On the relation between extremal dependence and concomitants

Amir Khorrami Chokami, Marie Kratz

ESSEC BUSINESS SCHOOL

Enlighten. Lead. Change.

On the relation between extremal dependence and concomitants

Amir Khorrami Chokami, Marie Kratz

ESSEC RESEARCH CENTER

WORKING PAPER 2301

JANUARY 16, 2023

On the relation between extremal dependence and concomitants

AMIR KHORRAMI CHOKAMI† AND MARIE KRATZ‡

† University of Torino and Collegio Carlo Alberto, Italy

‡ ESSEC Business School, CREAR, France

January 16, 2023

ABSTRACT

The study of concomitants has recently met a renewed interest due to its applications in selection procedures. For instance, concomitants are used in ranked-set sampling, to achieve efficiency and reduce cost when compared to the simple random sampling. In parallel, the search for new methods to provide a rich description of extremal dependence among multiple time series has rapidly grown, due also to its numerous practical implications and the lack of suitable models to assess it. Here, our aim is to investigate extremal dependence when choosing the concomitants approach. In this study, we show how the extremal dependence of a vector (X, Y) impacts the asymptotic behavior of the maxima over subsets of concomitants. Furthermore, discussing the various conditions and results, we investigate how transformations of the marginal distributions of X and Y influence the degeneracy of the limit.

Keywords: Asymptotic Theorems; Concomitants; Copula; (Tail) Dependence; Extremes; Gaussian; Logistic; Maxima; Order Statistics; Pareto; Slowly/Regularly Varying Functions; Tail Equivalence; Weak Convergence.

2010 AMS classification: 60G70; 62E20; 62G30.

CONTENTS

1 INTRODUCTION

Let (X_i, Y_i), $i = 1, \dots, n$, be a sequence of independent and identically distributed (i.i.d.) bivariate random vectors with common cumulative distribution (cdf) function $F(x, y)$. We define the *concomitants of order statistics* as follows. Arrange the X variates in ascending order, thus obtaining the order statistics $X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(n)}$ for the X variable. Then, the Y variable corresponding to the i-th order statistic $X_{(i)}$ is called the concomitant of the i-th order statistic, and is denoted by $Y_{[i]}$. For example, if the maximum of the X variates is $X_{(n)} \equiv X_3$, then its concomitant is $Y_{[n]} \equiv Y_3$. Concomitants of order statistics are of notable interest in practical applications: It is worth mentioning their utility in problems concerning the estimation of parameters for multivariate datasets affected by type II censoring, in the development of selection procedures (where it is more convenient to measure a characteristic linked to a measure of interest), in the setting of ranked-set sampling schemes. In particular, the latter deals with situations where the measurement of the variable of interest is time-consuming or expensive, but the ranking of a set of observations linked to the target variable is easy to be performed. For such cases, ranked-set sampling is a better alternative with respect to the simple random sampling, both in terms of efficiency and reduced cost, hence this renewed attention on the study of concomitants. We report the seminal works Dell and Clutter (1972), Stokes and Sager (1988), and more recently Wang et al. (2006), Shao (2010), Balci et al. (2013), Wang et al. (2016), Wang et al. (2017), Zamanzade and Wang (2018), Ozturk (2019) among others.
Our motivation is to investigate extremal dependence when choosing the concomitant approach. Looking at the literature, this topic has been tackled in the 90's. One can mention the seminal paper Nagaraja and David (1994) where, given a number k, they study the distribution of the maximum of the concomitants $V_1 = \max\left(Y_{[n]}, \dots, Y_{[n-k+1]}\right)$. Finite and asymptotic results are given in both cases k a fixed integer (*extreme case*) and $k = [np]$, $p \in (0, 1)$ (*quantile case*). This leads to the natural question of how much V_1 is close to the maximum $Y_{(n)}$, a fundamental problem in practical applications, tackled in Joshi and Nagaraja (1995). In their work, authors propose to link this question to the study of the joint distribution of two maxima of concomitants. More precisely, given an i.i.d. sequence of bivariate random vectors with parent random vector (X, Y), define, for $k \geqslant 1$,

$$V_1 = \max\left(Y_{[n]}, \dots, Y_{[n-k+1]}\right) \quad \text{and} \quad V_2 = \max\left(Y_{[n-k]}, \dots, Y_{[1]}\right). \tag{1}$$

Joshi and Nagaraja (1995) introduced the random variable $W_k := V_1/Y_{(n)}$ and proved that its cdf satisfies

$$\mathbb{P}(W_k \leqslant w) = \mathbb{P}(V_1 \leqslant wV_2,\, V_2 > 0)\,,$$

for which they need to compute the joint distribution of (V_1, V_2).
Before briefly reporting their result, let us introduce some notations that will be used throughout the paper.

Let $G(z;\mu,\sigma,\xi)$, for $\mu,\xi\in\mathbb{R},\sigma>0$, denote a Generalised Extreme Value distribution (GEV). A distribution function F is said to belong to the *Maximum Domain of Attraction* (MDA) of G, written as $F\in\mathscr{D}(G)$, if there exist normalizing sequences of constants $a_n>0$ and $b_n\in\mathbb{R}$ such that $F^n(a_nz+b_n)\to G(z)$, as $n\to\infty$ for any $z\in\Gamma_G=\{z\colon \xi(z-\mu)/\sigma>0\}$. The structure of a MDA is studied in Fisher and Tippett (1928), Gnedenko (1943), among others, and more recently, with another view, in Leonetti and Khorrami Chokami (2022). Sufficient conditions for a distribution F to belong to the MDA of a certain GEV G are the so called Von Mises conditions: We refer to Resnick (1987, Propositions 1.15 - 1.17) for a thorough description and provide a brief recall in Appendix C for the paper to be self-contained.

A measurable function $h\colon\mathbb{R}_+\mapsto\mathbb{R}_+$ is regularly varying at infinity with index a (written $h\in RV_a$) if for $x>0$

$$\lim_{t\to\infty}\frac{h(tx)}{h(t)}=x^a.$$

The index a is called the *exponent of variation*. When a is 0, h is said *slowly varying* (SV).

The *upper-tail dependence coefficient* of a random vector (X,Y) with copula C and marginal distributions F_X and F_Y, respectively, is defined as (Coles et al. (1999))

$$\lambda_u:=\lim_{\alpha\to1}\mathbb{P}\Big(Y>F_Y^{-1}(\alpha)\mid X>F_X^{-1}(\alpha)\Big)=\lim_{\alpha\to1}\frac{1-2\alpha-C(\alpha,\alpha)}{1-\alpha},\tag{2}$$

F^{-1} denoting the generalized inverse function of a cdf F.
Note that $\lambda_u\in[0,1]$, where $\lambda_u=0$ means extremal independence between X and Y.

Let us also introduce the following notation, borrowed from Joshi and Nagaraja (1995). Given a vector (X,Y), we define

$$F_1(y\mid x):=\mathbb{P}(Y\leqslant y\mid X>x),\tag{3}$$
$$F_2(y\mid x):=\mathbb{P}(Y\leqslant y\mid X\leqslant x),\tag{4}$$
$$F_3(y\mid x):=\mathbb{P}(Y\leqslant y\mid X=x).\tag{5}$$

The conditional distributions F_1, F_2 and F_3 play a fundamental role in computing the distributions of the concomitants of the order statistics as well as the distributions of their transformations.

Theorem 2 in Joshi and Nagaraja (1995): Suppose F_X satisfies one of the Von Mises conditions and, for all x and y, assume $F_X\in\mathscr{D}(G_X)$, $F_Y\in\mathscr{D}(G_Y)$, and

$$n\mathbb{P}(X>a_nx+b_n,Y>c_ny+d_n)\underset{n\to\infty}{\to}0,\quad\forall x,y\in\mathbb{R},\tag{6}$$

where the norming constants $(a_n)_n$, $(b_n)_n$, $(c_n)_n$ and $(d_n)_n$ are such that

$$F_3(c_ny+d_n\mid a_nx+b_n)\to1\qquad\text{as } n\to\infty.\tag{7}$$

Further, suppose there exist constants $\tilde{a}_n > 0$ *and* $\tilde{b}_n \in \mathbb{R}$ *such that*

$$F_1(\tilde{a}_n y + \tilde{b}_n \mid a_n x + b_n) \underset{n\to\infty}{\to} H_1(x, y). \tag{8}$$

Then, the cdf of (V_1, V_2) *satisfies*

$$F_{(V_1,V_2)}(\tilde{a}_n v_1 + \tilde{b}_n,\, c_n v_2 + d_n) \underset{n\to\infty}{\to} H(v_1)G_Y(v_2), \quad \forall v_1, v_2 \in \mathbb{R}, \tag{9}$$

where $H(v_1) = (k!)^{-1} \int H_1^k(x, v_1)(-\log G_X(x))^k g_X(x)\, dx$.

We begin our analysis by focusing on Condition (7). We first show in Lemma 1 the relation between (6) and (7), which proof is given in Appendix A.

Lemma 1. *Let* $x, y \in \mathbb{R}$. *Then, we have*

$$\begin{aligned} & n\mathbb{P}(X > a_n x + b_n, Y > c_n y + d_n) \underset{n\to\infty}{\to} 0 \\ \Leftrightarrow \quad & F_3\left(c_n y + d_n \mid a_n z + b_n\right) \underset{n\to\infty}{\to} 1, \ \forall z \geqslant x. \end{aligned} \tag{10}$$

Now, we can see that (7) can be directly understood in terms of the upper-tail dependence coefficient, namely:

Proposition 1. *Consider a bivariate random vector* (X, Y) *with copula* C. *Then,*

$$F_3\left(c_n y + d_n \mid a_n x + b_n\right) \underset{n\to\infty}{\to} 1, \quad \forall x \geqslant y \quad \Leftrightarrow \quad \lambda_u = 0$$

where the normalizing constants are chosen depending on the MDA to which X *and* Y *belong.*

Note that Proposition 1 is then simply another way to express Proposition 5.27 in Resnick (1987, p. 296), which states that

$$\lambda_u = 0 \quad \Leftrightarrow \quad n\mathbb{P}(X > a_n x + b_n, Y > c_n y + d_n) \underset{n\to\infty}{\to} 0.$$

Using Lemma 1 and Proposition 1, we can now rewrite Theorem 2 in Joshi and Nagaraja (1995) in terms of λ_u. This reformulation, given in Corollary 1, is helpful since it clarifies the connection between the asymptotic independence of (V_1, V_2) and that of (X, Y).

Corollary 1. *Suppose* F_X *satisfies one of the Von Mises conditions and, for all* x, y, $F_X \in \mathscr{D}(G_X)$ *and* $F_Y \in \mathscr{D}(G_Y)$. *Moreover, assume that the upper tail dependence coefficient between* X *and* Y *is* 0: $\lambda_u = 0$. *If there exist constants* $\tilde{a}_n > 0$, $\tilde{b}_n$ *s.t.* $F_1(\tilde{a}_n y + \tilde{b}_n \mid a_n x + b_n) \underset{n\to\infty}{\to} H(x, y)$, *then the cdf* $F_{(V_1,V_2)}$ *of* (V_1, V_2) *satisfies* (9).

The assumption of asymptotic independence between X and Y is generally too restrictive in practical applications, especially in this highly interconnected world. For instance, over the past 15 years, we have had to live through a major global financial crisis, then a pandemic, combined with an increase in cyber attacks. The presence of systemic risk, with a strong dependence among extremes, is one characteristic of all those events (*e.g.* Dacorogna and Kratz (2015)). Hence, investigating this (asymptotic) dependence aspect is of primary interest. This is the main objective of the next section.

2 JOINT ASYMPTOTICS FOR MAXIMA OVER SUBSETS OF CONCOMITANTS

We aim at studying the limiting distribution of the concomitants (V_1, V_2) when X and Y may exhibit some asymptotic dependence. To do so, we take advantage of the joint-tail model representation given in Ledford and Tawn (1998), then we state the main result (in Subsection 2.2) and discuss the choice of the normalizing constants to avoid degeneracy in the limit.

2.1 The LT joint-tail model

First, let us recall some definitions related to the slowly varying notion in the bivariate case. It will be needed to define the joint-tail model.

Definition 1. A function $\mathcal{L}$ that satisfies

$$\lim_{n\to\infty} \frac{\mathcal{L}(nx, ny)}{\mathcal{L}(n, n)} = r(x, y), \qquad \text{with } r(ax, ay) = r(x, y), \tag{11}$$

for all $a > 0$ and $(x, y) \in \mathbb{R}^2_+$, is said to be *bivariate slowly varying* (BSV).

It can be shown that there exists a univariate function r_* such that

$$r(x, y) = r_*(w), \quad \text{with} \quad w = \frac{x}{x+y}. \tag{12}$$

Definition 2. A BSV function $\mathcal{L}$ for which the function r_* defined in (12) satisfies

$$\frac{r_*(w)}{r_*(1-w)} \text{ is SV at } w = 0 \text{ and } w = 1,$$

is said to be *quasi-symmetric*.

Let us now present the joint-tail model, named the LT joint-tail model, introduced in Ledford and Tawn (1998) to study the asymptotic distribution of the concomitant of the max, $Y_{[n]}$, *i.e.* the limit distribution of the marginal distribution of V_1 (case $k = 1$), and to compute $\mathbb{P}\left(Y_{(n)} = Y_{[n]}\right)$.
Let (X, Y) be a bivariate random vector with unit Fréchet marginals (*i.e.* $F_X(x) = e^{-1/x}$, $x > 0$) and joint distribution function $F(x, y)$. Suppose that there exist functions $c(t)$ and $\psi(x, y)$ for which the joint survival function

$$\bar{F}(x, y) = 1 - e^{-1/x} - e^{-1/y} + F(x, y)$$

satisfies

$$\lim_{t\to\infty} \frac{t\bar{F}(tx, ty)}{c(t)} = \psi(x, y) \text{ for all } (x, y) \in \mathbb{R}^2_+. \tag{13}$$

Additionally suppose that there exists $\gamma \in \mathbb{R}$ such that

$$\begin{cases} \bar{F}(0, t)/\bar{F}(t, 0) & \text{is } RV_\gamma \text{ at infinity and} \\ \bar{F}(t, 0)/\bar{F}(0, t) & \text{is } RV_{-\gamma} \text{ at infinity.} \end{cases} \tag{14}$$

Then the following holds.

