Y)} > t/y\right\}}{\alpha(t)} - c(x, y)}{\beta(t)} =: K(x, y) \tag{13}$$

exists for all $x, y \geq 0$, $x + y > 0$, with $\alpha(t) \to 0$, as $t \to \infty$, and β ultimately of constant sign and tending to 0, and with a limiting function K which is neither constant nor a multiple of c. We further assume that the convergence is uniform on $\{(x, y) \in [0, \infty)^2 \,|\, x^2 + y^2 = 1\}$. In keeping with the prevailing theory for asymptotic dependence of bivariate extremes [see e.g. 13, 35] and linking with the results expounded in Section 1.1, we set $c(1, 1) = 1$ again and employ $\alpha(t)$ defined in (6), so that (13) does render a suitable second order refinement of (5). This implies that the function α is regularly varying of second order with auxiliary function β whereby we can assume that $c(x, x) = x^{1/\eta}$ and

$$K(x, x) = x^{1/\eta}\frac{1 - x^{\tau/\eta}}{\eta\tau}, \tag{14}$$

for all $x > 0$. Thus, $-1/\eta$ stands for the (negative) index of regular variation of α (i.e. $\alpha \in RV_{-1/\eta}$) and $\tau \geq 0$ is the second order parameter governing the speed of convergence in

(8). Moreover, we shall assume that $\lambda := \lim_{t\to\infty} t\alpha(t)$ exists, which would otherwise not be guaranteed if $\tau = 0$ and $\eta = 1$. We note that if $\tau > 0$, then the second order relation (13) holds locally uniformly on $[0,\infty)^2$ [cf. 13].

In order to derive the asymptotic distribution of the class of estimators set out in (12), it is useful to carry on with the above-delineated second order condition whose limiting function is (14), but formulated in terms of inverses. Defining the tail quantile function pertaining to F_Z as the generalised inverse $U(t) := \big(1/(1-F_Z)\big)^{\leftarrow}(t)$, we obtain the equivalent second order condition: for $x > 0$,

$$\lim_{t\to\infty} \frac{\frac{U(tx)}{U(t)} - x^{\eta}}{A(t)} = x^{\eta}\,\frac{x^{-\tau}-1}{-\tau}, \tag{15}$$

with the same parameters $\eta \in (0,1]$ and $\tau \geq 0$ as before. Furthermore, $A(t) = \beta(1/(1-F_Z)) \to 0$, as $t\to\infty$, and $|A| \in RV_{-\tau}$ [cf. 19, Theorem 3.2.5].

Assuming the second order condition (15) for our purposes, however, comes with a caveat: the gradient estimators stem from the standardisation to shifted-unit Fréchet marginals, under which transform the limit $c(x,x) = x^{1/\eta}$ is never obtained in the exact way that would warrant the convention $\tau = -\infty$. Heeding Section 1.1, we see that the cases of exact independence and complete dependence respectively yield the exact equalities $U(t) = t^{1/2}$ and $U(t) = t$, only if $U = U_T$ with T the minimum of the standard Pareto margins featuring in (13). When specialising on the distribution of Z, these exact boundary cases are automatically excluded; the best we can attain is $U_Z(t) \sim t^{\eta}$, for every $\eta \in (0,1]$. Therefore, there has to be some kind of repositioning in the second order condition to reflect the shifted-Fréchet marginals in the exact manner too. This is, in broad strokes, the theoretical justification for adopting instead the second order condition that follows from (13) in a principled way: we define the quantile function $V^{\star} := 1/2 + \big(-1/\log F\big)^{\leftarrow}$ and assume that there exists an eventually decreasing function $B(t) \to 0$, as $t\to\infty$, such that

$$\lim_{t\to\infty} \frac{\frac{V^{\star}(tx)}{V^{\star}(t)} - x^{\eta}}{B(t)} = x^{\eta}\,\frac{x^{-\tau_\star}-1}{\tau_\star} =: D_{\eta,\tau^\star}(x), \tag{16}$$

for all $x > 0$, with $\tau_\star \geq 0$. The auxiliary function B satisfies $|B| \in RV_{-\tau_\star}$. We note that a unit Fréchet random variable E has $V_E^{\star}(t) = t - 1/2$ and the quantile $V^{\star}$ for the shifted Fréchet Z is the aimed identity function. The equivalence between the above second order conditions pertaining to U and $V^{\star}$ is formally established later on, in Theorem 4.2. The most consequential aspect to the estimation which this theorem unveils is in the relationship between the respective second order parameters τ and $\tau_\star$ through its interplay with η. It results that $\tau_\star := \eta \wedge \tau$, meaning that the correction by 1/2, although appropriate for the Pareto-type fit to the Fréchet exceedances, it has a detrimental effect on that convergence rate which, if left unabated, can hinder the performance of the proposed unifying estimation of the residual dependence index η and potentially discredit its usefulness. From this realisation, devising a bias reduction procedure amenable to our semi-parametric setting is no longer an after-thought but rather a necessity, and indeed one at the heart of this paper. The results in the remainder of this section are aimed at making such critical second order component explicit. The next two theorems initiate the pathway for harnessing the approximation bias whose modelling will culminate in the reduced-bias gradient estimators introduced in Section 3.

2.3. Consistency and asymptotic normality of the gradient estimators

The proposed estimation of the residual dependence index $\eta \in (0, 1]$, primarily aimed at closing the gap between standard Pareto and Fréchet marginal transforms, builds on the idea that a functional representation for the estimators of the residual tail index relying on the tail quantile processes $\{Z_{n,n-[ms]}\}_{0<s<n/m}$ of (10) (as usual $[a]$ denotes the largest integer less than or equal to $a \in \mathbb{R}$) can transfer seamlessly to the analogous formulation in terms of common standard Pareto marginals. The $\sqrt{m}$-consistency [in the sense of 28, Definition 4.2] of the gradient estimator $\hat{\eta}_q^{(S)}$ defined in (12) is actually established by showing that $\hat{\eta}_q^{(S)}$ is asymptotically equivalent to its counterpart based on the pseudo-observables

$$T_i^{(n)} := \frac{n+1}{n+1-R(X_i)} \wedge \frac{n+1}{n+1-R(Y_i)}, \quad i = 1, 2, \ldots, n. \tag{17}$$

In particular, we write the Hill estimator of η in terms of the functional representation:

$$\hat{\eta}^{(H)} \equiv \hat{\eta}_1 = \frac{n}{m} \int_{T_{n,n-m}}^{\infty} \left(\log x - \log T_{n,n-m}\right) dF_T^{(n)}(x), \tag{18}$$

where, for each x, $F_T^{(n)}(x)$ is the empirical distribution for the sequence of the identically distributed random variables $T_i^{(n)}$, whose $(m+1)$-th descending order statistic is $T_{n,n-m}$.

Theorem 2.1. *Suppose condition* (13) *holds with* $\lim_{t\to\infty} t\alpha(t) = \lambda$ *finite. Let* $r(n) = n\alpha\left(n/k\right)$ *be a positive sequence such that* $r(n) \to \infty$, *and* $n/k \to \infty$, *as* $n \to \infty$. *For positive integers* $m := [r(n)]$, *assume that* $\sqrt{m}\, B(n/m) \to 0$, *with* B *from the equivalent relation* (16) *to* (13) *whilst* $x = y$. *Then, the gradient estimators* $\hat{\eta}_q^{(S)}$ *belonging to the class* (12) *are asymptotically equivalent to their respective standard Pareto analogue drawing on* (17) *which are here denoted by* $\hat{\eta}_q$. *Concisely: as* $n \to \infty$,

$$\sqrt{m}\left|\hat{\eta}_q^{(S)} - \hat{\eta}_q\right| \xrightarrow{P} 0.$$

Remark 1. The Hill estimator does not escape the grasp of Theorem 2.1 since it forms an integral part of the gradient configuration (12) when $q \to 1$.

Remark 2. In the asymptotic independence setting, the primary assumptions in the theorem concerning the relative growth of the sequences $r(n)$ and $k = k(n)$, as $n \to \infty$, imply $k \to \infty$ and $r(n)/k \to 0$. As a result, the most reliable estimates of η are those yielded by only a few of the larger observations from the sample $\{Z_i^{(n)}\}_{i\geq 1}$. For this reason, we shall use the terminology *ultra-intermediate sequence* when referring to $r(n)$ (or to m, interchangeably).

The stated condition $\sqrt{m}\, B(n/m) = o(1)$ imposes an (additional) upper bound on the intermediate sequence $r(n)$. Theorem 2.1 reinforces the point that, so long as the intermediate number $m := [r(n)]$ is not too large, both types of gradient estimators $\hat{\eta}_q^{(S)}$ and $\hat{\eta}_q$ are $\sqrt{m}$-consistent and eventually coincide in their asymptotic distribution with probability tending to one. Any remnant differences between these two classes of estimators can be disregarded via an appropriate choice of m that fulfils both sets of conditions in Theorem 2.1.

The asymptotic normality of the gradient estimator is established in the next theorem and in greater generality to the $\sqrt{m}$-consistency issued in Theorem 2.1. This refers in particular to the lower bound on the ultra-intermediate sequence $m = [r(n)] \to \infty$ that is now afforded through the assumption of a moderate convergence rate given by $\sqrt{m}\, B(n/m) = O(1)$ in lieu of the previous vanishingly small $\sqrt{m}\, B(n/m) = o(1)$, as $n \to \infty$.

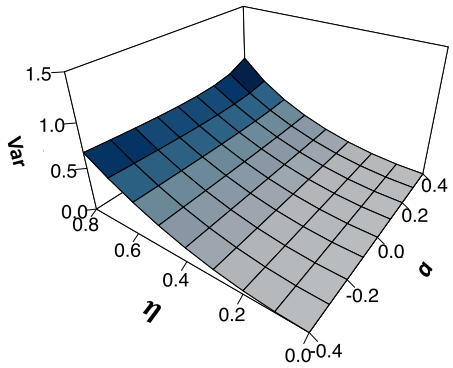

Fig 1. *Case of $b = -a$: asymptotic variance of $\hat{\eta}^{S}_{a,b}$ for true $\eta \in (0, 0.8]$, with $a \in [-0.4, 0.4]$.*

Theorem 2.2. *Under the conditions of Theorem 2.1 albeit with $\sqrt{m}\, B(n/m) \to \vartheta \in \mathbb{R}$, as $n \to \infty$, for every $a < 1/(2\eta)$ and $b/a \to -1$ embedded in*

$$\hat{\eta}^{(S)}_{a,b} = \frac{1}{b}\left[\left\{\int_0^1 \left(\frac{Z_{n,n-[ms]}}{Z_{n,n-m}}\right)^a ds\right\}^{b/a} - 1\right], \tag{19}$$

the following convergence in distribution holds:

$$\sqrt{m}\left(\hat{\eta}^{(S)}_{a,b} - \eta\right) \xrightarrow{d} N\left(\vartheta\, \mu_a, \sigma_a^2\right),$$

where $\mu_a = \mu_a(\eta, \tau_\star) = (1 - a\eta)/(1 - a\eta + \tau_\star)$ and $\sigma_a^2 = \sigma_a^2(\eta) = \eta^2(1 - a\eta)^2/(1 - 2a\eta)$.

Remark 3. By allowing $a \to 0$ in Theorem 2.2, the familiar asymptotic normality of the Hill estimator is retrieved in a straightforward way.

Theorem 2.2 brings to light the role of the second parameter $\tau_\star \geq 0$: the greater the $\tau_\star$, the smaller the second order (dominant) component of the asymptotic bias. This is due to the increased rate of convergence in the regularly varying part of (13). The constraint $a < 1/(2\eta)$, $\eta \in (0, 1]$ is a requirement for the finite asymptotic variance. Through simple algebra one can show that $\sigma_a^2(\eta) \geq \eta^2$, for all a, η, where equality holds for the Hill estimator (i.e., $\sigma^2_{a=0}(\eta) = \eta^2$). Moreover, the asymptotic variance increases with increasing $|a|$ but by much less than it does with η. Figure 1 aims to depicts these findings relating the asymptotic variance with varying $|a| \leq 0.4$.

3. Reduced-bias estimation

The most straightforward, and potentially the most effective way of removing the second order bias inherent to the estimator $\hat{\eta}^{(S)}_{a,b}$ without increasing the asymptotic variance, is to capitalise on a Hall-Welsh representation that will imply relation (13) [see 21].

The Hall-Welsh-type model, here architected so as to flesh out the approximation bias in the gradient estimation of the residual dependence index η, arises as a far-reaching class of regularly varying functions satisfying condition (16). It refers to the catch-all for distributions whose extremal quantile function $V^\star$ admits the expansion, as $t \to \infty$,

$$V^\star(t) = C^\eta t^\eta \left(1 + \eta D_1 C^{-\tau} t^{-\tau} + \eta D_2 C^{-\eta} t^{-\eta}\right), \tag{20}$$

for $C > 0$, $D_1, D_2 \neq 0$, whereby $B(t) = \eta\tau_\star D_j (Ct)^{-\tau_\star}$ with $j = 1$ if $\tau_\star = \tau > 0$, $j = 2$ if $\tau_\star = \eta > 0$. Importantly, this sub-model for $V^\star \in 2RV(\eta, \tau_\star)$ implies that the speed of convergence to the desired canonical η-varying function in the limit is tampered by the $1/2$-correction [cf. 29, p.218]. The reduced-bias gradient estimation will leverage on this altered rate of convergence. Unusually, the key implication of the shift by $1/2$ in standardisation to unit Fréchet marginals is that the inherent rate convergence of second order condition in (13) becomes slower, not faster [19, p.83]. For a great majority of extremal dependence models it means that the second order parameter $\tau_\star$ gets trapped at the value η. This makes bias-reduction all the more worth our while because, if the parameters $\tau_\star$ and η coincide, then both can be estimated in one stroke even by employing the same estimator.

In this setting, the reduced-bias version of estimator (12) that arises from Theorem 2.2 is given by:

$$\widetilde{\eta}_q(m, m^*) := \hat{\eta}_q^{(S)} \left\{ 1 - \left(\hat{\beta}\left(\frac{n}{m}\right) + \frac{1}{2Z_{n,n-m^*}} \right) \frac{1 - \hat{\eta}_q^{(S)}/p}{1 - \hat{\eta}_q^{(S)}/p + \widehat{\tau_\star}} \right\}, \tag{21}$$

with m, m^* both intermediate sequences of positive integers, i.e., $m = m(n) \to \infty$, $m^* = m^*(n) \to \infty$, $m/n \to 0$, as $n \to \infty$, and $m^* \leq m$ (implying that $m^*/n \to 0$), and where $\hat{\beta}(n/m)$ and $\widehat{\tau_\star}$ respectively denote consistent estimators for $\eta^{-1}B(n/m)$ and $\tau_\star > 0$. We remark that any such consistent estimators can be plugged in and that $\widetilde{\eta}_q(m, m^*)$ depends on m^* exclusively through $Z_{n,n-m^*}$.

