Contents

Introduction

Context

Management of environmental resources often requires the analysis of multivariate extreme values. In climate studies, extreme events such as heavy precipitation and record temperatures represent a major challenges due to their consequences. In the classical statistical theory, one is often interested in the behavior of the mean or average of a random variable X defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. This average will then be described through the expected value $\mathbb{E}[X]$ of the distribution. But in case of extreme events, it can be just as important to estimate tails probabilities. Furthermore, what if the second moment $\mathbb{E}[X^2]$ or even the mean is not finite? Then the central limit theorem does not apply and the classical theory, carried by the normal distribution, is no longer relevant.

Also, inference methods for assessing dependence have been increasingly in demand. The most popular approach is based on second moment of the underlying random variables, the covariance. It is well known that only linear dependence can be captured by the covariance that it is characterizing only for a few special classes of distribution. As a beneficial alternative of dependence modeling, the concept of copulas going back to [Sklar, 1959]. The copula $C : [0,1]^2 \rightarrow [0,1]$ of a random vector (X, Y) allows us to separate the effect of dependence from the effects of the marginal distribution such as :

$$\mathbb{P}\left\{X \leq x, Y \leq y\right\} = C\left(\mathbb{P}\left\{X \leq x\right\}, \mathbb{P}\left\{Y \leq y\right\}\right), \quad \forall (x,y) \in \mathbb{R}^2.$$

The main consequence of this identity is that the copula completely characterizes the stochastic dependence between X and Y. Investigating the notion of copulas within the framework of multivariate extreme value theory leads to the so called extreme value copulas.

Some extreme events, such as heavy precipitation or wind speed has also spatial characteristics and geostatisticians are striving to better understand the physical process in hands. In geostatistic, we often consider $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space, S a set of locations and $(E, \mathcal{E})$ a measurable state space. One can define on this probability space a stochastic process $X = \{X(s), s \in S\}$ with values on $(E, \mathcal{E})$. It is classical to define the following second-order statistic as the variogram (see [Gaetan and Guyon, 2008] Chapter 1.3 for definition and basic properties) :

$$2\gamma(h) = \mathbb{E}[|X(s+h) - X(s)|^2],$$

where $\{X(s), s \in S\}$ represents a spatial and stationary process with a well-defined covariance function. The function $\gamma(\cdot)$ is called the semi-variogram of X. With respect to extremes, this definition is not well adapted because a second order statistic is difficult to interpret inside the framework of extreme value theory or may not even be defined. To ensure that we always work with finite moments quantities, the following type of first-order variogram is introduced by [Cooley et al., 2006]

$$\nu(h) = \frac{1}{2}\mathbb{E}\left[|F(X(s+h)) - F(X(s))|\right], \tag{1}$$

where $F(u) = \mathbb{P}(X(s) \leq u)$ is named as the FMadogram. His link to the pairwise extremal dependence function (Section 4.3 of [Coles et al., 1999]) or the Pickands dependence function ([Pickands, 1981]) make it an interesting quantity to capture the dependence between the extrema of stochastic processes or random variables. Furthermore, this quantity may be seen as a dissimilarity measure among bivariate maxima to be used for clustering time-series as shown by [Bernard et al., 2013] or [Bador et al., 2015].

The main drawback of this quantity is that it only focus on the value of the diagonal section of the pairwise extremal dependence function. In the bivariate case, the FMadogram characterize solely the extremal coefficient for random variables X and Y (see Section 8.2.7 of [Beirlant et al., 2004]). To overpass this drawback, [Naveau et al., 2009] introduce the λ-FMadogram defined as,

$$\nu(h, \lambda) = \frac{1}{2}\mathbb{E}\left[\left|F^{\lambda}(X(s+h)) - F^{1-\lambda}(X(s))\right|\right], \tag{2}$$

for every $\lambda \in [0, 1]$. This quantity characterizes the pairwise extremal dependence function outside the diagonal section but also the whole Pickands dependence function ([Marcon et al., 2017]) and contribute to the vast literature of the estimation of the Pickands dependence function for bivariate extreme value copulas (see for example [Pickands, 1981], [Deheuvels, 1991], [Capéraà et al., 1997] or [Hall and Tajvidi, 2000]). Statisticians may estimate these quantities, but the classical results such as strong consistency or weak convergence may only apply if the data in hands are clean as possible. This induces that the process of data collection has not been corrupted such as the dataset is complete, each observation is independent from others and that the implicit law of the observations is still the same.

Nevertheless, as the volume of data expands, the problem of missing or contaminated data has been increasingly present in many fields. It frequently happens that some individuals of a sample from a multivariate population are not observed. If a sample be represented in matrix form by allowing rows to represent the individuals and columns the variables, then the matrix of the type of sample with which we are concerned is sparse. In dealing with fragmentary samples, it is important to have at hand techniques which will enable the statistician to extract as much information as possible from the data. A useful reference for general parametric statistical inferences with missing data was provided by [Little R.J.A., 1987].

Considering a sample from a random vector (X, Y) of incomplete data,

$$(X_t, Y_t, \delta_t), \quad t \in \{1, \dots, T\}, \tag{3}$$

where all the X_t's are observed and $\delta_t = 0$ if Y_t is missing, otherwise $\delta_t = 1$. The simple missing data pattern describes by (3) is basically created by the double sampling or two phase sampling (see Chapter 12 of [Cochran, 2007]). Samples like (3) may arise in survival analysis : The study of the duration time preceding an event of interest is considered with a series of random censors, which might prevent the capture of the whole survival time. This is known as the censoring mechanism and it arises from restrictions depending of the nature of the study. Typically, they may occur in medicine, with studies of the survival times before the recovery / decease from a specific disease. Another important example is often realized in comparing treatment effects of two educational programs. Individuals with lower scores on a preliminary test are more likely to receive the experimental treatment (*i.e.*, a compensatory study program), whereas those with higher preliminary scores are more inclined to take the standard control. This phenomenon is well-known as the selection problem and we refer to Chapter 2 of [Angrist and Pischke, 2008] for more details. Beside of missing observations, the process of data might be disturb in a way that innerly deteriorate the quality of some data and one may ask that the estimation process of the Madograms (Equation (1) and (2)) should be robust.

The topic of robustness in estimation has known an important research activity developed in the 60's and 70's resulting in a large number of publications. For a summary, the interested reader is referred to [Huber, 2011]. Robustness can be seen as an estimation procedure in which both stochastic and approximation errors are low (see Section 1.1 from [Baraud et al., 2016]). In other words, an estimator is robust if our model provides a reasonable approximation of the true one and derive an estimator which remains close to the true distribution. In this report, we mean by *robust* as *robust against outlier*, *e.g* the ϵ-contamination model (see [Huber, 1964]), or *robust again heavy-tailed data* where only low-order moments are assumed to be finite for the data distribution. There is no simple relation between the two definitions and the first framework of robustness that we have depicted. It is a main goal of this report to develop estimates of the (λ)-FMadogram in these sticky situations. To achieve our goal, we make of use of the hybrid copula estimator ([Segers, 2014]) for the missing data framework and we leverage the idea of Median-of-meaNs (MoN) for the contaminated data scheme.

Definitions and Notation

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and (X, Y) be a bivariate random vector with values in $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$. This random vector has a joint distribution function H and marginal distribution function F and G. A function $C : [0,1]^2 \rightarrow [0,1]$ is called a bivariate copula if it is the restriction to $[0,1]^2$ of a bivariate distribution function whose margins are given by the uniform distribution on the interval $[0,1]$. Since the work of [Sklar, 1959], it is well known that every distribution function H can be decomposed as $H(x,y) = C(F(x), G(y))$, for all $(x,y) \in \mathbb{R}^2$.

Definition 1 ([Gudendorf and Segers, 2009]). *A bivariate copula C is an extreme-value copula if and only if it admits a representation of the form*

$$C(u,v) = (uv)^{A\left(\frac{log(v)}{log(uv)}\right)}, \tag{4}$$

for all $u, v \in [0,1]$ and where $A(\cdot)$ is the Pickands dependence function, i.e., *$A : [0,1] \longrightarrow [1/2, 1]$ is convex and satisfies $t \vee (1-t) \leq A(t) \leq 1$, $\forall t \in [0,1]$.*

The upper and lower bound of A has special meanings, the upper bound $A(t) = 1$ corresponds to independence, whereas the lower bound $A(t) = t \vee (1-t)$ corresponds to the perfect dependence (comonotonicity). Notice that, on sections, the extreme value copula is of the form

$$C(u^t, u^{1-t}) = u^{A(t)}. \tag{5}$$

Let $(X_t, Y_t)_{t=1,\dots,T}$ be an *i.i.d.* sample of a bivariate random vector whose underlying copula is denoted by C and whose margins by F, G. For $x, y \in \mathbb{R}$, let $x \wedge y = \min(x,y)$ and $x \vee y = \max(x,y)$. Let $(b_{t,j})_{t \geq 1, j \in \{1,2\}}$ and $(a_{t,j})_{t \geq 1, j \in \{1,2\}}$ be respectively sequences of numbers and sequences of positive reals. We say that the sequence $(a_{T,1}^{-1}(\bigvee_{t=1}^T X_t - b_{T,1}), a_{T,2}^{-1}(\bigvee_{t=1}^T Y_t - b_{T,2}))$ belongs to the domain of attraction of H, if for all real values x, y (at which the limit is continuous and non-degenerate)

$$\mathbb{P}\left(\frac{\bigvee_{t=1}^T X_t - b_{T,1}}{a_{T,1}} \leq x, \frac{\bigvee_{t=1}^T Y_t - b_{T,2}}{a_{T,2}} \leq y \right) \underset{T \to \infty}{\rightarrow} H(x,y).$$

If this relationship hold, H is said to be a multivariate extreme value distribution. The FMadogram is one of the quantity used in dependence modeling to characterize the extremal coefficient, *i.e.* in the bivariate case $2A(2^{-1})$. We will call by FMadogram the following quantity

$$\nu = \frac{1}{2}\mathbb{E}\left[|F(X) - G(Y)|\right], \tag{6}$$

and the λ-FMadogram by the expression

$$\nu(\lambda) = \frac{1}{2}\mathbb{E}\left[\left|F^\lambda(X) - G^{1-\lambda}(Y)\right|\right]. \tag{7}$$

We suppose that we observe sequentially a quadruple defined by

$$(I_t X_t, J_t Y_t, I_t, J_t), \quad t \in \{1, \dots, T\}, \tag{8}$$

where $I_t = 0$ (resp. $J_t = 0$) if X_t (resp. Y_t) is missing, otherwise $I_t = 1$ (resp. $J_t = 1$), *i.e.* at each $t \in \{1, \dots, T\}$, one or both entries may be missing. The probability of observing a realization partially or completely, is denoted by $p_X = \mathbb{P}(I_t = 1) > 0$, $p_Y = \mathbb{P}(J_t = 1) > 0$, $p_{XY} = \mathbb{P}(I_t = 1, J_t = 1) > 0$ and we note by $\mathbf{p} = (p_X, p_Y, p_{XY})$. Let us now define the empirical cumulative distribution of X (resp. Y and (X,Y)) in case of missing data,

$$\hat{F}_T(u) = \frac{\sum_{t=1}^T 1_{\{X_t \leq u\}} I_t}{\sum_{t=1}^T I_t}, \quad \hat{G}_T(v) = \frac{\sum_{t=1}^T 1_{\{Y_t \leq v\}} J_t}{\sum_{t=1}^T J_t}, \quad \hat{H}_T(u,v) = \frac{\sum_{t=1}^T 1_{\{X_t \leq u, Y_t \leq v\}} I_t J_t}{\sum_{t=1}^T I_t J_t}. \quad (9)$$

Here, we weight the estimator by the number of observed data which is a natural estimator if divided by T of the probabilities of missing. We have all tools in hand to recall the definition of the *hybrid copula estimator* introduced by [Segers, 2014],

$$\hat{C}_T^{\mathcal{H}}(u,v) = \hat{H}_T(\hat{F}_T(u), \hat{G}_T(v)). \quad (10)$$

The normalized estimation error of the hybrid copula estimator is

$$\mathbb{C}_T^{\mathcal{H}}(u,v) = \sqrt{T}\left(\hat{C}_T^{\mathcal{H}}(u,v) - C(u,v)\right). \quad (11)$$

We will write the generalized inverse function of F (respectively G) as $F^{\leftarrow}(u) = inf\{v \in \mathbb{R} | F(v) \geq u\}$ (respectively $G^{\leftarrow}(u) = inf\{v \in \mathbb{R} | G(v) \geq u\}$) where $0 < u, v < 1$. Given $\mathcal{X} \subset \mathbb{R}^2$, let $\ell^\infty(\mathcal{X})$ denote the spaces of bounded real-valued function on $\mathcal{X}$. For $f : \mathcal{X} \rightarrow \mathbb{R}$, let $||f||_\infty = sup_{x \in \mathcal{X}} |f(x)|$. Here, we use the abbreviation $Qf = \int f dQ$ for a given measurable function f and signed measure Q. The arrows $\overset{a.s.}{\rightarrow}$, $\overset{d}{\rightarrow}$ denote almost sure convergence and convergence in distribution of random vectors. Weak convergence of a sequence of maps will be understood in the sense of J.Hoffman-Jørgensen (see Part 1 in the monograph by [van der Vaart and Wellner, 1996]). Given that $X, X_t, t \in \mathbb{N}^*$ are maps from $(\Omega, \mathcal{A}, \mathbb{P})$ into a metric space $\mathcal{X}$ and that X is Borel measurable, $(X_t)_{t \geq 1}$ is said to converge weakly to X if $\mathbb{E}^* f(X_t) \rightarrow \mathbb{E} f(X)$ for every bounded continuous real-valued function f defined on $\mathcal{X}$, where $\mathbb{E}^*$ denotes outer expectation in the event that X_t may not be Borel measurable. In what follows, weak convergence is denoted by $X_t \rightsquigarrow X$.

This work is organized as follows. In Chapter 1 we state some results on the weak convergence of the estimator of the λ-FMadogram with missing data. We also propose a closed formula for the asymptotic variance of the λ-FMadogram for a fixed $\lambda \in [0,1]$. To propose a robust estimator of the FMadogram, we leverage the idea of Median-of-meaNs (MoN) and state a concentration inequality that this estimator does verify.

Chapter 2 will present the performance of our estimator in a finite-sample framework. The asymptotic variance of the normalized estimation error of several models will be drawn with their empirical counterpart obtained through simulation. We also propose a reproduction of the experiment of the λ-FMadogram with a Smith's process as in [Naveau et al., 2009] and we will explain the augmentation of the Mean Squared Error while h is close to zero, *i.e.* where two locations are strongly dependent. This phenomenon will be also thoroughly explained through simulations and a counterexample.

For the ease of reading, we postponed all technical arguments and proofs in Chapter 3.

Chapter 1

Non parametric estimation of the Madogram with missing data

1.1 Considered estimator

For the rest of this report, we will assume that the copula C is of extreme value type as defined in Definition 1. Following [Segers, 2012], to guarantee the weak convergence of our empirical copula process, we introduce the following smoothness assumptions (see example 3.5 for a specific discussion on extreme value copula).

Assumption A.

(i) The bivariate distribution function H has continuous margins F and G.

(ii) The derivative of the Pickands dependence function $A'(t)$ exists and is continuous on $(0,1)$.

The Assumption A *(i)* guarantees the uniqueness of the representation $H(x,y) = C(F(x), G(y))$ on the range of (F,G). Under the Assumption A *(ii)*, the first-order partial derivatives of C with respect to u (*resp.* with respect to v) exist and is continuous on the set $\{(u,v) \in [0,1]^2 : 0 < u < 1\}$ (*resp.* on the set $\{(u,v) \in [0,1]^2 : 0 < v < 1\}$). We discuss that Assumption A *(ii)* is verified for extreme value copulas in Appendix A.

Definition 2. *Let $(I_1 X_1, J_1 Y_1, I_1, J_1), \dots, (I_T X_T, J_T Y_T, I_T, J_T)$ be a sample given by Equation (8), we defined the hybrid estimator of the λ-FMadogram*

$$\hat{\nu}_T^{\mathcal{H}}(\lambda) = \frac{1}{2\sum_{t=1}^T I_t J_t} \sum_{t=1}^T \left| \hat{F}_T^{\lambda}(X_t) - \hat{G}_T^{1-\lambda}(Y_t) \right| I_t J_t. \tag{1.1}$$

One may verify that in the complete data framework, *i.e.* with $p_X = p_Y = p_{XY} = 1$ we retrieve the λ-FMadogram such as defined in [Naveau et al., 2009], namely

$$\hat{\nu}_T(\lambda) = \frac{1}{2T} \sum_{t=1}^T \left| \hat{F}_T^{\lambda}(X_t) - \hat{G}_T^{1-\lambda}(Y_t) \right|, \tag{1.2}$$

with $\hat{F}_T$ (resp. $\hat{G}_T$) the empirical cumulative distribution function of X (resp. Y).

Remark 1. *Our estimator defined in (1.1) does not verify $\hat{\nu}_T^{\mathcal{H}}(0) = \hat{\nu}_T^{\mathcal{H}}(1) = 0.25$. In addition, the variance at $\lambda = 0$ or $\lambda = 1$ does not equal 0. Indeed, suppose that we evaluate this statistic at $\lambda = 0$, we thus obtain the following quantity :*

$$\hat{\nu}_T^{\mathcal{H}}(0) = \frac{1}{2\sum_{t=1}^T I_t J_t} \sum_{t=1}^T \left(1 - \hat{G}_T(Y_t)\right) I_t J_t.$$

In this situation, the sample $(X_t)_{t=1}^T$ is taken account through the indicator's sequence $(I_t)_{t=1}^T$ and induce a supplementary variance when estimating.

We can force our estimator as in [Naveau et al., 2009] to satisfy these endpoint corrections. This leads to the following corrected estimator :

$$\begin{aligned}\hat{\nu}_T^{\mathcal{H}*}(\lambda) = & \frac{1}{2\sum_{t=1}^T I_t J_t} \sum_{t=1}^T \left| \hat{F}_T^{\lambda}(X_t) - \hat{G}_T^{1-\lambda}(Y_t) \right| I_t J_t - \frac{\lambda}{2\sum_{t=1}^T I_t J_t} \sum_{t=1}^T \{1 - \hat{F}_T^{\lambda}(X_t)\} I_t J_t \\ & - \frac{1-\lambda}{2\sum_{t=1}^T I_t J_t} \sum_{t=1}^T \{1 - \hat{G}_T^{1-\lambda}(Y_t)\} I_t J_t + \frac{1-\lambda+\lambda^2}{2(2-\lambda)(1+\lambda)} \end{aligned} \tag{1.3}$$

Nevertheless, in the missing data framework, the asymptotic behaviour of $\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}*}(\lambda) - \nu(\lambda)\right)$ is not the same as $\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}}(\lambda) - \nu(\lambda)\right)$ and they should be studied apart. Now, we assume some conditions on the missing mechanism,

Assumption B. *We suppose for all $t \in \{1, \dots, T\}$, the pairs (I_t, J_t) and (X_t, Y_t) are independent, the data are missing completely at random (**MCAR**). Furthermore, we suppose that there exists at least one $t \in \{1, \dots, T\}$ such that $I_t J_t \neq 0$.*

Under this assumption, we state the strong consistency of our hybrid estimator of the λ-FMadogram.