Definition 3 (The LT joint-tail model)**.** Assume (X, Y) have unit Fréchet marginals. Let $\eta \in (0, 1]$ be the *coefficient of tail dependence* used to determine the decay rate of $F^{-1}(t, t)$, as $t \to \infty$. Under Assumptions (13) and (14), the joint survival distribution of (X, Y) is given by

$$\mathbb{P}(X > x, Y > y) = \mathcal{L}(x, y)x^{-\alpha}y^{-\beta}, \quad \forall x, y, \tag{15}$$

where $\alpha, \beta > 0$, $\alpha + \beta = \eta^{-1}$ and the function $\mathcal{L}(x, y)$ is a quasi-symmetric BSV function (see Definition 2).

The coefficient η describes the type of limiting dependence between X and Y, while the function $\mathcal{L}$ its relative strength given a particular value of η; see Ledford and Tawn (1998) and Heffernan (2000) for further comments. As pointed out in de Haan and Zhou (2011), the $\eta = 1/2$ corresponds to the case where X and Y are independent, while the case $\eta > 1/2$ indicates a positive association of the extremes of (X, Y). Values of η lower than $1/2$ represent cases where the extremes of X and Y are negatively associated. The bounding cases of perfect negative and positive dependence correspond respectively to $\eta \to 0$ and $\eta = 1$ with $\mathcal{L}(\cdot) = 1$. Coles et al. (1999) introduce an elementary measure of dependence: Since the survival copula $\bar{C}$ is given by $\bar{C}(u, v) = 1 - u - v + C(u, v)$, they consider the following indices to describe the upper-tail dependence:

$$\bar{\chi} := \lim_{u \to 1} \frac{2\log(1-u)}{\log \bar{C}(u, u)} - 1 = 2\eta - 1 \tag{16}$$

$$\lambda_u := 2 - \lim_{u \to 1} \frac{\log C(u, u)}{\log u} = \begin{cases} c & \text{if } \bar{\chi} = 1,\ \mathcal{L}(t) \to c > 0, \text{ as } t \to \infty, \\ 0 & \text{if } \bar{\chi} = 1,\ \mathcal{L}(t) \to 0, \text{ as } t \to \infty, \\ 0 & \text{if } \bar{\chi} < 1. \end{cases} \tag{17}$$

Since $\bar{\chi} = 1$ if and only if $\eta = 1$, we have a link between λ_u and η as well. In other words,

- X and Y are asymptotically independent when $\alpha + \beta = 1$ and $\mathcal{L}(n, n) \underset{n \to \infty}{\to} 0$, or when $\alpha + \beta \geqslant 1$ (with no condition on the limit as $n \to \infty$ of $\mathcal{L}(n, n)$);
- X and Y are asymptotically dependent when $\alpha + \beta = 1$ and $\mathcal{L}(n, n) \underset{n \to \infty}{\to} c > 0$. Note that, in this case, $c = \lambda_u$.

This turns out to be useful when considering extensions of Theorem 2 in Joshi and Nagaraja (1995), relaxing Condition (7).
For the sake of simplicity, we will use the notation $C_{\not\perp}$ to indicate the asymptotic dependence case, that is $\alpha + \beta = 1$ and $\mathcal{L}(n, n) \underset{n \to \infty}{\to} \lambda_u > 0$. Thus, the indicator $\mathbb{1}_{C_{\not\perp}} = 1$ points out the presence of asymptotic dependence.

2.2 Main result

Let us state our main result on the joint asymptotic behaviour of maxima of concomitants, allowing for asymptotic dependence of X and Y:

Theorem 1.
Let (X, Y) *be a bivariate random vector with unit Fréchet marginals and upper-tail dependence coefficient* λ_u*. Assume that* (X, Y) *follows the LT joint-tail model given in Definition 3, with survival cdf (15), and that the BSV function* $\mathcal{L}$ *is such that*

$$\mathcal{L}(nx, \tilde{a}_n y + \tilde{b}_n) \underset{n\to\infty}{\to} \tilde{c}(x, y) \quad \textit{where} \tag{18}$$
$$\tilde{c}(x, y)\, x^{1-\alpha} < y^{\beta}, \quad \tilde{c}(x, y) \underset{y\to 0}{\sim} y^{-\beta}, \quad \tilde{c}(x, y) = o\left(y^{-\beta}\right) \textit{ for } y \to \infty,$$
$$\textit{with} \quad \tilde{a}_n = \mathcal{O}\left(n^{\frac{1-\alpha}{\beta}}\right) \quad \textit{and} \quad \tilde{b}_n = O\left(\tilde{a}_n\right) \textit{ or } o\left(\tilde{a}_n\right).$$

Then, the joint distribution of the concomitants maxima (V_1, V_2) *defined in (1) satisfies:*

$$F_{(V_1,V_2)}(\tilde{a}_n v_1 + \tilde{b}_n,\, n v_2) \xrightarrow[n\to\infty]{} \int_0^{+\infty} H_1^k(v_1 \mid x)\, H_2(v_2 \mid x)\, \frac{x^{-k-2}}{k!}\, e^{-1/x}\, dx, \quad \forall v_1, v_2 \in \mathbb{R}, \tag{19}$$

where $H_1(y \mid x) := 1 - \tilde{c}(x, y)\, x^{1-\alpha} y^{-\beta}$ *and* $H_2(y \mid x)$ *is defined by the product of the limits given below in (20) and (21), respectively, with* $r(x, y) = \lambda_u^{-1} \tilde{c}(x, y)$*.*

To prove Theorem 1, we need to evaluate the asymptotic behavior of the three conditional distributions F_1, F_2 and F_3 defined in (3), (4) and (5), respectively, each adequately transformed to obtain a non-degenerated limit distribution. This is what is presented in the following lemma.

Lemma 2. *Assume the LT joint-tail model (15) holds and call* r *the limit function corresponding to* $\mathcal{L}$*. Then, as* $n \to \infty$*,*

$$F_2^n(ny \mid nx) \to \exp\left\{-\frac{1}{y}\left(1 - \lambda_u r(x, y)\left(\frac{y}{x}\right)^{\alpha} \mathbb{1}_{C_{\not\perp}}\right)\right\}, \tag{20}$$

$$F_3(ny \mid nx) \to 1 + \lambda_u x^{-\alpha+2} y^{\alpha-1}\left(\frac{\partial}{\partial x} r(x, y) - \frac{\alpha}{x} r(x, y)\right) \mathbb{1}_{C_{\not\perp}}. \tag{21}$$

Moreover, under Condition (18), we have

$$F_1(\tilde{a}_n y + \tilde{b}_n \mid nx) \underset{n\to\infty}{\to} \tilde{c}(x, y). \tag{22}$$

Note that taking $\mathbb{1}_{C_{\not\perp}} = 0$ in (20) and (21) provides $\lim_{n\to\infty} F_2^n(ny \mid nx) = e^{-1/y}$ and $\lim_{n\to\infty} F_3(ny \mid nx) = 1$, which characterise copulas with asymptotic independence, as considered in Joshi and Nagaraja (1995). The proofs of Theorem 1 and Lemma 2 are developed in Appendix A.
It is worth noticing that Condition (18) is needed to ensure that $H_1(v_1 \mid x)$ is a proper cdf for $v_1 \in (0, +\infty)$. Moreover, our Theorem 1 gives precise formulations on the choice of the normalizing constants $\tilde{a}_n$ and $\tilde{b}_n$. This is in fact a delicate point, reason why we dedicate a comprehensive discussion on it: Explanations relative to Condition (18) are given in the next subsection, while Section 4 addresses the fundamental example of the Gaussian bivariate distribution, for which Condition (18) does not hold.

2.3 Discussion on the choice of $\tilde{a}_n$ and $\tilde{b}_n$

In the following, we carefully deduce which are the only possible asymptotic behaviours of the constants $\tilde{a}_n$ and $\tilde{b}_n$. We highlight the importance of the present subsection, as finding suitable constants which lead to non degenerate limits represents generally an issue in asymptotic theories.
We start by noticing that the limit of $F_1(\tilde{a}_n y + \tilde{b}_n \mid nx)$ strongly depends on that of $\mathcal{L}(nx, \tilde{a}_n y + \tilde{b}_n)$. Indeed, we can write

$$F_1(\tilde{a}_n y + \tilde{b}_n \mid nx) = 1 - \frac{\mathcal{L}(nx, \tilde{a}_n y + \tilde{b}_n)(nx)^{-\alpha}(\tilde{a}_n y + \tilde{b}_n)^{-\beta}}{1 - e^{-1/(nx)}} \underset{n\to\infty}{\sim} 1 - \frac{\mathcal{L}(nx, \tilde{a}_n y + \tilde{b}_n)}{n^{\alpha-1}x^{\alpha-1}(\tilde{a}_n y + \tilde{b}_n)^{\beta}} \tag{23}$$

using the approximation $1 - e^{-1/(nx)} \underset{n\to\infty}{\sim} (nx)^{-1}$.
In the following, we discuss which sequences $\tilde{a}_n$ and $\tilde{b}_n$ should be chosen to have $F_1(\tilde{a}_n y + \tilde{b}_n \mid nx) \underset{n\to\infty}{\to} 1$, *i.e.* V_1 not degenerate.
Given a couple (x, y), we study the limit of (23) depending on the behavior of $\mathcal{L}$, namely:

(i) Assume $\mathcal{L}(nx, \tilde{a}_n y + \tilde{b}_n) \to 0$ as $n \to \infty$. To have $F_1(\tilde{a}_n y + \tilde{b}_n \mid nx) \not\to 1$, we need $\frac{\mathcal{L}(nx, \tilde{a}_n y + \tilde{b}_n)}{n^{\alpha-1}x^{\alpha-1}(\tilde{a}_n y + \tilde{b}_n)^{\beta}} \not\to 0$. But, since we are in the case $\mathcal{L}(nx, \tilde{a}_n y + \tilde{b}_n) \underset{n\to\infty}{\to} 0$, this would mean, either

- $n^{\alpha-1}x^{\alpha-1}(\tilde{a}_n y + \tilde{b}_n)^{\beta} \underset{n\to\infty}{\to} 0$ at the same rate of $\mathcal{L}$, which is impossible because of the assumptions of the model ($\mathcal{L} \sim \mathrm{SV}$), or
- $n^{\alpha-1}x^{\alpha-1}(\tilde{a}_n y + \tilde{b}_n)^{\beta} \underset{n\to\infty}{\to} 0$ faster than $\mathcal{L}$, which is impossible because we would have $\frac{\mathcal{L}(nx, \tilde{a}_n y + \tilde{b}_n)}{n^{\alpha-1}x^{\alpha-1}(\tilde{a}_n y + \tilde{b}_n)^{\beta}} \underset{n\to\infty}{\to} \infty$, that denies F_1 is a cdf.

Hence case (i) is not possible.

(ii) Assume $\mathcal{L}(nx, \tilde{a}_n y + \tilde{b}_n) \to \infty$ as $n \to \infty$. So, as $n \to \infty$, the term $n^{\alpha-1}x^{\alpha-1}(\tilde{a}_n y + \tilde{b}_n)^{\beta}$ tends to infinity faster than $\mathcal{L}(nx, \tilde{a}_n y + \tilde{b}_n)$ does, which yields that $F_1(\tilde{a}_n y + \tilde{b}_n \mid nx) \underset{n\to\infty}{\to} 1$. Hence, this case (ii) has also to be discarded.

(iii) The previous reasoning implies that we can only have, as $n \to \infty$,

$$\mathcal{L}(nx, \tilde{a}_n y + \tilde{b}_n) \underset{n\to\infty}{\to} \tilde{c}(x, y) \in (0, \infty).$$

Which conditions on $\tilde{a}_n$ and $\tilde{b}_n$ can be given? Here, we have to distinguish whether $\tilde{c}(x, y)$ depends on y or not.

(A) *Suppose* $\tilde{c}(x, y)$ *depends on* y. Then, $\frac{\mathcal{L}(nx, \tilde{a}_n y + \tilde{b}_n)}{n^{\alpha-1}x^{\alpha-1}(\tilde{a}_n y + \tilde{b}_n)^{\beta}} \underset{n\to\infty}{\sim} \frac{\tilde{c}(x, y)}{n^{\alpha-1}x^{\alpha-1}(\tilde{a}_n y + \tilde{b}_n)^{\beta}}$, so we have to consider the relative behaviour of $\tilde{a}_n$ and $\tilde{b}_n$.

(a) If $\tilde{a}_n = o(\tilde{b}_n)$, then $\tilde{a}_n y + \tilde{b}_n \sim \tilde{b}_n$, as $n \to \infty$. This implies $\frac{\tilde{c}(x,y)}{n^{\alpha-1}x^{\alpha-1}(\tilde{a}_n y+\tilde{b}_n)^{\beta}} \underset{n\to\infty}{\sim} \frac{\tilde{c}(x,y)}{n^{\alpha-1}x^{\alpha-1}\tilde{b}_n^{\beta}}$, so, in order to maintain the dependence on y, we need $n^{\alpha-1}\tilde{b}_n^{\beta} \underset{n\to\infty}{\not\to} 0, \infty$. This condition obviously depends on the value of α: Call $\gamma = (1-\alpha)/\beta$, then

- If $0 < \alpha < 1$, $n^{\alpha-1} \underset{n\to\infty}{\to} 0$, the only possibility is $\tilde{b}_n = \mathcal{O}(n^{\gamma})$ with $\gamma \in (0,1)$, since $\alpha + \beta \geqslant 1$.
- If $\alpha = 1$, then $\tilde{b}_n$ would be independent of n; it is then impossible to propose $(\tilde{a}_n, \tilde{b}_n)$ in such a case.
- If $\alpha > 1$, then $\tilde{b}_n = \mathcal{O}(n^{\gamma})$ with $\gamma < 0$. However, since we are interested in the dependence in the extremes, we need to consider

$$n\mathbb{P}(X > nx, Y > n^{\gamma}y) \leqslant n^{1-\gamma} \cdot n^{\gamma}\mathbb{P}(X > n^{\gamma}x, Y > n^{\gamma}y) \underset{n\to\infty}{\not\to} 0,$$

which implies $\gamma \in (0,1)$. So, the case $\alpha > 1$ cannot be considered.