Theorem 3.1. *Assume conditions of Theorem 2.2 hold. Let m^* be another intermediate sequence of positive integers such that $m^* \leq m$. With $a = 1 - 1/q$, as $n \to \infty$,*

(i) the estimator $\widetilde{\eta}_q := \widetilde{\eta}_q(m, m^*)$ defined in (21) is a $\sqrt{m}$-consistent reduced-bias estimator obtained from $\hat{\eta}_{a,b}^{(S)}$ in (19) for conjugate constants satisfying $a/b \to -1$.

(ii) If $m = [r(n)]$ satisfies $\sqrt{m}\, B(n/m) = O(1)$ then, for every $0 < q < 1 + |2\eta - 1|^{-1}$,

$$\sqrt{m}\left(\widetilde{\eta}_q - \eta\right) \xrightarrow{d} N_q, \tag{22}$$

where N_q is a normal variable with mean zero and variance $\sigma^2_{1-1/q}$ given in Theorem 2.2.

Remark 4. Confidence bounds for η based on (21) follow readily from Theorem 3.1(ii).

4. Basic asymptotic results

Semi-parametric inference can be viewed as the statistically meaningful distribution-free process that results from piecing together two key components: the study of a statistical functional, which typically serves as an estimator of the relevant parameters, and the study of the underlying empirical process. In Sections 2 and 3, we investigated suitable statistical functionals against the backdrop of asymptotic independence exhibited by a bivariate distribution $F \in \mathcal{D}(G)$. We now lay the groundwork for the study of the related empirical process.

For every x, consider the tail empirical distribution function of the random variables $Z_i^{(n)}$

$$\overline{F}_Z^{(n)}(x) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\left\{\frac{k}{n+1} Z_{n,n-i+1} \geq x\right\}}$$

and its pertaining tail quantile process defined as

$$Q_n(s) := Z^{(n)}_{n,n-[ms]}, \quad 0 < s < \frac{n}{m}. \tag{23}$$

Theorem 4.1. *Under the conditions of Theorem 2.2, there exits a probability space* $(\Omega, \mathcal{A}, P)$ *with the random variables in* (23) *and a sequence of standard Brownian motions* $\{W_n(s)\}_{s\geq 0}$ *such that, for all* $s_0 > 0$*, with arbitrarily small* $\varepsilon > 0$*,*

$$\sup_{0<s\leq s_0} s^{\eta+\frac{1}{2}+\epsilon} \left| \sqrt{m}\left(\frac{Q_n(s)}{V^*(n/m)} - s^{-\eta}\right) - \eta s^{-(\eta+1)} W_n(s) - \sqrt{m}\, B\Big(\frac{n}{m}\Big) s^{-\eta} \frac{s^{\tau_\star} - 1}{\tau_\star} \right| = o_p(1),$$

as $n \to \infty$.

Let V be the quantile function that reflects a unit Fréchet-type transform, also translated into the mainstream tail quantile function $U := (1/(1-F))^{\leftarrow}$:

$$V(t) := \Big(-\frac{1}{\log F}\Big)^{\leftarrow}(t) = U\Big(\frac{1}{1-e^{-1/t}}\Big).$$

By writing concisely $V^\star(t) := V(t) + 1/2$, we obtain the formulation for the relevant condition of second order regular variation heralding the deterministic approximation bias captured in Theorem 4.1. This refers in particular to the term of order $\sqrt{m}B(n/m)$. We contend that Theorem 4.2 furnishes a bedrock result for interchanging between Pareto and Fréchet-borne domains of attraction, in the sense that it provides actionable insight to being able to convert prior results on peaks-over-threshold estimation for heavy tails to corresponding inference that draws on block maxima' exceedances.

Theorem 4.2. *Let* U *be a tail quantile function satisfying the second order condition* (15) *with second order parameter* $\tau \geq 0$*. Then,* $V^\star$ *is also regularly varying with the same index* η *and is such that, for* $x > 0$*,*

$$\lim_{t\to\infty} \frac{\frac{V^\star(tx)}{V^\star(t)} - x^\eta}{B(t)} = \begin{cases} x^\eta \, \frac{x^{-\tau}-1}{\tau}, & \tau < \eta, \\ x^\eta \, \frac{x^{-\tau}-1}{\eta}, & \tau \geq \eta, \end{cases}$$

where

$$B(t) = \begin{cases} A(t), & \tau < \eta, \\ A(t) + \frac{\eta}{2}\frac{1}{V^\star(t)}, & \tau = \eta, \\ \frac{\eta}{2}\frac{1}{V^\star(t)}, & \tau > \eta. \end{cases}$$

Moreover, $|B| \in RV_{-\tau_\star}$ *with second order parameter governing the speed of convergence given by*

$$\tau_\star = \begin{cases} \tau, & 0 < \tau < \eta, \\ \eta, & 0 < \eta \leq \tau. \end{cases}$$

Remark 5. The reciprocal of the above-stated result is also true for all $\eta \in (0, 1]$ and $\tau > 0$.

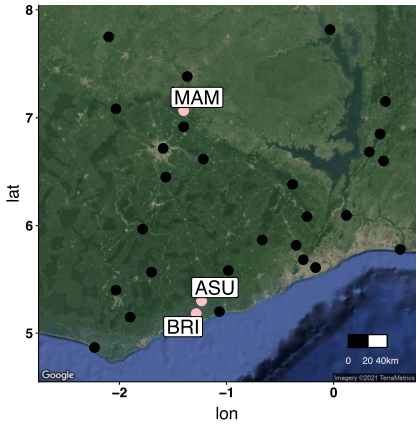

Fig 2. *Map of southern Ghana highlighting the three selected stations.*

5. Residual dependence in tropical extreme rainfall

The data studied in this section consist of daily rainfall measurements recorded between the years 1950 and 2017 at 591 irregularly spaced stations across Ghana. These data were collected, processed and quality controlled by the Ghana Meteorological Agency [see also 24]. Ghana has a strong seasonal rainfall cycle regulated by the West African monsoon. For the purposes of this illustrative analysis, we have singled out a pair of nearby gauging stations (within the range 5-15km) and a pair of well-separated stations distancing 190-200km from each other. We direct the focus of our inference to daily rainfall measurements recorded every June in the 68 years worth of available data because we wish to include the main rainy season. The obvious advantage is that this sample contains the largest proportion of rainy days and likely the highest frequency of extreme occurrences as well.

In order to ensure that the data are identically distributed, we have only considered stations in the southern part of the country (south of 8° lat) given the significant contrasts in rainfall regimes across the country: the north of Ghana is semi-arid, exhibiting a uni-model seasonal cycle, peaking in July/August; the south is more humid, with a bi-modal seasonal cycle, with peaks in June and October, and a break in August. None of the individual time series are without missing values, and the proportion of these was found to range between 5% to 95%. As an initial sifting of the data, we removed all stations with more than 3000 missing values, roughly equating to 12% of the full time series, which left us with 40 stations from which to select pairs. To minimise distributional variations due to systematic features in the spatial domain, the two selected pairs of stations have the station ASU in common, mapped out in Figure 2, with BRI located 5-10km away from ASU, and MAM 190-200km apart. The data have been pre-processed in order to remove inconsistencies, resulting in 1830 bivariate observations validated for our statistical analysis.

Even during the monsoon season there are many dry days (i.e., daily rainfall amounts $< 1mm$), and because statistics of extremes concerns data at the edge of the sample, we have chosen to conduct estimation of the residual dependence index by drawing only on the rainy days above the observed 90%-quantile for daily rainfall. Some evidence was found of tenuous dependence between very large rainfall values (i.e. of magnitude $> 50mm$) in June, including

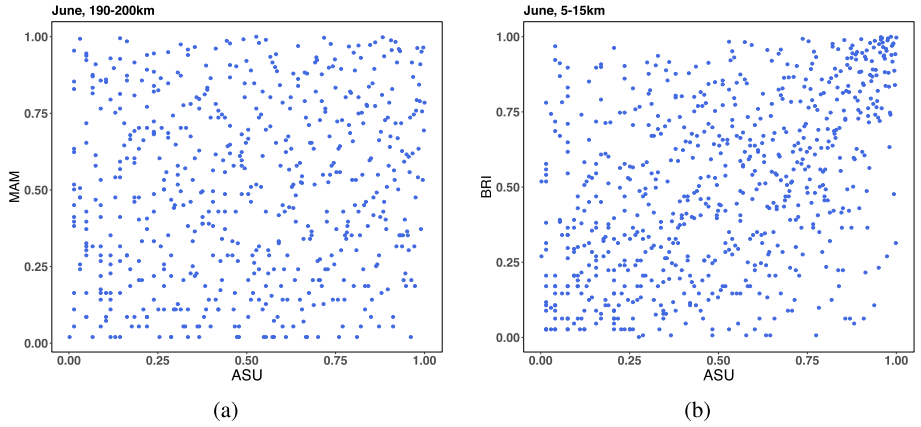

Fig 3. *Scatter plots of positive daily rainfall transformed to the unit uniform scale for the two pairs of stations: (a) (ASU, MAM); (b) (ASU, BRI).*

Fig 4. *Well-separated stations (ASU, MAM): The blue line depicts the sample path for the smaller* $q = 0.5$*, whereas the red line corresponds to the largest* $q = 1.5$*, both encased in their respective* 95% *confidence bands. In yellow are the Hill estimates* ($q = 1$)*. Plot (a) represents reduced-bias estimation of the residual dependence index* $\eta \in (0, 1]$ *and plot (b) relates the plain estimator with no bias reduction.*

pairs of stations situated at such long distances as $150km$ apart. Building on these findings, graphically portrayed in the two scatter-plots presented in Figure 3, we have benchmarked the range 190-200km for the spatial distance at which we expect to detect a weakening dependence in extreme tropical rainfall. Indeed, in [24], it is concluded that for moderate values of rainfall, stations located more than $150km$ apart appear to no longer exhibit significant dependence between simultaneously large rainfall occurrences.

We now seek to quantify the strength of pairwise association amongst observations dotting a squared area on the very the top right corner of the plots in Figure 3. This sort of residual dependence could have not been captured with the models used in [24]. Figures 4 and 5 display the estimates' paths yielded by the reduced-bias q-gradient estimator $\widetilde{\eta}_q(m, m^*)$ defined in (21). We clarify at this point that both scale function B and second order shape parameter $\tau_\star$ have been estimated externally via an adapted methodology to that of [8], whereby we fixed $m = [n^{0.999}]$ for calculating the relevant plug-in estimates that incorporate $\widetilde{\eta}_q(m, m^*)$.

Depicted in Figures 4 and 5 are the trajectories yielded by $\widetilde{\eta}_q(m, m^*)$ with assigned values

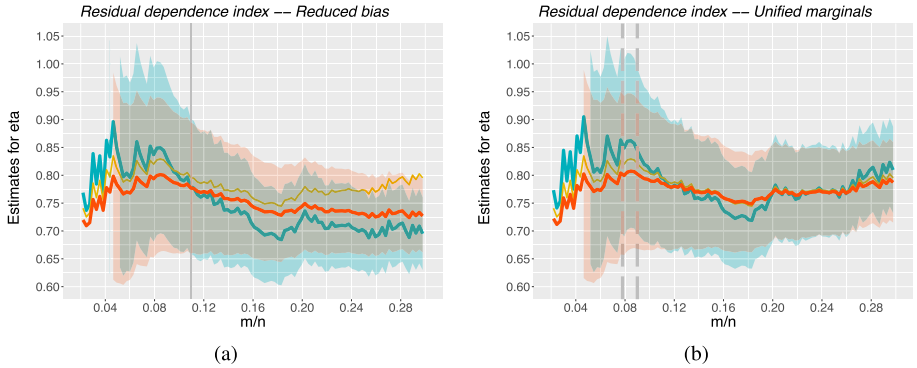

Fig 5. *Nearby stations (ASU, BRI): The blue line depicts the sample path for the smaller* $q = 0.5$*, whereas the red line corresponds to the largest* $q = 1.5$*, both encased in their respective* 95% *confidence bands. In yellow are the Hill estimates* ($q = 1$)*. Plot (a) represents reduced-bias estimation of the residual dependence index* $\eta \in (0, 1]$ *and plot (b) relates the plain estimator with no bias reduction.*

$q = 0.5, 1, 1.5$ (panel(a)) and those stemming from their plain gradient version $\hat{\eta}_q^{(S)}$ (panel (b)), cladded in their respective 95% confidence bands ($q = 0.5, 1.5$). We note that estimation by $\hat{\eta}_q^{(S)}$ is still endowed with the second order dominant component of the bias governed by $\sqrt{m}B(n/m) = O(1)$ (hence the upwards trend with increasing m). The central value $q = 1$ corresponds to the Hill estimator, which was not subjected to bias reduction in this instance.

The sample paths emanating from $\widetilde{\eta}_q(m, m^*)$ as well as those from $\hat{\eta}_q^{(S)}$ are more erratic for small values of m, reflecting the greater variance when only a small fraction m/n of exceedances above the threshold $z_{n,n-m}$ is used in the estimation. Although an expected finding, it still poses a practical challenge given the admonition in Remark 2 that $m = [r(n)]$ should emulate the ultra-intermediate sequence $r(n)$ in the hierarchy $m \ll k(n) \ll n$. A cursory look at the plots (b) of both Figures 4 and 5 reveals that, on the other hand, too large values of m/n yield an upwards trend in the gradient. It is this approximation bias that asserts itself on the rise in Figure 4(b) that we endeavour to curtail. In this instance, it reflects the type bias infused by using too many intermediate data points. Best practices are to scan for a region in the plot where the estimates' trajectories plateau and are relatively insensitive to a changing m/n. As a result of this standard practice, we found it reasonable to cast focus on the range of values limited by the dashed vertical lines in both Figures 4(b) and 5(b). While a stable region, where the gradient lines almost coincide around the estimated value $\eta = 0.55$ before breaking out again, clearly emerges in Figure 4, a similar desirable pattern is not as clear-cut in Figure 5. What is there to be revealed is that the proposed reduced-bias estimator also allows to overcome this challenge. The q-gradient trajectories of $\widetilde{\eta}_q(m, m^*)$ have been seen to work collectively towards delivering a unique inflection point, but rarely has it been the case with the plain estimator $\hat{\eta}_q^{(S)}$. The simulation study as part of Section 6 shows that even if we decide to stick with the Hill estimator $\hat{\eta}_1^{(S)}$, the reduced-bias gradient estimator $\widetilde{\eta}_q(m, m^*)$ can still prove of advantageous use as a threshold selection procedure because the knot-value m, which ties the gradient's paths together and upends them, points to admissible minimisers of the mean squared error of the plain q-gradient estimators too.

In relation to the pair of stations located further apart, the η-estimates displayed in Figure 4(a) show an appreciable stability region that stretches from $m/n = 6\%$ to $m/n = 10\%$,

approximately, with the gradient estimates coalescing around $\hat{\eta} = 0.5$ rather tightly, down from the estimated $\eta = 0.55$ in Figure 4(b). The Hill estimator $\hat{\eta}_1^{(S)}$ exhibits a steadily increasing bias making it hard to discern a reliable estimate for η in the absence of a formal procedure for threshold selection. We thus use the knot-value $m_0/n = 0.069$ from plot (a) to obtain the optimal estimate $\widetilde{\eta} = 0.51$. This estimate suggests clear asymptotic independence in tropical rainfall between the two locations ASU and MAM.