Proposition 1 (strong consistency). *Let $(I_1 X_1, J_1 Y_1, I_1, J_1), \dots, (I_T X_T, J_T Y_T, I_T, J_T)$ a i.i.d sample given by Equation (8). We have, under Assumption B for a fixed $\lambda \in [0,1]$, as $T \to \infty$*

$$\hat{\nu}_T^{\mathcal{H}}(\lambda) \xrightarrow{a.s.} \nu(\lambda).$$

Details on the proof are given in Section 3.1 in Chapter 3.

1.2 Functional central limit theorem with possible missing data

Without missing data, the weak convergence of the normalized estimation error of the empirical copula process has already been proved by [Fermanian et al., 2004] under the sole Assumption A. The difference being that C is continuously differentiable on the closed cube. This statement make use of previous results on the Hadamard differentiability of the map $\phi : D([0,1]^2) \to$

$\ell^\infty([0,1]^2)$ which transforms the cumulative distribution function H into its copula function C (see Lemma 3.9.28 from [van der Vaart and Wellner, 1996]). With the hybrid copula estimator, we need a technical assumption in order to guarantee the weak convergence of the process $\mathbb{C}_T^{\mathcal{H}}$ (see [Segers, 2014]),

Assumption C. *In the space $l^\infty(\mathbb{R}^2) \otimes (l^\infty(\mathbb{R}), l^\infty(\mathbb{R}))$ equipped with the topology of uniform convergence, we have the joint weak convergence*

$$\left(\sqrt{T}(\hat{H}_T - H); \sqrt{T}(\hat{F}_t - F), \sqrt{T}(\hat{G}_t - G)\right) \rightsquigarrow (\alpha \circ (F,G), \beta_1 \circ F, \beta_2 \circ G).$$

The stochastic processes α and β_j take values in $l^\infty([0,1]^2)$ and $l^\infty([0,1])$ respectively, and are such that $\alpha \circ F$ and $\beta_j \circ F_j$ have continuous trajectories on $[-\infty,\infty]^2$ and $[-\infty,\infty]$ almost surely.

Under Conditions A and C (see Theorem 1 in Appendix B), the stochastic process $\mathbb{C}_T^{\mathcal{H}}$ converges weakly to the tight Gaussian process S_C defined by,

$$S_C(u,v) = \alpha(u,v) - \frac{\partial C(u,v)}{\partial u}\beta_1(u) - \frac{\partial C(u,v)}{\partial v}\beta_2(v), \quad \forall (u,v) \in [0,1]^2. \tag{1.4}$$

Considering our statistical framework and missing mechanism, [Segers, 2014] shows (in Example 3.5) that the processes α, β_1 and β_2 take the following closed form :

$$\begin{aligned}
\beta_1(u) &= p_X^{-1}\mathbb{G}\left(1_{X \leq F^{\leftarrow}(u), I=1} - u 1_{I=1}\right), \\
\beta_2(v) &= p_Y^{-1}\mathbb{G}\left(1_{Y \leq G^{\leftarrow}(v), J=1} - v 1_{J=1}\right), \\
\alpha(u,v) &= p_{XY}^{-1}\mathbb{G}\left(1_{X \leq F^{\leftarrow}(u)} 1_{Y \leq G^{\leftarrow}(v), I=1, J=1} - C(u,v) 1_{I=1,J=1}\right).
\end{aligned}$$

Furthermore, we are able to compute their covariance functions given in the following lemma.

Lemma 1. *The covariance function of the process $\beta_1(u)$, $\beta_2(v)$ and $\alpha(u,v)$ are, for $(u, u_1, u_2, v, v_1, v_2) \in [0,1]^6$,*

$$\begin{aligned}
cov\left(\beta_1(u_1), \beta_1(u_2)\right) &= p_X^{-1}\left(u_1 \wedge u_2 - u_1 u_2\right), \\
cov\left(\beta_2(v_1), \beta_2(v_2)\right) &= p_Y^{-1}\left(v_1 \wedge v_2 - v_1 v_2\right), \\
cov\left(\beta_1(u), \beta_2(v)\right) &= \frac{p_{XY}}{p_X p_Y}\left(C(u,v) - uv\right),
\end{aligned}$$

and

$$\begin{aligned}
cov\left(\alpha(u_1,v_1), \alpha(u_2,v_2)\right) &= p_{XY}^{-1}\left(C(u_1 \wedge u_2, v_1 \wedge v_2) - C(u_1,v_1)C(u_2,v_2)\right), \\
cov\left(\alpha(u_1,v), \beta_1(u_2)\right) &= p_X^{-1}\left(C(u_1 \wedge u_2, v) - C(u_1,v)u_2\right), \\
cov\left(\alpha(u,v_1), \beta_2(v_2)\right) &= p_Y^{-1}\left(C(u, v_1 \wedge v_2) - C(u,v_1)v_2\right).
\end{aligned}$$

Some technical details are available in Section 3.2 in Chapter 3. We have all tools in hand to consider the weak convergence of the stochastic processes $(\sqrt{T}(\hat{\nu}^{\mathcal{H}}(\lambda) - \nu(\lambda)))_{\lambda \in [0,1]}$ and

$(\sqrt{T}(\hat{\nu}^{\mathcal{H}*}(\lambda)-\nu(\lambda)))_{\lambda\in[0,1]}$. To establish such a result, we use empirical process arguments formulated in [van der Vaart and Wellner, 1996]. This allows us to show the following theorem.

Theorem 1 (limit theorem with missing data). *Under Assumptions A, B, C we have the weak convergence in $\ell^\infty([0,1])$ for the hybrid estimator defined in (1.1) and (1.3), as $T\to\infty$,*

$$\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}}(\lambda)-\nu(\lambda)\right) \rightsquigarrow \Bigg(\frac{1}{2}\int_{[0,1]}\alpha(x^{\frac{1}{\lambda}},1)-\beta_1(x^{\frac{1}{\lambda}})dx+\frac{1}{2}\int_{[0,1]}\alpha(1,x^{\frac{1}{1-\lambda}})-\beta_2(x^{\frac{1}{1-\lambda}})dx \\ -\int_{[0,1]}S_C(x^{\frac{1}{\lambda}},x^{\frac{1}{1-\lambda}})dx\Bigg)_{\lambda\in[0,1]}, \tag{1.5}$$

$$\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}*}(\lambda)-\nu(\lambda)\right) \rightsquigarrow \Bigg(\frac{1-\lambda}{2}\int_{[0,1]}\alpha(x^{\frac{1}{\lambda}},1)-\beta_1(x^{\frac{1}{\lambda}})dx+\frac{\lambda}{2}\int_{[0,1]}\alpha(1,x^{\frac{1}{1-\lambda}})-\beta_2(x^{\frac{1}{1-\lambda}})dx \\ -\int_{[0,1]}S_C(x^{\frac{1}{\lambda}},x^{\frac{1}{1-\lambda}})dx\Bigg)_{\lambda\in[0,1]}. \tag{1.6}$$

Details of the proof are given in Section 3.3. As an integral of a tight Gaussian process, we know that the two normalized estimation errors follows a centered Gaussian variable for a given $\lambda\in[0,1]$. Furthermore, some computations are able to give a closed form of the variance of the limiting Gaussian law as an integral of the Pickands dependence function. This is summarized with the following corollary.

Proposition 2 (asymptotic variance closed formula). *For $\lambda\in[0,1]$, let $A_1(\lambda)=A(\lambda)/\lambda$, $A_2(\lambda)=A(\lambda)/(1-\lambda)$. Then, under the framework of Theorem 1, $\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}}(\lambda)-\nu(\lambda)\right)$ and $\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}*}(\lambda)-\nu(\lambda)\right)$ converges in distribution toward a centered Gaussian random variable with variance which has the following closed form*

$$Var\left(\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}}(\lambda)-\nu(\lambda)\right)\right)\underset{T\to\infty}{\longrightarrow}\frac{1}{4}\sigma_1^2(p_X,p_{XY})+\frac{1}{4}\sigma_2^2(p_Y,p_{XY})+\sigma_3^2(\boldsymbol{p}) \\ +\frac{1}{2}\sigma_{12}(\boldsymbol{p})-\sigma_{13}(\boldsymbol{p})-\sigma_{23}(\boldsymbol{p}),$$

$$Var\left(\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}*}(\lambda)-\nu(\lambda)\right)\right)\underset{T\to\infty}{\longrightarrow}\frac{(1-\lambda)^2}{4}\sigma_1^2(p_X,p_{XY})+\frac{\lambda^2}{4}\sigma_2^2(p_Y,p_{XY})+\sigma_3^2(\boldsymbol{p}) \\ +\lambda(1-\lambda)\frac{1}{2}\sigma_{12}(\boldsymbol{p})-(1-\lambda)\sigma_{13}(\boldsymbol{p})-\lambda\sigma_{23}(\boldsymbol{p}).$$

The expression of the functions $(\sigma_i^2)_{i\in\{1,2,3\}}$ and $(\sigma_{ij})_{i,j\in\{1,2,3\},i<j}$ are detailed in Section 3.4 of Chapter 3. Note that the variance of the limiting process of the normalized estimation error of $\hat{\nu}_T^{\mathcal{H}*}(\lambda)$ for a given λ is not always lower than that of $\hat{\nu}_T^{\mathcal{H}}(\lambda)$.

Corollary 1 (independence closed formula). *In the framework of Theorem 1 and if $C(u,v)=$*

uv, then the functions σ_3^2 *and* $(\sigma_{ij})_{i,j\in\{1,2,3\},i<j}$ *has the following form, for* $\lambda \in [0,1]$

$$\sigma_3^2(\boldsymbol{p}) = \left(\frac{\lambda(1-\lambda)}{1+\lambda(1-\lambda)}\right)^2 \left(\frac{p_{XY}^{-1}}{1+2\lambda(1-\lambda)} - \frac{p_X^{-1}(1-\lambda)}{2-(1-\lambda)+2\lambda(1-\lambda)} - \frac{p_Y^{-1}\lambda}{2-\lambda+2\lambda(1-\lambda)}\right),$$

$$\sigma_{13}(\boldsymbol{p}) = \left(p_{XY}^{-1} - p_X^{-1}\right)\lambda^2(1-\lambda)\left(\frac{1}{(1+\lambda)(1-\lambda)+\lambda+\lambda(1-\lambda)} + \frac{1}{1+\lambda(1-\lambda)}\left[\frac{1-\lambda}{1+2\lambda(1-\lambda)} - \frac{1}{1+\lambda}\right]\right),$$

$$\sigma_{23}(\boldsymbol{p}) = \left(p_{XY}^{-1} - p_Y^{-1}\right)\lambda(1-\lambda)^2\left(\frac{1}{\lambda(1+1-\lambda)+1-\lambda+\lambda(1-\lambda)} + \frac{1}{1+\lambda(1-\lambda)}\left[\frac{\lambda}{1+2\lambda(1-\lambda)} - \frac{1}{1+1-\lambda}\right]\right),$$

and $\sigma_{12}(\boldsymbol{p}) = 0$*.*

Details are given in Section 3.5 in Chapter 3.

1.3 Complete data framework

In the complete data framework when $p_X = p_Y = p_{XY} = 1$, *i.e.* when $\mathbf{p} = \mathbf{1}$, the hybrid copula estimator becomes the empirical copula process. The limiting Gaussian process $\mathbb{C}_T$ of the normalized error of the empirical copula process is given by

$$N_C(u,v) = B_C(u,v) - \frac{\partial C}{\partial u}(u,v)B_C(u,1) - \frac{\partial C}{\partial u}(u,v)B_C(1,v), \tag{1.7}$$

where B_C is a Brownian bridge in $[0,1]^2$ with covariance function

$$\mathbb{E}[B_C(u,v)B_C(u',v')] = C(u\wedge u', v\wedge v') - C(u,v)C(u',v').$$

In this setting, the asymptotic behaviour of $\sqrt{T}(\hat{\nu}_T(\lambda) - \nu(\lambda))$ and $\sqrt{T}(\hat{\nu}_T^{\mathcal{H}*}(\lambda) - \nu(\lambda))$ is the same. We refer the reader to Section 3.6 for details. Proceeding as in the proof of Theorem 1, we can show the following statement.

Proposition 3 (**limit theorem with complete data**)**.** *Under Assumption A and complete data framework we have the weak convergence in* $\ell^\infty([0,1])$ *for the* λ*-FMadogram defined in Equation (7) , as* $T \to \infty$*,*

$$\sqrt{T}\left(\hat{\nu}_T(\lambda) - \nu(\lambda)\right) \rightsquigarrow \left(-\int_{[0,1]} N_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})du\right)_{\lambda\in[0,1]}.$$

Indeed, in this setup, we can show that almost surely

$$B_C(u,1) - \frac{\partial C}{\partial u}(u,v)B_C(u,1) = 1, \quad \forall u \in [0,1], \quad B_C(1,v) - \frac{\partial C}{\partial v}(u,v)B_C(1,v) = 1, \quad \forall v \in [0,1].$$

And we retrieve the asymptotic limit in law of the normalized estimation error of the λ-

FMadogram as studied in [Marcon et al., 2017]. Furthermore, as the integral of a tight Gaussian process, we know that for a fixed $\lambda \in [0,1]$, the asymptotic law is a Gaussian random variable.

Proposition 4 (asymptotic variance closed formula). *Let $\lambda \in [0,1]$. Under Assumptions A and complete data framework we have the convergence in distribution for the λ-FMadogram defined in Equation (7), as $T \to \infty$,*

$$\sqrt{T}\left(\hat{\nu}_T(\lambda) - \nu(\lambda)\right) \xrightarrow{d} \mathcal{N}(0, \sigma_3^2(\boldsymbol{1})),$$

where $\sigma_3^2(\boldsymbol{1})$ is defined in Equation (3.6) in Section 3.4.

For a fixed $\lambda \in (0,1)$, [Naveau et al., 2009] proved that the asymptotic law of the normalized estimation error of the λ-FMadogram can be written as

$$\int_{[0,1]^2} f(x,y)dJ(x,y) = \frac{1}{2}\int_{[0,1]} f(1, y^{1/(1-\lambda)})dy + \frac{1}{2}\int_{[0,1]} f(x^{1/\lambda}, 1)dx - \int_{[0,1]} f(x^{1/\lambda}, x^{1/(1-\lambda)})dx, \tag{1.8}$$

for every measurable and bounded function $f : [0,1]^2 \to \mathbb{R}$. Some details explaining Equation (1.8) are given in Section C.1 in Appendix C. The special case $J(x,y) = 2^{-1}|x^{\lambda} - y^{1-\lambda}|$ satisfies the conditions, then some computations gives that :

$$\int_{[0,1]^2} N_C(u,v)dJ(u,v) = -\int_{[0,1]} N_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})du.$$

We hence retrieve the result of Corollary 4 where we give a closed expression of the variance.

We are able to infer the closed form without integral of the Pickands of the λ-Madogram's variance in the case of an independent Copula, *i.e.* when $C(u,v) = uv$. This result is summarised in the following statement:

Corollary 2 (independence closed formula). *Under Assumption A and if $C(u,v) = uv$, then the asymptotic variance of $\sqrt{T}(\hat{\nu}_T(\lambda) - \nu(\lambda))$ has the following form, for $\lambda \in [0,1]$*

$$\sigma_3^2(\boldsymbol{1}) = \left(\frac{\lambda(1-\lambda)}{1+\lambda(1-\lambda)}\right)^2 \left(\frac{1}{1+2\lambda(1-\lambda)} - \frac{1-\lambda}{2-(1-\lambda)+2\lambda(1-\lambda)} - \frac{\lambda}{2-\lambda+2\lambda(1-\lambda)}\right).$$

1.4 FMadogram with outliers and complete data

In order to propose a robust estimator we will assume that the sample is partitioned into K disjoint subsets $B_1, \dots, B_K$ of cardinalities $n_j := card(B_j)$ respectively, where the partitioning scheme is independent of the data. Let f be a measurable function from $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ to $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$, we define the following estimator of $\mathbb{E}[f(X,Y)]$ by

$$\mathbb{P}_{n_j} f = \frac{1}{n_j}\sum_{j \in B_j} f(X_j, Y_j).$$

We define the MoN estimator of f as solutions of the optimization problem

$$\hat{f}_{MoN} = \underset{z \in \mathbb{R}}{argmin} \sum_{j=1}^{K} \left| \mathbb{P}_{n_j} f - z \right|, \tag{1.9}$$

which, if we note $med(\cdot)$ the usual univariate median

$$\hat{f}_{MoN} = med(\mathbb{P}_{n_1} f, \dots, \mathbb{P}_{n_K} f), \tag{1.10}$$

is a solution of Equation (1.9). We are restricting our analysis to the FMadogram in Equation (6) to avoid technical difficulties, but the proof will be similar with a discussion according to the value of the bound and λ using that $||x|^\lambda - |y|^\lambda| \leq |x-y|^\lambda$.

Intuitively, we replace the linear operator of expectation with the median of averages taken over non-overlapping blocks of the data, in order to get a robust estimate thanks to the median step (see [Lerasle et al., 2019] for a similar idea applied to Kernel). The MoN is one of the mean estimators that achieve a sub-Gaussian behavior under mild conditions. Introduced during the 1980s [Nemirovsky and Yudin, 1983] for the estimation of the mean of real-valued random variables, that is easy to compute, while exhibiting attractive robustness properties by the median step.

Definition 3. *Let $B_1, \dots, B_K$ a partition of the set $\{1, \dots, T\}$. Denote by $\hat{F}_{n_j}$ (resp. $\hat{G}_{n_j}$) the empirical cumulative distribution for the cumulative distribution of X (resp. Y) computed within block B_j. The MoN-based FMadogram estimator is defined by*

$$\hat{\nu}_{MoN} = \underset{z \in \mathbb{R}}{argmin} \sum_{j=1}^{K} \left| \hat{\nu}_{n_j} - z \right|, \tag{1.11}$$

where $\hat{\nu}_{n_j} = \frac{1}{2n_j} \sum_{t \in B_j} |\hat{F}_{n_j}(X_t) - \hat{G}_{n_j}(Y_t)|$.

That is, in Equation (1.9), we take $f(x,y) = |x-y|$ and $\mathbb{P}_{n_j} = C_{n_j}$ the empirical copula constructed on the block B_j.

Assumption 1. *The sample $((X_1, Y_1), \dots, (X_T, Y_t))$ contains $T - T_o$ outliers drawn according to distribution H, and T_o outliers, upon which no assumption is made.*

In the presence of outliers, the key point is to focus on sane blocks, $i.e$ on blocks that does not contain a single outlier, since no inference can be made about blocks hit by an outlier. One way to ensure that sane blocks to be in majority is to consider twice more blocks than outliers. Indeed, in the worst case scenario, each outlier contaminates one block, but the sane blocks remain more numerous. Let K_s denote the total number of sane block containing no outliers. In other words, there exists $\delta \in (0, 1/2]$ such that $K_s \geq K(1/2 + \delta)$. If the data are free from contaminations, then $K_s = K$ and $\delta = 1/2$.

We present also a concentration inequality that the MoN-based estimator of the FMadogram may verified. We suppose without loss of generality that $n_j = \lceil T/K \rceil$ for every $j \in \{1, \dots, K\}$. We can prove the following deviation bounds for our MoN-based estimator.