(b) If $\tilde{b}_n = o(\tilde{a}_n)$ or $\tilde{b}_n = \mathcal{O}(\tilde{a}_n)$, then $\tilde{a}_n y + \tilde{b}_n \underset{n\to\infty}{\sim} \tilde{a}_n y$ and

$$\frac{\tilde{c}(x,y)}{n^{\alpha-1}x^{\alpha-1}(\tilde{a}_n y + \tilde{b}_n)^{\beta}} \underset{n\to\infty}{\sim} \frac{1}{n^{\alpha-1}\tilde{a}_n^{\beta}}\,\tilde{c}(x,y)\,x^{1-\alpha}.$$

Hence, we need to consider $\tilde{a}_n = \mathcal{O}(n^{\gamma})$ and $\alpha \in (0,1)$.

(B) *Suppose* $\tilde{c}(x,y)$ *does not depend on* y. In this case,

$$\frac{\mathcal{L}(nx, \tilde{a}_n y + \tilde{b}_n)}{n^{\alpha-1}x^{\alpha-1}(\tilde{a}_n y + \tilde{b}_n)^{\beta}} \underset{n\to\infty}{\sim} \frac{n^{1-\alpha}}{\max\left(\tilde{a}_n^{\beta}y^{\beta}, \tilde{b}_n^{\beta}\right)}\tilde{c}x^{1-\alpha},$$

and since we do not want to loose the dependence on y, we need to assume $\max\left(\tilde{a}_n^{\beta}y^{\beta}, \tilde{b}_n^{\beta}\right) = \tilde{a}_n^{\beta}y^{\beta}$. Moreover, we need to ask $n^{1-\alpha}\tilde{a}_n^{-\beta} \not\to 0$ (otherwise $F_1(\tilde{a}_n y + \tilde{b}_n \mid nx) \underset{n\to\infty}{\to} 1$), which implies $\tilde{a}_n = \mathcal{O}(n^{\gamma})$ and $\alpha \in (0,1)$.

Therefore, to obtain a non degenerate limit for $F_1(\tilde{a}_n y + \tilde{b}_n \mid nx)$, we need to assume $\alpha \in (0,1)$, $\tilde{a}_n = \mathcal{O}(n^{\gamma})$ and $\tilde{b}_n = o(\tilde{a}_n)$ or $\tilde{b}_n = \mathcal{O}(\tilde{a}_n)$. Now that the main result is provided with a justification of the constants involved, let us illustrate the theoretical results on standard examples of the risk literature.

3 ILLUSTRATION OF THE RESULTS

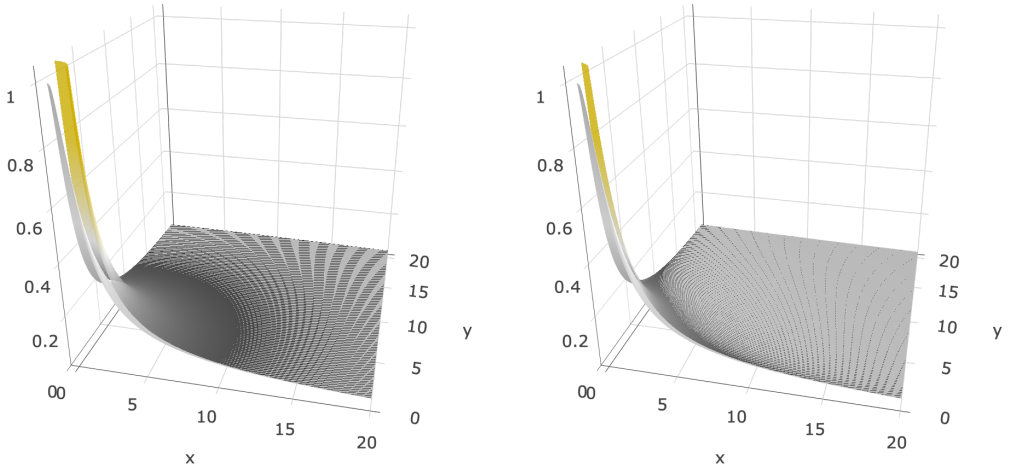
In this section, we consider two examples: the Pareto - Lomax with Survival Clayton copula and the Bivariate Symmetric Logistic Extremal copula. We apply our asymptotic approximations given in Theorem 1 and compare it with

the empirical results obtained via simulation. The choice of such examples lies in the fact that the extremal dependence is ruled by parameters, so easy to control. Note also that for the following examples we can achieve the same results by direct computation (with a considerable amount of time and tedious calculations), thus it is possible to make comparisons by means of the mathematical expressions obtained.
We simulate 5000 replicates of 100000 bivariate samples. The dimension k of the subsample used to compute V_1 is set to 10 (Appendix B contains simulations relative to other values of k, to have insights of its impact on the asymptotic result).

3.1 Pareto - Lomax marginals with Survival Clayton copula

Let (X_P, Y_P) be a random vector with Survival Clayton(θ) copula and Pareto - Lomax$(\nu, 1)$ marginals (the subscript P indicates the Pareto marginals):

$$\bar{F}_{X_P}(x) = (1+x)^{-\nu},\ \forall x > 0, \quad \bar{F}_{Y_P}(y) = (1+y)^{-\nu},\ \forall y > 0, \quad \nu > 0,$$
$$\bar{F}_P(x,y) = \mathbb{P}(X_P > x, Y_P > y) = \left((1+x)^{\nu\theta} + (1+y)^{\nu\theta} - 1\right)^{-1/\theta},\ \theta > 0.$$

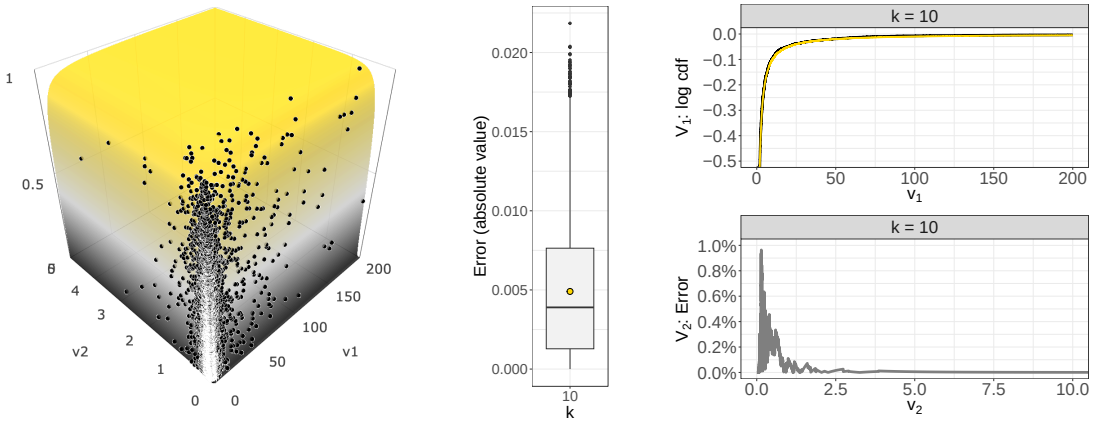


(a) *Survival Clayton copula* $(\theta = 2)$ **(b)** *Logistic copula* $(\gamma = 1/2)$

Figure 1: Joint tail $\bar{F}(x,y)$ (gray) with unit Fréchet marginals, as expressed in (24), and its approximation (black and gold) given by Equation (15). The parameters α and β are both set to 0.5 in the two examples.

First, we transform te Pareto marginals into the unit Fréchet ones, to match the assumptions of the model given in Theorem 1. We can write

$$F(x,y) = F_P\left(F_{X_P}^{-1}\left(e^{-1/x}\right), F_{Y_P}^{-1}\left(e^{-1/y}\right)\right),$$



(a) *Bivariate distribution of* (V_1, V_2). **(b)** *Error evaluation.* **(c)** *Marginal distributions diagnostics.*

Figure 2: Survival Clayton($\theta = 2$), with $k = 10$. Panel (a) shows the empirical univariate log cdf of V_1 (in black), together with the asymptotic theoretical log values (in gold) obtained by numerically integrating the asymptotic joint cdf (19) to obtain the marginal. Panel (b) shows the absolute values of the errors between empirical and theoretical cdf of V_2. The upper plot of panel (c) shows the empirical univariate cdf of V_1 (in black), together with the asymptotic theoretical values (in gold), in log scale, obtained by numerically integrating the asymptotic joint cdf (19). The lower plot of panel (c) represents the L_1 error between the marginal empirical and theoretical cdf.

where, for the Pareto distribution, $F^{-1}_{X_P}(q) = F^{-1}_{Y_P}(q) = (1-q)^{-1/\nu} - 1$. We obtain

$$\bar{F}(x,y) = \left(\left(1-e^{-1/x}\right)^{-\theta} + \left(1-e^{-1/y}\right)^{-\theta} - 1\right)^{-1/\theta}. \tag{24}$$

Now, observe that the joint survival (24) can be rewritten, for $x > x_0 > 0$ and $y > y_0 > 0$, as

$$\begin{aligned}\bar{F}(x,y) &\sim \left(x^\theta + y^\theta - 1\right)^{-1/\theta} = \left(x^\theta + y^\theta - 1\right)^{-1/\theta} (xy)^{1/2}(xy)^{-1/2} \\ &= \mathcal{L}(x,y)x^{-1/2}y^{-1/2},\end{aligned}$$

where $\mathcal{L}(x,y) = \left(x^\theta + y^\theta - 1\right)^{-1/\theta}(xy)^{1/2}(xy)$ and clearly $\alpha = \beta = 1/2$.
Note that

$$\mathcal{L}(n,n) = \left(2n^\theta - 1\right)^{-1/\theta} n \to 2^{-1/\theta} \quad \text{as } n \to \infty.$$

This is coherent with the value of the upper-tail dependence coefficient, since

$$\lambda_U = \lim_{q\to 1} \mathbb{P}\left(Y > F^{-1}_Y(q) \mid F^{-1}_X(q)\right) = \lim_{q\to 1} (1-q)^{-1}\left(2(1-q)^{-\theta} - 1\right)^{-1/\theta} = 2^{-1/\theta}.$$

According to Theorem 1, since $1 - \alpha = \beta$, we have $\tilde{a}_n = a_n = n$. Hence,

$$\mathcal{L}(nx, ny) = \left(n^\theta x^\theta + n^\theta y^\theta - 1\right)^{-1/\theta} n(xy)^{1/2} \underset{n\to\infty}{\to} \left(x^\theta + y^\theta\right)^{-1/\theta}(xy)^{1/2} = \tilde{c}(x,y).$$

Note that, for each $a > 0$,

$$\tilde{c}(ax, ay) = \left((ax)^\theta + (ay)^\theta\right)^{-1/\theta} (a^2xy)^{1/2} = \tilde{c}(x, y),$$

which means that $\tilde{c}(x, y)$ and also $r(x, y) := 2^{1/\theta}\tilde{c}(x, y)$ satisfy (11). Moreover it is straightforward to show that $\tilde{c}(x, y)$ satisfies Condition (18) in Theorem 1. Since it holds

$$\frac{\partial}{\partial x} r(x, y) = 2^{1/\theta} \frac{\partial}{\partial x} \tilde{c}(x, y) = 2^{1/\theta} x^{-1} \tilde{c}(x, y) \left(\frac{1}{2} - \left(1 + \left(\frac{y}{x}\right)^\theta\right)^{-1}\right),$$

applying Lemma 2 and Theorem 1 gives that

$$H_1(y \mid x) = 1 - \left(1 + \left(\frac{y}{x}\right)^\theta\right)^{-1/\theta}, \tag{25a}$$

$$H_2(y \mid x) = \left(1 - \left(1 + \left(\frac{y}{x}\right)^\theta\right)^{-1/\theta - 1}\right) e^{-\frac{1}{y} + \left(1 + \left(\frac{y}{x}\right)^\theta\right)^{-1/\theta}}. \tag{25b}$$

The plot in Figure 2a shows the bivariate surface computed by applying Theorem 1 (the integration is performed numerically) and the values of the bivariate empirical cdf computed in the simulated V_1 and V_2. These points lie on the surface, thus displaying the very good quality of the asymptotic approximation of Theorem 1. It can also be assessed from Figures 2b and 2c, showing the boxplot of the absolute value of the errors between the black cloud and the surface in Figure 2a, the asymptotic approximation of the marginal distributions of V_1 and the errors in V_2, respectively. For what concerns the error (lower plot of Figure 2c), we observe that it is small for low values of v_2 (the maximum error is close to 1%) and converges very quickly to 0. In particular, it is 0% for all the values of v_2 we are considering.