The two plots in Figure 5 reveal a consensus between the q-gradient estimators. A stable region in the gradient emerges along the range $0.078 \leq m/n \leq 0.09$. It is advisable to examine the q-gradients is tandem, rather than zero in on a particular value q, in order to dissipate doubts around the value m that will likely lead to an efficient estimation of the residual dependence index η. In this sense, the knot-value $m_0/n = 0.112$, where the inflection of the reduced-bias $q-$gradient occurs and its trajectories are upended, yields the estimate $\widetilde{\eta} = 0.77$. This means that concomitantly high rainfall values in stations BRI and ASU tend to occur much more frequently than under the near pairwise extremal independence regime inferred for the pair of stations ASU and MAM. To not observe a value of η very close to one, even at such a short distance, is to be expected for tropical rainfall, where sporadic but intense storms tend to be very localised. For example, when a convective storm initiated over one station travels in the opposite direction to that of the other station in the pairwise analysis.

6. Simulation results

The simulation results encompassing this section draw on $N = 1000$ random samples, each of which consisting of $n = 500$ i.i.d. realisations of the random pair (X, Y) with continuous distribution function F. Through the uniform copula vector (U, V), uniquely determined by the copula function $C : [0,1]^2 \to [0,1]$ such that $C(u,v) = F\big(F_1^{\leftarrow}(u), F_2^{\leftarrow}(v)\big)$ [39], we simulated from the parent distributions listed below in the domain of attraction of a max-stable extreme value distribution G while satisfying condition (5). We recall that (5) renders a pre-limit characterisation of (1) when $G(x,y) = G_1(x)G_2(y)$. We have chosen the copulae:

(i) **Frank distribution** with copula function

$$C_\theta(u,v) = -\frac{1}{\theta}\log\Big(1 - \frac{(1-e^{-\theta u})(1-e^{-\theta v})}{1-e^{-\theta}}\Big), \quad (u,v) \in [0,1]^2,\ \theta > 0,$$

satisfying the second order condition (15) with $\eta = \tau = 1/2$. These values indicate near independence.

(ii) **Ali-Mikhail-Had distribution**, whose copula function is given by

$$C_\theta(u,v) = \frac{u\,v}{1-\theta(1-u)(1-v)}, \quad (u,v) \in [0,1]^2,\ \theta \in [-1,1].$$

For $\theta = -1$ the second order condition (15) holds with $\eta = 1/3$ and $\tau = 2/3$. A value $\eta < 1/2$ indicates that joint exceedances of the same probability-threshold occur less frequently than they would have if X and Y were independent random variables. The convergence rate in (16) determines $\tau_\star = 1/3$, i.e. slower convergence than that of (15).

(iii) **Farlie-Gumbel-Morgenstern**, defined by the copula function

$$C_\theta(u,v) = uv\big(1+\theta(1-u)(1-v)\big), \quad (u,v) \in [0,1]^2,\ \theta \in [-1,1].$$

For $\theta = -1$ the second order condition (15) holds with $\eta = 1/3 = \tau$. There is no change in the convergence rate from applying the shifted-Fréchet framework, i.e. $\tau = \tau_\star = 1/3$.

(iv) **Bivariate Normal**, with Gaussian copula $C_\theta(u,v) = \Phi_\theta\big(\Phi^{-1}(u), \Phi^{-1}(v)\big)$ presented in Example 1, where it is shown that, for some correlation coefficient $\theta < 1$, condition (5) is satisfied with $\eta = (1+\theta)/2$. A value $\eta > 1/2$ indicates that (X, Y) exceed the same probability-threshold more frequently than if X and Y were independent random variables. A second order parameter $\tau = 0$ in condition (13) compounded with $t\alpha(t) \to 0$ deems the Gaussian copula within the scope of the estimation methods in this paper. We recall Example 1 in this respect.

These parent distributions F (equiv. C) have been selected for their flexibility and with the aim of illustrating performance of our proposed gradient estimators, both individually and by comparison, while operating on varying strength of association between components X and Y (equiv. U and V). Not only have we sought to cover the two asymptotic independence regimes around the near independence case of $\eta = 1/2$, we have also aimed to address the type of asymptotic independence found in tropical rainfall extremes studied in Section 5 thus furthering prior results from [24].

In order to better evaluate how the proposed gradient estimation fares in the landscape of general estimators for an index of regular variation (in the sense of (8)), the estimator for heavy and super-heavy-tailed distributions introduced in [1] is here canvassed for comparison. If drawing on the shifted Fréchet pseudo-observables, it turns out that the referred estimator is made up of two members of the class of gradient estimators, both with assigned $q = 1/2,\ 1/3$:

$$\hat{\eta}^{SH} := \frac{2\hat{\eta}^{(S)}_{1/3} - \hat{\eta}^{(S)}_{1/2}}{1 + 2(\hat{\eta}^{(S)}_{1/2} - \hat{\eta}^{(S)}_{1/3})} = \frac{\sum_{i=0}^{m-1} Z_{n,n-m}/Z_{n,n-i} - \sum_{i=0}^{m-1} \big(Z_{n,n-m}/Z_{n,n-i}\big)^2}{2\sum_{i=0}^{m-1} \big(Z_{n,n-m}/Z_{n,n-i}\big)^2 - \sum_{i=0}^{m-1} Z_{n,n-m}/Z_{n,n-i}}\,. \tag{24}$$

The estimates' trajectories yielded by $\hat{\eta}^{SH}$, and displayed as part of Figures 6–9, add to this finding since they emerge closely aligned with the sample paths trodden by the $q-$gradient estimator $\hat{\eta}^{(S)}_q$. Theorem 4.1 can be deployed in order to prove consistency and asymptotic normality of $\hat{\eta}^{SH}$ in a straightforward manner. Our simulation results are directly comparable with the findings in [3]. Moving on to the bias reduction (BR) of gradient estimators, again in the interest of comparison, we adopt the maximum likelihood (ML) estimator of η from [2] that uses Fréchet marginals with $w = 2/3$. This was the value that yielded uniformly most favourable results across all the listed copulae (i)–(iv). We note that the ML estimator is not closed-form and relies on a numerical maximisation method. The alternative linear reduced-bias estimator investigated in [2, Eq. 16] also enters this simulation study.

All Figures 6 to 9 show averages of the N estimates of η obtained with the $q-$gradient estimator $\hat{\eta}^{(S)}_q$ defined in (12), including its dual definition that takes on the standard Pareto pseudo-random variables (17). In particular, the Pareto counterpart of $\hat{\eta}^{(S)}_1$ is the Hill estimator given in (18). The respective simulated mean squared errors are shown on the right hand-side panels of these figures. The purpose of the pair of plots (a)–(b) and (c)–(d) is to demonstrate that the estimation of η rooted in the standard Pareto marginal transform is very much in tune with that drawing on the corrected unit Fréchet transform advocated for the $q-$gradient estimator $\hat{\eta}^{(S)}_q$. The shifted Fréchet parent slightly accentuates the role of the q-gradient estimation as it widens its range in the average as well as in the empirical mean squared error, a consistent aspect to many more parent models used in our comprehensive simulation study, from which study these four copulae have been selected.

The third row of plots (i.e., (d)–(e) throughout) displays the bias-reduction results yielded by the RB gradient estimator $\tilde{\eta}_q(m, m^*)$ from Section (3), with accompanying ML estimator. The SH estimator still features in these plots for benchmarking against the plain estimation just above. The attribution of an optimal value m in the sense of minimising the mean squared error is given to the region (shadowy vertical stripe) where the gradient loses its swell and eventually trickles down to a single line of undistinguishable trajectories beyond which the lines for $q < 1$ and $q > 1$ trade places. This short interval of values of m/n, that in some cases, narrows to a single point of "all on one" intersection appears to be unique across m/n. Importantly, the simulation study highlights that the knot-value m such that the reduced-bias gradient $\widetilde{\eta}_q(m)$ coalesces can equally lead to improved inference via alternative estimators of the residual dependence index η, noticeably the ML estimator, since it seems to persistently indicate the optimal number m for which both gradient and ML estimator's empirical mean squared error's trajectories are found at their lowest ebb. This is especially pertinent in relation to the ML estimator because, unlike the reduced-bias gradient estimator, its optimal number m tends to instil large bias at the expense of minimising the variance.

Figure 10 marks the point of difference between the previously addressed RB estimators and the linearised estimator presented in Eq.16 of [2]. The substantive bias stemming from the Linear estimator obviously impacts its empirical mean squared error significantly and therefore these plots have been omitted.

Finally, regarding the choice of m^* specific to threshold in the reduced-bias estimation, any value in the order of $1/\sqrt{m}$ or $m^* = [n^{0.3}]$, borne by the proof of Theorem 3.1, is fairly suitable. Either of these suggestions can be viewed as a feasible and appropriate choice since the estimation of η remains virtually unaffected when alternating between them.

7. Proofs

The purpose of this section is to collate all the proofs and auxiliary lemmas to the proofs of the asymptotic results encompassing Sections 2 and 3.

Throughout this section we will follow the same notation and assume conditions from the earlier sections in their relevant instances. For improved understanding, we begin with foundational results, underpinning those in Section 4, and then proceed incrementally towards covering the key results in Sections 2 and 3. The initial lemma is of independent interest in the sense that it offers a weak invariance principle for the sequence of normalised tail empirical processes generated by $\overline{F}_Z^{(n)}$ (cf. (11)). No extreme value theory is invoked at this point although this lemma drives an integral part of the proof of Theorem 4.1 to establishing the gaussian limit of the suitably normalised tail quantile process $\{Q_n(s)\}$ that engages the second order condition (16).

Lemma 7.1. *[**A weak invariance principle**] Assume there exists a positive function α such that*

$$\lim_{t\to\infty} \frac{P\left\{\frac{1}{1-F_1(X)} > t/x,\ \frac{1}{1-F_2(Y)} > t/x\right\}}{\alpha(t)} = x^{1/\eta}, \tag{25}$$

for all $x > 0$. Let $F_j^{(n)}$, $j = 1, 2$ be the marginal empirical distribution functions associated with X and Y, respectively (cf. (9)). Then, for an intermediate sequence $k = k(n)$ and $r(n) = n\alpha(n/k) \to \infty$, as $n \to \infty$,

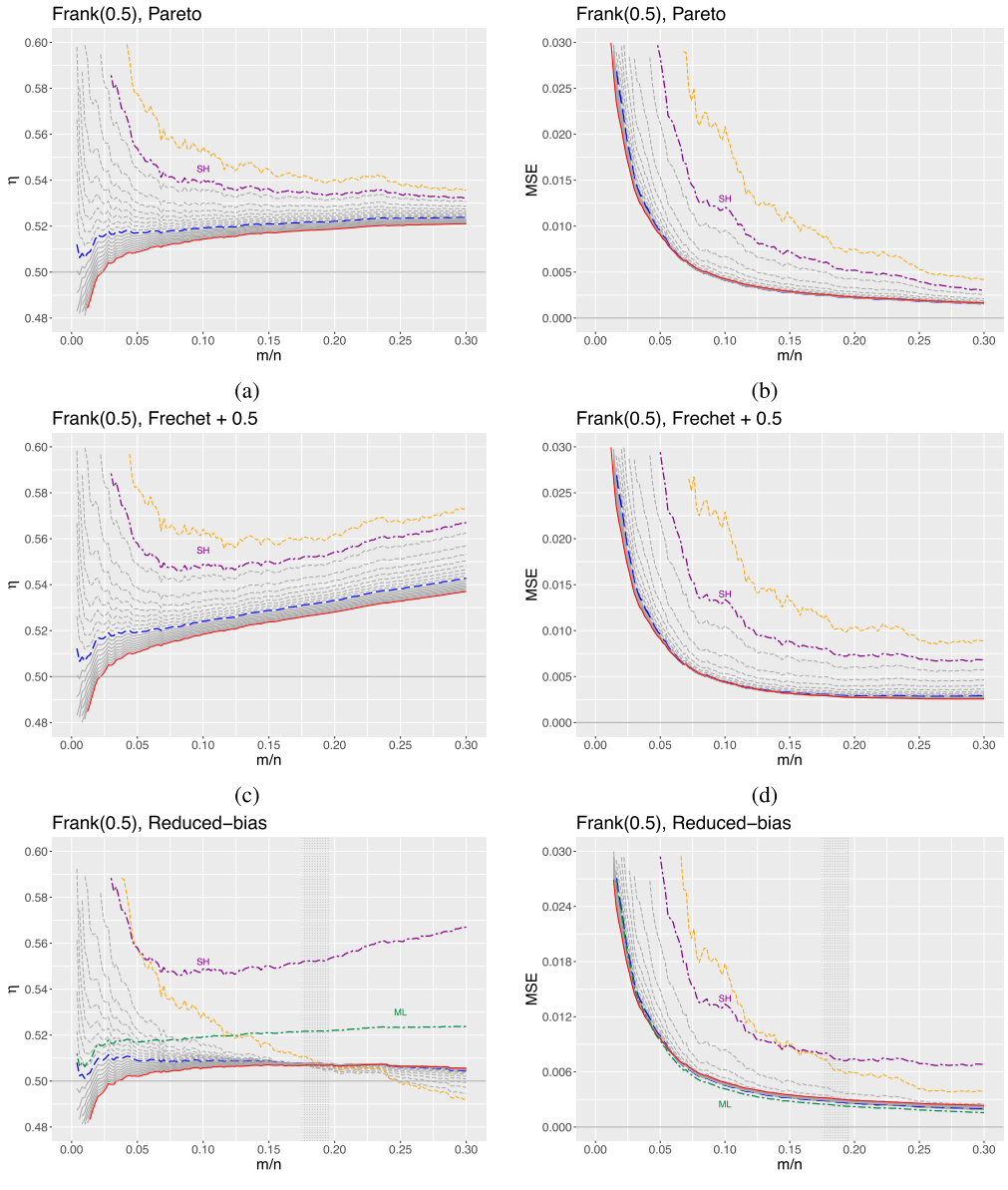

Fig 6. ***Frank copula with*** $\theta = 0.5$ *(*$\eta = \tau = 1/2$*): Identical plots (a)–(b) and (c)–(d) show the gradient estimator* $\hat{\eta}_q^{(S)}$ *and super-heavy (SH) dedicated estimator* $\hat{\eta}^{(SH)}$ *with either standard Pareto or corrected Fréchet marginals. Plots (e)–(f) display reduced-bias gradient estimation* $\tilde{\eta}_q^{(S)}(m, [\sqrt{m}])$ *and the numerical ML estimation from [2]. All plots: dashed lines represent* $q < 1$ *with orange for the lower bound* $q = 0.1$*; solid lines pertain to* $q > 1$ *with red for the upper bound* $q = 1.9$*; The Hill estimator (*$q = 1$*) is shown in the long-dashed blue line.*