Theorem 2 (Consistency and outlier-robustness). *Under Assumption 1, for any $\eta \in]0,1[$ such that $K = \delta^{-1} log(1/\eta)$ it holds that with probability $1-\eta$,*

$$|\hat{\nu}_{MoN} - \nu| \leq \frac{3}{\sqrt{2}} \frac{log\left(6e2^{\frac{1}{\delta}}\right)}{\delta} \sqrt{\frac{log\,(1/\eta)}{T}}.$$

Details of the proof are available in Section 3.7 in Chapter 3. We thus add some remark on the concentration bound.

Remark 2.

- *Dependence on T : These finite-sample guarantees show that the MoN-based estimator is robust to outliers, providing consistent estimates with high probability even under arbitrary contamination (affecting less than half of the samples).*
- *Dependence on δ : Recall that higher δ corresponds to less outliers in which case the bounds above become tighter.*
- *Dependence on η : A lower η gives a greater bound for which the estimator hold with a greater probability.*

Chapter 2

Numerical results

2.1 Considered parametric models

We present several models that will be used in the simulation section in order to assess our findings remains in finite-sample settings.

1. The asymmetric logistic model [Tawn, 1988] defined by the following dependence function:

$$A(t) = (1-\psi_1)t + (1-\psi_2)(1-t) + [(\psi_1 t)^\theta + (\psi_2(1-t))^\theta]^{\frac{1}{\theta}},$$

   with parameters $\theta \in [1, \infty[$, ψ_1, $\psi_2 \in [0,1]$. The special case $\psi_1 = \psi_2 = 1$ gives us the symmetric model of Gumbel. In the symmetric model, we retrieve the independent case when $\theta = 1$, the dependence between the two variables is stronger as θ goes to infinity.

2. The asymmetric negative logistic model [Joe, 1990], namely,

$$A(t) = 1 - [(\psi_1(1-t))^{-\theta} + (\psi_2 t)^{-\theta}]^{-\frac{1}{\theta}},$$

   with parameters $\theta \in (0,\infty)$, $\psi_1, \psi_2 \in (0,1]$. The special case $\psi_1 = \psi_2 = 1$ returns the symmetric negative logistic of [Oliveira and Galambos, 1977].

3. The asymmetric mixel model [Tawn, 1988] :

$$A(t) = 1 - (\theta + \kappa)t + \theta t^2 + \kappa t^3,$$

   with parameters θ and κ satisfying $\theta \geq 0$, $\theta + 3\kappa \geq 0$, $\theta + \kappa \leq 1$, $\theta + 2\kappa \leq 1$. The special case $\kappa = 0$ and $\theta \in [0,1]$ yields the symmetric mixed model. In the symmetric mixed model, when $\theta = 0$, we go back to the independent copula.

4. The model of Hüsler and Reiss [Hüsler and Reiss, 1989],

$$A(t) = (1-t)\Phi\left(\theta + \frac{1}{2\theta} log\left(\frac{1-t}{t}\right)\right) + t\Phi\left(\theta + \frac{1}{2\theta} log\left(\frac{t}{1-t}\right)\right),$$

   where $\theta \in (0,\infty)$ and Φ is the standard normal distribution function. As θ goes to 0^+, the dependence between the two variables is stronger. When θ goes to infinity, we are near independence.

5. The t-EV model [Demarta and McNeil, 2005], in which

$$A(w) = wt_{\nu+1}(z_w) + (1-w)t_{\nu+1}(z_{1-w}),$$
$$\text{with } z_w = (1+\nu)^{1/2}[w/(1-w)^{\frac{1}{\nu}} - \theta](1-\theta^2)^{-1/2},$$

and parameters $\nu > 0$, and $\theta \in (-1,1)$, where $t_{\nu+1}$ is the distribution function of a Student-t random variable with $\nu+1$ degrees of freedom.

2.2 Complete data framework

2.2.1 Extreme value copulas

A vast Monte Carlo study is used here to illustrate Theorem 1 of Chapter 1 in finite-sample settings while none data are missing. In Figure 1, for each $\lambda \in [0,1]$, 500 random samples of size $T = 256$ were generated from the Gumbel copula (see Model 1 in Section 2.1) with $\theta \in \{1.0, 1.5, 2.0, 2.5, 3.0\}$. For each sample, the λ-FMadogram estimators were computed where the margins are unknown. For each estimator we estimate the empirical version of the normalized estimation error variance, namely

$$\mathcal{E}_T(\lambda) := \widehat{Var}\left(\sqrt{T}\left(\hat{\nu}_T^*(\lambda) - \nu(\lambda)\right)\right), \tag{2.1}$$

over 500 samples. Then we compare $\mathcal{E}_T(\lambda)$ with its associated theoretical asymptotic variance using the form exhibits in Corollary 4. Similar results were obtained for many other extreme-value dependence models (see Figure 2). We can notice the following :

1. When A is symmetric, one would expect the asymptotic variance of the estimator to reach its maximum at $\lambda = 1/2$. Such is not always the case, however, as illustrated by the t-EV model (see Figure 2 Panel 2f).
2. In the asymmetric negative logistic model (see Model 2 in Section 2.1), the asymptotic of the λ-FMadogram is close to zero for all $\lambda \in [0, 0.3]$. This is due to the fact that $A(\lambda) \approx 1-t$ for this model.

Remarks 1 and 2 are also observed in [Genest and Segers, 2009]. We propose in Figure 3 the theoretical asymptotic variance depending of θ and λ for six model taken in Section 2.1. The parameters of Figure 2 and Figure 3 are chosen accordingly to [Genest and Segers, 2009].

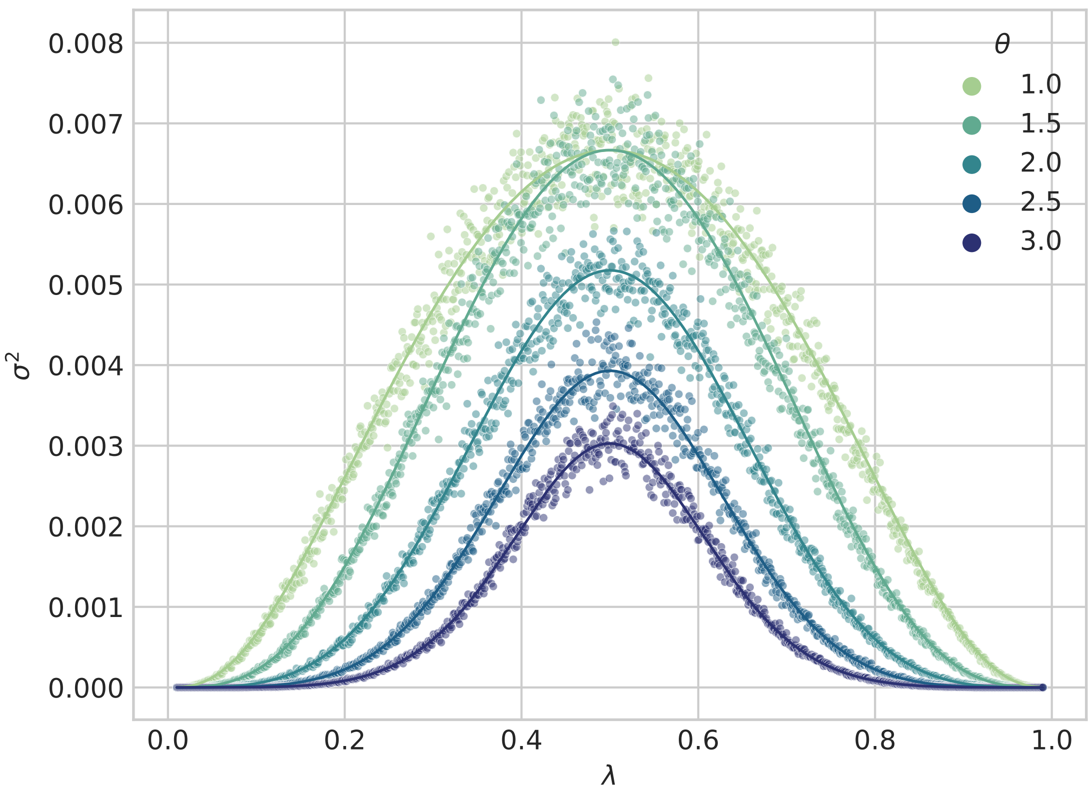
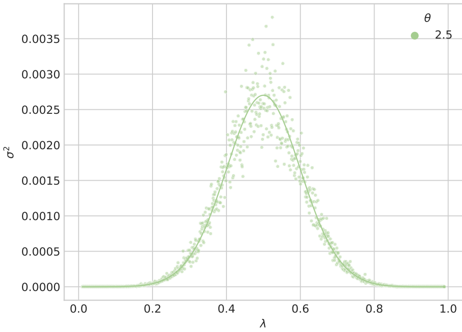
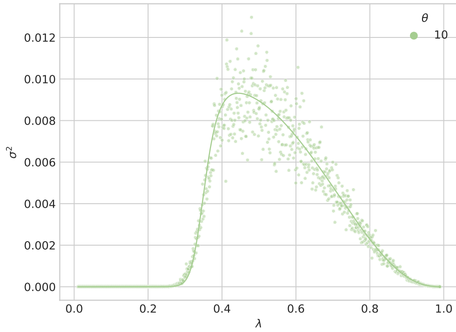


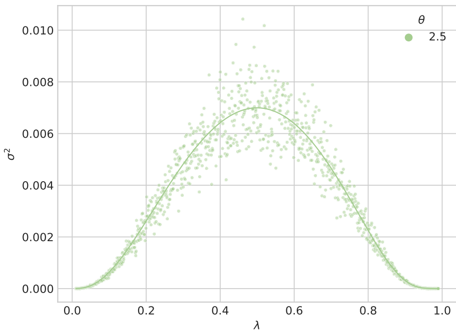
Figure 1: $\mathcal{E}_T(\lambda)$ in Equation (2.1), as a function of λ, based on 500 samples of size $T = 256$ from the Gumbel copula with $\theta = \{1.0, 1.5, 2.0, 2.5, 3.0\}$ chosen in such a way that $\lambda \in \{i/1000 : i = 10, \ldots, 990\}$. Solid lines are obtained using Proposition 4.



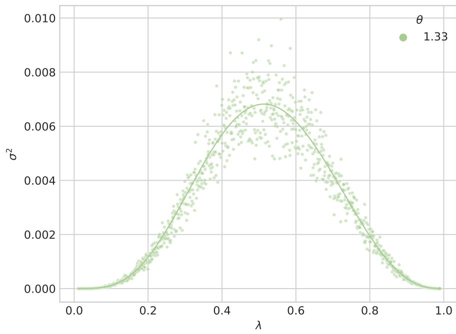
(a) **Galambos** ($\theta = 2.5$)



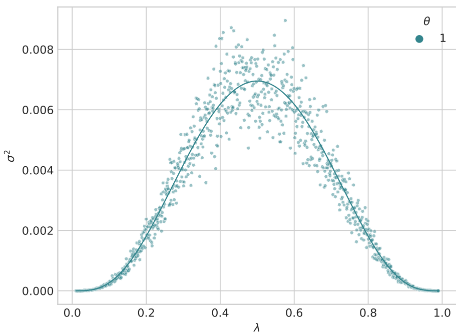
(b) **Asy. Neg. Log.** ($\theta = 10$, $\psi_1 = 0.5$, $\psi_2 = 1.0$)



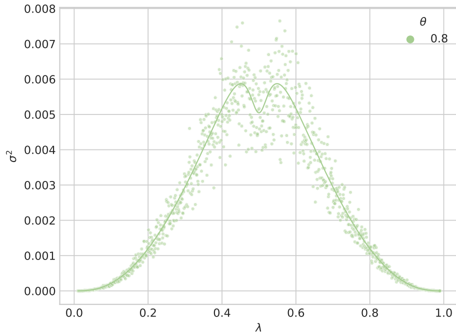
(c) **Asy. Log.** ($\theta = \frac{5}{2}$, $\psi_1 = 0.1$, $\psi_2 = 1.0$)



(d) **Asy. Mixed** ($\theta = \frac{4}{3}$, $\kappa = -\frac{1}{3}$)

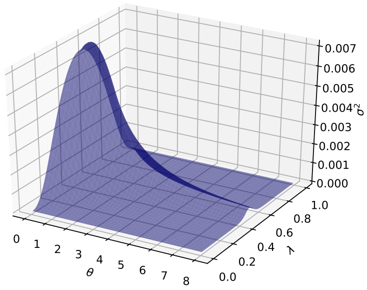


(e) **Hüsler-Reiss** ($\theta = 1$)

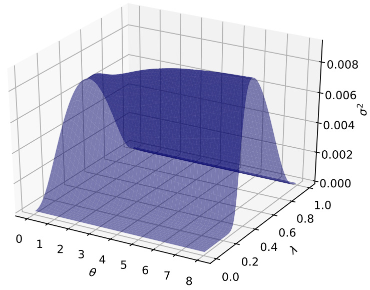


(f) **t-EV** ($\theta = 0.8$, $\nu = 0.2$)

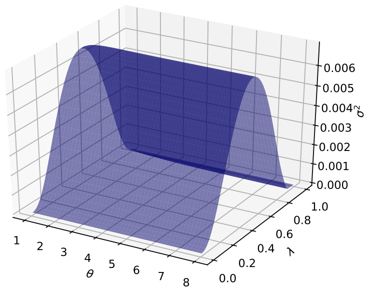
Figure 2: $\mathcal{E}_T(\lambda)$ in (2.1), as a function of λ, based on 500 samples of size $T = 256$ of the λ-FMadogram. Solid lines are obtained using Proposition 4.



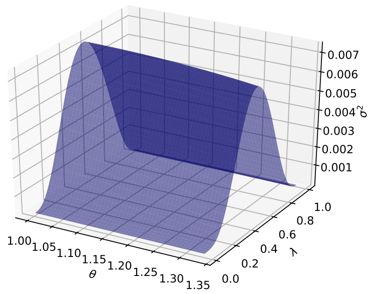
(a) **Galambos**



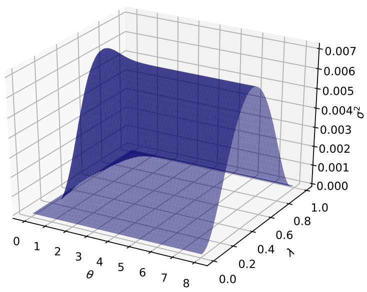
(b) **Asy. Neg. Log. ($\psi_1 = 0.5$, $\psi_2 = 1.0$)**



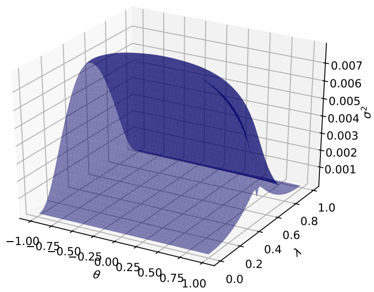
(c) **Asy. log. ($\psi_1 = 0.1, \psi_2 = 1.0$)**



(d) **Asy. Mixed ($\kappa = -\frac{1}{3}$)**



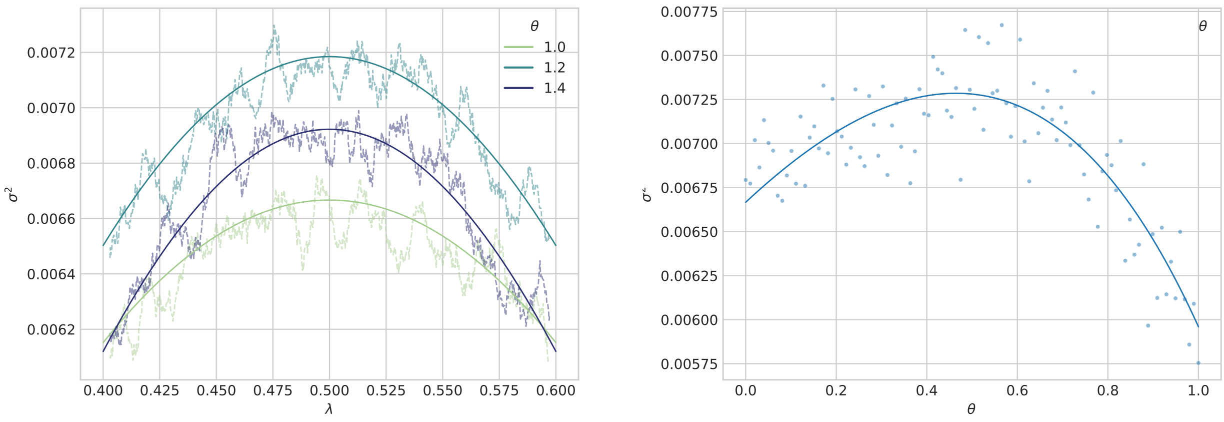
(e) **Hüsler-Reiss**



(f) **t-EV ($\nu = 0.2$)**

Figure 3: Theoretical value of $\mathcal{E}_T(\lambda)$ using Proposition 4 depending on λ and the parameter θ of the chosen Pickands dependence function.

2.2.2 Non-monotonicity of the variance with respect to the dependence parameter



(a) (b)

Figure 4: Panel (a) depicts $\mathcal{E}_T(\lambda)$ based on 500 samples of size $T = 256$ from the Gumbel copula with $\theta = \{1.0, 1.2, 1.4\}$ chosen in such a way that $\lambda \in \{i/1000 : i = 400, \dots, 600\}$. The dotted lines are moving averages made of the 1000 empirical estimates of the variance. Panel (b) shows $\mathcal{E}_T(\lambda)$ based on 2000 sample of size $T = 512$ from the symmetric mixed model with $\lambda = 0.5$ chosen in such a way that $\theta \in \{i/100 : i = 0, \dots 100\}$. Solid lines are the theoretical asymptotic variance computed using Proposition 4.

Looking at Figure 1, one may make the third remark :

3. There is no strict dominance between the asymptotic variance of the normalized estimator and the parametric extreme value copula. In other words, let A be a Pickands dependence function then we have the following

$$1 \geq A(t), \quad \forall t \in [0,1] \not\Longrightarrow \sigma^2_{3_\pi} \geq \sigma^2_{3_A}, \quad \forall \lambda \in [0,1], \tag{2.2}$$

with $\sigma^2_{3_\pi}$ (resp. $\sigma^2_{3_A}$) denotes the theoretical asymptotic variance of the normalized estimation error when $A(t) = 1$ (resp. for $1 \geq A(t)$), $\forall t \in [0,1]$.

Figure 4a shows the same model with different values of θ and with a reduced scale for λ. The moving average is computed for 10 empirical variances for each θ. As the dependency parameter θ increases, we can find some λ for which the asymptotic variance is greater than the asymptotic variance in the case of independence. That figure supports our counterexample (see Section C.2 in Appendix C for details) that draws the same conclusion. Also, Figure 4b depicts the asymptotic variance for a fixed $\lambda = 0.5$ for the symmetric mixed model (see Model 3 in Section 2.1) with $\theta \in [0,1]$. When $\theta = 0$, we are turning back to the independent copula $C(u,v) = uv$ and its asymptotic variance is given by $1/150$ for this value of λ. When the random variables X and Y are becoming positively dependent, *i.e.* when θ increase in this model, the asymptotic variance for this given λ increase also, but after a certain threshold which depends on the chosen model, the variance starts to decrease.