3.2 Unit Fréchet marginals with Logistic Extremal copula

Let (X, Y) be a random vector with unit Fréchet marginals and a bivariate symmetric logistic extremal dependence, *i.e.*

$$F(x, y) = e^{-V(x,y)},$$

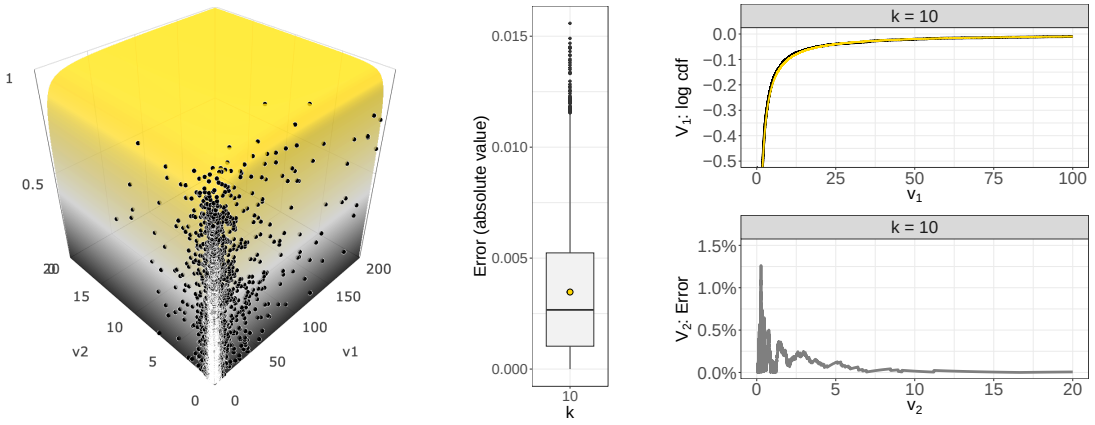
where

$$V(x, y) = \left(x^{-1/\gamma} + y^{-1/\gamma}\right)^\gamma, \; x, y > 0, \; \gamma \in (0, 1).$$

By considering the bivariate joint tail approximation (15) proposed in Ledford and Tawn (1997) characterised by a general extremal dependence function V, we can write in our case

$$\mathcal{L}(x, y) = (xy)^{-1/2}(x + y) - (xy)^{1/2} V(x, y). \tag{26}$$

Note that



(a) *Bivariate distribution of* (V_1, V_2).

(b) *Error evaluation.*

(c) *Marginal distributions diagnostics.*

Figure 3: Logistic($\gamma = 1/2$), with $k = 10$. Panel (a) shows the empirical univariate log cdf of V_1 (in black), together with the asymptotic theoretical log values (in gold) obtained by numerically integrating the asymptotic joint cdf (19) to obtain the marginal one. Panel (b) shows the absolute values of the errors between the empirical and the theoretical cdf of V_2. The upper plot of panel (c) shows the empirical univariate cdf of V_1 (in black), together with the asymptotic theoretical values (in gold), in log scale, obtained by numerically integrating the asymptotic joint cdf (19). The lower plot of panel (c) represents the L_1 error between the marginal empirical and theoretical cdf.

$$\mathcal{L}(n, n) = 2 - V(1, 1),$$

since V is homogeneous of order -1 (*i.e.* $V(nx, ny) = n^{-1}V(x, y)$). Moreover, $1 - \alpha = \beta$, so that $\tilde{a}_n = a_n = n$. Hence, we can write

$$\begin{aligned}\mathcal{L}(nx, ny) &= n^{-1}(xy)^{-1/2}n(x + y) - n(xy)^{1/2}n^{-1}V(x, y)\\ &= (xy)^{-1/2}(x + y) - (xy)^{1/2}V(x, y) = \tilde{c}(x, y).\end{aligned}$$

Note that, for each $a > 0$,

$$\tilde{c}(ax, ay) = \tilde{c}(x, y),$$

so that the candidate $\mathcal{L}$ in (26) is a proper BSV function. Simple algebra helps to prove that $\tilde{c}(x, y)$ satisfies Condition (18) in Theorem 1. Since it holds

$$\frac{\partial}{\partial x}\tilde{c}(x, y) = \frac{(xy)^{-1/2}}{2}\left(1 - \frac{y}{x} - yV(x, y)\left(1 - 2(xV(x, y))^{-1/\gamma}\right)\right),$$

by applying Lemma 2 and Theorem 1, we obtain

$$H_1(y \mid x) = x\left(V(x, y) - y^{-1}\right), \tag{27a}$$

$$H_2(y \mid x) = \left(1 + \left(\frac{y}{x}\right)^{-\frac{1}{\gamma}}\right)^{\gamma-1} e^{-\frac{1}{x}\left(\left(1+\left(\frac{y}{x}\right)^{-\frac{1}{\gamma}}\right)^{\gamma}-1\right)}. \tag{27b}$$

After having obtained the functions H_1 and H_2, we can now apply a numerical integration to compute the joint cdf (19).

4 DISCUSSION ON THE GAUSSIAN CASE

Whilst our focus so far was on obtaining the joint asymptotic cdf of (V_1, V_2) when (X, Y) exhibits asymptotic or extremal dependence, we question if Condition (18) of Theorem 1 is strongly related to the extremal dependence assumption. In other words, what does occur if the BSV function $\mathcal{L}$ goes to 0 or to infinity? This is why we return to the case when (X, Y) is Gaussian, as it is well known that (X, Y) is then asymptotically independent. This case has already been treated in the literature, due both to practical interest of the distribution and to its known manageability. Although direct computations in this case become simple, problems arise in applying models involving approximations with SV functions. In fact, even if such approximations are provided, they are then not used to compute for instance the conditional distributions we encountered above (as far as we could observe in the literature). The objective of the present section is to question this apparent paradox and provide tools to overcome this issue, thus bridging the gap between direct computations and those made through asymptotic models, with some attention put on the asymptotic behaviour of the BSV $\mathcal{L}$.

4.1 Existing literature

Let us start to review the existing literature on the Gaussian case.

(a) Joshi and Nagaraja (1995): By direct computations, the authors are able to provide constants $\tilde{a}_n$, $\tilde{b}_n$ such that the rescaled distribution of (V_1, V_2) converges to a product of a standard normal cdf to the power k and a Gumbel cdf. When applying those constants in our case, after the transformation of the Gaussian marginals into unit Fréchet, Condition (18) in Theorem 1 is not satisfied. In fact, we obtain that F_1 degenerates for $n \to \infty$.

Hence, our question: Could we find other transformations of the marginals that let us circumvent the case $\mathcal{L} \to 0$, and possibly find a new method or solution to handle this case? The relevance of this problem lies in the fact that relations between asymptotic dependence and the behavior of $\mathcal{L}$ are generally unknown: We know by formula (16) that the behaviour of $\mathcal{L}(n, n)$ and the values η takes are linked to the upper-tail dependence coefficient λ_u, but the case $\mathcal{L} \not\to 0$ and $\eta < 1$ is not tackled. Moreover, we also need to consider $\mathcal{L}(n, \tilde{a}_n)$ instead of $\mathcal{L}(n, n)$.

(b) Ledford and Tawn (1998): The focus here is only on the asymptotic distribution of the concomitant of the maximum $X_{(n)}$, that is the asymptotic distribution of V_1 when $k = 1$ (in our notation), meaning to study F_1. The authors explain that there can be cases where unit Fréchet marginals are not appropriate and suggest a standard Gumbel transformation for the variable Y, arriving to a new BSV function $\mathcal{L}_1$ (in fact the first order approximation of $\mathcal{L}$) that avoids the degeneracy of the limit of H_1 (with our

notations). However, to compare this approach with the case where both the marginals are unit Fréchet, they need to assume $y \to \infty$.

Let (X_N, Y_N) be a bivariate normal random vector with standard normal marginals and with correlation $0 < \rho < 1$. Theorem 3 in Joshi and Nagaraja (1995) states that, in the extreme case (k fixed, $n \to \infty$), it holds that

$$\mathbb{P}\left(\frac{V_1 - \tilde{a}_n}{\tilde{b}_n} \leqslant v_1, \frac{V_2 - a_n}{b_n} \leqslant v_2\right) \underset{n\to\infty}{\to} \Phi^k(v_1) e^{-e^{-v_2}}, \tag{28}$$

where

$$a_n = (2\log n)^{-1/2} \qquad b_n = \sqrt{2\log n} - \frac{1}{2}\frac{\log(4\pi\log n)}{\sqrt{2\log n}} \tag{29a}$$

$$\tilde{a}_n = \sqrt{1-\rho^2} \qquad \tilde{b}_n = \rho b_n. \tag{29b}$$

Formula (28) represents an asymptotic independence property of V_1 and V_2, strongly connected to the fact that

$$F_1(\tilde{a}_n y + \tilde{b}_n \mid a_n x + b_n) = \mathbb{P}\big(Y_N \leqslant \tilde{a}_n y + \tilde{b}_n \mid X_N > a_n x + b_n\big) \underset{n\to\infty}{\to} \Phi(y). \tag{30}$$

Recall that our model in Theorem 1 is not applicable here. In fact, Condition (18) is not satisfied as $\mathcal{L}(n, \tilde{a}_n) \to 0$ for any choice of the normalizing constants. Thus, further study would be needed in such a case.

A way to look at the problem may be through Wadsworth and Tawn (2013). The authors study the effect of letting the components of a bivariate random vector grow at different rates, with the result of stating a new class of regular variation conditions and providing a new characterization of the link between the multivariate tail decay and the considered marginal growth rates.

More precisely, the authors transform the marginals of the original vector into standard exponentials, called (X_E, Y_E), and assume that, for all $(\zeta_1, \zeta_2) \in \mathbb{R}^2_+ \setminus \{0\}$,

$$\mathbb{P}(X_E > \zeta_1 \log n, Y_E > \zeta_2 \log n) \underset{n\to\infty}{\sim} L(n; \zeta_1, \zeta_2) n^{-\kappa(\zeta_1,\zeta_2)} \tag{31}$$

where L is a univariate SV function in n as $n \to \infty$, and the function $\kappa(\zeta_1, \zeta_2)$ *maps the different marginal growth rates to the joint tail decay rate* with the role of providing information *about the level of dependence between variables at sub-asymptotic levels*. Assumption (31) is then used to describe models for conditional probabilities of extreme exceedances, *i.e.*

$$\mathbb{P}(X_E > \zeta_1 \log n \mid Y_E > \log n) \underset{n\to\infty}{\sim} L(n; \zeta_1, 1) e^{(1-\kappa(\zeta_1,\zeta_2))\log n}. \tag{32}$$

Studying (32) is proven by the authors to be equivalent to the search for functions a^*, b^* and h such that

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{X_E - b^*(\log n)}{a^*(\log n)} > x \,\middle|\, Y_E > \log n\right) = e^{-h(x)}. \tag{33}$$

The use of the exponential marginals is helpful to identify the normalizing constants for the bivariate Gaussian example, thus partially filling the gap

left open in Ledford and Tawn (1998), where, as previously explained, a way to tackle the problem is given in the case of Fréchet-Gumbel marginals, albeit that solution can be used only for some high values of Y. However, this discussion does not yet answer our main question: Why, when using unit Fréchet marginals, aren't we able to find normalizing constants that lead to a non-degenerate limit for $\mathcal{L}(nx, \tilde{a}_n y + \tilde{b}_n)$? Our understanding of the problem is made more complete by the insightful paper by Heffernan and Resnick (2007). There, the authors explore the implications of assuming the existence of a scaling function $c_1^*(\cdot)$, a centering function $c_2^*(\cdot)$, and a non-null Radon measure μ on Borel subsets of $[-\infty, \infty] \times (-\infty, \infty]$ such that, for each fixed $y \in E_\xi := \{y \in \mathbb{R} : 1 + \xi y > 0\}$ (where ξ is the shape parameter of the extreme value distribution relative to F_Y), it holds

a) $\mu([-\infty, x] \times (y, +\infty))$ is a finite and non-degenerate distribution function in x

b) the limit of the cdf of the normalized X given that Y is extreme exists and

$$\mathbb{P}\left(\frac{X - c_2^* \circ b^{\leftarrow}(t)}{c_1^* \circ b^{\leftarrow}(t)} \leqslant x \,\middle|\, Y > t \right) \underset{t \to y^+}{\to} \mu([-\infty, x] \times (0, +\infty)), \tag{34}$$

where y^+ is the right-end point of F_Y. It is worth noticing the fact that we are free to change the marginal of Y without disturbing the convergence in (34), but this is not true for the X variable. In the paper it is proven that, in the case of the bivariate standard Gaussian, it is not possible to obtain a non-degenerate expression for μ in (34) if the marginals of X and Y are both transformed into the Pareto type, while an asymptotic form is still reachable if the marginals are transformed into exponentials. We refer to Heffernan and Resnick (2007, Sec. 7) for a thorough discussion of this aspect.

4.2 Revisiting the Gaussian case

Now that we have a better and more complete understanding on the issues underlying the treatment made in the literature of the Gaussian case, we provide a way to find suitable constants through the application of model (32) and formulas (29a) and (29b). We know that, if (X_N, Y_N) is a bivariate normal random vector and correlation $\rho \in (0, 1)$,

$$\mathbb{P}\left(Y_N > \tilde{a}_n y + \tilde{b}_n \mid X_N > a_n x + b_n\right) \underset{n \to \infty}{\to} 1 - \Phi(y), \tag{35}$$

where the constants are given in (29a) and (29b). The following results hold:

Theorem 2. *Let (X_E, Y_E) be a bivariate random vector with unit exponential marginals and correlation $\rho \in (0, 1)$, such that (32) holds, with $\zeta_1 := \zeta_1(n, y)$. Then, we have*

$$\mathbb{P}\left(Y_E > \tilde{a}_{n,E} y + \tilde{b}_{n,E} \mid X_E > \log n\right) \to \frac{1}{y\sqrt{2\pi}} e^{-(\rho y)^2/2}, \quad \text{as } n \to \infty, \tag{36}$$

where

$$\tilde{a}_{n,E} = \rho\sqrt{1-\rho^2}\left(\sqrt{2\log n} - \frac{1}{2}\frac{\log(4\pi\log n)}{\sqrt{2\log n}}\right) \tag{37a}$$

and $$\tilde{b}_{n,E} = \rho^2\log n - \frac{\rho^2}{2}\log(4\pi\log n) + \frac{\rho^2}{16}\frac{(\log(4\pi\log n))^2}{\log n}. \tag{37b}$$

Corollary 2. *Note that, if we assume* y *large enough (as in Ledford and Tawn (1998)), then we can use the Mill's ratio to obtain*

$$\mathbb{P}\left(Y_E > \tilde{a}_{n,E}y + \tilde{b}_{n,E} \mid X_E > \log n\right) \underset{n\to\infty}{\sim} \rho\left(1-\Phi(\rho y)\right). \tag{38}$$

Refer to Appendix A for the proofs.
The limit distribution given in (38) is similar to those in Joshi and Nagaraja (1995) and Wadsworth and Tawn (2013). The differences among the final formulas are due to the different chosen orders of the Taylor approximations.