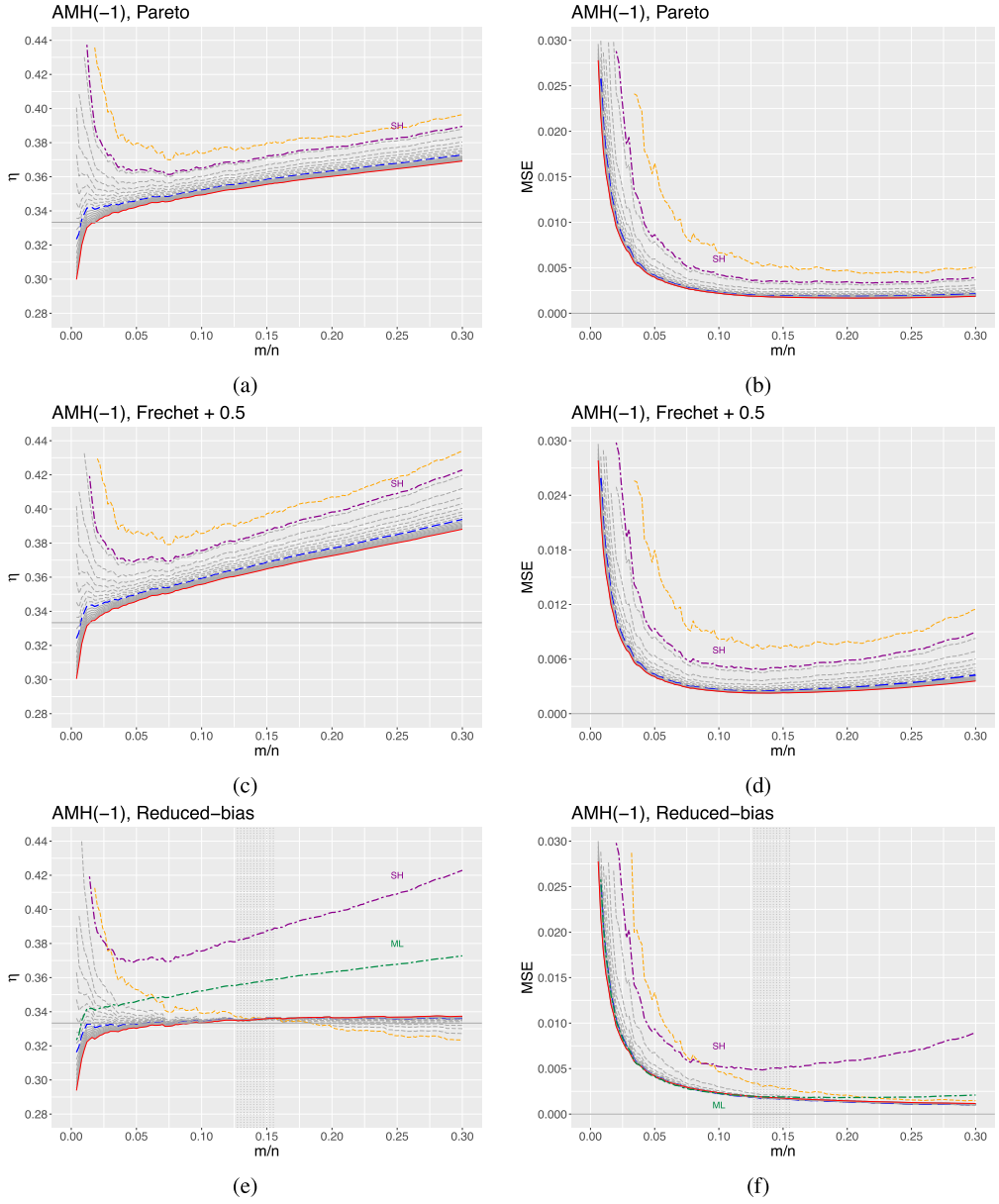

Fig 7. ***Ali-Mikhail-Haq copula with*** $\theta = -1$ $(\eta = 1/3, \tau = 2/3)$*: Identical plots (a)–(b) and (c)–(d) show the gradient estimator* $\hat{\eta}_q^{(S)}$ *and super-heavy (SH) dedicated estimator* $\hat{\eta}^{(SH)}$ *with either standard Pareto or corrected Fréchet marginals. Plots (e)–(f) display reduced-bias gradient estimation* $\widetilde{\eta}_q^{(S)}(m, [\sqrt{m}])$ *and the numerical ML estimation from [2]. All plots: dashed lines represent* $q < 1$ *with orange for the lower bound* $q = 0.1$*; solid lines pertain to* $q > 1$ *with red for the upper bound* $q = 1.9$*; The Hill estimator* $(q = 1)$ *is shown in the long-dashed blue line.*

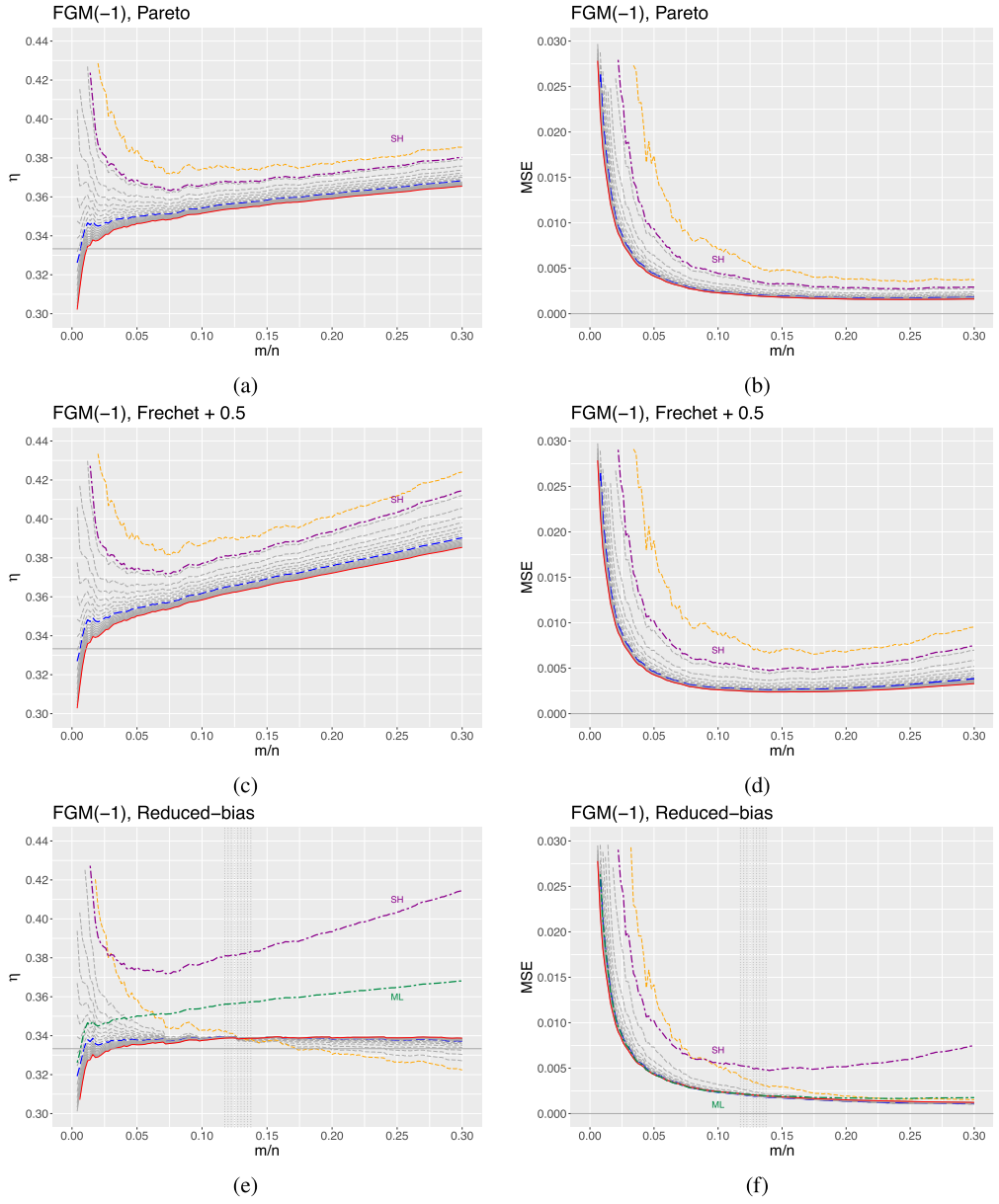

Fig 8. ***Farlie-Gumbel-Morgenstern copula with*** $\theta = -1$ ($\eta = \tau = 1/3$)*: Identical plots (a)–(b) and (c)–(d) show the gradient estimator* $\hat{\eta}_q^{(S)}$ *and super-heavy (SH) dedicated estimator* $\hat{\eta}^{(SH)}$ *with either standard Pareto or corrected Fréchet marginals. Plots (e)–(f) display reduced-bias gradient estimation* $\widetilde{\eta}_q^{(S)}(m, [\sqrt{m}])$ *and the numerical ML estimation from [2]. All plots: dashed lines represent* $q < 1$ *with orange for the lower bound* $q = 0.1$*; solid lines pertain to* $q > 1$ *with red for the upper bound* $q = 1.9$*; The Hill estimator* ($q = 1$) *is shown in the long-dashed blue line.*

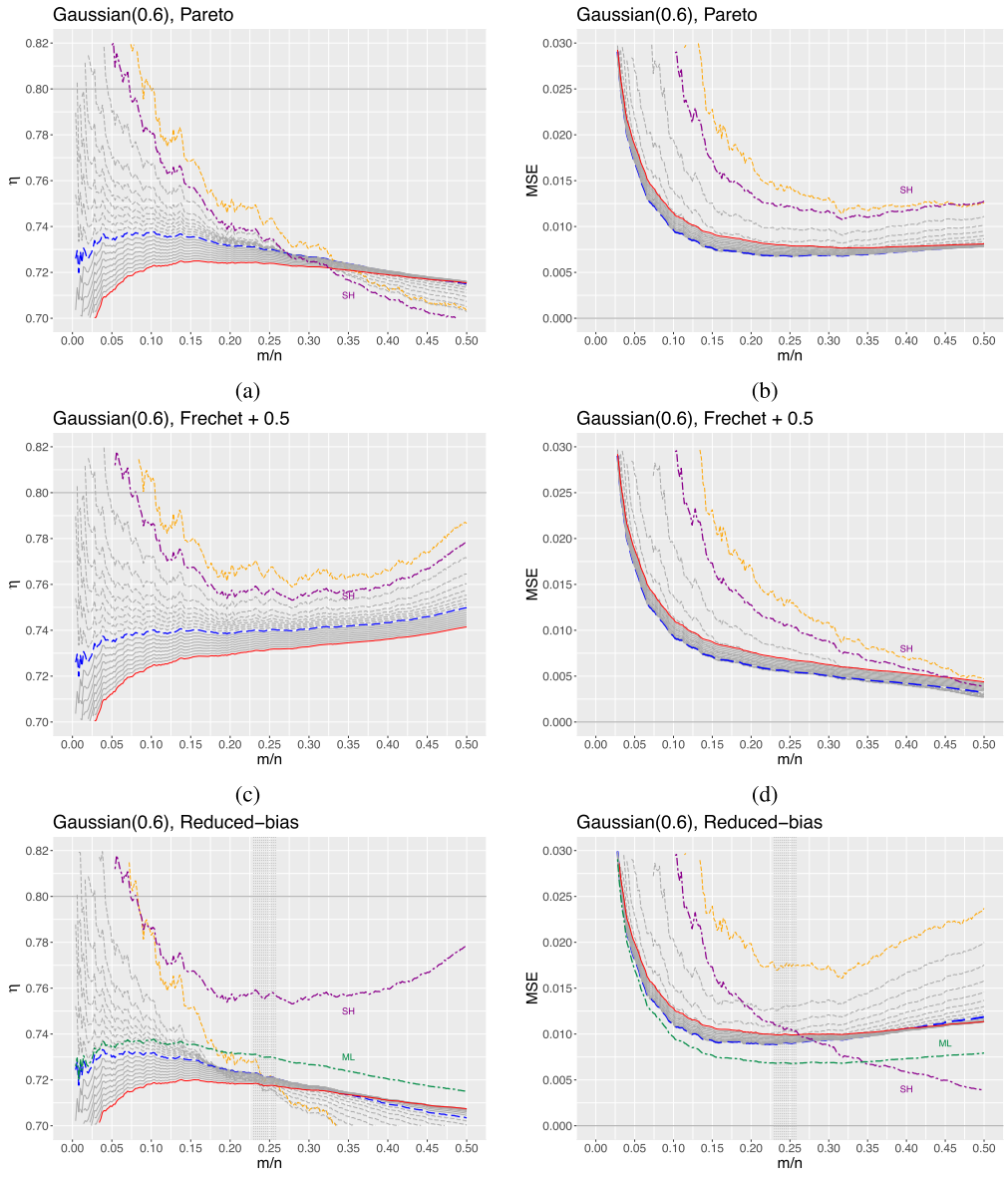

Fig 9. ***Gaussian copula with*** $\theta = 0.6$ $(\eta = 0.8, \tau = 0)$*: Identical plots (a)–(b) and (c)–(d) show the gradient estimator* $\hat{\eta}_q^{(S)}$ *and super-heavy (SH) dedicated estimator* $\hat{\eta}^{(SH)}$ *with either standard Pareto or corrected Fréchet marginals. Plots (e)–(f) display reduced-bias gradient estimation* $\tilde{\eta}_q^{(S)}(m, [\sqrt{m}])$ *and the numerical ML estimation from [2]. All plots: dashed lines represent* $q < 1$ *with orange for the lower bound* $q = 0.1$*; solid lines pertain to* $q > 1$ *with red for the upper bound* $q = 1.9$*; The Hill estimator* $(q = 1)$ *is shown in the long-dashed blue line.*