2.2.3 Max-Stable processes

To determine the quality of the λ-FMadogram for estimating the pairwise dependence of maxima in space, [Naveau et al., 2009] computes on a particular class of simulated max-stable random fields. They focus on the Smith's max-stable process ([Smith, 2005]). We recall the bivariate distribution for the max-stable process model proposed by Smith is equal to :

$$\mathbb{P}\left(X(s) \leq u, X(s+h) \leq v\right) = exp\left[-\frac{1}{u}\Phi\left(\frac{a}{2}+\frac{1}{a}log(\frac{v}{u})\right) - \frac{1}{v}\Phi\left(\frac{a}{2}+\frac{1}{a}log(\frac{u}{v})\right)\right], \quad (2.3)$$

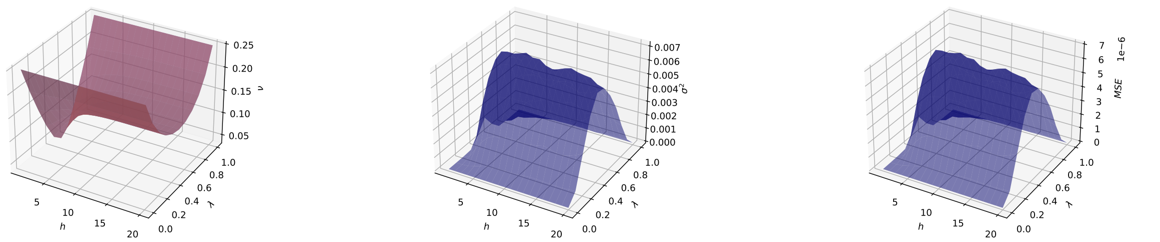
where Φ denotes the standard normal distribution function, with $a^2 = (h^\top \Sigma^{-1} h)$ and Σ is a covariance matrix. In case of isotropic field, we set $\Sigma = \sigma^2 I_2$. For this kind of process, the pairwise extremal dependence function $V(\cdot,\cdot)$ (see section 4.3 of [Coles et al., 1999] for a definition) is given by :

$$V(u,v) = \frac{1}{u}\Phi\left(\frac{a}{2}+\frac{1}{a}log(\frac{v}{u})\right) + \frac{1}{v}\Phi\left(\frac{a}{2}+\frac{1}{a}log(\frac{u}{v})\right). \quad (2.4)$$

Furthermore, for a max-stable process, the theoretical value of the λ-FMadogram is given by

$$\nu(h,\lambda) = \frac{V(\lambda, 1-\lambda)}{1+V(\lambda,1-\lambda)} - \frac{3}{2(1+\lambda)(1+1-\lambda)}, \quad (2.5)$$

for any $\lambda \in [0,1]$ (see Proposition 1 of [Naveau et al., 2009]). The λ-FMadogram was estimated independently for each simulated field with $T = 1024$. The xy-space $[0,20] \times [0,1]$, represent the distance h and parameter λ. In Smith's model, the pairwise dependence function between two locations s and $s+h$ decrease as the distance h between these two points increases. The



(a) **Estimated value** (b) **variance** (c) **Mean Squared Error**

Figure 5: Simulation results obtained by generating 300 independently and identically distributed Smith random fields. The dependence structure is characterized by (2.3) with $\Sigma = 25 I_2$. Panel (5a) shows the estimated and the true λ-FMadogram. Panel (5b) the $\mathcal{E}_T(\lambda)$ in Equation (2.1). Panel (5c) depicts the mean squared error between the true and estimated λ-FMadogram for all h and λ.

surface in Figure 5a provides the mean value of the estimated λ-FMadogram in blue, the true quantity is given by the surface in red. Figure 5c indicates the mean squared error between the estimated λ-FMadogram and the true value. As expected, the error is close to zero at the two boundary planes $\lambda = 0$ and $\lambda = 1$ by construction of the estimator. The largest mean squared errors are obtained where $\lambda = 0.5$, especially for very small distances, *i.e.* near $h = 0$. This behaviour is now well known from Section 2.2.2.

2.2.4 Block maxima model

In this Section, we derive the behavior of the asymptotic variance of componentwise maxima of i.i.d random vectors having a t copula distribution. A bivariate t copula is defined as :

$$C_{\nu,\theta}(u,v) = \int_{]-\infty, t_\nu^{\leftarrow}(u)]} \int_{]-\infty, t_\nu^{\leftarrow}(v)]} \frac{1}{2\pi(1-\theta^2)^{1/2}} \left(1 + \frac{x^2 - 2\theta xy + y^2}{\nu(1-\rho^2)} \right)^{-(\nu+2)/2} dydx, \qquad (2.6)$$

where $\nu > 0$ is the number of degrees of freedom, $\theta \in [-1, 1]$ is the linear correlation coefficient, t_ν is the distribution function of a t-distribution with ν degrees of freedom. According to [Demarta and McNeil, 2005] the bivariate t copula $C_{\nu,\theta}$ is attracted to the t extreme value copula. Hence, we simulate $X_{1j}, \dots, X_{Mj}$, $j \in \{1, \dots, T\}$, a block of M variables from a t copula and we take the maximum in this block. This step is repeated several times in order to form a sample $(\bigvee_{i=1}^M X_{i1}, \dots, \bigvee_{i=1}^M X_{iT})$ of length T. The result depicts on Figure 6 is what we

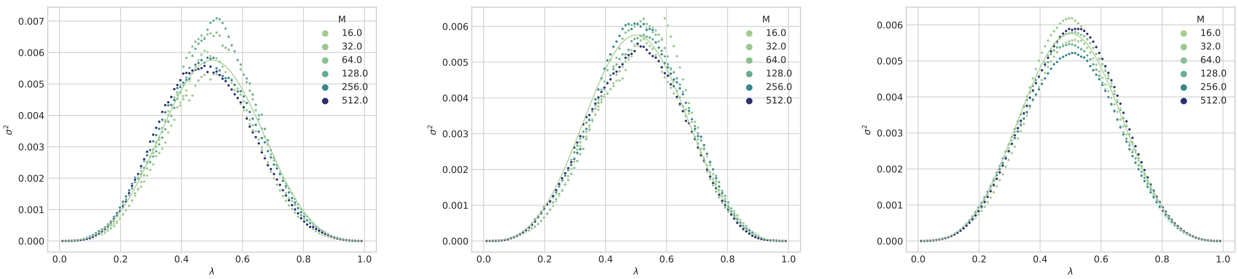


Figure 6: Simulation results obtained by generating $T \in \{128, 256, 512\}$ blocks maxima's of length $M \in \{16, 32, 64, 128, 256, 512\}$ from t-copula with parameters $\theta = 0.8$ and $\nu = 3$. Each T is associated, in increasing order, to the left, middle and right panel. For each $\lambda \in \{i/100, i = 1, \dots, 99\}$, we estimate $\mathcal{E}_T(\lambda)$ in Equation (2.1) on 100 estimator of λ-FMadogram. The solid line is the theoretical asymptotic variance of t-EV copula with $\theta = 0.8$ and $\nu = 3$ given by Proposition 4.

waited for. As we expected, when the number of observations in block maxima, the empirical variance converge towards the asymptotic variance for a t-EV copula. Furthermore, as the sample's length increases, more the empirical variance fits the theoretical one.

2.3 Missing data framework

In each section, we estimate the empirical variance on several Monte Carlo simulations of the normalized estimation error for both hybrid estimator. These errors are define respectively for the hybrid and the corrected estimators by

$$\mathcal{E}_T^{\mathcal{H}}(\lambda) := \widehat{Var}\left(\sqrt{T} \left(\hat{\nu}_T^{\mathcal{H}}(\lambda) - \nu(\lambda) \right) \right), \qquad (2.7)$$

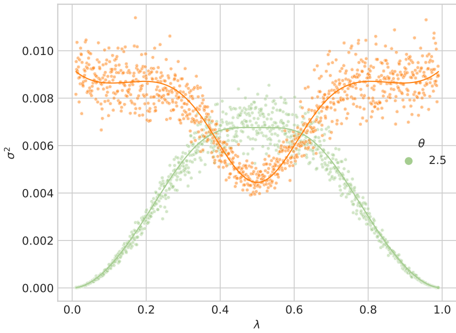
$$\mathcal{E}_T^{\mathcal{H}*}(\lambda) := \widehat{Var}\left(\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}*}(\lambda) - \nu(\lambda)\right)\right). \tag{2.8}$$

For $p_X \in]0,1]$ and $p_Y \in]0,1]$ the indicator of missing of variables X and Y is generated according to a Bernoulli distribution

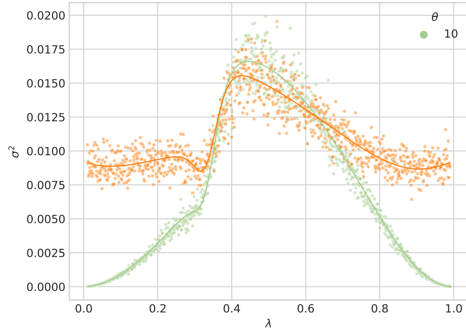
$$I \sim \mathcal{B}(p_X), \quad J \sim \mathcal{B}(p_Y).$$

We also set that $p_{XY} = p_X p_Y$, *i.e.* I and J are independent.

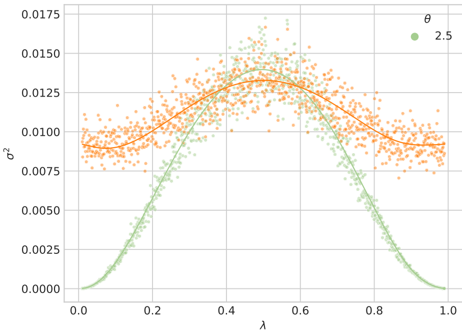
Figure 7 presents the results for the six models defined in Chapter 1. For each model, the green (*resp.* red) line presents the asymptotic variance $\mathcal{E}_T^{\mathcal{H}}(\lambda)$ (*resp.* $\mathcal{E}_T^{\mathcal{H}*}(\lambda)$) for $\hat{\nu}_T^{\mathcal{H}}$ (*resp.* $\hat{\nu}_T^{\mathcal{H}*}$) given by Corollary 2. For each $\lambda \in \{\frac{i}{1000}, i = \frac{10}{1000}, \ldots \frac{990}{1000}\}$, we estimate its empirical counterpart. Here, we take $p_X = p_Y = 0.75$. As waited, we directly see that both empirical and theoretical values of the variance of the normalized error of $\hat{\nu}_T^{\mathcal{H}}$ is different from zero for each extremity of λ. Furthermore, in some models, we also lose the "parabolic" shape of the curve (see Figure (7a)). The introduction of the corrected estimator may us to recover the same pattern as noticed in Figure 2 of Section 2.2.1. Notice that, in terms of variance, we do not have a strict dominance from one estimator to another as it was mentioned before.



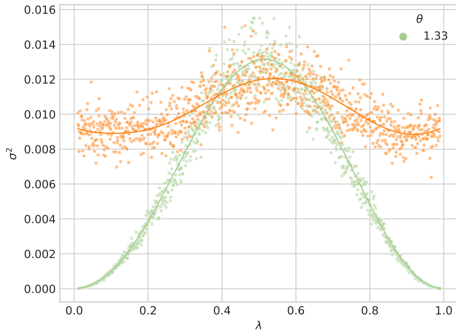
(a) **Galambos** **(**$\theta = 2.5$**)**



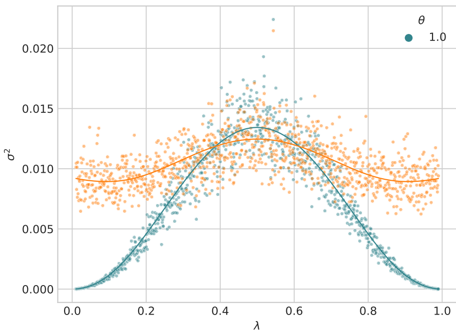
(b) **Asym. Neg. Log.** **(**$\theta = 10$**,** $\psi_1 = 0.5$**,** $\psi_2 = 1.0$**)**



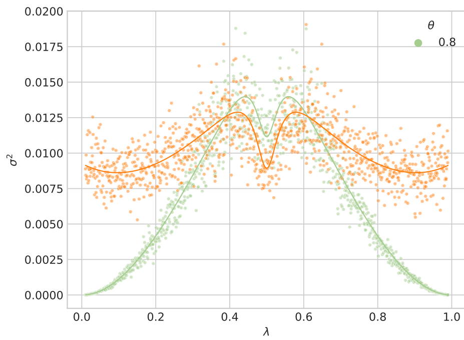
(c) **Asy. Log.** **(**$\theta = \frac{5}{2}$**,** $\psi_1 = 0.1$**,** $\psi_2 = 1.0$**)**



(d) **Asy. Mixed** **(**$\theta = \frac{4}{3}$**,** $\kappa = -\frac{1}{3}$**)**



(e) **Husler-Reiss** **(**$\theta = 1.0$**)**



(f) **t-EV** **(**$\theta = 0.8$**,** $\nu = 0.2$**)**

Figure 7: Equation (2.7) (in red) and Equation (2.8) (in green), as a function of λ, of the asymptotic variances of the estimators of the λ-FMadogram for six extreme-value copula models. The empirical variances are based on 500 samples of size $T = 256$. Solides lines are the theoretical value given by Proposition 2.

2.4 Robustness and outliers with complete data

We now compare our MoN-based estimator in Equation (1.11) defined in Section 1.4 to the FMadogram. We designed two types of outliers :

- *top-left*: the outliers are drawn $i.i.d$ from a uniform distribution $\mathcal{U}([0, 0.05] \times [0.95, 1])$.
- *bottom-right*: the outliers are drawn $i.i.d$ from a uniform distribution $\mathcal{U}([0.95, 1] \times [0, 0.05])$.

In each case, sane data are sampled from the desired copula model. Then all data are inverted by the quantile function of a standard Gaussian distribution. Both types of outliers are considered in the two frameworks of contamination, *i.e. à la* Huber and adversarial one. In the adversarial contamination, our rule is to sample again points which are closer to the point $(0.5, 0.5)$ for the Gumbel model and the Asymmetric Logistic (see Model 1). For the Negative Asymmetric Logistic (see 2), we sample again points which are closer to $(1.0, 1.0)$.

Figure 8 present an illustration of contaminated data for three selected copula models for all types of contaminations and outliers we consider. We report the squared bias of each estimator for all the models considered in Figure 9 for three extreme value copula, two types of outliers and contaminations. When there are no outliers, the MoN-based estimator and the FMadogram has the same bias, this is due that $\hat{\nu}_{MoN} = \hat{\nu}_T$ when $K = 1$.

Otherwise, our MoN-based estimator is not always better than the FMadogram. This is the case when the Pickands dependence function is symmetric at $\lambda = 0.5$. When it is not, we can see that our MoN-based estimator become much more reliable when the Pickands is asymmetric. Indeed, when the asymmetry of the dependence model induces a concentration of the points on the bottom right, we see that the MoN-based estimator is better than the FMadogram for the *bottom-right* types of outliers. That is shown in Figure 9 for the asymmetric Negative Logistic Model for the second and the fourth rows.

The same observation is drawn in the Asymmetric Logistic Model which performs comparatively for the Huber's contamination and better for the adversarial one. For these asymmetric models, the MoN-based estimator is more robust than the FMadogram because the *bottom-right* types of outliers (resp. *top-left*) upsets ranks of sane data for the asymmetric negative logistic model (resp. for the asymmetric logistic model).

(a) **Gumbel** (b) **Asy. Log.** (c) **Asy. Neg. Log.**

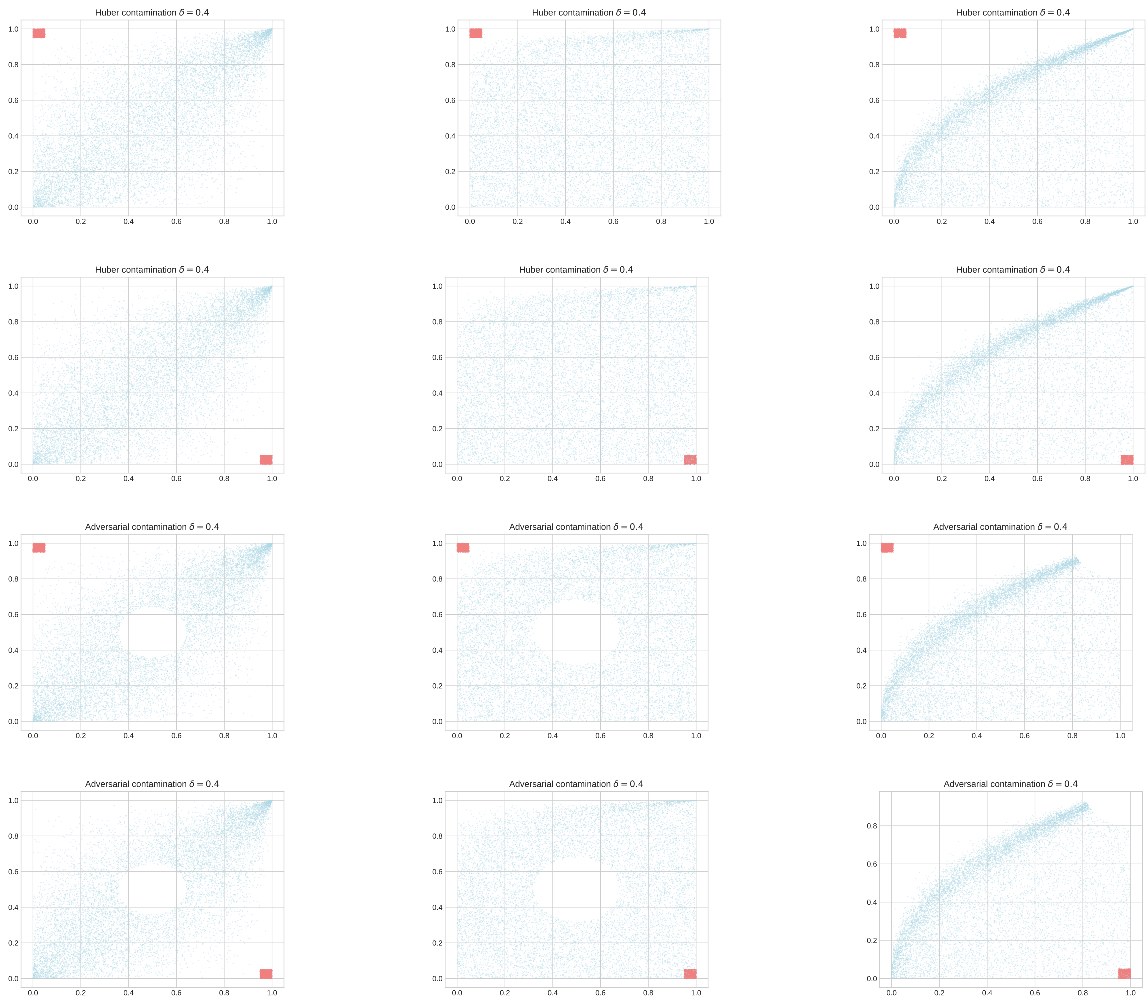


Figure 8: Sample of 10000 data with a fraction of 10% of outliers. Sane points are depicted in blue while contaminated ones are in red. The two first rows are respectively the *top-left* and *bottom-right* types of outliers for the Huber's contamination model while the two next are for the adversarial contamination. For Gumbel's model, we took $\theta = 1.5$, for the asymmetric logistic model, we consider $\theta = 2.5$, $\psi_1 = 0.1$, $\psi_2 = 1.0$ while the asymmetric negative logistic model is defined with $\theta = 10$, $\psi_1 = 0.5$, $\psi_2 = 1.0$.