5 DISCUSSION

The work presented has provided a deep understanding of the impact of the extremal dependence of a bivariate random vector (X, Y) on the joint asymptotic behavior of suitably rescaled maxima over subsets of concomitants. Studying the framework of asymptotically dependent random variables is actually of great interest in practical applications and completes the literature.
Our main result addresses this problem, providing the expression of the asymptotic joint cdf of the two rescaled maxima of concomitants. In particular, we take advantage of the joint-tail model proposed in Ledford and Tawn (1997) to provide the asymptotic form of all the conditional distributions describing Y given the behaviour of X, which contribute to the asymptotic joint cdf of the two maxima of concomitants. We also explain in detail how to choose suitable normalizing constants needed to avoid the degeneracy of such distributions (thus preventing the joint cdf of the maxima of concomitants to be degenerate as well). Our theoretical results are illustrated through two examples of interest in the risk analysis literature: We compute the analytical expressions of the conditional distribution and perform a numerical integration to obtain the joint cdf. We show the goodness of the asymptotic approximation in all the examined cases through a numerical simulation.
Furthermore, we reconsider Theorem 2 in Joshi and Nagaraja (1995) and show how some hypotheses are connected to the upper-tail dependence coefficient characterizing the copula of the parent vector (X, Y), thus providing an intuitive interpretation of the formulas therein and a new reformulation of the cited Theorem.
More importantly, we revisit the example of the bivariate Gaussian distribution in our case, furnishing a comprehensive discussion on it. We deeply study the reasons why the joint-tail model in Ledford and Tawn (1997) fails in providing

a non-degenerate asymptotic distribution and fill the gap in the extension of the joint-tail model proposed in Wadsworth and Tawn (2013), by showing how to compute suitable normalization constants in this case. Our results are coherent with those obtained by direct computation in Joshi and Nagaraja (1995).

ACKNOWLEDGEMENTS

The authors are grateful for funding to Labex MME DII (ANR-11-LABX-0023-01) and ESOMAS department (University of Turin). This paper was mainly written during the first author's visiting period at the ESSEC Business School, Paris, France and during the second author's visiting period at the ESOMAS Department, University of Turin, Turin, Italy. Both authors are grateful to the hosting institutions.

REFERENCES

Balci, S., A. D. Akkaya, and B. E. Ulgen (2013). Modified maximum likelihood estimators using ranked set sampling. *Journal of Computational and Applied Mathematics 238*, 171–179.

Coles, S., J. Heffernan, and J. Tawn (1999). Dependence measures for extreme value analyses. *Extremes 2*(4), 339–365.

Dacorogna, M. M. and M. Kratz (2015). Living in a stochastic world and managing complex risks. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2668468.

de Haan, L. and C. Zhou (2011). Extreme residual dependence for random vectors and processes. *Advances in Applied Probability 43*(1), 217 – 242.

Dell, T. R. and J. L. Clutter (1972). Ranked set sampling theory with order statistics background. *Biometrics 28*(2), 545–555.

Fisher, R. A. and L. H. C. Tippett (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society 24*(2), 180?190.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics 44*(3), 423–453.

Heffernan, J. E. (2000). A directory of coefficients of tail dependence. *Extremes 3*(3), 279–290.

Heffernan, J. E. and S. I. Resnick (2007). Limit laws for random vectors with an extreme component. *The Annals of Applied Probability 17*(2), 537 – 571.

Joshi, S. and H. Nagaraja (1995). Joint distribution of maxima of concomitants of subsets of order statistics. *Bernoulli 1*(3), 245 – 255.

Kaufmann, E. and R.-D. Reiss (1992). On conditional distributions of nearest neighbors. *Journal of Multivariate Analysis 42*(1), 67–76.

Ledford, A. W. and J. A. Tawn (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society. Series B (Methodological) 59*(2), 475–499.

Ledford, A. W. and J. A. Tawn (1998). Concomitant tail behaviour for extremes. *Advances in Applied Probability 30*(1), 197–215.

Leonetti, P. and A. Khorrami Chokami (2022). The maximum domain of attraction of multivariate extreme value distributions is small. *Electronic Communications in Probability 27*, 1 – 8.

Nagaraja, H. N. and H. A. David (1994). Distribution of the Maximum of Concomitants of Selected Order Statistics. *The Annals of Statistics 22*(1), 478–494.

Ozturk, O. (2019). Two-stage cluster samples with ranked set sampling designs. *Annals of the Institute of Statistical Mathematics 71*(1), 63–91.

Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*, Volume 4 of *Applied Probability*. New York: Springer. First Printing.

Shao, X. (2010). The dependent wild bootstrap. *Journal of the American Statistical Association 105*(489), 218–235.

Stokes, S. L. and T. W. Sager (1988). Characterization of a ranked-set sample with application to estimating distribution functions. *Journal of the American Statistical Association 83*(402), 374–381.

Sweeting, T. J. (1985). On Domains of Uniform Local Attraction in Extreme Value Theory. *The Annals of Probability 13*(1), 196 – 205.

Wadsworth, J. and J. Tawn (2013). A new representation for multivariate tail probabilities. *Bernoulli 19*(5B), 2689 – 2714.

Wang, X., S. Ahn, and J. Lim (2017). Unbalanced ranked set sampling in cluster randomized studies. *Journal of Statistical Planning and Inference 187*, 1–16.

Wang, X., J. Lim, and L. Stokes (2016). Using ranked set sampling with cluster randomized designs for improved inference on treatment effects. *Journal of the American Statistical Association 111*(516), 1576–1590.

Wang, X., L. Stokes, J. Lim, and M. Chen (2006). Concomitants of multivariate order statistics with application to judgment poststratification. *Journal of the American Statistical Association 101*(476), 1693–1704.

Zamanzade, E. and X. Wang (2018). Proportion estimation in ranked set sampling in the presence of tie information. *Computational Statistics* 33(3), 1349–1366.

A APPENDIX – PROOFS

Proof of Lemma 1. Using the notations of Lemma 1, and indicating with $\mathscr{D}(G_{\gamma_1})$, $\mathscr{D}(G_{\gamma_2})$, $\mathscr{D}(G_0)$ the MDAs of the Fréchet, the Weibull and the Gumbel distributions, respectively, we can write, for $x, y \in \mathbb{R}$ and x^+ denoting the upper end-point of the distribution of X,

$$\begin{aligned}
&n\mathbb{P}(X > a_n x + b_n, Y > c_n y + d_n) \\
&= \int_{a_n x + b_n}^{x^+} n\mathbb{P}(Y > c_n y + d_n \mid X = t)\, f_X(t)\, dt \\
&\leqslant n \int_{a_n x + b_n}^{x^+} f_X(t)\, dt = n(1 - F_X(a_n x + b_n)) < \infty, \forall n \in \mathbb{N}.
\end{aligned}$$

As shown in the proof of Proposition 2.5 in Resnick (1987), it holds

$$n a_n f_X(a_n z + b_n) \mathbb{1}\left(x \leqslant z < a_n^{-1}(x^+ - b_n)\right) \underset{n\to\infty}{\to} \mathbb{1}(z \geqslant x)\, g(z)$$

with

$$g(z) := \begin{cases} \gamma_1 z^{-(\gamma_1+1)},\ z > 0,\ \gamma_1 > 0, & \text{if } F_X \in \mathscr{D}(G_{\gamma_1}), \\ \gamma_2(-z)^{\gamma_2 - 1},\ z < 0,\ \gamma_2 > 0, & \text{if } F_X \in \mathscr{D}(G_{\gamma_2}), \\ e^{-z},\ z \in \mathbb{R}, & \text{if } F_X \in \mathscr{D}(G_0). \end{cases}$$

Note the role of the indicator $\mathbb{1}\left(x \leqslant z < a_n^{-1}(x^+ - b_n)\right)$, needed to ensure that we are integrating over the support of X. In order to apply the Extended Dominated Convergence Theorem, we need to check that

$$\int_{\mathbb{R}} n a_n f_X(a_n z + x^+) \mathbb{1}(z \geqslant x)\, dz = n\,(1 - F_X(a_n x + b_n)) \underset{n\to\infty}{\to} \int_{\mathbb{R}} g(z) \mathbb{1}(z \geqslant x)\, dz.$$

For the Gumbel case, we have

$$n\,(1 - F_X(a_n x + b_n)) \to e^{-x} \text{ and } \int_{\mathbb{R}} g(z) \mathbb{1}(z \geqslant x)\, dz = e^{-x}.$$

The Fréchet and Weibull cases follow the same reasoning. Since the two quantities are equal, we can exchange limit and integral, as follows:

$$\begin{aligned}
&\lim_{n\to\infty} \int_{\mathbb{R}} n\mathbb{P}(Y > c_n y + d_n \mid X = t)\, f_X(t) \mathbb{1}(t \geqslant a_n x + b_n)\, dt \\
&= \int_{\mathbb{R}} \lim_{n\to\infty} \mathbb{P}(Y > c_n y + d_n \mid X = a_n z + b_n) \cdot \lim_{n\to\infty} n a_n f_X(a_n z + b_n) \\
&\qquad \cdot \mathbb{1}\left(z < a_n^{-1}(x^+ - b_n)\right) \cdot \mathbb{1}(z \geqslant x)\, dz.
\end{aligned}$$

Since the integrand is positive, then the integral tends to 0 if and only if the latter tends to 0, that is, if and only if

$$\mathbb{P}(Y > c_n y + d_n \mid X = a_n z + b_n) = 1 - F_3\left(c_n y + d_n \mid a_n z + b_n\right) \underset{n\to\infty}{\to} 0$$

for all $z \geqslant x$ (note that $\lim_{n\to\infty} n a_n f_X(a_n z + b_n) \neq 0$). Hence the result. □

Proof of Lemma 2. In order to make use of (15), we write all the conditionals in terms of $\bar{F}$ and recall the notation $C_{\not\perp}$ for the case $\alpha + \beta = 1$ and $\lambda_U > 0$. For what concerns $F_2(y \mid x)$, we have

$$F_2(y \mid x) = \mathbb{P}(Y \leqslant y \mid X < x) = 1 - \mathbb{P}(Y > y \mid X < x) = 1 - \frac{\bar{F}_Y(y) - \bar{F}(x,y)}{F_X(x)}.$$

Since

$$\begin{aligned}
\frac{\bar{F}_Y(ny) - \bar{F}(nx,ny)}{F_X(nx)} &= e^{\frac{1}{nx}}\left(1 - e^{-\frac{1}{ny}} - \frac{\mathcal{L}(nx,ny)}{n^{\alpha+\beta}x^{\alpha}y^{\beta}}\right) \\
&\underset{n\to\infty}{\sim} \frac{1}{n}\left(1 + \frac{1}{nx}\right)\left(\frac{1}{y} - \frac{\mathcal{L}(nx,ny)}{n^{\alpha+\beta-1}x^{\alpha}y^{\beta}}\right) = \frac{1}{ny}\left(1 + \frac{1}{nx}\right)\left(1 - \frac{\mathcal{L}(nx,ny)}{n^{\alpha+\beta-1}x^{\alpha}y^{\beta-1}}\right) \\
&\underset{n\to\infty}{\sim} \frac{1}{y}\left(1 + \frac{1}{nx}\right)\left(1 - \frac{\mathcal{L}(n,n)}{n^{\alpha+\beta-1}} r(x,ny) x^{-\alpha} y^{1-\beta}\right) \underset{n\to\infty}{\sim} \frac{1}{ny}\left(1 - \lambda_U r(x,y)\left(\frac{y}{x}\right)^{\alpha} \mathbb{1}_{C_{\not\perp}}\right),
\end{aligned}$$

we obtain

$$F_2^n(ny \mid nx) = \left[1 - \left(\frac{\bar{F}_Y(ny)}{F_X(nx)} - \frac{\bar{F}(nx,ny)}{F_X(nx)}\right)\right]^n \underset{n\to\infty}{\sim} \left[1 - \frac{1}{ny}\left(1 - \lambda_U r(x,y)\left(\frac{y}{x}\right)^{\alpha} \mathbb{1}_{C_{\not\perp}}\right)\right]^n.$$

Thus

$$F_2^n(ny \mid nx) \underset{n\to\infty}{\to} e^{-\frac{1}{y}\left(1 - \lambda_U r(x,y)\left(\frac{y}{x}\right)^{\alpha} \mathbb{1}_{C_{\not\perp}}\right)}.$$

For $F_3(y \mid x)$, note that

$$\begin{aligned}
F_3(y \mid x) &= \frac{1}{f_X(x)}\frac{\partial}{\partial x} F(x,y) = \frac{1}{f_X(x)}\frac{\partial}{\partial x}\left(F_X(x) - \bar{F}(y) + \bar{F}(x,y)\right) \\
&= 1 + \frac{1}{f_X(x)}\frac{\partial}{\partial x}\bar{F}(x,y) = 1 + x^{2-\alpha} y^{-\beta} e^{\frac{1}{x}}\left(\frac{\partial}{\partial x}\mathcal{L}(x,y) - \frac{\alpha}{x}\mathcal{L}(x,y)\right).
\end{aligned}$$