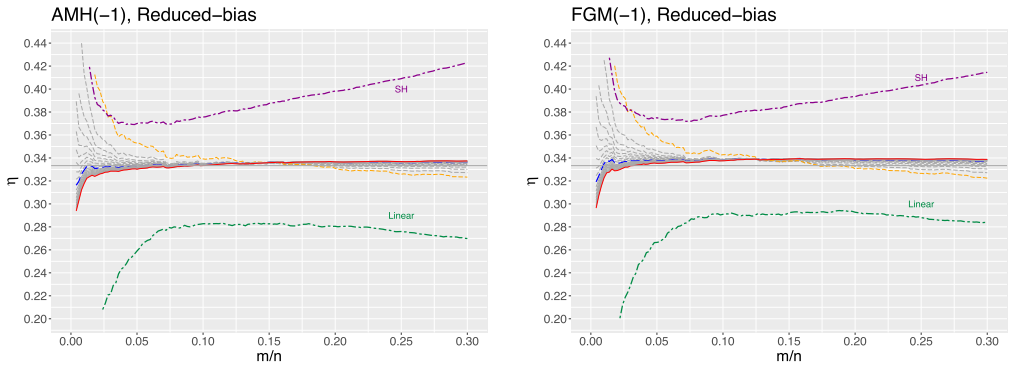

Fig 10. *Analogous plots to (e) from Figures 7 and 8 with the (also reduced-bias) Linear estimator from [2] in lieu of the ML estimator.*

$$\sqrt{r(n)}\left(\frac{1}{r(n)}\sum_{i=1}^{n}\mathbb{1}\Big\{\frac{k}{n}\Big(\frac{1}{-\log F_1^{(n)}(X_i)\vee -\log F_2^{(n)}(Y_i)}+\frac{1}{2}\Big)\geq x\Big\} - \frac{P\big\{X>V_1\big(\frac{n}{kx}-\frac{1}{2}\big),\, Y>V_2\big(\frac{n}{kx}-\frac{1}{2}\big)\big)}{\alpha(n/k)}\right)\longrightarrow W\big(x^{1/\eta}\big),$$

weakly in $D(0,\infty)$*, where* W *denotes standard Brownian motion.*

Proof. Consider the random pairs (X_i,Y_i), $i=1,\ldots,n$, of i.i.d. copies of (X,Y). The primary focus is on the direct empirical analogue to the copula survival function $C(1-x,1-y)$ such that $C(x,y)=P\big(1-F_1(X)<x,1-F_2(Y)<y\big)$. We firstly aim at the development for the sample counterpart of $C(x,x)$, i.e. the joint tail empirical process for $x>0$ in terms of

$$\begin{aligned}
&\sum_{i=1}^{n}\mathbb{1}\{X_i\geq X_{n-[kx]+1,n},\, Y_i\geq Y_{n-[kx]+1,n}\}\\
&=\sum_{i=1}^{n}\mathbb{1}\Big\{\big(-\log F_1^{(n)}(X_i)\big)^{-1}\wedge\big(-\log F_2^{(n)}(Y_i)\big)^{-1}\geq\big(-\log(1-[kx]/n)\big)^{-1}\Big\}\\
&=\sum_{i=1}^{n}\mathbb{1}\Big\{\frac{k}{n}E_i^{(n)}\geq-\frac{k/n}{\log\big(1-\frac{[kx]}{n}\big)}\Big\}=:\sum_{i=1}^{n}I_i^{(n)}(x),
\end{aligned}\tag{26}$$

with identically distributed random variables $E_i^{(n)}:=Z_i^{(n)}-1/2$ arising from those defined in (10).

Owing to the power-series $\sum_{n\geq0}\frac{t^n}{n+1}=-\log(1-t)/t$, for $|t|<1$, we can write the following stochastic inequalities for the partial sequence of indicator random variables in (26): there exists $\varepsilon>0$ such that

$$\begin{aligned}
\mathbb{1}\Big\{\frac{k}{n+1}E_i^{(n)}\geq\frac{1}{x}\Big(1-\frac{k}{n}\frac{x}{2}\big(1-\varepsilon(\frac{k}{n}x)^{\varepsilon}\big)\Big)\Big\}&\leq I_i^{(n)}(x)\\
&\leq\mathbb{1}\Big\{\frac{k}{n+1}E_i^{(n)}\geq\frac{1}{x}\Big(1-\frac{k}{n}\frac{x}{2}\big(1+\varepsilon(\frac{k}{n}x)^{\varepsilon}\big)\Big)\Big\},
\end{aligned}$$

uniformly in x on a compact set bounded away from zero. This gives information about total

boundedness in the desired approximation to $\overline{F}_Z^{(n)}$. Specifically, for every $\delta > 0$, we have that

$$\lim_{\varepsilon' \downarrow 0} \limsup_{n\to\infty} P\Big(\max_{1\le i\le n} \big| I_i^{(n)}(x) - I_{i,\varepsilon'}^{(n)}(x) \big| > 1-\delta, \ \text{ for } 0 \le x \le T \Big) = 0,$$

with intervening

$$I_{i,\varepsilon'}^{(n)}(x) := \mathbb{1}\Big\{ \frac{k}{n} E_i^{(n)} \ge \frac{1}{x}\Big(1 - \frac{k}{n}\frac{x}{2}\big(1 \pm \varepsilon'(\frac{k}{n}x)^{\varepsilon'}\big)\Big)\Big\}.$$

This implies in turn that, for each $n \in \mathbb{N}$ and for every $\delta > 0$, there exists an arbitrarily small $\varepsilon' > 0$ as before, such that

$$P\Big(\big| I_i^{(n)}(x) - I_{i,\varepsilon'}^{(n)}(x) \big| \le 1-\delta, \ \text{ for all } i = 1,\dots,m; \ \text{ for } 0 \le x \le T \Big) > 1-\delta,$$

i.e.,

$$P\Big(\max_{1\le i\le n} \big| I_i^{(n)}(x) - I_{i,\varepsilon'}^{(n)}(x) \big| = 0 \ \text{ for } \ 0 \le x \le T \Big) > 1-\delta.$$

The latter configures an equicontinuity property. In relation to (26), we note that

$$\frac{1}{n}\Big| \sum_{i=1}^n \big(I_i^{(n)}(x) - I_{i,\varepsilon'}^{(n)}(x) \big) \Big| \le \frac{1}{n}\sum_{i=1}^n \big| I_i^{(n)}(x) - I_{i,\varepsilon'}^{(n)}(x) \big| \le \max_{1\le i\le n} \big| I_i^{(n)}(x) - I_{i,\varepsilon'}^{(n)}(x) \big|, \tag{27}$$

whereupon a Skorokhod construction is invoked in order to ascertain that, as $n \to \infty$

$$\sup_{x\in[0,s_0]} \Big| \frac{1}{n}\sum_{i=1}^n I_i^{(n)}(x) - \frac{1}{n}\sum_{i=1}^n \mathbb{1}\Big\{ \frac{k}{n} Z_i^{(n)} \ge \frac{1}{x} \Big\} \Big| = o(1), \quad a.s. \tag{28}$$

Next, we deal with the relevant joint tail empirical process on the basis of the intermediate sequence $r(n)$ associated with the desired transform to common corrected unit-Fréchet margins. To this end, we define the tail empirical function:

$$1 - F_Z^{(n)}(x) := \frac{1}{n}\sum_{i=1}^n \mathbb{1}\Big\{ \alpha^{\leftarrow}\Big(\frac{r(n)}{n+1} \Big) Z_{n,n-i+1} \ge x \Big\}.$$

We proceed to evaluating its asymptotic mean and variance in two parts. Firstly, we employ the series expansion $t/(-\log(1-t)) = 1 - \sum_{n\ge1} \frac{t^n}{n+1}$, with $t := (k/n)x \to 0$, as $n \to \infty$, yielding:

$$\begin{aligned} \alpha\Big(\frac{n}{k} x^{-1} \Big) &= P\Big\{ -\log F_1(X) < -\log\Big(1 - \frac{k}{n}x\Big), \ -\log F_2(Y) < -\log\Big(1 - \frac{k}{n}x\Big) \Big\} \\ &= P\Big\{ \frac{1}{-\log F_1(X)} \wedge \frac{1}{-\log F_2(Y)} \ge \frac{n}{k} x^{-1} \Big(1 - \frac{k}{n}\frac{x}{2} - \big(\frac{k}{2n}\big)^2 \frac{x^2}{3}\big(1 + o(1)\big) \Big) \Big\}. \end{aligned}$$

Due to the absolute continuity of the marginal distributions F_i, $i = 1, 2$, the above implies with $Z := 1/2 + \big[(-\log F_1(X)) \vee (-\log F_2(Y)) \big]^{-1}$ (in the spirit of (9)-(10)) that, for $\varepsilon > 0$,

$$\lim_{n\to\infty} \sum_{i=1}^{[n/k]} P\Big\{ Z_i > \frac{n}{k} x^{-1} \Big(1 - \big(\frac{k}{n}\big)^2 (1+\varepsilon) \frac{x^2}{12} \Big) \Big\} < \infty\,.$$

Fixing $\delta = x(1 - \vartheta x^{-2+\vartheta}) > 0$, for some $\vartheta \in (0,1)$, we find that the corresponding truncated expected value of the identically distributed random variables Z_i is derived from the expansion (7), leading to

$$\int_\delta^\infty \overline{F}_Z(t)\,dt = \int_\delta^\infty \mathcal{L}(t)\Big(1-\exp\{-\frac{1}{t-1/2}\}\Big)^{1/\eta}\,dt = \int_\delta^\infty \mathcal{L}(t)\,t^{-1/\eta}\big(1+O(t^{-2})\big)^{-1/\eta}\,dt.$$

The latter integral exists and is finite for $0 < \eta < 1$ as a result of Karamata's theorem which states that the slowly varying function $\mathcal{L}$ (i.e. $\lim_{t\to\infty} \mathcal{L}(tx)/\mathcal{L}(t) = 1$, for all $x > 0$) eventually falters to the role of a mere constant in the integral [19, Appendix B]. We notice that this representation also removes the term of order t^{-1} on the right hand-side, whereby

$$\lim_{n\to\infty} \sup_{x>0} \Big|\alpha\Big(\frac{n}{k}x\Big) - \overline{F}_Z\Big(\frac{n}{k}x\Big)\Big| = 0. \tag{29}$$

The remainder of the proof is concerned with the regular variation of $\overline{\alpha}$ embedded in (25), equivalently the regular variation α defined in (6), as we draw near the concrete goal of attributing the limiting Gaussian process W^* to the properly normalised $1 - F_Z^{(n)}(x)$. From the approximation (29), we have tightness for each $x > 0$ in that

$$\begin{aligned}\mathbb{E}\big[1-F_Z^{(n)}(x)\big] &= \frac{1}{n}\sum_{i=1}^n P\Big(Z_i^{(n)} \geq \frac{n+1}{k}x\Big)\\ &\approx P\Big(1-F_1(X) \leq \frac{k}{nx},\, 1-F_2(Y) \leq \frac{k}{nx}\Big) = C\Big(\frac{k}{n}x^{-1}, \frac{k}{n}x^{-1}\Big)\end{aligned}$$

and, owing to asymptotic independence of X and Y, that

$$\begin{aligned}Var\Big(1-F_Z^{(n)}(x)\Big) &\approx \frac{1}{n^2}\sum_{i=1}^n P\Big(Z_i^{(n)} \geq \frac{n+1}{k}x\Big)\Big\{1-P\Big(Z_i^{(n)} \geq \frac{n+1}{k}x\Big)\Big\}\\ &\approx \frac{1}{n+1}C\Big(\frac{k}{n}x^{-1}, \frac{k}{n}x^{-1}\Big) = x^{-1/\eta}\,O\big(\frac{\alpha(k/n)}{n+1}\big),\end{aligned} \tag{30}$$

as $n \to \infty$. Secondly, we write the sum of the two building blocks for attaining the Gaussian process approximating (26) provided suitable normalisation in view of the convergence rate identified above:

$$\begin{aligned}\sqrt{r(n)}\Bigg\{\frac{1}{r(n)}\sum_{i=1}^{[r(n)]} \mathbb{1}\{X_i \geq X_{n-[kx]+1,n},\, Y_i \geq Y_{n-[kx]+1,n}\} - \frac{1-F_Z^{(n)}(x^{-1})}{\alpha(n/k)}\\ + \Bigg(\frac{1-F_Z^{(n)}(x^{-1})}{\alpha(k/n)} - \frac{P\big(1-F_1(X) \leq \frac{k}{n}x,\, 1-F_2(Y) \leq \frac{k}{n}x\big)}{\alpha(n/k)}\Bigg)\Bigg\}.\end{aligned}$$

Continuing along similar lines to the reasoning followed in the proof of Lemma 6.2 of [13] (see also [19], p. 268), we now handle the reduced uniform process embedded in (26) and apply Theorem 3.1 in [15] for establishing the weak convergence:

$$\sqrt{r(n)}\Bigg\{\frac{1}{r(n)}\sum_{i=1}^{[r(n)]} I_i^{(n)}(x) - \frac{P\big(1-F_1(X) < \frac{k}{n}x,\, 1-F_2(Y) < \frac{k}{n}x\big)}{\alpha(n/k)}\Bigg\}_{0<x<s_0} \xrightarrow{d} \Big\{W^*(x)\Big\}_{x>0}, \tag{31}$$

uniformly for every x bounded away from zero, where W^* is a Gaussian process with mean zero and covariance given by $E[W^*(x)W^*(y)] = c_1(x \wedge y)$, with $c_1(x) = \lim_{t\to\infty} \alpha(tx^{-1})/\alpha(t)$. Given (30), we are able to establish that $W^* \equiv \{W(x^{1/\eta})\}_{x>0}$ is indeed the residual process in

$$\sqrt{r(n)}\left\{\frac{1-F_Z^{(n)}(x)}{\alpha(n/k)} - \frac{P\big(1-F_1(X)<\frac{k}{n}x,\ 1-F_2(Y)<\frac{k}{n}x\big)}{\alpha(n/k)}\right\}_{0<x<\infty} \xrightarrow{d} \{W^*(x)\}_{x>0}. \quad (32)$$

The above converge holds weakly in $D(0,\infty)$. The two equalities in (27) conclude the proof. □

The invariance principle in Lemma 7.1 posits its flexibility for wider usages since it is not strictly encompassed by extreme value theory [cf. 4, p.346]. We have not yet really employed a proper domain-of-attraction condition at this point because the characterisation in (25) is not a sufficient condition for (2). The suite of worked examples in [35] is however a good testament to the significance placed on (25) within the context of asymptotic independence for bivariate extremes, where such a (dis)connection is expounded.

Dictated by the adopted marginal transform (*inter alia*, the unit Fréchet), the critical role of the slowly varying part of the relevant function α unravels. When delving into the process of obtaining Sibuya's coefficient $R(1,1) = \lambda$ in the limit of (3), we find via the expansion (7):

$$\begin{aligned}\lambda &= \lim_{t\to\infty} tP\{X > V_1(t),\ Y > V_2(t)\} = \lim_{t\to\infty} tP\Big\{\frac{1}{-\log F_1(X)} \wedge \frac{1}{-\log F_2(Y)} > t\Big\}\\ &= \lim_{t\to\infty} tP\{X > U_1\big(\frac{1}{1-e^{-1/t}}\big),\ Y > U_2\big(\frac{1}{1-e^{-1/t}}\big)\} = \lim_{t\to\infty}\big(t-\frac{1}{2}\big)\overline{q}(t),\end{aligned}$$

where $\overline{q}(t) := P\{1/(1-F_1(X)) > t,\ 1/(1-F_2(Y)) > t\}$. While addressing the Gaussian copula as part of Example 1 – and indeed the main point in the matter – the above ascertains $\lambda = 0$ and the identification $\eta = (1+\theta)/2$, both coincident with those yielded by (5). But it also surfaces that the slowly varying part has been tampered: the significance of extra term $\overline{q}(t)/2$ in the above beckons, as it can become asymptotically proportional to the slowly varying part of $t\overline{q}(t)$ for θ close to 1, i.e. none of these being dominant at governing the actual rate of convergence to $\lambda = 0$ (cf. Example 1). Keeping track of the marginal distributions, especially through the characterisation encompassing (2)–(3), is an aspect that is accentuated in Remark 6.2.2 of [19], emphasised in [10, 30] and alluded to in [14], but somewhat suppressed in [31].