(a) **Gumbel** (b) **Asy. Log.** (c) **Asy. Neg. Log.**

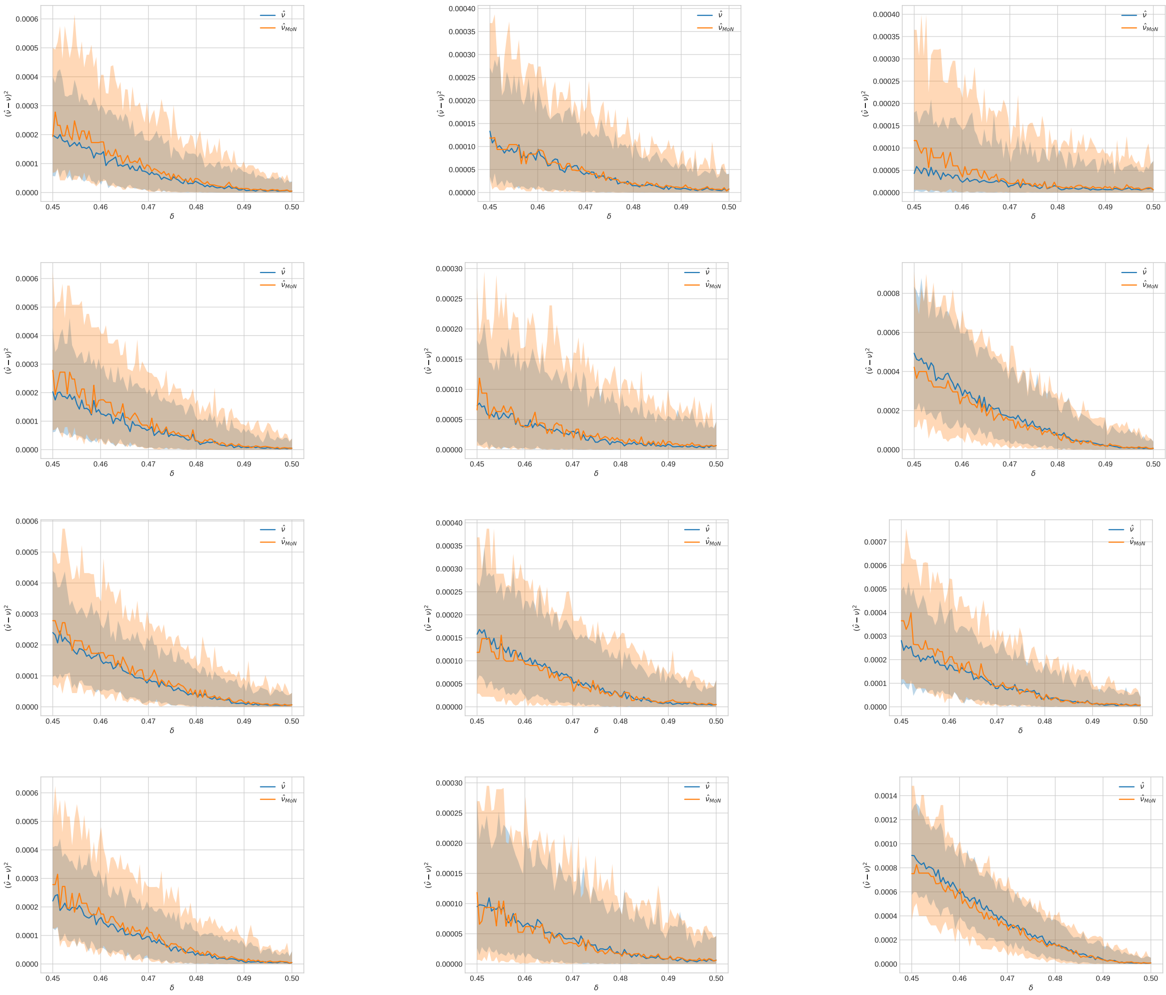


Figure 9: Squared bias of the MoN-based estimator (orange) and the FMadogram (blue) and the 90% confidence band computed on 100 estimators build on a sample of length $T = 1000$. The two first row is respectively the *top-left* and *bottom-right* types of outliers for the Huber's contamination model while the two next are for the adversarial contamination. For Gumbel's model, we took $\theta = 1.5$, for the asymmetric logistic model, we consider $\theta = 2.5$, $\psi_1 = 0.1$, $\psi_2 = 1.0$ while the asymmetric negative logistic model is defined with $\theta = 10$, $\psi_1 = 0.5$, $\psi_2 = 1.0$.

Chapter 3

Proofs

3.1 Proof of Proposition 1

The estimator $\hat{\nu}(\lambda)$ is strongly consistent since it holds

$$\left| \frac{1}{2\sum_{t=1}^{T} I_t J_t} \sum_{t=1}^{T} \left| \hat{F}_T^{\lambda}(X_t) - \hat{G}_T^{1-\lambda}(Y_t) \right| I_t J_t - \frac{1}{2}\mathbb{E} \left| F^{\lambda}(X) - G^{1-\lambda}(Y) \right| \right|$$
$$\leq \left| \frac{1}{2\sum_{t=1}^{T} I_t J_t} \sum_{t=1}^{T} \left| \hat{F}_T^{\lambda}(X_t) - \hat{G}_T^{1-\lambda}(Y_t) \right| I_t J_t - \frac{1}{2\sum_{t=1}^{T} I_t J_t} \sum_{t=1}^{T} \left| F^{\lambda}(X_t) - G^{1-\lambda}(Y_t) \right| I_t J_t \right|$$
$$+ \left| \frac{1}{2\sum_{t=1}^{T} I_t J_t} \sum_{t=1}^{T} \left| F^{\lambda}(X_t) - G^{1-\lambda}(Y_t) \right| I_t J_t - \frac{1}{2}\mathbb{E} \left| F^{\lambda}(X) - G^{1-\lambda}(Y) \right| \right|.$$

The second term converges almost surely to zero by the strong Law of Large Numbers and Assumption B. For the first term, we have

$$\left| \frac{1}{2\sum_{t=1}^{T} I_t J_t} \sum_{t=1}^{T} \left| \hat{F}_T^{\lambda}(X_t) - \hat{G}_T^{1-\lambda}(Y_t) \right| I_t J_t - \frac{1}{2\sum_{t=1}^{T} I_t J_t} \sum_{t=1}^{T} \left| F^{\lambda}(X_t) - G^{1-\lambda}(Y_t) \right| I_t J_t \right|$$
$$\leq \frac{1}{2\sum_{t=1}^{T} I_t J_t} \sum_{t=1}^{T} \left| |\hat{F}_T^{\lambda}(X_t) - \hat{G}_T^{1-\lambda}(Y_t)| - \left| F^{\lambda}(X_t) - G^{1-\lambda}(Y_t) \right| \right| I_t J_t,$$
$$\leq \frac{1}{2\sum_{t=1}^{T} I_t J_t} \sum_{t=1}^{T} \left| \hat{F}_T^{\lambda}(X_t) - F^{\lambda}(X_t) - \left(\hat{G}_T^{1-\lambda}(Y_t) - G^{1-\lambda}(Y_t) \right) \right| I_t J_t,$$
$$\leq \frac{1}{2\sum_{t=1}^{T} I_t J_t} \sum_{t=1}^{T} \left| \hat{F}_T^{\lambda}(X_t) - F^{\lambda}(X_t) \right| I_t J_t + \left| \hat{G}_T^{1-\lambda}(Y_t) - G^{1-\lambda}(Y_t) \right| I_t J_t.$$
$$\leq \frac{1}{2} \sup_{t\in\{1,\ldots,T\}} \left| \hat{F}_T^{\lambda}(X_t) - F^{\lambda}(X_t) \right| + \frac{1}{2} \sup_{t\in\{1,\ldots,T\}} \left| \hat{G}_T^{1-\lambda}(Y_t) - G^{1-\lambda}(Y_t) \right|$$

which converges almost surely to zero, according to the Glivencko-Cantelli theorem.

□

3.2 Proof of Lemma 1

Consider the following functions from $\{0,1\}^2 \times \mathbb{R}^2$ into $\mathbb{R}$: for $(x, y) \in \mathbb{R}^2$,

$$\begin{aligned} f_1(I,J,X,Y) &= \mathbb{1}_{\{I=1\}}, \quad g_{1,x} = \mathbb{1}_{\{X\leq x, I=1\}}, \\ f_2(I,J,X,Y) &= \mathbb{1}_{\{I=1\}}, \quad g_{2,x} = \mathbb{1}_{\{X\leq x, I=1\}}, \\ f_3 &= f_1 f_2, \quad g_{3,x,y} = g_{1,x} g_{2,y}. \end{aligned}$$

Let P denote the common distribution of the quadruples (I, J, X, Y). Consider the collection of functions

$$\mathcal{F} = \{f_1, f_2, f_3\} \cup \{g_{1,x} : x \in \mathbb{R}\} \cup \{g_{2,y} : y \in \mathbb{R}\} \cup \{g_{3,x,y} : (x, y) \in \mathbb{R}^2\}.$$

The empirical process $\mathbb{G}_T$ defined by

$$\mathbb{G}_T(f) = \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} f(I_t, J_t, X_t, Y_t) - \mathbb{E}[f(I_t, J_t, X_t, Y_t)]\right), \quad f \in \mathcal{F},$$

converge in $\ell^\infty(\mathcal{F})$ to a P-Brownian bridge $\mathbb{G}$ (see [Segers, 2014]). To establish such a statement, results on empirical processes based on the theory of Vapnik-Cervonenkis classes (VC-classes) of functions as formulated in [van der Vaart and Wellner, 1996] were used. We now add some lines of algebra to establish the weak convergence of the processes $\hat{F}_T(x)$, $\hat{G}_T(y)$ and $\hat{H}_T(x, y)$. These lines are made in the first process as the method is similar to the others. For $x \in \mathbb{R}$, we may write

$$\hat{F}_T(x) = \frac{p_X F(x) + T^{-1/2}\mathbb{G}_T g_{1,x}}{p_X + T^{-1/2}\mathbb{G}_T f_1}.$$

We may obtain :

$$\begin{aligned} p_X(\hat{F}_T(x) - F(x)) &= T^{-1/2}(\mathbb{G}_T(g_{1,x}) - \mathbb{G}_T(f_1)\hat{F}_T(x)), \\ &= T^{-1/2}(\mathbb{G}_T(g_{1,x} - F(x)f_1)) + T^{-1/2}\mathbb{G}_T(f_1)(F(x) - \hat{F}_T(x)). \end{aligned}$$

Multiplying by $\sqrt{T}$ and dividing by p_X gives :

$$\sqrt{T}(\hat{F}_T(x) - F(x)) = p_X^{-1}(\mathbb{G}_T(g_{1,x} - F(x)f_1)) + p_X^{-1}\mathbb{G}_T(f_1)(F(x) - \hat{F}_T(x)).$$

By the central limit theorem, we have that $\mathbb{G}_T(f_1) \xrightarrow{d} \mathcal{N}(0, \mathbb{P}(f_1 - \mathbb{P}f_1)^2)$, applying the law of the large number gives us that $(F(x) - \hat{F}_T(x)) = o_{\mathbb{P}}(1)$. With the help of Slutksy theorem, the second term in the right hand side is thus $o_{\mathbb{P}}(1)$, so

$$\sqrt{T}(\hat{F}_T(x) - F(x)) = p_X^{-1}(\mathbb{G}_T(g_{1,x} - F(x)f_1)) + o_{\mathbb{P}}(1).$$

As a consequence, we obtain the following limiting process of the Lemma :

$$\beta_1(u) = p_X^{-1}\mathbb{G}(1_{X\leq F^{\leftarrow}(u), I=1} - u1_{I=1}).$$

We know that the covariance of a $\mathbb{P}$-Gaussian process is given by $\mathbb{E}[\mathbb{G}(f)\mathbb{G}(g)] = \mathbb{P}(fg) - \mathbb{P}(f)\mathbb{P}(g)$ where f, g are measurable functions. Now, using that, we have

$$\begin{aligned}
cov[\beta_1(u_1), \beta_1(u_2)] &= p_X^{-2}\mathbb{E}\left[\mathbb{G}(1_{X\leq F^{\leftarrow}(u_1), I=1} - u_1 1_{I=1})\mathbb{G}(1_{X\leq F^{\leftarrow}(u_2), I=1} - u_2 1_{I=1})\right], \\
&= p_X^{-2}(\mathbb{P}\left[(1_{X\leq F^{\leftarrow}(u_1), I=1} - u_1 1_{I=1})(1_{X\leq F^{\leftarrow}(u_2), I=1} - u_2 1_{I=1})\right]), \\
&= p_X^{-2}(\mathbb{P}(I=1)\mathbb{P}(X \leq F^{\leftarrow}(u_1), X \leq F^{\leftarrow}(u_2)) - u_1 u_2 \mathbb{P}(I=1)), \\
&= p_X^{-1}(u_1 \wedge u_2 - u_1 u_2).
\end{aligned}$$

That is our statement.

3.3 Proof of Theorem 1

We do the proof only for the normalized error of $\hat{\nu}^{\mathcal{H}*}$ as the proof of $\hat{\nu}^{\mathcal{H}}$ is clearly similar. Using that $\mathbb{E}[F(X)^\alpha] = \frac{1}{1+\alpha}$ $(\alpha \neq 1)$, we can write $\nu(\lambda)$ as :

$$\begin{aligned}
\nu(\lambda) = &\frac{1}{2}\mathbb{E}\left[|F^\lambda(X) - G^{1-\lambda}(Y)|\right] - \frac{\lambda}{2}\mathbb{E}\left[1 - F^\lambda(X)\right] - \frac{1-\lambda}{2}\mathbb{E}\left[1 - G^{1-\lambda}(Y)\right] \\
&+ \frac{1}{2}\frac{1-\lambda-\lambda^2}{(1+\lambda)(1+1-\lambda)}.
\end{aligned}$$

Let us note, by g_λ the function defined as:

$$g_\lambda \colon [0,1]^2 \to [0,1], \quad (u,v) \mapsto u^\lambda \vee v^{1-\lambda} - \frac{1}{2}\left((1-\lambda)u^\lambda + \lambda v^{1-\lambda}\right).$$

We are able to write our estimator of the λ-FMadogram (resp. the λ-FMadogram) in missing data framework as an integral with respect to the hybrid copula estimator (resp. the copula function). We then have :

$$\begin{aligned}
\hat{\nu}_T^{\mathcal{H}*}(\lambda) &= \frac{1}{\sum_{t=1}^T I_t J_t}\sum_{t=1}^T g_\lambda(\hat{F}_T(X_t), \hat{G}_T(Y_t)) I_T J_t + c_\lambda = \int_{[0,1]^2} g_\lambda(u,v) d\hat{C}_T^{\mathcal{H}}(u,v) + c_\lambda, \\
\nu(\lambda) &= \int_{[0,1]^2} g_\lambda(u,v) dC(u,v) + c_\lambda.
\end{aligned}$$

Where c_λ a constant depending on λ. Using the same tools introduced to prove Lemma B.1, we are able to show that :

$$\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}*}(\lambda) - \nu(\lambda)\right) = \frac{1}{2}\left((1-\lambda)\int_{[0,1]} \mathbb{C}_T^{\mathcal{H}}(x^{\frac{1}{\lambda}}, 1)dx + \lambda\int_{[0,1]} \mathbb{C}_T^{\mathcal{H}}(1, x^{\frac{1}{1-\lambda}})dx\right) - \int_{[0,1]} \mathbb{C}_T^{\mathcal{H}}(x^{\frac{1}{\lambda}}, x^{\frac{1}{1-\lambda}})dx.$$

Consider the function $\phi\colon \ell^\infty([0,1]^2) \to \ell^\infty([0,1]), f \mapsto \phi(f)$, defined by

$$(\phi(f))(\lambda) = \frac{1}{2}\left((1-\lambda)\int_{[0,1]} f(x^{\frac{1}{\lambda}},1)dx + \lambda \int_{[0,1]} f(1, x^{\frac{1}{1-\lambda}})dx\right) - \int_{[0,1]} f(x^{\frac{1}{\lambda}}, x^{\frac{1}{1-\lambda}})dx.$$

This function is linear and bounded thus continuous. The continuous mapping theorem (Theorem 1.3.6 of [van der Vaart and Wellner, 1996]) implies, as $T \to \infty$

$$\sqrt{T}(\hat{\nu}_T - \nu) = \phi(\mathbb{C}_T^{\mathcal{H}}) \rightsquigarrow \phi(S_C),$$

in $\ell^\infty([0,1])$. We note that $S_C(u,1) = \alpha(u,1) - \beta_1(u)$ and $S_C(1,v) = \alpha(1,v) - \beta_2(v)$. Indeed, just remark that for the first one we have $\beta_2(1) = 0$ and $\partial C/\partial u(u,1) = 1$ a.s. We thus obtain our statement.