It follows that

$$\begin{aligned}
F_3(ny \mid nx) &\underset{n\to\infty}{\sim} 1 + n^{2-\alpha-\beta} e^{\frac{1}{nx}} x^{2-\alpha} y^{-\beta}\left(\frac{\partial}{\partial(nx)}\mathcal{L}(nx,ny) - \frac{\alpha}{nx}\mathcal{L}(nx,ny)\right) \\
&= 1 + e^{\frac{1}{nx}} x^{2-\alpha} y^{-\beta} n^{2-\alpha-\beta}\left(\frac{1}{n}\frac{\partial}{\partial x}\mathcal{L}(nx,ny) - \frac{\alpha}{nx}\mathcal{L}(nx,ny)\right) \\
&\underset{n\to\infty}{\sim} 1 + e^{\frac{1}{nx}} n^{1-\alpha-\beta}\mathcal{L}(n,n) x^{2-\alpha} y^{-\beta}\left(\frac{\partial}{\partial x} r(x,y) - \frac{\alpha}{x} r(x,y)\right) \\
&\underset{n\to\infty}{\to} 1 + \lambda_U x^{2-\alpha} y^{-\beta}\left(\frac{\partial}{\partial x} r(x,y) - \frac{\alpha}{x} r(x,y)\right) \mathbb{1}_{C_{\not\perp}},
\end{aligned}$$

which concludes the proof. □

Proof of Theorem 1. The proof is based on Lemma 2, and Joshi and Nagaraja (1995). To begin with, we report the finite-sample bivariate cumulative distribution of (V_1, V_2) (Joshi and Nagaraja (1995, Equation (2.1))),

$$F_{(V_1,V_2)}(v_1, v_2) = \mathbb{E}\left[h\left(v_1, v_2, X_{(n-k)}\right)\right] \tag{39}$$

where the function h is defined as

$$h(v_1, v_2, x) = F_1^k(v_1 \mid x) F_2^{n-k-1}(v_2 \mid x)\, F_3(v_2 \mid x).$$

The rational under Equation (39) lies in Kaufmann and Reiss (1992). Conditioned on the event $\{X_{(n-k)} = x\}$, the maxima V_1 and V_2 have the following behaviour: V_1 can be considered as the sample maximum of a random sample of size k from $F_1(\cdot \mid x)$, while V_2 behaves like the sample maximum of a sample of size $n-k$, composed by $n-k-1$ i.i.d. variables from $F_2(\cdot \mid x)$ and a single random variable from $F_3(\cdot \mid x)$.
Following Joshi and Nagaraja (1995), from Equation (39), we can write

$$F_{(V_1,V_2)}(\tilde{a}_n v_1 + \tilde{b}_n,\, c_n v_2 + d_n) = \int F_1(\tilde{a}_n v_1 + \tilde{b}_n \mid a_n x + b_n) F_2^{n-k-1}(c_n v_2 + d_n \mid a_n x + b_n)$$
$$F_3(c_n v_2 + d_n \mid a_n x + b_n)\, a_n f_X(a_n x + b_n)\, dx.$$

Since our goal is to compute the joint asymptotic distribution of V_1 and V_2 as $n \to \infty$, we refer to Joshi and Nagaraja (1995) and Nagaraja and David (1994) for what concerns the swap between the limit operation and the integral. Since X is unit Fréchet distributed, it is clear that the marginal F_X is in the domain of attraction of $G_X(x) = e^{-1/x}$ and that $g_X(x) = x^{-2}e^{-1/x}$. We refer to Lemma 1 and Result 1 in Nagaraja and David (1994), for the convergence as $n \to \infty$ of the term $a_n f_X(a_n x + b_n)$ to $g_{(k)}(x) := \frac{(-\log G_X(x))^k}{k!} g_X(x)$. Theorem 1 then follows by Condition (18) and Lemma 2. □

Proof of Theorem 2. Note that $X_N > a_n x + b_n$ is equivalent to $X_N > b_n$, as $a_n \to 0$ when $n \to \infty$. Let us now report the model in Wadsworth and Tawn (2013) in terms of standard normal marginals, namely

$$\begin{aligned}
&\mathbb{P}(Y_E > \zeta_1(n,y) \log n \mid X_E > \log n) \\
&= \mathbb{P}\left(Y_N > \Phi^{-1}\left(1 - n^{-\zeta_1(n,y)}\right) \,\middle|\, X_N > \Phi^{-1}\left(1 - n^{-1}\right)\right) \\
&\sim L_1\left(n; \zeta_1(n,y), 1\right) n^{1-\kappa(\zeta_1(n,y),1)}
\end{aligned} \tag{40}$$

where

$$\kappa(\zeta_1(n,y), 1) = \frac{\zeta_1(n,y) + 1 - 2\rho\sqrt{\zeta_1(n,y)}}{1-\rho^2}, \qquad \text{and} \tag{41}$$

$$L_1(n; \zeta_1(n,y), 1) = (4\pi \log n)^{\frac{2\rho^2 - \rho\left(\sqrt{\zeta_1(n,y)} + 1/\sqrt{\zeta_1(n,y)}\right)}{2(1-\rho^2)}} \frac{\zeta_1(n,y)^{\frac{1-\rho/\sqrt{\zeta_1(n,y)}}{2(1-\rho^2)}} (1-\rho^2)^{3/2}}{\left(\sqrt{\zeta_1(n,y)} - \rho\right)\left(1 - \rho\sqrt{\zeta_1(n,y)}\right)}, \tag{42}$$

whenever $\rho^2 < \min\{\zeta_1(n,y), 1\}$. The authors explain that the SV function L, which we have in the case of the standard exponential marginals (from Equation (32)), is such that $L/L_1 \sim 1$ as $n \to \infty$.
To ease the notation, in the following, we use ζ_1 for $\zeta_1(n,y)$. We want to find some constants $\tilde{a}_{n,E}$ and $\tilde{b}_{n,E}$ such that $\{Y_E > \zeta_1(n,y)\log n\}$ is equivalent to $\{Y_E > \tilde{a}_{n,E}y + \tilde{b}_{n,E}\}$, so we set

$$\zeta_1 = \frac{\tilde{a}_{n,E}y + \tilde{b}_{n,E}}{\log n}. \tag{43}$$

By using the first equality in (40) and knowing that if $Y_N \sim \mathcal{N}(0,1)$ then $-\log(1-\Phi(Y_N)) \sim \text{Exp}(1)$, it follows that

$$\zeta_1 = \frac{-\log\left(1-\Phi(\tilde{a}_n y + \tilde{b}_n\right)}{\log n}. \tag{44}$$

This means that we can rewrite (35) by transforming the marginals into standard exponential and then apply the model (40). Combining (43) and (44), we obtain

$$\begin{aligned}
\tilde{a}_{n,E}y + \tilde{b}_{n,E} &= -\log\left(1-\Phi(\tilde{a}_n y + \tilde{b}_n\right) \\
&\underset{n\to\infty}{\sim} -\log\left(\frac{\phi(\tilde{a}_n y + \tilde{b}_n)}{\tilde{a}_n y + \tilde{b}_n}\right) \text{ through the Mill's ratio, if } \tilde{a}_n \to \infty \text{ or } \tilde{b}_n \to \infty \\
&= -\log\left(\frac{1}{\sqrt{2\pi}}e^{-(\tilde{a}_n y+\tilde{b}_n)^2/2}\right) + \log\left(\tilde{a}_n y + \tilde{b}_n\right) \\
&= \frac{1}{2}\log(2\pi) + \frac{1}{2}\tilde{a}_n^2 y^2 + \tilde{a}_n\tilde{b}_n y + \frac{1}{2}\tilde{b}_n^2 + \log\tilde{b}_n + \log\left(1 + \frac{\tilde{a}_n}{\tilde{b}_n}y\right) \\
&\underset{n\to\infty}{\sim} \frac{1}{2}\log(2\pi) + \frac{1}{2}\tilde{a}_n^2 y^2 + \tilde{a}_n\tilde{b}_n y + \frac{1}{2}\tilde{b}_n^2 + \log\tilde{b}_n + \frac{\tilde{a}_n}{\tilde{b}_n}y \quad \text{if } \tilde{a}_n = o\left(\tilde{b}_n\right) \\
&= \tilde{a}_n\left(\tilde{b}_n + \frac{1}{\tilde{b}_n}\right)y + \frac{1}{2}\log(2\pi) + \frac{1}{2}\tilde{a}_n^2 y^2 + \frac{1}{2}\tilde{b}_n^2 + \log\tilde{b}_n \\
&\underset{n\to\infty}{\sim} \tilde{a}_n\tilde{b}_n y + \frac{1}{2}\tilde{b}_n^2 \quad \text{(the discarded terms are } o\left(\tilde{a}_n\tilde{b}_n\right)\text{)}.
\end{aligned}$$

So, we can identify the terms, which gives

$$\tilde{a}_{n,E} = \tilde{a}_n\tilde{b}_n, \qquad \tilde{b}_{n,E} = \frac{1}{2}\tilde{b}_n^2.$$

Note that, up to now, we have not used the constants in (29b), but found some general conditions on the their growth as n increases. By using (29b), we have

$$\begin{aligned}
\tilde{a}_{n,E} &= \rho\sqrt{1-\rho^2}\left(\sqrt{2\log n} - \frac{1}{2}\frac{\log(4\pi\log n)}{\sqrt{2\log n}}\right) \\
\tilde{b}_{n,E} &= \rho^2\log n - \frac{\rho^2}{2}\log(4\pi\log n) + \frac{\rho^2}{16}\frac{(\log(4\pi\log n))^2}{\log n}.
\end{aligned}$$

Replacing the constants in (43) it yields

$$\zeta_1 = \rho^2\left(1+\frac{1}{\rho}\sqrt{\frac{2(1-\rho^2)}{\log n}}y-\frac{1}{2}\frac{\log(4\pi\log n)}{\log n}+o\left((\log n)^{-1}\right)\right). \tag{46}$$

Now, notice that with ζ_1 defined in (46), the condition $\rho^2 < \min\{\zeta_1(n,y),1\}$ is always satisfied for large n and for $\rho \in (0,1)$, as it yields $\rho^2 < \rho^2 + \mathcal{O}\left((\log n)^{-1/2}\right)$. Hence, since

$$\begin{aligned}\sqrt{\zeta_1} &= \rho\left(1+\frac{1}{\rho}\sqrt{\frac{2(1-\rho^2)}{\log n}}y-\frac{1}{2}\frac{\log(4\pi\log n)}{\log n}+o\left((\log n)^{-1}\right)\right)^{1/2}\\ &\underset{n\to\infty}{\sim} \rho\left(1+\frac{1}{2\rho}\sqrt{\frac{2(1-\rho^2)}{\log n}}y-\frac{1}{4}\frac{\log(4\pi\log n)}{\log n}-\frac{1-\rho^2}{4\log n}y^2+o\left((\log n)^{-1}\right)\right),\end{aligned}$$

the expression of $\kappa(\zeta_1,1)$ becomes

$$\kappa(\zeta_1,1) = 1+\frac{\rho^2}{2\log n}y^2+o\left((\log n)^{-1}\right).$$

By using these expressions into Equation (40), we have

$$\begin{aligned}&\mathbb{P}(Y_E > \zeta_1(n,y)\log n \mid X_E > \log n)\\ &\underset{n\to\infty}{\sim} \frac{\exp\left\{\frac{2\rho^2-\rho\left(\sqrt{\zeta_1}+1/\sqrt{\zeta_1}\right)}{2(1-\rho^2)}\log(4\pi\log n)+\frac{1-\rho/\sqrt{\zeta_1}}{2(1-\rho^2)}\log\zeta_1+(1-\kappa(\zeta_1,1))\log n\right\}}{(1-\rho^2)^{-3/2}(\sqrt{\zeta_1}-\rho)(1-\rho\sqrt{\zeta_1})}.\end{aligned}$$

Since

$$\begin{aligned}\left(1-\frac{\rho}{\sqrt{\zeta_1}}\right)\log\zeta_1 &\underset{n\to\infty}{\sim} \frac{\log\rho}{\rho}\sqrt{\frac{2(1-\rho^2)}{\log n}}y+\frac{1}{2\rho^2}\frac{2\left(1-\rho^2\right)}{\log n}y^2\\ &\quad-\frac{\log\rho}{2}\frac{\log(4\pi\log n)}{\log n}-\frac{1-\rho^2}{2}\log\rho\frac{y^2}{\log n}+o\left((\log n)^{-1}\right)\\ &=\frac{\log\rho}{\rho}\sqrt{\frac{2(1-\rho^2)}{\log n}}y-\frac{\log\rho}{2}\frac{\log(4\pi\log n)}{\log n}\\ &\quad+\left(\log\rho-\frac{2}{\rho^2}\right)\frac{1-\rho^2}{2\log n}y^2+o\left((\log n)^{-1}\right),\end{aligned}$$

$$\begin{aligned}(\sqrt{\zeta_1}-\rho)(1-\rho\sqrt{\zeta_1}) &\underset{n\to\infty}{\sim} \frac{(1-\rho^2)^{3/2}}{\sqrt{2\log n}}y-\rho\frac{1-\rho^2}{4}\frac{\log(4\pi\log n)}{\log n}\\ &\quad-\rho\frac{1-\rho^4}{4\log n}y^2+o\left((\log n)^{-1}\right)\end{aligned}$$

and

$$\begin{aligned}\frac{2\rho^2-\rho\left(\sqrt{\zeta_1}+1/\sqrt{\zeta_1}\right)}{2(1-\rho^2)} &\underset{n\to\infty}{\sim} -\frac{1}{2}+\frac{1}{4\rho}\sqrt{\frac{2(1-\rho^2)}{\log n}}y-\frac{1}{8}\frac{\log(4\pi\log n)}{\log n}\\ &\quad-\frac{1-\rho^2}{8\log n}y^2+o\left((\log n)^{-1}\right),\end{aligned}$$

where we neglected all the terms characterized by a faster decay with respect to $(\log n)^{-1}$, we obtain

$$\mathbb{P}(Y_E > \zeta_1(n,y)\log n \mid X_E > \log n) \underset{n\to\infty}{\sim} \frac{1}{y\sqrt{2\pi}} e^{-(\rho y)^2/2}.$$