Before getting underway with the proof of Theorem 4.1 which brings together Pareto and Fréchet marginal transforms to overcome the above caveats, we state and prove a purely analytic lemma.

Lemma 7.2. *Define* $V^\star(t) := V(t) + 1/2$. *Suppose* (15) *holds for some* $\eta \in (0,1]$ *and* $\tau \geq 0$. *Then,*

$$\lim_{t\to\infty} \frac{U(t)}{V^\star(t)} = 1.$$

Proof. With the already defined V and U, and with $U^\star(t) := U(t) + 1/2$, it holds that

$$\frac{U(t)}{V^\star(t)} = \frac{U^*(t)}{U^*\Big(\frac{1}{1-e^{-1/t}}\Big)} - \frac{1/2}{U^*\Big(\frac{1}{1-e^{-1/t}}\Big)}. \quad (33)$$

The latter term vanishes as $t \to \infty$.

With $\beta_\star(t) = t^{-\eta}A(t)U(t)$, whereby $|\beta_\star| \in RV_{-\tau}$, the second order condition (15) writes as

$$\lim_{t\to\infty} \frac{(tx)^{-\eta}U(tx) - t^{-\eta}U(t)}{\beta_\star(t)} = \frac{x^{-\tau}-1}{\tau}, \tag{34}$$

for all $x > 0$. The case of $\tau = 0$ in the limit reads in the continuity sense as $-\log x$. Defining $U^\star := U + 1/2$, this implies in turn that

$$\frac{(tx)^{-\eta}U^*(tx) - t^{-\eta}U^*(t)}{\beta_\star(t)} = \frac{x^{-\tau}-1}{\tau}\big(1+o(1)\big) + \frac{x^{-\eta}-1}{2U(t)A(t)}, \quad |U^*(t)A(t)| \in RV_{-\tau+\eta},$$

whence,

$$x^{-\eta}\frac{U^*(tx)}{U^*(t)} - 1 = \left\{\frac{x^{-\tau}-1}{\tau}\,A(t)\frac{U(t)}{U^*(t)} + \frac{x^{-\eta}-1}{2}\,\frac{1}{U^*(t)}\right\}\big(1+o(1)\big). \tag{35}$$

Additionally, we note that because $U \in RV_\eta$, $\eta > 0$, we have for any constant $c \in \mathbb{R}$,

$$\frac{U(tx)-c}{U(t)-c} = \frac{U(tx)}{U(t)}\Big(1 - \frac{c}{U(t)}\Big)^{-1}\big(1+o(1)\big) = \frac{U(tx)}{U(t)}\big(1+o(1)\big).$$

Setting $c = 1/2$ and taking $y = U^*(tx)/U^*(t)$ in the equality $1/(1+y) = 1 - y + y^2/(1+y)$, for $y \neq -1$, relation (35) gives, as $t \to \infty$,

$$x^{\eta}\frac{U^*(t)}{U^*(tx)} - 1 = -\frac{x^{-\eta}\frac{U^*(tx)}{U^*(t)} - 1}{x^{-\eta}\frac{U^*(tx)}{U^*(t)}} = \frac{x^{-\tau}-1}{-\tau}A(t)\big(1+o(1)\big) - \frac{x^{-\eta}-1}{2}\,\frac{1}{U^*(t)}\big(1+o(1)\big).$$

Finally, Taylor's expansion of length two for $y/(1-e^{-y})$ about zero ascertains the result:

$$\frac{U^*(t)}{U^*\Big(\frac{1}{1-e^{-1/t}}\Big)} = \Big(\frac{1/t}{1-e^{-1/t}}\Big)^{-\eta} = 1 - \frac{\eta}{2}\frac{1}{t} + \frac{\eta(1+3\eta)}{24}\frac{1}{t^2} + o\big(\max(A(t), 1/U^*(t))\big),$$

which, coupled with (33), enables to conclude that $U(t) \sim V^\star(t)$, as $t \to \infty$. □

Proof of Theorem 4.1: We start by recalling the pseudo-observables mirroring a sample of $Z_i^{(n)}$, $i = 1, \ldots, n$ independent copies of Z as those defined in (10),

$$Z_i^{(n)} := \left\{\Big(-\log\frac{R(X_i)}{n+1}\Big) \vee \Big(-\log\frac{R(Y_i)}{n+1}\Big)\right\}^{-1} + \frac{1}{2}.$$

Next, we define their associated empirical process, suitable normalised (outside) by $k/(n+1)$ and (inside) $\alpha(n/k) = r(n)/n$ that casts its focus on the tail region: for every $x > 0$,

$$\frac{\Gamma_{n,r(n)}(x)}{\alpha(k/n)} := \frac{1}{r(n)}\sum_{i=0}^{n-1} \mathbb{1}\Big\{\alpha^{\leftarrow}\Big(\frac{r(n)}{n+1}\Big)\,Z_i^{(n)} > x\Big\}.$$

Under the conditions of the theorem linking the ultra-intermediate sequence $r(n)$ with the intermediate $k = k(n)$, $k \to \infty$, we have that

$$\sqrt{r(n)}\left\{\frac{\Gamma_{n,r(n)}(x^{-1})}{\alpha(k/n)} - \frac{\alpha(k/n)}{\alpha(k/(nx))}\right\} \longrightarrow W(x,x),$$

weakly in $D\big((0,1]^2\big)$, where $W(x,x)$ is a Gaussian process with mean zero and covariance structure (note that $\lambda = 0$ is attached to asymptotic independence) given by

$$E[W(x,x)W(y,y)] = (x \wedge y)^{1/\eta} = x^{1/\eta} \wedge y^{1/\eta}.$$

[cf. 15, 30, 13]. On account of simplicity, we assume for now that

$$\sqrt{r(n)}\Big(\frac{\alpha(n/k\,x^{-1})}{\alpha(n/k)} - x^{1/\eta}\Big) \to 0,$$

as $n \to \infty$. We shall come back to this at the next stage in the proof since this where the deterministic approximation bias will stem from. Through application of Cramér's delta-method, we obtain that

$$\sqrt{r(n)}\,\Big\{\frac{\alpha\big(\frac{n+1}{k}\big)}{\Gamma_{n,r(n)}(x^{-1})} - x^{-1/\eta}\Big\} \longrightarrow -x^{-2/\eta}\,W(x,x),$$

weakly in $D\big((1,\infty]^2\big)$. By virtue of Vervaat's Lemma (cf. [19], Appendix A), the next asymptotic statement in terms of generalised inverses—including $\alpha^{\leftarrow}(r(n)/(n+1)) = k/(n+1)$ (cf. relation (25))—is proved to hold almost surely and uniformly on compact intervals: for every $\vartheta \le s \le s_0$ and all $\vartheta, s_0 > 0$, as $n \to \infty$,

$$\left|\sqrt{r(n)}\,\left\{\frac{U_n^Z\big(s/\alpha(\frac{n+1}{k})\big)}{\alpha^{\leftarrow}(\frac{r(n)}{n+1})} - s^{\eta} - \Big(\frac{U_Z\big(\frac{n+1}{r(n)}s\big)}{\alpha^{\leftarrow}(\frac{r(n)}{n+1})} - s^{\eta}\Big)\right\} - \eta s^{\eta+1}\overline{W}(s^{-\eta})\right| = o_p(1), \tag{36}$$

with $\overline{W}(s^{-\eta}) := W(s^{-\eta}, s^{-\eta})$. The second term in (36) is a deterministic approximation term specific to the original underlying distribution F (or $\overline{F}_Z$ in turn).

Now we give the tail quantile process a closer inspection: it develops as

$$\frac{U_n^Z\big(\alpha(\frac{n+1}{k})s^{-1}\big)}{\alpha^{\leftarrow}(\frac{r(n)}{n+1})} = \Big(\frac{1}{1-\Gamma_{n,r(n)}}\Big)^{\leftarrow}\Big(\frac{s}{\alpha((n+1)/k)}\Big) = \frac{k}{n+1}Z^{(n)}_{n,n-[r(n)s]} = \frac{k}{n+1}Q_n\Big(\frac{r(n)}{m}s\Big).$$

We refer the reader to definition (23) for the latter equality. It is thus fruitful to work with $U = \alpha^{\leftarrow}$ and V^* in tandem. Theorem 4.2 enables to write:

$$\begin{aligned} s^{\eta}\,\frac{k}{n+1}Q_n\Big(\frac{r(n)}{m}s\Big) &= \left\{\frac{V^{\star}\Big(\frac{n}{r(n)}s\Big)}{V^{\star}\Big(\frac{n}{r(n)}\Big)} + O\Big(B\big(\frac{n}{r(n)}\big)\Big)\right\}\frac{Z^{(n)}_{n,n-[r(n)s]}}{U\Big(\frac{n+1}{r(n)}\Big)} \\ &= \frac{Z^{(n)}_{n,n-[r(n)s]}}{V^{\star}\Big(\frac{n}{r(n)}\Big)}\,\frac{V^{\star}\Big(\frac{n}{r(n)}s\Big)}{U\Big(\frac{n+1}{r(n)}\Big)} + O\Big(B\big(\frac{n}{r(n)}\big)\Big). \end{aligned} \tag{37}$$

From Lemma 7.2, conforming to Thereom 4.2 with $x = x(t) \to 1$, we derive that $V^*\big(n/r(n)\big) = U\big((n+1)/r(n)\big)\big(1 + o(B(n/r(n)))\big)$. Hence, it follows immediately from Theorem 2.1 in [11] for a.s. of empirical quantiles to their theoretical analogues coupled with Theorem 4.2, that

$$s^{\eta}\,\frac{k}{n+1}Q_n\Big(\frac{r(n)}{m}s\Big) = 1 + O_p\Big(\frac{1}{\sqrt{r(n)}}\Big), \tag{38}$$

as $n \to \infty$. We now work towards strengthening (36) through the strong approximation for a weighted quantile process provided in [12], p.381 (see also Theorem 5.1.1 of [11]). To this

end, we note the following straightforward variant of Theorem B.3.10 of [19]: under condition (16), there exists a function $B_0(t) \sim B(t)$ (and so without loss of generality $B_0 \equiv B$) such that for any $\delta > 0$, there exits $t_0 = t_0(\delta)$ such that for all $tx > t_0$ and $x > 0$,

$$\left| \frac{V^\star(tx)/V^\star(t) - x^\eta}{B(t)} - x^\eta \frac{x^{-\tau_\star} - 1}{\tau_\star} \right| < \delta x^{-\tau_\star + \delta}. \tag{39}$$

The above is employed to deal with the second (deterministic) term in (36) by substituting $t = n/r(n) \to \infty$ and $x = s^{-1} > 0$. Additionally, we note that as part of Lemma 7.2 we showed that $U_Z(t)/V^\star(t) = 1 + o\big(B(t)\big)$.

Therefore, application of the triangle inequality upon (36) leads to the Gaussian limit of the desired tail quantile process: for all $0 < \varepsilon < 1/2$ and arbitrarily small $\vartheta > 0$,

$$\sup_{\vartheta \le s \le s_0} s^{1/2+\varepsilon} \left| \sqrt{r(n)} \left\{ s^\eta \, \frac{k}{n} U_n^Z\Big(\frac{s^{-1}}{\alpha\big(\frac{n+1}{k}\big)} \Big) - 1 \right\} + O\left(\sqrt{r(n)}\, B\Big(\frac{n}{r(n)} \Big) \right) - \eta s^{-1} \overline{W}(s^\eta) \right| = o_p(1), \tag{40}$$

as $n \to \infty$, with $\overline{W}(s^\eta) \stackrel{d}{=} W^*(s)$ where W^* is standard Brownian motion.

Finally, in order to obtain a total boundedness result that out-spans the natural support $[1/(n+1), 1 - 1/(n+1)]$ of the quantile process while covering the entire domain of the limit s^η, $s \in (0, 1]$ relevant for extremes, we borrow the arguments from [13, pp.271-272]. In order to bound the stochastic variation in the tails, a proper weight function $q(s) = s^{1/2+\varepsilon}$ which is essentially Chibisov-O'Reilly's specialised into the tail empirical processes [see 12, 15] is required. Since the law of iterated logarithm entails that

$$\lim_{\vartheta \downarrow 0} \sup_{0 < s \le \vartheta} s^{-1/2+\varepsilon} W^*(s) = 0 \quad a.s.,$$

the boundedness of the supremum of the relevant weighted process around zero comes from affording $\vartheta \downarrow 0$ in the above. Under the assumptions of theorem, we have for $m := [r(n)]$ that $V^\star(n/m)/V^\star\big(n/r(n)\big) = 1 + o\big(B(n/r(n))\big)$ (cf. (39) specialised in $x = x(t) \to 1$), thus ensuring that s_0 in (40) can run on such subsequence of $n/r(n)$ by a contiguity argument due to the continuous paths of Brownian motion. Therefore (40) holds weakly on $D(0, 1]$ and a.s. uniformly on a compact interval bounded away from zero. □

7.1. Proofs for Section 2

Proof of Theorem 2.1: The proof is made at slightly greater generality than the result in the theorem warrants. It starts off from a functional representation of the proposed estimators amenable to the tail quantile process encapsulated in (23). Namely, with $Z_{n,n-m} = Q_n(1)$ and $a/b \to -1$ (but not necessarily $a/b = -1$), we write:

$$\begin{aligned} & \sqrt{m} \left| \hat{\eta}_{a,b}^{(S)} - \hat{\eta}_{a,b} \right| \\ = \; & \frac{\sqrt{m}}{a} \left| - \left\{ \frac{n}{m} \int_{Z_{n,n-m}}^{\infty} \Big(\frac{x}{Z_{n,n-m}} \Big)^a \, dF_Z^{(n)}(x) \right\}^{-1} + \left\{ \frac{n}{m} \int_{T_{n,n-m}}^{\infty} \Big(\frac{x}{T_{n,n-m}} \Big)^a \, dF_T^{(n)}(x) \right\}^{-1} \right|. \end{aligned}$$

Because the asymptotic behaviour of the functional with respect to T is relatively well-known [see e.g. 13, 17], this pathway towards consistency, where the aim is to prove that the estimator