3.4 Proof of Proposition 2

We are able to compute the variance for each process and they are given by the following expressions :

$$\sigma_1^2(p_X, p_{XY}) := Var\left(\int_{[0,1]} \alpha(u^{\frac{1}{\lambda}},1) - \beta_1(u^{\frac{1}{\lambda}})du\right),$$

$$\sigma_2^2(p_Y, p_{XY}) := Var\left(\int_{[0,1]} \alpha(1,u^{\frac{1}{1-\lambda}}) - \beta_2(u^{\frac{1}{1-\lambda}})du\right),$$

$$\sigma_3^2(\mathbf{p}) := Var\left(\int_{[0,1]} S_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})du\right).$$

For $\sigma_1^2(p_X, p_{XY})$, we may compute

$$\begin{aligned}
\sigma_1^2(p_X, p_{XY}) &= \mathbb{E}\left[\int_{[0,1]} \alpha(u^{\frac{1}{\lambda}},1) - \beta_1(u^{\frac{1}{\lambda}})du \int_{[0,1]} \alpha(v^{\frac{1}{\lambda}},1) - \beta_1(v^{\frac{1}{\lambda}})dv\right], \\
&= \int_{[0,1]^2} \mathbb{E}\left[\alpha(u^{\frac{1}{\lambda}},1)\alpha(v^{\frac{1}{\lambda}},1)duv\right] - 2\int_{[0,1]^2} \mathbb{E}\left[\alpha(u^{\frac{1}{\lambda}},1)\beta_1(v^{\frac{1}{\lambda}})duv\right] \\
&+ \int_{[0,1]^2} \mathbb{E}\left[\beta_1(u^{\frac{1}{\lambda}})\beta_1(v^{\frac{1}{\lambda}})duv\right], \\
&= \left(p_{XY}^{-1} - p_X^{-1}\right)\int_{[0,1]^2} (u \wedge v)^{\frac{1}{\lambda}} - u^{\frac{1}{\lambda}} v^{\frac{1}{\lambda}} duv.
\end{aligned}$$

Using the same techniques, we have for $\sigma_2^2(p_Y, p_{XY})$

$$\sigma_2^2(p_Y, p_{XY}) := \left(p_{XY}^{-1} - p_Y^{-1}\right)\int_{[0,1]^2} (u \wedge v)^{\frac{1}{1-\lambda}} - u^{\frac{1}{1-\lambda}} v^{\frac{1}{1-\lambda}} duv.$$

We compute directly

$$\int_{[0,1]^2} (u \wedge v)^{\frac{1}{\lambda}} - u^{\frac{1}{\lambda}} v^{\frac{1}{\lambda}} duv = \left(\frac{\lambda}{1+\lambda}\right)^2 \frac{1}{1+2\lambda},$$

and similarly

$$\int_{[0,1]^2} (u \wedge v)^{\frac{1}{1-\lambda}} - u^{\frac{1}{1-\lambda}} v^{\frac{1}{1-\lambda}} = \left(\frac{1-\lambda}{1+1-\lambda}\right)^2 \frac{1}{1+2(1-\lambda)}.$$

For the covariances, we note

$$\begin{aligned}
\sigma_{12}(\mathbf{p}) &:= \mathbb{E}\left[\int_{[0,1]} \alpha(u^{\frac{1}{\lambda}}, 1) - \beta_1(u^{\frac{1}{\lambda}})du \int_{[0,1]} \alpha(1, v^{\frac{1}{1-\lambda}}) - \beta_2(v^{\frac{1}{1-\lambda}})dv\right], \\
\sigma_{13}(\mathbf{p}) &:= \mathbb{E}\left[\int_{[0,1]} \alpha(u^{\frac{1}{\lambda}}, 1), -\beta_1(u^{\frac{1}{\lambda}})du \int_{[0,1]} S_C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})dv\right], \\
\sigma_{23}(\mathbf{p}) &:= \mathbb{E}\left[\int_{[0,1]} \alpha(1, u^{\frac{1}{1-\lambda}}) - \beta_2(u^{\frac{1}{1-\lambda}})du \int_{[0,1]} S_C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})dv\right].
\end{aligned}$$

Some algebra gives for the first one

$$\sigma_{12}(\mathbf{p}) = \left(p_{XY}^{-1} - p_X^{-1} - p_Y^{-1} + \frac{p_{XY}}{p_X p_Y}\right) \int_{[0,1]^2} C(u^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}}) - u^{\frac{1}{\lambda}} v^{\frac{1}{1-\lambda}}. \tag{3.1}$$

For $\sigma_{13}(\mathbf{p})$, using the definition of the process $S_C(u,v)$ some lines gives

$$\begin{aligned}
\sigma_{13}(\mathbf{p}) &= \mathbb{E}\left[\int_{[0,1]} \alpha(u^{\frac{1}{\lambda}}, 1) - \beta_1(u^{\frac{1}{\lambda}})du \int_{[0,1]} \left(\alpha(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})dv - \frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial v}\beta_2(v^{\frac{1}{\lambda}})\right)\right], \\
&= \left(p_{XY}^{-1} - p_X^{-1}\right) \int_{[0,1]^2} C((u \wedge v)^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}}) - u^{\frac{1}{\lambda}} C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})duv \qquad (3.2) \\
&- \left(p_Y^{-1} - \frac{p_{XY}}{p_X p_Y}\right) \int_{[0,1]^2} \frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial v} \left(C(u^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}}) - u^{\frac{1}{\lambda}} v^{\frac{1}{1-\lambda}}\right) duv, \qquad (3.3)
\end{aligned}$$

where we use, in the first line that

$$\mathbb{E}\left[\int_{[0,1]} \alpha(u^{\frac{1}{\lambda}}, 1) - \beta_1(u^{\frac{1}{\lambda}})du \int_{[0,1]} \frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial u}\beta_1(v^{\frac{1}{\lambda}})dv\right] = 0.$$

Similarly,

$$\begin{aligned}
\sigma_{23}(\mathbf{p}) &= \left(p_{XY}^{-1} - p_Y^{-1}\right) \int_{[0,1]^2} C(v^{\frac{1}{\lambda}}, (u \wedge v)^{\frac{1}{1-\lambda}}) - u^{\frac{1}{1-\lambda}} C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})duv \qquad (3.4) \\
&- \left(p_X^{-1} - \frac{p_{XY}}{p_X p_Y}\right) \int_{[0,1]^2} \frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial u} \left(C(v^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}}) - v^{\frac{1}{\lambda}} u^{\frac{1}{1-\lambda}}\right) duv. \qquad (3.5)
\end{aligned}$$

We go back to the quantity σ_3^2. But before, we introduce some notations for convenience. Let $\lambda \in [0,1]$, using the property exhibited in Equation (5). We may find a similar pattern for partial derivatives,

$$\frac{\partial C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})}{\partial u} = \frac{C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})}{u^{\frac{1}{\lambda}}} \left(A(\lambda) - A'(\lambda)\lambda\right),$$
$$\frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial v} = \frac{C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{v^{\frac{1}{1-\lambda}}} \left(A(\lambda) + A'(\lambda)(1-\lambda)\right).$$

Furthermore, the integral $\int_{[0,1]^2} C(u,v)duv$ does not admit, in general, a closed form. But we are able to express it with respect to a simple integral of the Pickands dependence function. We note, for notational convenience the following functional

$$f\colon [0,1] \times \mathcal{A} \to [0,1], \quad (\lambda, A) \mapsto \left(\frac{\lambda(1-\lambda)}{A(\lambda) + \lambda(1-\lambda)}\right)^2.$$

Now, we can write σ_3^2 in the following way

$$\sigma_3^2(\mathbf{p}) = \left(p_{XY}^{-1}\gamma_1^2 + p_X^{-1}\gamma_2^2 + p_Y^{-1}\gamma_3^2\right) - 2p_X^{-1}\gamma_{12} - 2p_Y^{-1}\gamma_{13} + 2\frac{p_{XY}}{p_X p_Y}\gamma_{23}, \tag{3.6}$$

with

$$\begin{aligned}
Var\left(\int_{[0,1]} \alpha(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})du\right) &= \int_{[0,1]^2} C((u\wedge v)^{\frac{1}{\lambda}}, (u\wedge v)^{\frac{1}{1-\lambda}}) - C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})duv \\
&= p_{XY}^{-1}\gamma_1^2, \\
Var\left(\int_{[0,1]} \frac{\partial C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})}{\partial u}\beta_1(u^{\frac{1}{\lambda}})du\right) &= p_X^{-1}\int_{[0,1]^2} (u\wedge v)^{\frac{1}{\lambda}} - u^{\frac{1}{\lambda}}v^{\frac{1}{\lambda}}duv \\
&= p_X^{-1}\gamma_2^2 \\
Var\left(\int_{[0,1]} \frac{\partial C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})}{\partial v}\beta_2(u^{\frac{1}{1-\lambda}})du\right) &= p_Y^{-1}\int_{[0,1]^2} (u\wedge v)^{\frac{1}{1-\lambda}} - u^{\frac{1}{1-\lambda}}v^{\frac{1}{1-\lambda}}duv \\
&= p_Y^{-1}\gamma_3^2.
\end{aligned}$$

These integrals are tractable and we compute

$$\gamma_1^2 = f(\lambda, A)\left(\frac{A(\lambda)}{A(\lambda) + 2\lambda(1-\lambda)}\right),$$

$$\gamma_2^2 = f(\lambda, A)\left(\frac{\kappa^2(\lambda, A)(1-\lambda)}{2A(\lambda) - (1-\lambda) + 2\lambda(1-\lambda)}\right),$$
$$\gamma_3^2 = f(\lambda, A)\left(\frac{\zeta^2(\lambda, A)\lambda}{2A(\lambda) - \lambda + 2\lambda(1-\lambda)}\right).$$

We now compute the covariance :

$$
\begin{aligned}
p_X^{-1}\gamma_{12} &:= cov\left(\int_{[0,1]} S_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})du, \int_{[0,1]} \beta_1(v^{\frac{1}{\lambda}})\frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial u}dv\right) \\
&= \int_{[0,1]}\int_{[0,1]} \mathbb{E}[S_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})\beta_1(v^{\frac{1}{\lambda}})]\frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial u}dudv, \\
&= p_X^{-1}\int_{[0,1]^2}\left(C((u\wedge v)^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}}) - C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})v^{\frac{1}{\lambda}}\right)\frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial u}dudv.
\end{aligned}
$$

Let us decompose the integrals into two parts, one under the segment $[0,v]$, the other under $[v,1]$. The first one gives

$$
\int_{[0,1]}\int_{[0,v]}(C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}}) - C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})v^{\frac{1}{\lambda}})\frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial u}dudv = \\
\frac{\kappa(\lambda,A)}{2}f(\lambda,A)\left(\frac{1-\lambda}{2A(\lambda)+(2\lambda-1)(1-\lambda)}\right).
$$

For the second part, using Fubini, we have :

$$
\int_{[0,1]}\int_{[0,u]}(C(v^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}}) - C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})v^{\frac{1}{\lambda}})\frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial u}dvdu.
$$

For the right hand side of the minus sign, we may compute :

$$
\int_{[0,1]}\int_{[0,u]} C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})v^{\frac{1}{\lambda}}\frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial u}dvdu = \frac{\kappa(\lambda,A)}{2}f(\lambda,A).
$$

For the last one, some substitutions may be considered.

$$
\int_{[0,1]}\int_{[0,u]} C(v^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})\frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial u}dvdu. \tag{3.7}
$$

Following the proof of Proposition 3.3 from [Genest and Segers, 2009], the substitution $v^{\frac{1}{\lambda}} = x$ and $u^{\frac{1}{1-\lambda}} = y$ yields

$$
\begin{aligned}
&\int_{[0,1]}\int_{[0,u]} C(v^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})\frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial u}dvdu \\
&= \lambda(1-\lambda)\int_{[0,1]}\int_{[0,y^{\frac{1-\lambda}{\lambda}}]} C(x,y)\frac{\partial C(x, x^{\frac{\lambda}{1-\lambda}})}{\partial u}x^{\lambda-1}y^{-\lambda}dxdy \\
&= \lambda(1-\lambda)\kappa(\lambda,A)\int_{[0,1]}\int_{[0,y^{\frac{1-\lambda}{\lambda}}]} C(x,y)x^{\frac{A(\lambda)}{1-\lambda}-(1-\lambda)-1}y^{-\lambda}dxdy.
\end{aligned}
$$

Next, use the substitution $x = w^{1-s}$ and $y = w^s$. Note that $w = xy \in [0,1]$, $s = log(y)/log(xy) \in [0,1]$, $C(x,y) = w^{A(s)}$ and the Jacobian of the transformation is $-log(w)$. As the constraint

$x < y^{-1+1/\lambda}$ reduces to $s < \lambda$, the integral becomes:

$$\begin{aligned}
&-\lambda(1-\lambda)\kappa(\lambda, A) \int_{[0,\lambda]} \int_{[0,1]} w^{A(s)+(1-s)(A_2(\lambda)-(1-\lambda)-1)-s\lambda} log(w) dw ds \\
&= \lambda(1-\lambda)\kappa(\lambda, A) \int_{[0,\lambda]} \left[A(s) + (1-s)(A_2(\lambda) - 1 - (1-\lambda)) - s\lambda + 1\right]^{-2} ds.
\end{aligned}$$

and we thus obtain γ_{12} by doing the sum. Next we compute the following integral :

$$\begin{aligned}
\frac{p_{XY}}{p_X p_Y}\gamma_{23} &:= cov\left(\int_{[0,1]} \beta_1(u^{\frac{1}{\lambda}}) \frac{\partial C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})}{\partial u} du, \int_{[0,1]} \beta_2(v^{\frac{1}{1-\lambda}}) \frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial v} dv\right) \\
&= \frac{p_{XY}}{p_X p_Y} \mathbb{E}\left[\int_{[0,1]} \int_{[0,1]} \beta_1(u^{\frac{1}{\lambda}})\beta_2(v^{\frac{1}{1-\lambda}}) \frac{\partial C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})}{\partial u} \frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial v} du dv\right] \\
&= \int_{[0,1]} \int_{[0,1]} \left(C(u^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}}) - u^{\frac{1}{\lambda}} v^{\frac{1}{1-\lambda}}\right) \frac{\partial C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})}{\partial u} \frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial v} du dv.
\end{aligned}$$

The second term can be easily handled and its value is given by :

$$\int_{[0,1]} \int_{[0,1]} u^{\frac{1}{\lambda}} v^{\frac{1}{1-\lambda}} \frac{\partial C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})}{\partial u} \frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial v} du dv = f(\lambda, A)\kappa(\lambda, A)\zeta(\lambda, A).$$

For the first, use the substitutions $u^{\frac{1}{\lambda}} = x$ and $v^{\frac{1}{1-\lambda}} = y$. This yields :

$$\lambda(1-\lambda) \int_{[0,1]} \int_{[0,1]} C(x, y) \frac{\partial C(x, x^{\frac{\lambda}{1-\lambda}})}{\partial u} \frac{\partial C(y^{\frac{1-\lambda}{\lambda}}, y)}{\partial v} x^{\lambda-1} y^{-\lambda} dx dy.$$

Then, make the substitutions $x = w^{1-s}$, $y = w^s$ that were used for the preceding integral gives :

$$\begin{aligned}
&-\lambda(1-\lambda)\kappa(\lambda, A)\zeta(\lambda, A) \int_{[0,1]} \int_{[0,1]} w^{A(s)+(1-s)(A_2(\lambda)-(1-\lambda)-1)+s(A_1(\lambda)-\lambda-1)} log(w) dw ds \\
&= \lambda(1-\lambda)\kappa(\lambda, A)\zeta(\lambda, A) \int_{[0,1]} \left[A(s) + (1-s)(A_2(\lambda) - (1-\lambda) - 1) + s(A_1(\lambda) - \lambda - 1) + 1\right]^{-2} ds.
\end{aligned}$$

Similarly, the last covariance requires the same tools as used before, it is left to the reader to find that

$$\zeta(\lambda, A)\Bigg(f(\lambda, A)\left(\frac{\lambda^2 - A(\lambda)}{2A(\lambda) - \lambda + 2\lambda(1-\lambda)}\right)\Bigg),$$

under the segment $[0, v]$ and,

$$\lambda(1-\lambda)\zeta(\lambda, A) \int_{[\lambda,1]} \left[A(s) + s(A_1(\lambda) - 1 - \lambda) - (1-s)(1-\lambda) + 1\right]^{-2} ds,$$

on the segment $[1,v]$. Using all the tools depicted, we compute, in the same manner, the integral in Equation (3.1), (3.2), (3.3), (3.4), (3.5) and their values are given by

$$(3.1) = \lambda(1-\lambda)\left(\int_{[0,1]} [A(s)-(1-s)(1-\lambda)-s\lambda+1]^{-2}\,ds - \frac{1}{(1+\lambda)(1+1-\lambda)}\right),$$

$$\begin{aligned}(3.2) = \lambda(1-\lambda)\Bigg(&\int_{[0,\lambda]} [A(s)-(1-s)(1-\lambda)-s\lambda+1]^{-2}\,ds \\ &+ \frac{\lambda}{A(\lambda)+\lambda(1-\lambda)}\left[\frac{1-\lambda}{A(\lambda)+2\lambda(1-\lambda)} - \frac{1}{1+\lambda}\right]\Bigg),\end{aligned}$$

$$\begin{aligned}(3.3) = \zeta(\lambda,A)\lambda(1-\lambda)\Bigg[&\int_{[0,1]} [A(s)+s(A_1(\lambda)-\lambda-1)-(1-s)(1-\lambda)+1]^{-2}ds \\ &- \frac{\lambda}{(1+\lambda)(A(\lambda)+\lambda(1-\lambda))}\Bigg],\end{aligned}$$

$$\begin{aligned}(3.4) = \lambda(1-\lambda)\Bigg(&\int_{[\lambda,1]} [A(s)-(1-s)(1-\lambda)-s\lambda+1]^{-2}\,ds \\ &+ \frac{(1-\lambda)}{A(\lambda)+\lambda(1-\lambda)}\left[\frac{\lambda}{A(\lambda)+2\lambda(1-\lambda)} - \frac{1}{1+1-\lambda}\right]\Bigg),\end{aligned}$$

$$\begin{aligned}(3.5) = \kappa(\lambda,A)\lambda(1-\lambda)\Bigg[&\int_{[0,1]} [A(s)+(1-s)(A_2(\lambda)-\lambda-1)-s(1-\lambda)+1]^{-2}ds \\ &- \frac{1-\lambda}{(1+1-\lambda)(A(\lambda)+\lambda(1-\lambda))}\Bigg].\end{aligned}$$

And that's the end of the proof.

3.5 Proof of Corollary 1

We use the same notations as in Section 3.4. As $C(u,v)=uv$, we have that

$$(3.1) = (3.3) = (3.5) = 0.$$

And the integral in Equation (3.2) become :

$$\begin{aligned}&\int_{[0,1]}\int_{[0,v]} u^{\frac{1}{\lambda}} v^{\frac{1}{1-\lambda}} du dv + \frac{\lambda^2(1-\lambda)}{1+\lambda(1-\lambda)}\left[\frac{1-\lambda}{1+2\lambda(1-\lambda)} - \frac{1}{1+\lambda}\right] = \\ &\lambda^2(1-\lambda)\left(\frac{1}{(1+\lambda)(1-\lambda)+\lambda+\lambda(1-\lambda)} + \frac{1}{1+\lambda(1-\lambda)}\left[\frac{1-\lambda}{1+2\lambda(1-\lambda)} - \frac{1}{1+\lambda}\right]\right).\end{aligned}$$

The multiplication by $(p_{XY}^{-1} - p_X^{-1})$ gives σ_{13}. Same is for Equation (3.4)

$$\int_{[0,1]}\int_{[0,u]} v^{\frac{1}{\lambda}} u^{\frac{1}{1-\lambda}} dvdu + \frac{\lambda(1-\lambda)^2}{1+\lambda(1-\lambda)}\left[\frac{\lambda}{1+2\lambda(1-\lambda)} - \frac{1}{1+1-\lambda}\right] =$$
$$\lambda(1-\lambda)^2\left(\frac{1}{\lambda(1+1-\lambda)+1-\lambda+\lambda(1-\lambda)} + \frac{1}{1+\lambda(1-\lambda)}\left[\frac{\lambda}{1+2\lambda(1-\lambda)} - \frac{1}{1+1-\lambda}\right]\right).$$

And we obtain σ_{23} with multiplying by $(p_{XY}^{-1} - p_Y^{-1})$. In the independent case $A(t) = 1$ for every $t \in [0,1]$ implies that that $\kappa(\lambda, A) = \zeta(\lambda, A) = 1$ for every $\lambda \in [0,1]$. Then for σ_3^2, Equation (3.7) equals :

$$f(\lambda,1)\left(\frac{1+\lambda(1-\lambda)}{2-(1-\lambda)+2\lambda(1-\lambda)}\right).$$

Summing all the elements gives

$$\int_{[0,1]^2}\left(C((u\wedge v)^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}}) - C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})v^{\frac{1}{\lambda}}\right)\frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial u} dudv = f(\lambda,1)\left(\frac{1-\lambda}{2-(1-\lambda)+2\lambda(1-\lambda)}\right).$$

Same computations gives

$$\int_{[0,1]^2}\left(C(u^{\frac{1}{\lambda}}, (u\wedge v)^{\frac{1}{1-\lambda}}) - C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})v^{\frac{1}{1-\lambda}}\right)\frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial v} dudv = f(\lambda,1)\left(\frac{\lambda}{2-\lambda+2\lambda(1-\lambda)}\right).$$

In independent case, we have the following equality :

$$\int_{[0,1]^2}\left(C(u^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}}) - u^{\frac{1}{\lambda}}v^{\frac{1}{1-\lambda}}\right)\frac{\partial C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})}{\partial u}\frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial v} dudv = 0.$$

We then obtain for σ_3^2 when we sum all the elements :

$$\sigma_3^2(\mathbf{p}) = \left(\frac{\lambda(1-\lambda)}{1+\lambda(1-\lambda)}\right)^2\left(\frac{p_{XY}^{-1}}{1+2\lambda(1-\lambda)} - \frac{p_X^{-1}(1-\lambda)}{2-(1-\lambda)+2\lambda(1-\lambda)} - \frac{p_Y^{-1}\lambda}{2-\lambda+2\lambda(1-\lambda)}\right)$$

3.6 Asymptotic Behavior of Equation (1.3) under complete data

It is readily verified that

$$\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}*}(\lambda) - \nu(\lambda)\right) =$$
$$\frac{1}{2}\left((1-\lambda)\int_{[0,1]}\mathbb{C}_T(x^{\frac{1}{\lambda}}, 1)dx + \lambda\int_{[0,1]}\mathbb{C}_T(1, x^{\frac{1}{1-\lambda}})dx\right) - \int_{[0,1]}\mathbb{C}_T(x^{\frac{1}{\lambda}}, x^{\frac{1}{1-\lambda}})dx.$$

Using the same argument as in the proof of Theorem 1, we can show, with complete data, that $(1-\lambda)\int_{[0,1]}\mathbb{C}_T(x^{\frac{1}{\lambda}}, 1)dx \rightsquigarrow \delta_{\{0\}}$ and $\lambda\int_{[0,1]}\mathbb{C}_T(1, x^{\frac{1}{1-\lambda}})dx \rightsquigarrow \delta_{\{0\}}$, where $\delta_{\{0\}}$ refers to the Dirac measure at 0. We thus obtain that, by extended Slutsky's lemma (example 1.4.7 of [van der Vaart and Wellner, 1996])

$$\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}*}(\lambda) - \nu(\lambda)\right) \rightsquigarrow -\int_{[0,1]} N_C(x^{\frac{1}{\lambda}}, x^{\frac{1}{1-\lambda}})dx.$$

That's what we wanted to prove.