□

Proof of Corollary 2. Notice that

$$\mathbb{P}(Y_E > \zeta_1(n,y)\log n \mid X_E > \log n) \underset{n\to\infty}{\sim} \frac{\rho}{\rho y\sqrt{2\pi}} e^{-(\rho y)^2/2} = \rho\frac{\varphi(\rho y)}{\rho y},$$

where φ denotes the probability density function of the standard normal distribution. Under the assumption that y is large, it is straightforward to show, by using the Mill's ratio, that

$$\mathbb{P}(Y_E > \zeta_1(n,y)\log n \mid X_E > \log n) \underset{n\to\infty}{\sim} \rho\,(1-\Phi(\rho y)),$$

which completes the proof. □

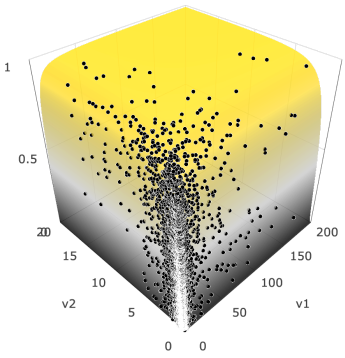
B APPENDIX – DISCUSSION ON k

The present section contains some graphical results of the examples considered in Section 3. We recall that each example runs a simulation of 5000 replicates of 100000 bivariate sample (X,Y). As a result, we obtain a bivariate sample for (V_1,V_2) of dimension 5000 (each replicate of the (X,Y) sample gives one couple of (V_1,V_2)), for each k. The values of k indicating the dimension of the extreme set are fixed to $1, 50, 100, 250, 500, 1000, 1500, 2000$. However, the figures do not show the results for each of the considered values: This choice is made to foster the readability, as a lot of almost similar plots on a single page may be redundant. From a quick look at Figures 4 to 7 for both examples, we can draw the same important conclusions:

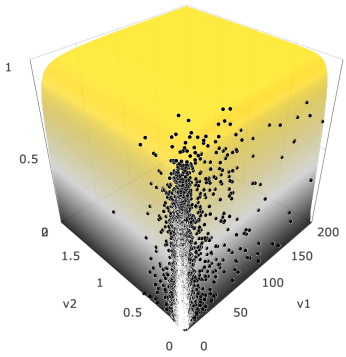
(a) The range of the entries relative to V_1 is much wider than the range of the entries corresponding to V_2;

(b) The cloud of black points in the surfaces in Figure 4 and Figure 6 becomes closer and closer to the v_1 axis as k increases.

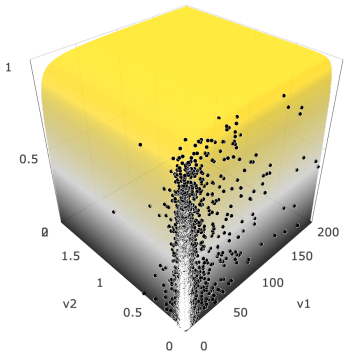
This is a clear effect of the positive dependence between (X,Y). To see this, think about the definition of (V_1,V_2): V_1 is the maximum of the concomitants of the largest k order statistics and V_2 is the maximum of the remaining set. The clear difference between the ranges of the two samples indicates that the highest concomitants contribute to the values of V_1 and the lowest concomitants are responsible of the relatively low values of V_2. Moreover, as k increases, almost all the values that V_2 takes are lower than 1, while V_1 can assume values which are greater than 100, for example. Explained differently, high values of concomitants generally correspond to high order statistics of X.



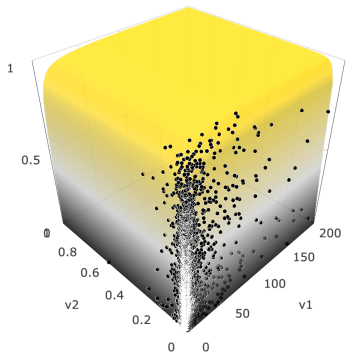
(a) $k = 1$.



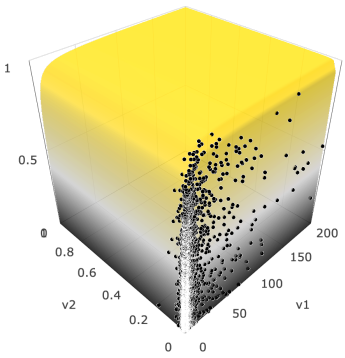
(b) $k = 50$.



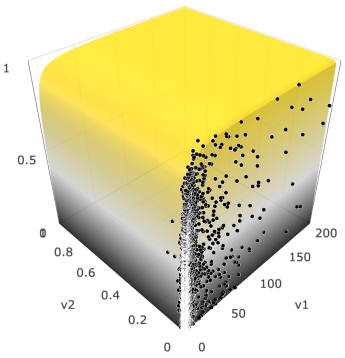
(c) $k = 100$.



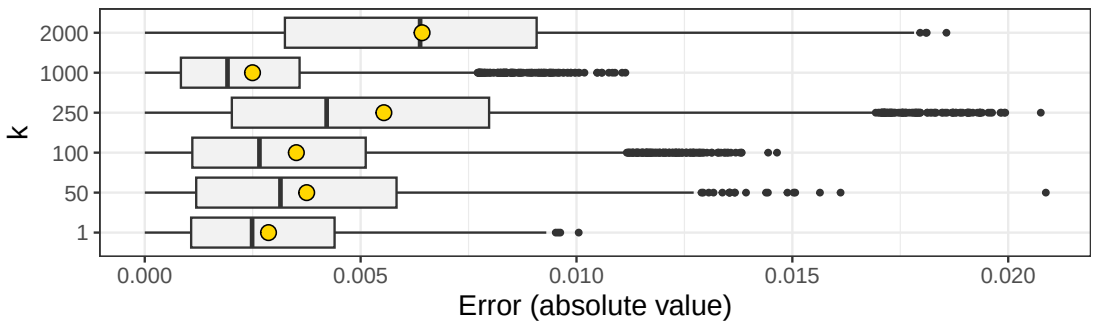
(d) $k = 250$.



(e) $k = 1000$.

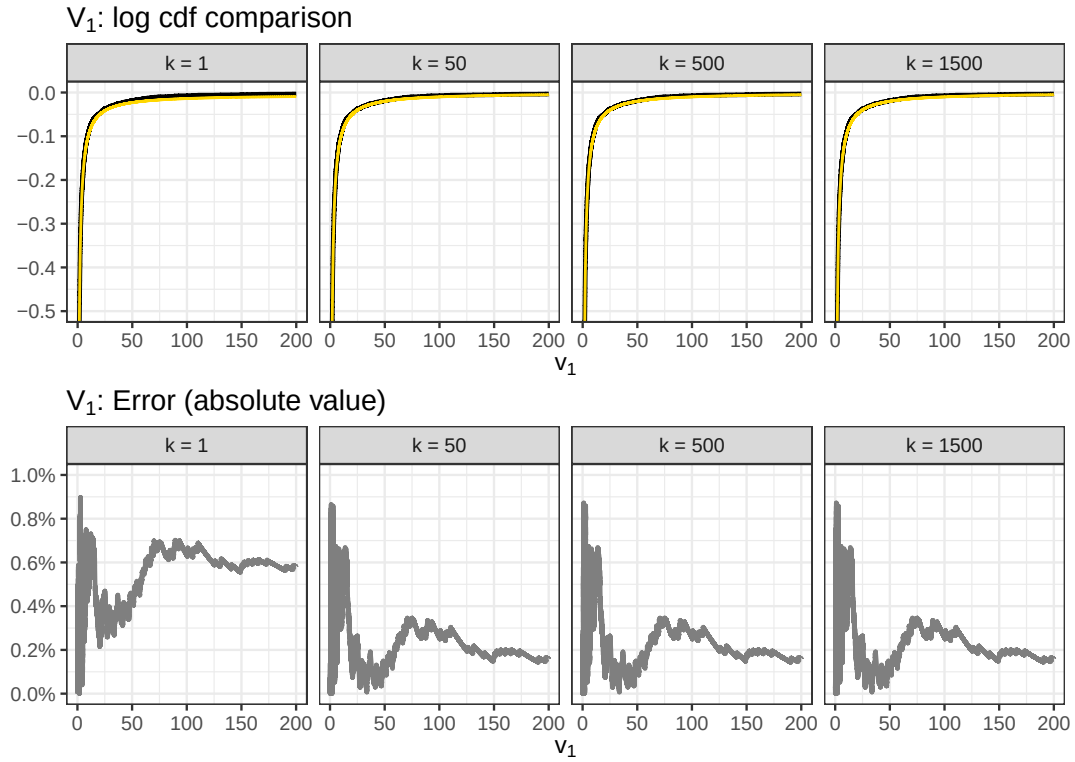


(f) $k = 2000$.

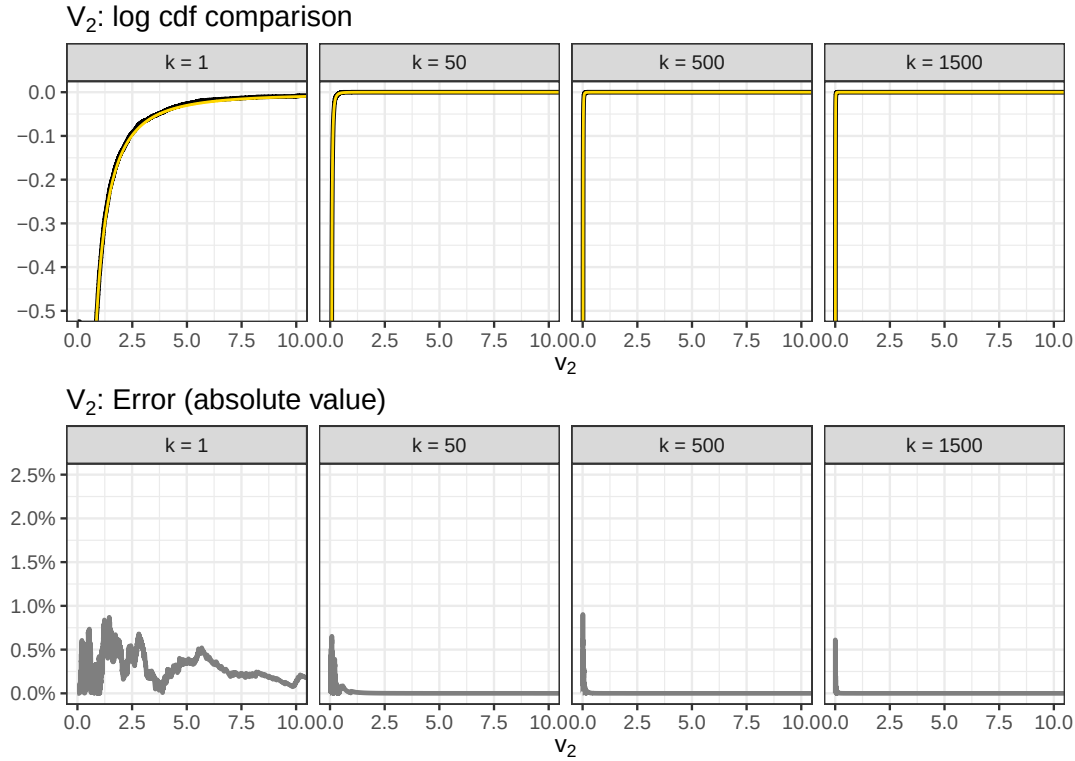


(g) *Boxplots of the errors.*

Figure 4: Survival Clayton($\theta = 2$). Bivariate cdf computed by numerically integrating equation (19), for various values of k. The black bullets correspond to the bivariate empirical cdf of the simulated samples. Panel g) shows the boxplots of the errors in absolute value between the bivariate asymptotic cdf computed in $(V_{1,i}, V_{2,i})$ for $i = 1, \ldots, 5000$ and the corresponding empirical bivariate cdf. The gold dots represent the error means.

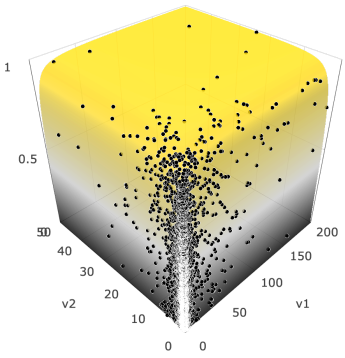


(a) V_1*: log marginal cdf comparison and marginal error behaviour with respect to* v_1.

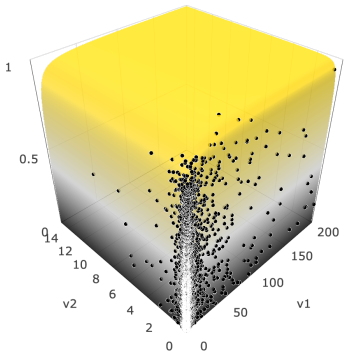


(b) V_2*: log marginal cdf comparison and marginal error behaviour with respect to* v_2.

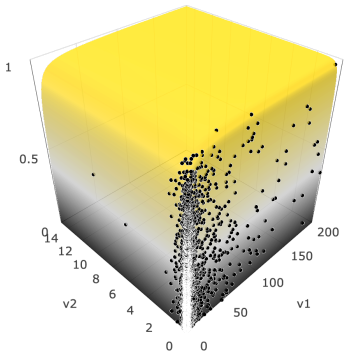
Figure 5: Survival Clayton($\theta = 2$). Empirical univariate cdf of V_1 and V_2, together with the asymptotic theoretical values obtained by numerically integrating the asymptotic joint cdf (19).



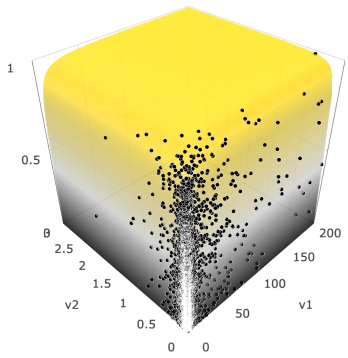
(a) $k = 1$.