$\hat{\eta}_{a,b}^{(S)}$ enjoys to a certain extent identical properties to those from an already purveyed estimator, was found to simplify matters. Without loss of generality, we have the following representation for every $a < 1/\eta$:

$$\begin{aligned}
&\sqrt{m}\left\{\frac{n}{m}\int_{Z_{n,n-m}}^{\infty}\Big(\frac{x}{Z_{n,n-m}}\Big)^{a}\,dF_Z^{(n)}(x)-\Big(1+\frac{1}{a-1/\eta}\Big)\right\} \qquad (41)\\
=\;&\sqrt{m}\left\{\frac{n}{m(Z_{n,n-m})^{a}}\int_{Z_{n,n-m}}^{\infty}\int_{0}^{x}at^{a-1}\,dt\,dF_Z^{(n)}(x)-\Big(1+\frac{1}{a-1/\eta}\Big)\right\}\\
=\;&\sqrt{m}\,\frac{\big(V^{\star}(\frac{n}{m})\big)^{a}}{(Z_{n,n-m})^{a}}\left\{\Big(\frac{V^{\star}(\frac{n}{m})}{Z_{n,n-m}}\Big)^{a-1/\eta}\frac{1}{a-1/\eta}+\int_{1}^{\infty}\frac{n}{m}\Big(1-F_Z^{(n)}\big(xV^{\star}(\frac{n}{m})\big)\Big)a\,\frac{dx}{x^{1-a}}\right.\\
&\qquad\left.+\int_{\frac{Z_{n,n-m}}{V^{\star}(m/n)}}^{1}\frac{n}{m}\Big(1-F_Z^{(n)}\big(xV^{\star}(\frac{n}{m})\big)\Big)a\,\frac{dx}{x^{1-a}}-\frac{1}{a-1/\eta}\right\}.
\end{aligned}$$

Theorem 4.1 with $s=1$, determines the relevant building blocks:

$$\sqrt{m}\left\{\Big(\frac{V^{\star}(\frac{n}{m})}{Z_{n,n-m}}\Big)^{1/\eta}-1\right\}\frac{1}{a-1/\eta}+\sqrt{m}\,\{I+II\}\big(1+O_p(1/\sqrt{m}\big),$$

with I and II defined as

$$\begin{aligned}
I &:= \int_{1}^{\infty}\frac{n}{m}\Big(1-F_Z^{(n)}\big(xV^{\star}(\frac{n}{m})\big)\Big)a\,\frac{dx}{x^{1-a}},\\
II &:= \int_{\frac{Z_{n,n-m}}{V^{\star}(m/n)}}^{1}\frac{n}{m}\Big(1-F_Z^{(n)}\big(xV^{\star}(\frac{n}{m})\big)\Big)a\,\frac{dx}{x^{1-a}}.
\end{aligned}$$

The next step is show that $\sqrt{m}\,I=o_p(1)$. Since condition (16) holds locally uniformly, then (32) coupled with Theorem B.3.14 of [19] (see also Theorem 2.3.8 therein) entails that $\sqrt{m}B\big(\alpha^{\leftarrow}(m/n)\big)=O(1)$ whereby

$$|\sqrt{m}\,I|\le\sup_{x>0}\sqrt{m}\left|\frac{n}{m}\overline{F}_Z^{(n)}\Big(xV^{\star}(\frac{n}{m})\Big)-x^{-1/\eta}\right|\times\left(\int_{\frac{Z_{n,n-m}}{V^{\star}(m/n)}}^{1}at^{a-1}\,dt\right),$$

with the first terms on the right hand-side bounded with probability one and the second vanishing to zero, also ensured by Theorem 4.1. Theorem 4.1 in conjunction with the dominated convergence theorem ensures that $\sqrt{m}\,II=O_p(1)$. We now turn to (41) where we put $h(\eta)=1+(a-1/\eta)^{-1}$. A direct application of the δ-method determines, as $n\to\infty$,

$$\begin{aligned}
&\sqrt{m}\left\{\left(\frac{n}{m}\int_{Z_{n,n-m}}^{\infty}\Big(\frac{x}{Z_{n,n-m}}\Big)^{a}\,dF_Z^{(n)}(x)\right)^{-1}-\frac{1}{h(\eta)}\right\}\\
=\;&\frac{a\eta-1}{(a\eta+1-\eta)^2}\sqrt{m}\left\{1-\Big(\frac{V^{\star}(\frac{n}{m})}{Z_{n,n-m}}\Big)^{1/\eta}\right\}-\Big(\frac{a\eta-1}{a\eta+1-\eta}\Big)^{2}\sqrt{m}\,II+o_p(1).
\end{aligned}$$

It follows from Lemma 7.1 in conjunction with the dual strong approximation provided by Theorem 2.4.8 of [19] for the tail quantile process ascribed to standard Pareto marginals that, for every $\varepsilon>0$,

$$P\left(\lim_{n\to\infty}\left|\Big(\frac{V^{\star}(\frac{n}{m})}{Z_{n,n-m}}\Big)^{1/\eta}-\Big(\frac{U(\frac{n}{m})}{T_{n,n-m}}\Big)^{1/\eta}\right|>\varepsilon\right)=0.$$

Finally, for some function h_0 in a one-to-one correspondence with h, we have

$$\lim_{n\to\infty} E\left[\sqrt{m}\,|\hat{\eta}_q^{(S)} - \hat{\eta}_q|\right] \le \lim_{n\to\infty} E\left[\sqrt{m}\,|\hat{\eta}_q^{(S)} - h_0(\eta)|\right] + \lim_{n\to\infty} E\left[\sqrt{m}\,|\hat{\eta}_q - h_0(\eta)|\right] = 0,$$

and the result in the theorem is ascertained by application of Markov's inequality. □

The proof of Theorem 2.2 which we lay out next, provides a glimpse of what the above-mentioned function g_0 must look like for general class (11). Theorem 2.2 is solely reliant on the weighted approximation for the quantile process established in Theorem 4.1.

Proof of Theorem 2.2: The starting point is the distributional representation for the underpinning quantile process that features in Theorem 4.1. Defining

$$\{Z_n(s)\}_{0<s\le 1} := \left\{\left(\frac{Q_n(s)}{Q_n(1)}\right)^a - s^{-a\eta}\right\}_{0<s\le 1}, \tag{42}$$

we obtain through an appropriate Taylor's expansion that

$$\frac{Z_n(s)}{a\eta s^{-(a\eta+1)}} = \left(\frac{1}{\sqrt{m}}\big(W_n(s) - sW_n(1)\big) + B\big(\frac{n}{m}\big)\frac{s}{\eta}\frac{s^{\tau_\star}-1}{\tau_\star} + o_p\big(\max(s^{-1/2-\varepsilon}, 1)\big)\right)\left(1 + O_p\big(\frac{1}{\sqrt{m}}\big)\right),$$

where the o_p-term is uniform on a compact interval bounded away from zero. Hence the residual process, for every $0 < \varepsilon < 1/2$,

$$\Gamma^*_{n,m}(s) := s^{a\eta+1/2+\varepsilon}\left(\sqrt{m}\,Z_n(s) - a\eta s^{-(a\eta+1)}B_n(s) - \sqrt{m}B\big(\frac{n}{m}\big)\,a\eta s^{-a\eta}\frac{s^{\tau_\star}-1}{\eta\tau_\star}\right),$$

with $\{B_n(s)\}\overset{d}{=}\{W_n(s) - sW_n(1)\}$, $s \in (0,\,1]$ a sequence of Brownian bridges. This process is employed as follows, for any $a < 1/(2\eta)$:

$$\left|\int_0^1 \sqrt{m}\left\{\left(\frac{Q_n(s)}{Q_n(1)}\right)^a - s^{-a\eta}\right\} ds - \int_0^1 a\eta s^{-(a\eta+1)}B_n(s)\,ds \right.$$
$$\left. - \sqrt{m}\,B\big(\frac{n}{m}\big)\int_0^1 a s^{-a\eta}\frac{s^{\tau_\star}-1}{\tau_\star}\,ds\right| \le \sup_{0<s\le 1}\left|\Gamma^*_{n,m}(s)\right|\int_0^1 s^{-a\eta-1/2-\varepsilon}\,ds = o_p(1),$$

as $n \to \infty$. The first part of the theorem arises in a straightforward way via Cramér's δ-method with $-b/a \sim 1$, not depending on n. Finally, given that increments of a Gaussian process are independent normal random variables, then for large n the integral $\int_0^1 a\eta s^{-(a\eta+1)}B_n(s)\,ds$ as limit resolves to a sum of normals and eventually a normal random variable in itself. Finally, in order to derive the variance of the limiting normal random variable, it suffices to work with $Z(s) := \eta s^{-a\eta-1}B(s)$, $0 \le s \le 1$ (B denotes Brownian bridge in here), for which process $Var\big(\int_0^1 Z(s)\,ds\big) = E\big(\int_0^1\int_0^1 Z(s)Z(t)\,ds\,dt\big) = \eta^2\big((1-a\eta)^2(1-2a\eta)\big)^{-1}$. □

7.2. Proofs for Section 3

Proof of Theorem 4.2: The crux of the proof consists in translating second order regular variation into extended regular variation of an appropriate function related to the former.

With the already defined quantile function V, such that $V(t) = U\big(1/(1-e^{-1/t})\big)$, and $B(t) = t^{-\eta}\beta(t)U(t)$ whereby $|B| \in RV_{-\tau}$, we have that

$$\begin{aligned}\frac{(tx)^{-\eta}V(tx) - t^{-\eta}V(t)}{B(t)} &= \frac{(tx)^{-\eta}V(tx) - t^{-\eta}U(t)}{B(t)} - \frac{t^{-\eta}V(t) - t^{-\eta}U(t)}{B(t)} \\ &= \frac{(tx)^{-\eta}U\big(1/(1-e^{-1/(tx)})\big) - t^{-\eta}U(t)}{B(t)} - t^{-\eta}\frac{U\big(1/(1-e^{-1/t})\big) - U(t)}{B(t)}.\end{aligned}$$

Owing to the second order regular variation for U with index $\eta > 0$ encapsulated in (34), which holds locally uniformly for $x > 0$, and by noting that $x(t) = t^{-1}/(1-e^{-1/t}) \to 1$, as $t \to \infty$, we find the representation:

$$\frac{(tx)^{-\eta}V(tx) - t^{-\eta}V(t)}{B(t)} = \frac{\left(\frac{1/t}{1-e^{-1/(tx)}}\right)^{-\tau} - 1}{\tau}\big(1 + o(1)\big) - t^{-\eta}\frac{\left(\frac{1/t}{1-e^{-1/t}}\right)^{-\tau} - 1}{\tau}\big(1 + o(1)\big).$$

Now it is only a matter of applying Taylor's expansion followed by judicious manipulation of the terms on the right hand-side so that, for all $x > 0$, a suitable representation of subsequent order is available:

$$\frac{x^{-\eta}\frac{V(tx)}{V(t)} - 1}{\beta(t)\,U(t)/V(t)} = \frac{x^{-\tau} - 1}{\tau} - \frac{1}{2t}x^{-\tau-1} + o\big(\frac{1}{t}\big). \tag{43}$$

For tackling $V^\star(t) = V(t) + 1/2$, we plug in the above into the extended regular variation development. Hence, as $t \to \infty$,

$$\begin{aligned}\frac{(tx)^{-\eta}V^\star(tx) - t^{-\eta}V^\star(t)}{B(t)} &= \frac{(tx)^{-\eta}V(tx) - t^{-\eta}V(t)}{B(t)} + \frac{t^{-\eta}}{2}\frac{x^{-\eta} - 1}{B(t)} \\ &= \frac{x^{-\tau} - 1}{\tau} + \frac{t^{-\eta}}{B(t)}\frac{x^{-\eta} - 1}{2} - \frac{1}{t}\frac{x^{-\tau-1}}{2} + o(t^{-1}) + o(1),\end{aligned}$$

for all $x > 0$. Given that in the present setting of asymptotic independence the range $\eta < 1$ is required, the third order term in (43) becomes negligible (note that $|t^\eta B(t)| \in RV_{-\tau+\eta}$ and $\eta < 1 + \tau$, $\tau > 0$), thus resulting in the following expansion for $V^\star$:

$$\frac{x^{-\eta}\frac{V^\star(tx)}{V^\star(t)} - 1}{\beta(t)\,U(t)/V^\star(t)} = \frac{x^{-\tau} - 1}{\tau} + \frac{1}{\beta(t)U(t)}\frac{x^{-\eta} - 1}{2} + o\Big(\frac{1}{\beta(t)U(t)}\Big).$$

Under the conditions of the theorem, Lemma 7.2 above now enables replacement of U with $V^\star$ everywhere and the desired result detailing second order regular variation for $V^\star$ in full thus arises. Specifically,

$$\frac{V^\star(tx)}{V^\star(t)} = \beta(t)\,x^\eta\frac{x^{-\tau} - 1}{\tau} + \frac{\eta}{2}\frac{1}{V^\star(t)}\,x^\eta\frac{x^{-\eta} - 1}{\eta} + o\big(\beta(t)\big) + o\Big(\frac{1}{V^\star(t)}\Big) \qquad (t \to \infty).$$

□

Proof of Theorem 3.1: The first part of the theorem follows from Theorem 2.1 in conjunction with Theorem 4.2. In particular, we note that (15) implies (16) through the appropriate adaptation of the auxiliary function of second order. In relation to the second part of the

theorem, with $\widetilde{\eta}_a \equiv \widetilde{\eta}_a(m, m^*)$ and $\hat{\eta}_a^{(S)} \equiv \hat{\eta}_a^{(S)}(m)$, we decompose $\sqrt{m}\,(\widetilde{\eta}_a - \eta)$ into the sum:

$$\sqrt{m}\left\{\hat{\eta}_a^{(S)}\left(1-\hat{\beta}\Big(\frac{n}{m}\Big)\frac{1-a\,\hat{\eta}_a^{(S)}}{1-a\,\hat{\eta}_a^{(S)}+\hat{\tau}}\right)-\eta\right\}+\sqrt{m}\,\frac{1}{E^{\star}_{n,n-m^*}}\,\frac{\hat{\eta}_a^{(S)}}{2}\,\frac{1-a\,\hat{\eta}_a^{(S)}}{1-a\,\hat{\eta}_a^{(S)}+\hat{\tau}}. \tag{44}$$

Subsequently, the methodology devised in [8] ascertains that the residual bias in the first $\sqrt{m}$-term is of lower order than that associated with the relevant assumption $\sqrt{m}\,B(n/m) = O(1)$, as $n \to \infty$, namely

$$\sqrt{m}\left\{\hat{\eta}_a^{(S)}\left(1-\hat{\beta}\Big(\frac{n}{m}\Big)\frac{1-a\,\hat{\eta}_a^{(S)}}{1-a\,\hat{\eta}_a^{(S)}+\hat{\tau}}\right)-\eta\right\} = N_a + o_p\big(\sqrt{m}\,B(n/m)\big),$$

where N_a is a Normal random variable with mean zero and variance $\sigma_a^2 > 0$, the same variance as in Theorem 2.2. For concluding the proof, it only remains to show that the second term in (44) becomes negligible with probability tending to one. Thus, we apply the equality $1/(x+1) = 1 - x + x^2/(1+x)$, valid for all $x \neq -1$, through the identification

$$x \equiv x(k,m) := \Big(\frac{k}{n+1}Q_n(1) - 1\Big) = O_p(1/\sqrt{m}), \tag{45}$$