3.7 Proof of Theorem 2

We will denote by S the index set of sane blocks. For the rest of the section, λ is a fixed constant between 0 and 1.

Lemma 2. *For every positive ϵ, it holds that*

$$\mathbb{P}\bigg\{\ |\hat{\nu}_{MoN} - \nu| > \epsilon\bigg\} \leq \mathbb{P}\bigg\{\ \left|\hat{\nu}_{n_j} - \nu\right| > \epsilon\bigg\}^{K\delta} 2^K, \quad j \in S. \tag{3.8}$$

Proof First, observe that the event

$$\{|\hat{\nu}_{MoN} - \nu| > \epsilon\},$$

implies that at least $K/2$ of $\hat{\nu}_{n_j}$ for $j \in \{1, \ldots, K\}$ has to be outside distance ϵ of ν. Namely,

$$\{|\hat{\nu}_{MoN} - \nu| > \epsilon\} \subset \left\{\left|\sum_{k\in S} \mathbb{1}_{\{|\hat{\nu}_{n_j} - \nu| \geq \epsilon\}}\right| \geq \frac{K}{2}\right\}.$$

Then for the first inequality, we have

$$\begin{aligned}
\mathbb{P}\bigg\{\ |\hat{\nu}_{MoN} - \nu| > \epsilon\bigg\} &\leq \mathbb{P}\left\{\left|\sum_{k\in S} \mathbb{1}_{\{|\hat{\nu}_{n_j} - \nu| \geq \epsilon\}}\right| \geq \frac{K}{2}\right\}, \\
&\leq \mathbb{P}\left\{\left|\sum_{j\in S} \mathbb{1}_{\{|\hat{\nu}_{n_j} - \nu| \geq \epsilon\}}\right| \geq K_s - \frac{K}{2}\right\}, \\
&\leq \mathbb{P}\left\{\left|\sum_{j\in S} \mathbb{1}_{\{|\hat{\nu}_{n_j} - \nu| \geq \epsilon\}}\right| \geq K_s\left(1 - \frac{1}{2}(\frac{1}{2} + \delta)^{-1}\right)\right\}.
\end{aligned}$$

All these inequalities results from $K \geq K_s \geq K(2^{-1} + \delta)$ and that $K_s + K_o = K$. Notice that the random variable $\sum_{j\in S} \mathbb{1}_{\{|\hat{\nu}_{n_j} - \nu|\}}$ is distributed according to a binomial random variable with K_s trials and probability p_ϵ with

$$p_\epsilon = \mathbb{P}\left\{\left|\hat{\nu}_{n_j} - \nu\right| > \epsilon\right\}.$$

It can thus be upper bounded by

$$\sum_{n=\lceil K_s(1-\frac{1}{2}(\frac{1}{2}+\delta)^{-1})\rceil}^{K_s} \binom{K_s}{n} p_\epsilon^n (1-p_\epsilon)^{n-K_s} \leq p_\epsilon^{K_s\left(1-\frac{1}{2}(\frac{1}{2}+\delta)^{-1}\right)} \sum_{n=1}^{K_s} \binom{K_s}{n},$$
$$\leq p_\epsilon^{K_s\left(1-\frac{1}{2}(\frac{1}{2}+\delta)^{-1}\right)} 2^{K_s},$$
$$\leq p_\epsilon^{K\delta} 2^K.$$

When we use that $K_s(1-2^{-1}(2^{-1}+\delta)^{-1}) \geq K\delta$ and $K_s \leq K$. That is our statement.

Lemma 3. *For every $j \in S$ and $\epsilon > 0$, we have*

$$p_\epsilon \leq \mathbb{P}\left\{\left|\frac{1}{2n_j}\sum_{t\in B_j}|F(X_t)-G(Y_t)| - \frac{1}{2}\mathbb{E}\,|F(X)-G(Y)|\right| > \frac{\epsilon}{3}\right\} \tag{3.9}$$
$$+\,\mathbb{P}\left\{\underset{t\in B_j}{sup}\left|\hat{F}_{n_j}(X_t)-F(X_t)\right| > \frac{2\epsilon}{3}\right\} + \mathbb{P}\left\{\underset{t\in B_j}{sup}\left|\hat{G}_{n_j}(Y_t)-G(Y_t)\right| > \frac{2\epsilon}{3}\right\}. \tag{3.10}$$

Proof First, notice that we can obtain the following upper bound

$$\left|\hat{\nu}_{n_j}-\nu\right| = \left|\frac{1}{2n_j}\sum_{t\in B_j}|\hat{F}_{n_j}(X_t)-\hat{G}_{n_j}(Y_t)| - \frac{1}{2}\mathbb{E}|F(X)-G(Y)|\right|,$$
$$\leq \left|\frac{1}{2n_j}\sum_{t\in B_j}\Big(|\hat{F}_{n_j}(X_t)-\hat{G}_{n_j}(Y_t)| - |F(X_t)-G(Y_t)|\Big)\right|$$
$$+\left|\frac{1}{2n_j}\sum_{t\in B_j}|F(X_t)-G(Y_t)| - \frac{1}{2}\mathbb{E}|F(X)-G(Y)|\right|.$$

The first expression can be bounded by

$$\left|\frac{1}{2n_j}\sum_{t\in B_j}\Big(|\hat{F}_{n_j}(X_t)-\hat{G}_{n_j}(Y_t)| - |F(X_t)-G(Y_t)|\Big)\right|,$$
$$\overset{(a)}{\leq} \frac{1}{2n_j}\sum_{t\in B_j}\Big||\hat{F}_{n_j}(X_t)-\hat{G}_{n_j}(Y_t)| - |F(X_t)-G(Y_t)|\Big|,$$
$$\overset{(b)}{\leq} \frac{1}{2n_j}\sum_{t\in B_j}\left|\hat{F}_{n_j}(X_t)-F(X_t)-\Big(\hat{G}_{n_j}(Y_t)-G(Y_t)\Big)\,|\right|,$$
$$\overset{(c)}{\leq} \frac{1}{2}\underset{t\in B_j}{sup}\left|\hat{F}_{n_j}(X_t)-F(X_t)\right| + \frac{1}{2}\underset{t\in B_j}{sup}\left|\hat{G}_{n_j}(Y_t)-G(Y_t)\right|.$$

We used triangle inequality in (a), $||x|-|y|| \leq |x-y|$ in (b) and both triangle inequality and that $\sum_{t=1}^T x_t \leq T \underset{t\in\{1,\ldots,T\}}{sup} x_t$ in (c). Since :

$$\left\{\left|\hat{\nu}_{n_j}-\nu\right|\le\epsilon\right\}\supseteq\left\{\frac{1}{2}\underset{t\in B_j}{sup}\left|\hat{F}_{n_j}(X_t)-F(X_t)\right|\le\frac{\epsilon}{3}\right\}\cap\left\{\frac{1}{2}\underset{t\in B_j}{sup}\left|\hat{G}_{n_j}(Y_t)-G(Y_t)\right|\le\frac{\epsilon}{3}\right\}$$
$$\cap\left\{\left|\frac{1}{2n_j}\sum_{t\in B_j}|F(X_t)-G(Y_t)|-\frac{1}{2}\mathbb{E}|F(X)-G(Y)|\right|\le\frac{\epsilon}{3}\right\}.$$

We thus obtain our lemma using union bound.

We can write Equation 3.8 such as

$$exp\left(K\delta log\left(p_\epsilon 2^{\frac{1}{\delta}}\right)\right). \tag{3.11}$$

The DKW inequality (see page 384 in [Boucheron et al., 2013], [Massart, 1990] or in the proof of Theorem 1 in [Alquier et al., 2020] for a similar application) gives us an upper bound for Equation (3.10) in the following form :

$$4exp\left(-\frac{8}{9}n_j\epsilon^2\right)\le 4exp\left(-\frac{2}{9}n_j\epsilon^2\right).$$

Clearly, we have that

$$\mathbb{E}\left[\frac{1}{2n_j}\sum_{t\in B_j}|F(X_t)-G(Y_t)|\right]=\frac{1}{2}\mathbb{E}|F(X)-G(Y)|,$$

and, for every $t\in B_j$

$$\frac{1}{2n_j}|F(X_t)-G(Y_t)|\le\frac{1}{n_j}.$$

Applying Hoeffding's inequality permits us to bound Equation (3.9) by

$$2exp\left(-\frac{2}{9}n_j\epsilon^2\right).$$

Summing all these components and the use of Lemma 3 yields to

$$p_\epsilon\le 6exp\left(-\frac{2}{9}n_j\epsilon^2\right).$$

Plugging this inequality in Equation (3.11) leads to

$$\mathbb{P}\left\{|\hat{\nu}_{MoN}-\nu|>\epsilon\right\}\le exp\left(K\delta log\left(6e^{-\frac{2\epsilon^2 n_j}{9}}2^{\frac{1}{\delta}}\right)\right).$$

It can be set to η by choosing $K = log(1/\eta)\delta^{-1}$ and ϵ such that $6e^{-\frac{2\epsilon^2 n_j}{9}} 2^{\frac{1}{\delta}} = 1/e$, or again

$$\epsilon = \frac{3}{\sqrt{n_j}} log\left(6e2^{\frac{1}{\delta}}\right) = \frac{3}{\sqrt{2}} \frac{log\left(6e2^{\frac{1}{\delta}}\right)}{\delta} \sqrt{\frac{log\left(1/\eta\right)}{T}}.$$

And we are done.

Bibliography

[Alquier et al., 2020] Alquier, P., Chérief-Abdellatif, B.-E., Derumigny, A., and Fermanian, J.-D. (2020). Estimation of copulas via maximum mean discrepancy.

[Angrist and Pischke, 2008] Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

[Bador et al., 2015] Bador, M., Naveau, P., Gilleland, E., Castellà, M., and Arivelo, T. (2015). Spatial clustering of summer temperature maxima from the cnrm-cm5 climate model ensembles & e-obs over europe. *Weather and Climate Extremes*, 9:17–24. The World Climate Research Program Grand Challenge on Extremes – WCRP-ICTP Summer School on Attribution and Prediction of Extreme Events.

[Baraud et al., 2016] Baraud, Y., Birgé, L., and Sart, M. (2016). A new method for estimation and model selection: ρ-estimation. *Inventiones Mathematicae*.

[Beirlant et al., 2004] Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley. Pagination: 522.

[Bernard et al., 2013] Bernard, E., Naveau, P., Vrac, M., and Mestre, O. (2013). Clustering of Maxima: Spatial Dependencies among Heavy Rainfall in France. *Journal of Climate*, 26(20):7929–7937.

[Boucheron et al., 2013] Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration inequalities. A nonasymptotic theory of independence*. Oxford University Press.

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

[Cochran, 2007] Cochran, W. (2007). *Sampling Techniques, 3Rd Edition*. A Wiley publication in applied statistics. Wiley India Pvt. Limited.

[Coles et al., 1999] Coles, S., Heffernan, J., and Tawn, J. (1999). Dependence measures for extreme value analyses. *Extremes*, 2:339 – 365.

[Cooley et al., 2006] Cooley, D., Naveau, P., and Poncet, P. (2006). *Variograms for spatial max-stable random fields*, pages 373–390. Springer New York, New York, NY.

[Deheuvels, 1991] Deheuvels, P. (1991). On the limiting behavior of the pickands estimator for bivariate extreme-value distributions. *Statistics & Probability Letters*, 12(5):429–439.

[Demarta and McNeil, 2005] Demarta, S. and McNeil, A. J. (2005). The t copula and related copulas. *International Statistical Review*, 73(1):111–129.

[Fermanian et al., 2004] Fermanian, J.-D., Radulovic, D., and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, 10(5):847 – 860.

[Gaetan and Guyon, 2008] Gaetan, C. and Guyon, X. (2008). *Modélisation et statistique spatiales*. Mathématiques & applications. Springer, Berlin Heidelberg New York.

[Genest and Segers, 2009] Genest, C. and Segers, J. (2009). Rank-based inference for bivariate extreme-value copulas. *The Annals of Statistics*, 37(5B):2990 – 3022.

[Gudendorf and Segers, 2009] Gudendorf, G. and Segers, J. (2009). Extreme-value copulas.

[Hall and Tajvidi, 2000] Hall, P. and Tajvidi, N. (2000). Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli*, 6(6):835–844.

[Huber, 1964] Huber, P. J. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1):73 – 101.

[Huber, 2011] Huber, P. J. (2011). *Robust Statistics*, pages 1248–1251. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Hüsler and Reiss, 1989] Hüsler, J. and Reiss, R.-D. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statistics & Probability Letters*, 7(4):283–286.

[Joe, 1990] Joe, H. (1990). Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statistics & Probability Letters*, 9:75–81.

[Lerasle et al., 2019] Lerasle, M., Szabó, Z., Mathieu, T., and Lecué, G. (2019). MONK – Outlier-Robust Mean Embedding Estimation by Median-of-Means. In *ICML 2019 - 36th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Long Beach, United States.

[Little R.J.A., 1987] Little R.J.A., R. D. (1987). *Statistical analysis with missing data*. Wiley Series in Probability and Statistics.

[Marcon et al., 2017] Marcon, G., Padoan, S., Naveau, P., Muliere, P., and Segers, J. (2017). Multivariate nonparametric estimation of the pickands dependence function using bernstein polynomials. *Journal of Statistical Planning and Inference*, 183:1–17.

[Massart, 1990] Massart, P. (1990). The Tight Constant in the Dvoretzky-Kiefer-Wolfowitz Inequality. *The Annals of Probability*, 18(3):1269 – 1283.

[Naveau et al., 2009] Naveau, P., Guillou, A., Cooley, D., and Diebolt, J. (2009). Modeling pairwise dependence of maxima in space. *Biometrika*, 96(1):1–17.

[Nemirovsky and Yudin, 1983] Nemirovsky, A. and Yudin, D. (1983). *Problem Complexity and Method Efficiency in Optimization*. John Wiley, New York.

[Oliveira and Galambos, 1977] Oliveira, J. D. T. and Galambos, J. (1977). The asymptotic theory of extreme order statistics. *International Statistical Review*, 47:230.

[Pickands, 1981] Pickands, J. (1981). Multivariate extreme value distribution. *Proceedings 43th, Session of International Statistical Institution, 1981.*

[Segers, 2012] Segers, J. (2012). Asymptotics of empirical copula processes under nonrestrictive smoothness assumptions. *Bernoulli*, 18(3):764 – 782.

[Segers, 2014] Segers, J. (2014). Hybrid copula estimators.

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

[Smith, 2005] Smith, R. L. (2005). Max-stable processes and spatial extremes. In *None*.

[Tawn, 1988] Tawn, J. A. (1988). Bivariate extreme value theory: Models and estimation. *Biometrika*, 75(3):397–415.

[van der Vaart and Wellner, 1996] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Process: With Applications to Statistics*. Springer.

Appendix A

Study of the Pickands dependence function

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and (X, Y) be a bivariate random vector with values in $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$. This random vector has a joint distribution function H and marginal distribution function F and G. We suppose the copula function of H is an extreme-value copula type, *i.e.* if and only if it admits a representation of the form

$$C(u,v) = (uv)^{A(log(v)/(log(uv)))} \tag{A.1}$$

for all $u, v \in [0,1]$ and where $A(\cdot)$ is the Pickands dependence function, *i.e.*, $A : [0,1] \longrightarrow [1/2, 1]$ is convex and satisfies $t \vee (1-t) \leq A(t) \leq 1$, $\forall t \in [0,1]$.

We will call by λ-FMadogram the following quantity

$$\nu(\lambda) = \frac{1}{2}\mathbb{E}\left[\left|F^{\lambda}(X) - G^{1-\lambda}(Y)\right|\right]. \tag{A.2}$$

In the following proposition, we establish some properties of the λ-FMadogram.

Proposition A.1. *Let (X,Y) a $\mathbb{R}^2$-valued random vector of distribution H. We have, for each $\lambda \in [0,1]$,*

(i) $\frac{\lambda \vee (1-\lambda)}{\lambda \vee (1-\lambda) + \lambda(1-\lambda)} - \frac{1}{2}\left(\frac{1}{1+\lambda} + \frac{1}{1+1-\lambda}\right) \leq \nu(\lambda) \leq \frac{1}{1+\lambda(1-\lambda)} - \frac{1}{2}\left(\frac{1}{1+\lambda} + \frac{1}{1+1-\lambda}\right)$,

(ii) $\nu(0) = \nu(1) = 0.25$, *and if* $\lambda \in (0,1)$,

$$\nu(\lambda) = \frac{A(\lambda)}{A(\lambda) + \lambda(1-\lambda)} - \frac{1}{2}\left(\frac{1}{1+\lambda} + \frac{1}{1+1-\lambda}\right). \tag{A.3}$$

Proof The first statement results directly from *(ii)*. To show *(ii)* we define the following function,

$$\nu_{\lambda} \colon [0,1]^2 \to [0,1], \quad (u,v) \mapsto u^{\lambda} \vee v^{1-\lambda} - \frac{1}{2}(u^{\lambda} + v^{1-\lambda}).$$

Using Lemma B.1 in Appendix and the equality $|u^\lambda - v^{1-\lambda}| = u^\lambda \vee v^{1-\lambda} - 2^{-1}(u^\lambda + v^{1-\lambda}) = \nu_\lambda(u,v)$ gives,

$$\begin{aligned}
\nu(\lambda) &= \frac{1}{2}\left(\int_{[0,1]} C(x^{\frac{1}{\lambda}},1)dx + \int_{[0,1]} C(1,x^{\frac{1}{1-\lambda}})dx\right) - \int_{[0,1]} C(x^{\frac{1}{\lambda}},x^{\frac{1}{1-\lambda}})dx, \\
&= \frac{1}{2}\left(\int_{[0,1]} x^{\frac{1}{\lambda}}dx + \int_{[0,1]} x^{\frac{1}{1-\lambda}}dx\right) - \int_{[0,1]} x^{\frac{A(\lambda)}{\lambda(1-\lambda)}}dx, \\
&= \frac{1}{2}\left(\frac{\lambda}{1+\lambda} + \frac{1-\lambda}{1+1-\lambda}\right) - \frac{\lambda(1-\lambda)}{A(\lambda)+\lambda(1-\lambda)}, \\
&= \frac{A(\lambda)}{A(\lambda)+\lambda(1-\lambda)} - \frac{1}{2}\left(\frac{1}{1+\lambda} + \frac{1}{1+1-\lambda}\right)
\end{aligned}$$

That is our statement.