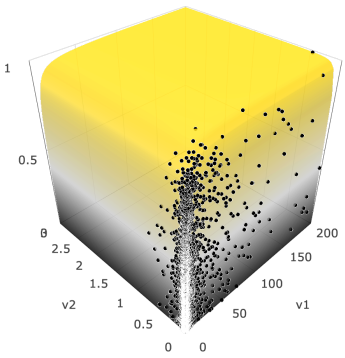
(b) $k = 50$.



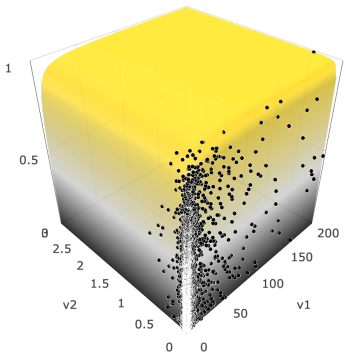
(c) $k = 100$.



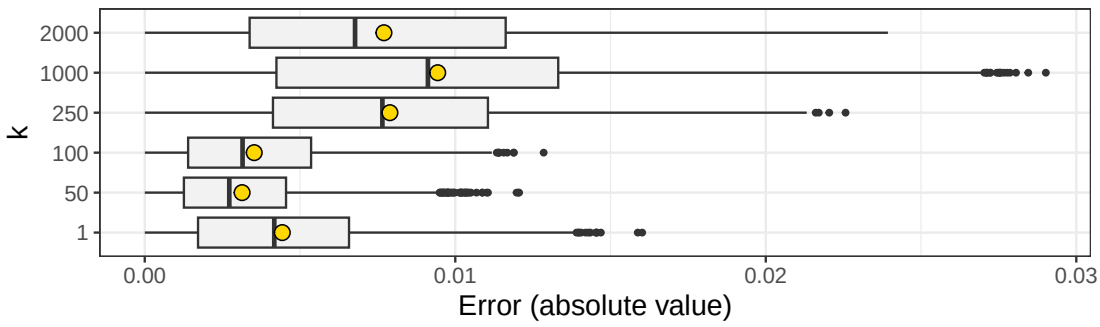
(d) $k = 250$.



(e) $k = 1000$.

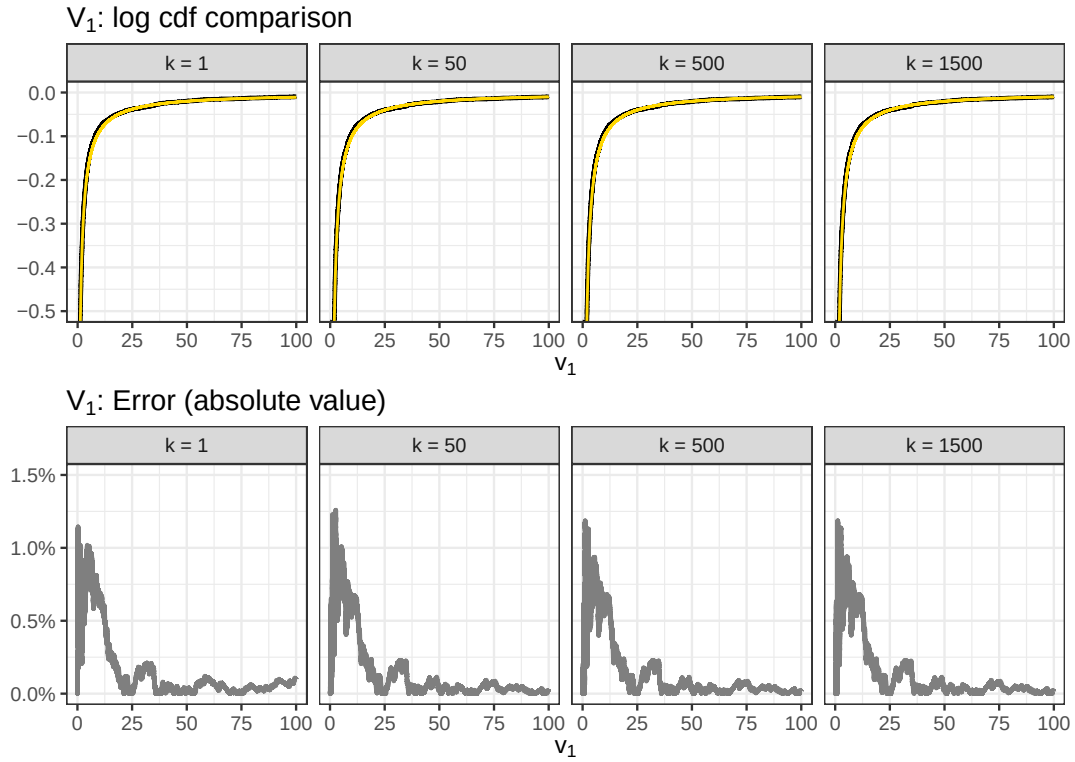


(f) $k = 2000$.

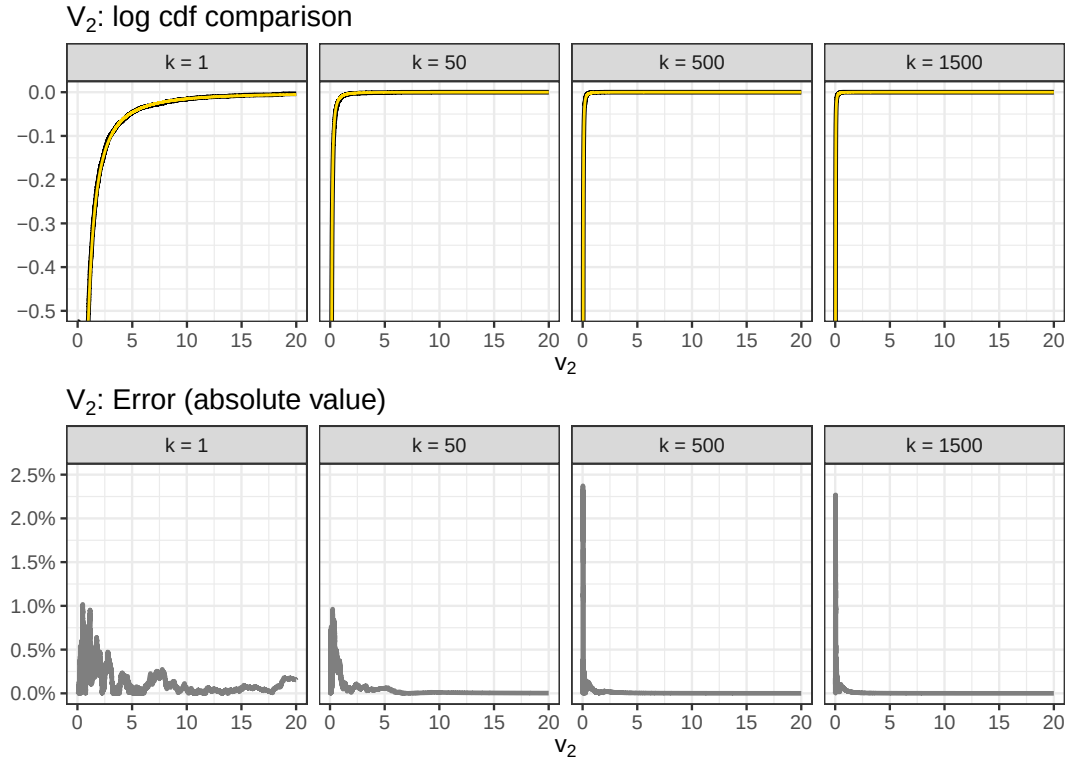


(g) *Boxplots of the errors.*

Figure 6: Logistic($\gamma = 1/2$). Bivariate cdf computed by numerically integrating equation (19), for various values of k. The black bullets correspond to the bivariate empirical cdf of the simulated samples. Panel g) shows the boxplots of the errors in absolute value between the bivariate asymptotic cdf computed in $(V_{1,i}, V_{2,i})$ for $i = 1, \ldots, 5000$ and the corresponding empirical bivariate cdf. The gold dots represent the error means.



(a) V_1*: log marginal cdf comparison and marginal error behaviour with respect to* v_1.



(b) V_2*: log marginal cdf comparison and marginal error behaviour with respect to* v_2.

Figure 7: Logistic($\gamma = 1/2$) example. Empirical univariate cdf of V_1 and V_2, together with the asymptotic theoretical values obtained by numerically integrating the asymptotic joint cdf (19).

It has to be said that the plotting range of the surfaces is chosen in order to enhance the visualization purposes, as extremely wide intervals would have been not suitable to observe the results. However, there can be points which fall outside the plotting area: Figure 4e contains more points in the middle-right part, while Figure 4f contains more points in the upper-right part. This indicates that when k is 1000, some concomitants high in value do not belong to the extreme set, but when k increases to 2000, the partition to which they belong changes. In other words, V_1 "steals" high values to V_2 when k increases. Plots relative to the marginal distributions support the presented conclusions: There is no noticeable difference among the log cdf shown in Figure 5a and Figure 7a, which means that the sample in V_1 does not change so much, while the upper end point of the distribution of V_2 appears to shrink towards 0 (Figure 5b and Figure 7b). This is again an indication of the "stealing" argument: Since high values of concomitants contributing to V_2 pass into the extreme set as k increases, only low values of concomitants are left to compute V_2, which becomes smaller and smaller. As previously said, this is the effect of the positive dependence between (X, Y).
Finally, the extremely low errors shown in the boxplots in Figure 4g and Figure 6g ensure that the asymptotic approximation is almost perfect. The same conclusion can be made from the plots sowing the marginal errors in Figure 5 and Figure 7. Despite some variability for low values of ν_1 and ν_2 (which is small anyway), the errors rapidly approach 0 as ν_1 and ν_2 increase, which is important for practical applications as we are usually interested in studying what happens over high thresholds.

C APPENDIX – THE VON MISES CONDITIONS

This section reports the Von Mises conditions, as written in Resnick (1987), borrowing his notation.

Proposition 1.15 (Resnick (1987)): Suppose F is absolutely continuous with positive density f in some neighborhood of ∞.

(a) If for some $\alpha > 0$

$$\lim_{x\to\infty} xf(x)/(1-F(x)) = \alpha \tag{47}$$

then $F \in \mathscr{D}(\Phi_\alpha)$. We may choose a_n to satisfy $a_n f(a_n) \sim \alpha/n$.

(b) If f is nonincreasing and $F \in \mathscr{D}(\Phi_\alpha)$ then (47) holds.

(c) Equation (47) holds iff for some z_0 and all $x > z_0$, we have

$$1 - F(x) = c \exp\left\{-\int_{z_0}^{x} t^{-1}\alpha(t)\, dt\right\}$$

where $\lim_{t\to\infty} \alpha(t) = \alpha$ and c is a positive constant.

Proposition 1.16 (Resnick (1987)): Suppose F *has finite right end-point* x_0 *and is absolutely continuous in a left neighborhood of* x_0 *with positive density* f.

(a) *If for some* $\alpha > 0$

$$\lim_{x \uparrow x_0} (x_0 - x)\, f(x)/(1 - F(x)) = \alpha \tag{48}$$

then $F \in \mathscr{D}(\Psi_\alpha)$.

(b) *If* f *is nonincreasing and* $F \in \mathscr{D}(\Psi_\alpha)$ *then* (48) *holds.*

(c) *Equation* (48) *holds iff* $c(x)$ *can be taken to be constant in some left neighborhood of* x_0 *in the representation*

$$1 - F(x) = c(x) \exp\left\{ -\int_{x_0 - 1}^{x} \delta(t)/(x_0 - t)\, dt \right\} \quad \text{and for } x < x_0$$

where $\lim_{t \uparrow x_0} \delta(t) = \alpha$, $\lim_{t \uparrow x_0} c(t) = c_0 > 0$.

Proposition 1.17 (Resnick (1987)): Let F *be absolutely continuous in a left neighborhood of* x_0 *with density* f.

(a) *If*

$$\lim_{t \uparrow x_0} f(x) \int_x^{x_0} (1 - F(t))\, dt/(1 - F(x))^2 = 1 \tag{49}$$

then $F \in D(\Lambda)$. *In this case we may take*

$$f(t) = \int_t^{x_0} (1 - F(s))\, ds/(1 - F(x))$$
$$b_n = (1/(1 - F))^{\leftarrow}(n), \quad a_n = f(b_n).$$

(b) *If* f *is nonincreasing and* $F \in \mathscr{D}(\Lambda)$ *then holds.*

(c) *Equation* (49) *holds iff*

$$1 - F(x) = c \exp\left\{ -\int_{z_0}^{x} (g(t)/f(t))\, dt \right\}, \quad z_0 < x < x_0, \tag{50}$$

where $\lim_{t \uparrow 0} g(x) = 1$ *and* f *is absolutely continuous with density* $f(x) \to 0$ *as* $x \uparrow x_0$.

(d) *Equation* (49) *or* (50) *are equivalent to* $tf((1/(1 - F))^{\leftarrow}(t)) \in RV_0$ *(Sweeting (1985)).*

ESSEC Business School
3 avenue Bernard-Hirsch
CS 50105 Cergy
95021 Cergy-Pontoise Cedex
France
Tél. +33 (0)1 34 43 30 00
www.essec.edu

ESSEC Executive Education
CNIT BP 230
92053 Paris-La Défense
France
Tél. +33 (0)1 46 92 49 00
www.executive-education.essec.edu

ESSEC Asia-Pacific
5 Nepal Park
Singapore 139408
Tél. +65 6884 9780
www.essec.edu/asia

ESSEC | CPE Registration number 200511927D
Period of registration: 30 June 2017 - 29 June 2023
Committee of Private Education (CPE) is part of SkillsFuture Singapore (SSG)

ESSEC Afrique
Plage des Nations - Golf City
Route de Kénitra - Sidi Bouknadel (Rabat-Salé)
Morocco
Tél. +212 (0)5 37 82 40 00
www.essec.edu

CONTACT

Research Center
research@essec.edu