(cf. (23), (36) and (38) with $s = 1$ and $m = [r(n)]$), specifically in that

$$\begin{aligned} 0 < \sqrt{m}\,\frac{1}{E^{\star}_{n,n-m}} &= \frac{k}{n+1}\left\{\sqrt{m}\Big(\frac{k}{n+1}Q_n(1)-1\Big)+\sqrt{m}\,\frac{\Big(\frac{k}{n+1}Q_n(1)-1\Big)^2}{\frac{k}{n+1}Q_n(1)}\right\} \\ &= \frac{k}{n+1}\big\{O_p(1)+o_p(1)\big\}, \end{aligned}$$

thus surrendering the already anticipated $o_p(1)$. This convergence can well be taken to a quicker rate by using $m^* = \sqrt{m}$ in place of m for the random threshold $Q_n(1)$ so that the associated k^* in view of (45) is made small enough to ensure faster convergence $k^*/(n+1) \to 0$. This accounts for the precise result in the theorem. □

Appendix: Considerations on the half-shift to the marginal rank transform

Let G_0 define the class of simple max-stable distributions, i.e., the class of distribution functions arising in the limit (1), given by

$$G_0(x,y) := G\Big(\frac{x^{\gamma_1}-1}{\gamma_1}, \frac{x^{\gamma_2}-1}{\gamma_2}\Big),$$

signifying that both marginal distributions are fixed as unit Fréchet: $G_0(x,\infty) = G_0(\infty,x) = \exp(-1/x)$, for $x > 0$. In order to address pre-limit extremal behaviour, in terms of the initial (unknown) distribution F, we consider its pertaining marginal quantile functions $V_i := \big(1/-\log F_i\big)^{\leftarrow}$, $i = 1, 2$, where $^{\leftarrow}$ indicates the generalised inverse function, and assume that, for every $(x,\ y)$ such that $0 < G_0(x,y) < 1$,

$$\lim_{t\to\infty}\frac{t\left\{1-F\big(V_1(tx-1/2), V_2(ty-1/2)\big)\right\}}{t\left\{1-F\big(V_1(t-1/2), V_2(t-1/2)\big)\right\}} = \frac{-\log G_0(x,y)}{-\log G_0(1,1)} =: S(x,y) \tag{46}$$

exists with $S(x, y)$ positive. If $1 - F\big(V_1(t - 1/2), V_2(t - 1/2)\big)$ is regularly varying at infinity with index -1, then $F \in \mathcal{D}(G)$, in which case the marginal distribution of G will have the previously stated standard von-Mises form G_i, $i = 1, 2$; we have convergence of the actual (unknown) marginal distributions F_i to their corresponding G_i, $i = 1, 2$, with respective extreme value indices γ_1 and γ_2, and (equivalently) the following condition of extended regular variation holds with their respective quantile functions V_i, $i = 1, 2$: there exist positive functions a_1, a_2 such that

$$\lim_{t\to\infty} \frac{V_i(tx) - V_i(t)}{a_i(t)} = \frac{x^{\gamma_i} - 1}{\gamma_i},$$

for all $x > 0$. For the purpose of our results, it is worth noting that the above extreme value condition on the marginal distributions of F remains valid through the composition $V_i(g)$, with any positive function $g \in RV_1$. By continuity of G and monotonicity of F, the convergence in (1) is ascertained to hold uniformly. Hence, it is possible to replace x and y with $x(n)$ and $y(n)$ respectively defined as

$$x(n) := \frac{V_1(nx - 1/2) - b(n)}{a(n)} \underset{n\to\infty}{\longrightarrow} \frac{x^{\gamma_1} - 1}{\gamma_1}$$
$$y(n) := \frac{V_2(ny - 1/2) - d(n)}{c(n)} \underset{n\to\infty}{\longrightarrow} \frac{y^{\gamma_2} - 1}{\gamma_2},$$

in the analogous uniform convergence that preserved while taking the logarithm of both hand-sides of (1), which dictates $b(n) = V_1(n - 1/2)$ and $d(n) = V_2(n - 1/2)$ such that:

$$\lim_{n\to\infty} n \left\{1 - F\big(V_1(nx - 1/2), V_2(ny - 1/2)\big)\right\} \tag{47}$$
$$= \lim_{n\to\infty} n \left\{1 - F\Big(V_1(n - 1/2) + a_1(n)\frac{x^{\gamma_1} - 1}{\gamma_1},\, V_2(n - 1/2) + a_2(n)\frac{y^{\gamma_2} - 1}{\gamma_2} \Big)\right\}$$
$$= -\log G_0(x, y). \tag{48}$$

In particular (and w.l.g.), the joint convergence above entails that the marginal distribution of $M_{X,n}$ converges to a unit Fréchet distribution provided the above marginal standardisation:

$$\lim_{n\to\infty} F^n\big(V_1(nx - 1/2), \infty\big) = G_0(x, \infty) = G\Big(\frac{x^{\gamma_1} - 1}{\gamma_1}, \infty\Big) = \exp\{-x^{-1}\}.$$

By the other hand, heeding the definition of the quantile function V_1, we find via Taylor's expansion (7) upon the marginal tail distribution that, for every $x > 0$,

$$\lim_{n\to\infty} F^n\big(V_1(nx - 1/2), \infty\big) = \lim_{n\to\infty} \Big[1 - P\Big\{\frac{1}{-\log F_1(X)} > \frac{x^{-1}}{n}\frac{nx}{-\log\big(1 - \frac{x^{-1}}{n}\big)}\Big\}\Big]^n$$
$$= \lim_{n\to\infty} \Big[1 - P\Big\{F_1(X) > 1 - \frac{x^{-1}}{n}\Big\}\Big]^n = \lim_{n\to\infty} \Big(1 - \frac{x^{-1}}{n}\Big)^n = e^{-x^{-1}},$$

thus uncovering the hallmark unit Fréchet marginals of the limiting simple max-stable distribution G_0. Therefore, the development in (47)–(48) strengthens the result that the transformation of the marginal distributions to unit Fréchet with a shift by $1/2$ makes it possible to eschew the influence of the margins and promptly draw extremal behaviour from the dependence structure on its own. Through this construct, we arrive at the prime extreme value condition:

$$\lim_{n\to\infty} F^n\big(V_1(nx - 1/2), V_2(ny - 1/2)\big) = G_0(x, y),$$

where the limit G_0 is the class of simple max-stable distributions.

Acknowledgements

We are thankful to the Editor, the Associate Editor and the anonymous referees for their constructive comments, which have helped improve the paper considerably. We are grateful to the Ghana Meteorological Authority for providing the daily precipitation dataset used in this study.

Funding

Cláudia Neves gratefully acknowledges support from UKRI-EPSRC Innovation Fellowship grant EP/S001263/1 and EP/S001263/2. Her work is also partly supported by FCT – Fundação para a Ciência e a Tecnologia, I.P., under CEAUL Research Unit, UID/00006/2025, DOI: https://doi.org/10.54499/UID/00006/2025, where she is Visiting Associate Professor. This work stems from the PhD research of Jennifer Israelsson, who received funding from the UKRI-EPSRC Centre for Doctoral Training in Mathematics of Planet Earth, grant EP/L016613/1.

References

[1] ALVES, I. F., DE HAAN, L. and NEVES, C. (2009). A test procedure for detecting super-heavy tails. *Journal of Statistical Planning and Inference* **139** 213–227. https://doi.org/10.1016/j.jspi.2008.04.026 MR2473999

[2] BEIRLANT, J., DIERCKX, G. and GUILLOU, A. (2011). Bias-reduced estimators for bivariate tail modelling. *Insurance: Mathematics and Economics* **49** 18–26. https://doi.org/10.1016/j.insmatheco.2011.01.010 MR2811890

[3] BEIRLANT, J. and VANDEWALLE, B. (2002). Some comments on the estimation of a dependence index in bivariate extreme value statistics. *Statistics & Probability Letters* **60** 265–278. https://doi.org/10.1016/S0167-7152(02)00281-X MR1942930

[4] BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J. and TEUGELS, J. (2004). *Statistics of Extremes: Theory and Application*. Wiley Series in Probability and Statistics. MR2108013

[5] DI BERNARDINO, E., MAUME-DESCHAMPS, V. and PRIEUR, C. (2013). Estimating a bivariate tail: A copula based approach. *Journal of Multivariate Analysis* **119** 81–100. https://doi.org/10.1016/j.jmva.2013.03.020 MR3061417

[6] BÜCHER, A. and ZHOU, C. (2021). A horse race between the block maxima method and the peak–over–threshold approach. *Statistical Science* **36** 360–378. https://doi.org/10.1214/20-STS795 MR4293095

[7] BÜCHER, A., VOLGUSHEV, S. and ZOU, N. (2019). On second order conditions in the multivariate block maxima and peak over threshold method. *Journal of Multivariate Analysis* **173** 604–619. https://doi.org/10.1016/j.jmva.2019.04.011 MR3959186

[8] CAEIRO, F., GOMES, M. I. and PESTANA, D. D. (2005). Direct reduction of bias of the classical Hill estimator. *Revstat* **3** 111–136. MR2259357

[9] COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer London, London. https://doi.org/10.1007/978-1-4471-3675-0_1 MR1932132

[10] COLES, S. G. and WALSHAW, D. (1994). Directional modelling of extreme wind speeds. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **43** 139–157.
[11] CSÖRGŐ, M. (1983). *Quantile Processes with Statistical Applications*. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611970289 MR0745130
[12] CSÖRGŐ, M. and HORVÁTH, L. (1993). *Weighted Approximations in Probability and Statistics*. Wiley Series in probability and mathematical statistics. MR1215046
[13] DRAISMA, G., DREES, H., FERREIRA, A. and DE HAAN, L. (2004). Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli* **10** 251–280. https://doi.org/10.3150/bj/1082380219 MR2046774
[14] EASTOE, E. F. and TAWN, J. A. (2012). Modelling the distribution of the cluster maxima of exceedances of subasymptotic thresholds. *Biometrika* **99** 43–55. https://doi.org/10.1093/biomet/asr078 MR2899662
[15] EINMAHL, J. H. J. (1997). Poisson and Gaussian approximation of weighted local empirical processes. *Stochastic Processes and their Applications* **70** 31–58. https://doi.org/10.1016/S0304-4149(97)00055-0 MR1472958
[16] EINMAHL, J. H. J., FERREIRA, A., DE HAAN, L., NEVES, C. and ZHOU, C. (2022). Spatial dependence and space–time trend in extreme events. *The Annals of Statistics* **50** 30–52. https://doi.org/10.1214/21-AOS2067 MR4383006
[17] GOEGEBEUR, Y. and GUILLOU, A. (2012). Asymptotically unbiased estimation of the coefficient of tail dependence. *Scandinavian Journal of Statistics* **40** 174–189. https://doi.org/10.1111/j.1467-9469.2012.00800.x MR3024038
[18] GONG, Y. and HUSER, R. (2022). Asymmetric tail dependence modeling, with application to cryptocurrency market data. *The Annals of Applied Statistics* **16** 1822–1847. https://doi.org/10.1214/21-AOAS1568 MR4455901
[19] DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory: An Introduction*. Springer. MR2234156
[20] DE HAAN, L. and ZHOU, C. (2011). Extreme residual dependence for random vectors and processes. *Advances in Applied Probability* **43** 217–242. https://doi.org/10.1239/aap/1300198520 MR2761155
[21] HALL, P. and WELSH, A. H. (1985). Adaptive estimates of parameters of regular variation. *The Annals of Statistics* **13** 331–341. https://doi.org/10.1214/aos/1176346596 MR0773171
[22] HILL, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3** 1163–1174. https://doi.org/10.1214/aos/1176343247 MR0378204
[23] HUSER, R., OPITZ, T. and THIBAUD, E. (2017). Bridging asymptotic independence and dependence in spatial extremes using Gaussian scale mixtures. *Spatial Statistics* **21** 166–186. https://doi.org/10.1016/j.spasta.2017.06.004 MR3692183
[24] ISRAELSSON, J., BLACK, E., NEVES, C., TORGBOR, F. F., GREATREX, H., TANU, M. and LAMPTEY, P. N. L. (2020). The spatial correlation structure of rainfall at the local scale over southern Ghana. *Journal of Hydrology: Regional Studies* **31**. https://doi.org/10.1016/j.ejrh.2020.100720
[25] KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer. MR2724368

[26] LEDFORD, A. W. and TAWN, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* **83** 169–187. https://doi.org/10.1093/biomet/83.1.169 MR1399163

[27] LEDFORD, A. W. and TAWN, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **59** 475–499. https://doi.org/10.1111/1467-9868.00080 MR1440592

[28] LEHMANN, E. L. and CASELLA, G. (1998). *Theory of Point Estimation*, 2 ed. Springer Science & Business Media, New York. MR1639875

[29] NEVES, C. (2009). From extended regular variation to regular variation with application in extreme value statistics. *Journal of Mathematical Analysis and Applications* **355** 216–230. https://doi.org/10.1016/j.jmaa.2009.01.058 MR2514463

[30] PENG, L. (1999). Estimation of the coefficient of tail dependence in bivariate extremes. *Statistics & Probability Letters* **43** 399–409. https://doi.org/10.1016/S0167-7152(98)00280-6

[31] POON, S.-H., ROCKINGER, M. and TAWN, J. (2003). Modelling extreme-value dependence in international stock markets. *Statistica Sinica* **13** 929–953. MR2026056

[32] RAMOS, A. and LEDFORD, A. (2009). A new class of models for bivariate joint tails. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71** 219–241. https://doi.org/10.1111/j.1467-9868.2008.00684.x MR2655531

[33] REISS, R.-D. and THOMAS, M. (2007). *Statistical Analysis of Extreme Values*. Birkhäuser Basel. https://doi.org/10.1007/978-3-7643-7399-3 MR1464696

[34] SANG, H. and GELFAND, A. E. (2009). Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics* **16** 407–426. https://doi.org/10.1007/s10651-007-0078-0 MR2749848

[35] SCHLATHER, M. (2001). Examples for the coefficient of tail dependence and the domain of attraction of a bivariate extreme value distribution. *Statistics & Probability Letters* **53** 325–329. https://doi.org/10.1016/S0167-7152(01)00090-6 MR1841635

[36] SHERMAN, M. (2010). *Spatial Statistics and Spatio-Temporal Data: Covariance Functions and Directional Properties*. Wiley Series in Probability and Statistics. MR2815783

[37] SIBUYA, M. (1960). Bivariate extreme statistics, I. *Annals of the Institute of Statistical Mathematics* **11** 195–210. https://doi.org/10.1007/BF01682329 MR0115241

[38] SIGAUKE, C., RAVELE, T. and JHAMBA, L. (2022). Extremal Dependence Modelling of Global Horizontal Irradiance with Temperature and Humidity: An Application Using South African Data. *Energies* **15**. https://doi.org/10.3390/en15165965

[39] SKLAR, A. (1959). Fonctions de répartition à *n* dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris* **8** 229–231. MR0125600

[40] SUBRAMANYAM, K. (1990). Some comments on positive quadrant dependence in three dimensions. In *IMS Lecture Notes Monogr. Ser* **16** 443–449. MR1193997

[41] TILLOY, A., MALAMUD, B. D., WINTER, H. and JOLY-LAUGEL, A. (2020). Evaluating the efficacy of bivariate extreme modelling approaches for multi-hazard scenarios. *Natural Hazards and Earth System Sciences* **20** 2091–2117. https://doi.org/10.5194/nhess-20-2091-2020