Remark 3. *The upper bound (resp. the lower bound) in (i) is exactly the value of the λ-FMadogram when X and Y are independent (resp. perfectly positive dependent),* i.e. *when $A(t) = 1$ (resp. $A(t) = t \vee (1-t)$).*

Let $\mathcal{A}$ be the space of Pickands dependence functions. We will denote by $\kappa(\lambda, A)$ and $\zeta(\lambda, A)$ the two functional such as :

$$\begin{aligned}
&\kappa\colon [0,1] \times \mathcal{A} \to [0,1], \quad (t,A) \mapsto A(t) - A'(t)t, \\
&\zeta\colon [0,1] \times \mathcal{A} \to [0,1], \quad (t,A) \mapsto A(t) + A'(t)(1-t).
\end{aligned}$$

We are able to prove the two following lemmas

Lemma A.1. *Using the properties of the Pickands dependence function, we have that*

$$0 \leq \kappa(t,A) \leq 1, \quad 0 \leq \zeta(t,A) \leq 1, \quad 0 < t < 1.$$

Furthermore, if A admits a second derivative on $]0,1[$, $\kappa(\cdot, A)$ (resp. $\zeta(\cdot, A)$) is a decreasing function (resp. an increasing function).

Proof First, using that the graph of a (differentiable) convex function lies above all of its tangents and using that $A(t) \geq t$ gives, for $0 < t < 1$:

$$A'(t) \leq \frac{A(1) - A(t)}{1-t} = \frac{1 - A(t)}{1-t} \leq 1.$$

Same reasoning using $A(t) \geq 1 - t$ leads to:

$$A'(t) \geq \frac{A(t) - A(0)}{t - 0} = \frac{A(t) - 1}{t} \geq -1.$$

Let's fall back to κ and ζ. If we suppose that A admits a second derivative, the derivative of κ (resp ζ) with respect to λ gives:

$$\kappa'(t,A) = -tA''(t) < 0, \quad \zeta'(t,A) = (1-t)A''(t) > 0, \quad \forall t \in]0,1[.$$

Using $\kappa(0) = 1$, $\kappa(1) = 1 - A'(1) \geq 0$ gives $0 \leq \kappa(\lambda, A) \leq 1$. As $\zeta(0) = 1 + A'(0) \geq 0$ and $\zeta(1) = 1$, we have $0 \leq \zeta(\lambda, A) \leq 1$. That is the statement.

Now, we can obtain the same result while removing the hypothesis of A admits a second derivative. As A is a convex function, for $x, y \in [0,1]$, we may have the following inequality:

$$A(x) \geq A(y) + A'(y)(x-y).$$

Take $x = 0$ and $y = t$ gives $1 \geq A(t) - tA'(t) = \kappa(t)$. Now, using that $-tA'(t) \geq -t$, clearly $A(t) - tA'(t) \geq A(t) - t \geq 0$. As $A(t) \geq max(t, 1-t)$. We thus obtain our statement.

Lemma A.2. *If A admits a derivative, then $\lim_{t \to 0^+} A'(t)$ and $\lim_{t \to 1^-} A'(t)$ exists and are finite.*

Proof As A is convex and derivable, it follows that $A'(\cdot)$ is increasing. Furthermore, in the proof of Lemma A.1, we showed that $-1 \leq A'(t) \leq 1$ for every $t \in (0,1)$ and therefore bounded. Then the two limits exist and are finite.

The partial derivatives of the extreme value copula are given by

$$\frac{\partial C(u,v)}{\partial u} = \begin{cases} \frac{C(u,v)}{u} \kappa(\frac{log(v)}{log(uv)}, A), & \text{if}\, 0 < u, v \leq 1, \\ 0, & \text{if}\, v = 0, \quad 0 \leq u \leq 1, \end{cases}$$

$$\frac{\partial C(u,v)}{\partial v} = \begin{cases} \frac{C(u,v)}{v} \zeta(\frac{log(v)}{log(uv)}, A), & \text{if}\, 0 < u, v \leq 1, \\ 0, & \text{if}\, u = 0, \quad 0 \leq v \leq 1. \end{cases}$$

The properties of A imply $0 \leq A(t) - tA'(t) \leq 1$ and $0 \leq A(t) + (1-t)A'(t) \leq 1$ where $t = log(v)/log(uv)$ (see Lemma A.1). Therefore, if $v \searrow 0$, then $\partial C(u,v)/\partial u \to 0$ as required. We also need that the functionals $t \mapsto \kappa(t, A)$ and $t \mapsto \zeta(t, A)$ be defined on 0 and 1 in order that the previous discussion holds. Such is always the case as stated by Lemma A.2. We set with extend by continuity

$$A'(0) = \lim_{t \to 0^-} A'(t), \quad A'(1) = \lim_{t \to 1^+} A'(t).$$

Thus, both functionals are defined for every $t \in [0,1]$. An extreme value copula verify smoothness condition 2.1 of [Segers, 2012]. This implies the weak convergence of the empirical copula process if C is of extreme value type.

Appendix B

Auxiliary results

Theorem B.1 (Theorem 3 of [Fermanian et al., 2004]). *Suppose that H has continuous marginal distribution functions and that the copula function $C(x,y)$ has continuous partial derivatives. Then the empirical copula process $\{\mathbb{C}_T(u,v), 0 \leq u, v \leq 1\}$ converges weakly to a Gaussian process $\{N_C(u,v), 0 \leq u,v \leq 1\}$ in $\ell^\infty([0,1]^2)$.*

Under the assumptions defined in Assumption A, the following proposition from [Naveau et al., 2009] hold.

Proposition B.1 (Proposition 3 of [Naveau et al., 2009]). *Suppose that Assumptions A holds and let J be a function of bounded variation, continuous from above and with discontinuities of the first kind. Then:*

$$T^{-1/2} \sum_{t=1}^{T} \Big(J(\hat{F}_T(X_t), \hat{G}_T(Y_t)) - \mathbb{E}\left[J(F(X), G(Y))\right] \Big)$$

converges in distribution to $\int_{[0,1]^2} N_C(u,v) dJ(u,v)$ where $N_C(u,v)$ and the integral is well defined as a Lebesgue-Stieltjes integral. The special case, $J(x,y) = 2^{-1}|x^\lambda - y^{1-\lambda}|$ provide the weak of convergence of the λ-FMadogram estimator :

$$T^{1/2} \left\{ \hat{\nu}_T(\lambda) - \frac{1}{2}\mathbb{E}\left[|F^\lambda(X) - G^{1-\lambda}(Y)|\right] \right\}$$

converge in distribution to $\int_{[0,1]^2} N_C(u,v) dJ(u,v)$ where the latter integral satisfies :

$$\int_{[0,1]^2} f(x,y) dJ(x,y) = \frac{1}{2}\int_{[0,1]} f(0, y^{1/(1-\lambda)}) dy + \frac{1}{2}\int_{[0,1]} f(x^{1/\lambda}, 0) dx - \int_{[0,1]} f(x^{1/\lambda}, x^{1/(1-\lambda)}) dx \tag{B.1}$$

for all bounded-measurable function $f : [0,1]^2 \mapsto \mathbb{R}$.

Lemma B.1. *(Lemma A.1 of [Marcon et al., 2017]) For $\lambda \in [0,1]$, let H be any distribution function in $[0,1]^2$, let ν_λ be the function defined by*

$$\nu_\lambda \colon [0,1]^2 \to [0,1], \quad (u,v) \mapsto u^\lambda \vee v^{1-\lambda} - \frac{1}{2}(u^\lambda + v^{1-\lambda}),$$

Then

$$\int_{[0,1]^2} \nu_\lambda(u,v) dH(u,v) = \frac{1}{2}\left(\int_{[0,1]} H(x^{\frac{1}{\lambda}}, 1)dx + \int_{[0,1]} H(1, x^{\frac{1}{1-\lambda}})dx\right) - \int_{[0,1]} H(x^{\frac{1}{\lambda}}, x^{\frac{1}{1-\lambda}})dx. \tag{B.2}$$

Proof We have,

$$u^\lambda \vee v^{1-\lambda} = 1 - \int_{[0,1]} \mathbb{1}_{\{u^\lambda \leq x, v^{1-\lambda} \leq x\}} dx = 1 - \int_{[0,1]} \mathbb{1}_{\{u \leq x^{\frac{1}{\lambda}}, v \leq x^{\frac{1}{1-\lambda}}\}} dx,$$

using the same technique, we may have,

$$\frac{1}{2}(u^\lambda + v^{1-\lambda}) = 1 - \frac{1}{2}\left(\int_{[0,1]} \mathbb{1}_{\{u \leq x^{\frac{1}{\lambda}}\}} + \mathbb{1}_{\{v \leq x^{\frac{1}{1-\lambda}}\}}\right) dx,$$

We obtain by substracting the two terms above and integration with respect to H,

$$\int_{[0,1]^2} v_\lambda(u,v) dH(u,v) = \frac{1}{2}\int_{[0,1]^2}\int_{[0,1]} \left(\mathbb{1}_{\{u \leq x^{\frac{1}{\lambda}}\}} + \mathbb{1}_{\{v \leq x^{\frac{1}{1-\lambda}}\}}\right) dH(u,v)dx - \int_{[0,1]^2}\int_{[0,1]} \mathbb{1}_{\{u \leq x^{\frac{1}{\lambda}}, v \leq x^{\frac{1}{1-\lambda}}\}} dH(u,v)dx$$

Applying Fubini lead us to the conclusion.

Theorem B.2 (Theorem 2.3 in [Segers, 2014])**.** *If conditions A and C holds, then uniformly in $u \in [0,1]^2$,*

$$\begin{aligned}\sqrt{T}\left\{\hat{C}_T(u,v) - C(u,v)\right\} = \sqrt{T}\left\{\hat{H}_T((F,G)^{\leftarrow}(u,v) - C(u,v)\right\} \\ - \frac{\partial C(u,v)}{\partial u}\sqrt{T}\left\{\hat{F}_T(F^{\leftarrow}(u)) - u\right\} 1_{(0,1)}(u) \\ - \frac{\partial C(u,v)}{\partial v}\sqrt{T}\left\{\hat{G}_T(G^{\leftarrow}(v)) - v\right\} 1_{(0,1)}(v) + o_{\mathbb{P}}(1)\end{aligned}$$

as $T \to \infty$. Hence in $l^\infty([0,1]^2)$ equipped with the supremum norm, as $T \to \infty$,

$$\left(\sqrt{T}\left\{\hat{C}_T(u,v) - C(u,v)\right\}\right)_{u,v \in [0,1]^2} \rightsquigarrow \left(\alpha(u,v) - \frac{\partial C(u,v)}{\partial u}\beta_1(u) - \frac{\partial C(u,v)}{\partial v}\beta_2(v)\right)_{u,v \in [0,1]^2} \tag{B.3}$$

The processes α, β_1 and β_2 have continuous trajectories almost surely. The right-hand side in (B.3) is well-defined because $\beta_j(0) = \beta_j(1) = 1$ almost surely with $j \in \{1,2\}$.

Appendix C

Supplementary results

C.1 A Lemma for Equation (1.8)

Lemma C.1. *For all bounded-measurable function $f : [0,1]^2 \mapsto \mathbb{R}$, if $J(s,t) = 2^{-1}|s^\lambda - t^{1-\lambda}|$, then the following integral satisfies:*

$$\int_{[0,1]^2} f(x,y)dJ(x,y) = \frac{1}{2}\int_{[0,1]} f(1, y^{1/(1-\lambda)})dy + \frac{1}{2}\int_{[0,1]} f(x^{1/\lambda}, 1)dx - \int_{[0,1]} f(x^{1/\lambda}, x^{1/(1-\lambda)})dx.$$

Proof Let $A = [0,s]\times[0,t]$, a closed pavement of $[0,1]^2$, where $s,t \in [0,1]$. Thus, $A \in \mathcal{B}([0,1])^2$. Let us introduce the following indicator function :

$$f_{s,t}(x,y) = \mathbb{1}_{\{(x,y)\in[0,1]^2, 0\le x\le s, 0\le y\le t\}}.$$

Then, for this function, we have in one hand :

$$\int_{[0,1]^2} f_{s,t}(x,y)dJ(x,y) = J(s,t) - J(0,0) = \frac{1}{2}|s^\lambda - t^{1-\lambda}|,$$

in other hand, using the equality $2^{-1}|x-y| = 2^{-1}(x+y) - x\wedge y$, one has to show

$$\begin{aligned}\frac{1}{2}\left|s^\lambda - t^{1-\lambda}\right| &= \frac{s^\lambda}{2} + \frac{t^{1-\lambda}}{2} - s^\lambda \wedge t^{1-\lambda}\\ &= \frac{1}{2}\int_{[0,1]} f_{s,t}(x^{\frac{1}{\lambda}}, 1)dx + \frac{1}{2}\int_{[0,1]} f_{s,t}(1, y^{\frac{1}{1-\lambda}})dy - \int_{[0,1]} f_{s,t}(x^{\frac{1}{\lambda}}, x^{\frac{1}{1-\lambda}})dx.\end{aligned}$$

Notice that the class

$$\mathcal{E} = \Big\{ A \in \mathcal{B}([0,1]^2) : \int_{[0,1]^2} \mathbb{1}_A(x,y) dJ(x,y) = \int_{[0,1]} \mathbb{1}_A(x^{\frac{1}{\lambda}}, 1) dx \\ + \int_{[0,1]} \mathbb{1}_A(1, y^{\frac{1}{1-\lambda}}) dy - \int_{[0,1]} \mathbb{1}_A(x^{\frac{1}{\lambda}}, x^{\frac{1}{1-\lambda}}) dx \Big\},$$

contains the class $\mathcal{P}$ of all closed pavements of $[0,1]^2$. It is otherwise a monotone class (or λ-system). Hence as the class $\mathcal{P}$ of closed pavement is a π-system, the class monotone theorem ensure that $\mathcal{E}$ contains the sigma-field generated by $\mathcal{P}$, that is $\mathcal{B}([0,1]^2)$.

This result holds for simple function $f(x,y) = \sum_{i=1}^n \lambda_i \mathbb{1}_{A_i}$ where $\lambda_i \in \mathbb{R}$ and $A_i \in \mathcal{B}([0,1]^2)$ for all $i \in \{1, \dots, n\}$. We then can prove the identity for positive measurable function by approximation with an increasing sequence of simple function using Beppo-Levy theorem. This identity is then extended to measurable function $f : [0,1]^2 \mapsto \mathbb{R}$ considering $f = f_+ - f_-$ with $f_+ = max(f, 0)$ and $f_- = min(-f, 0)$. We take the function bounded-measurable in order that the left hand size of the equality is well defined as a Lebesgue-Stieljes integral.

C.2 A counter example against variance's monotony with respect to an increasing positive dependence

First, notive that, under dependency condition, the variance of the λ-FMadogram evaluated in $\lambda = 0.5$ is equal to $1/150$.

Lemma C.2. *Let us consider $A(t) = 1 - \theta t + \theta t^2$ where $\theta \in [0,1]$. If we take $\lambda = 0.5$, there exist $\theta \in (0,1)$ such that*

$$Var\left(\int_{[0,1]} N_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}}) du \right) > \frac{1}{150}. \tag{C.1}$$

Proof For this dependence function, we have immediately :

$$\kappa(\lambda, A) = 1 - \theta\lambda^2, \quad \zeta(\lambda, A) = 1 - \theta(1-\lambda)^2.$$

For $\lambda = 0.5$, we notice that $\kappa(0.5, A) = \zeta(0.5, A)$. By a simple change of variable, we notice that :

$$\int_0^{0.5} [A(s)+(1-s)(2A(0.5)-0.5-1)-0.5s+1]^{-2} ds = \int_{0.5}^1 [A(s)+s(2A(0.5)-0.5-1)-0.5(1-s)+1]^{-2} ds.$$

By substitution, we have for the chosen copula that,

$$\int_{0.5}^1 [A(s) + s(2A(0.5) - 0.5 - 1) - 0.5(1-s) + 1]^{-2} ds = \int_{0.5}^1 [\frac{3}{2} - s(\theta + 1 - 2A(0.5)) + s^2\theta] ds.$$

Let us take $\theta = 2A(0.5) - 1$, which implies by direct computation that $\theta = 2/3 > 0$. Let us make of use of this lemma :

Lemma C.3. *Let a, b be two reals. Note $I_n = \int_{\mathbb{R}} (ax^2+b)^n dx$, then :*

$$I_n = \frac{2n-3}{2b(n-1)} I_{n-1} + \frac{x}{2b(n-1)(ax^2+b)}.$$

Proof An integration by parts gives and some algebra gives:

$$\begin{aligned} I_{n-1} &= \frac{x}{(ax^2+b)^{n-1}} + 2(n-1)\int_{\mathbb{R}} \frac{ax^2}{(ax^2+b)^n} dx, \\ &= \frac{x}{(ax^2+b)^{n-1}} + 2(n-1)I_{n-1} - 2b(n-1)I_n. \end{aligned}$$

Solving the equation for I_n gives the result.

We want to compute the following quantity :

$$\int_{0.5}^{1} [\frac{3}{2} + s^2\frac{2}{3}]^{-2} ds.$$

The lemma gives :

$$\begin{aligned} \int_{0.5}^{1} [\frac{3}{2} + s^2\frac{2}{3}]^{-2} ds &= 36 \int_{0.5}^{1} [4s^2+9]^{-2} ds \\ &= 2\left(\frac{7}{20} + \int_{0.5}^{1} (4s^2+9)^{-1} ds\right), \\ &= 2\left(\frac{7}{20} + \frac{1}{6}\int_{1/3}^{2/3} \frac{1}{u^2+1} du\right). \end{aligned}$$

Where we have made the substitution $u = 2s/3$ in the third line. Then :

$$\int_{0.5}^{1} [\frac{3}{2} + s^2\frac{2}{3}]^{-2} ds = 2\left[\frac{7}{20} + \frac{1}{6}(atan(2/3) - atan(1/3))\right] \approx 0.142596.$$

For the last integral, we have, by substitution for $\lambda = 0.5$ and $\theta = 2/3$:

$$\int_0^1 [A(s) + (1-s)(2A(0.5) - 0.5 - 1) + s(2A(0.5) - 0.5 - 1) + 1]^{-2} ds = \int_0^1 [\frac{13}{6} - \frac{2}{3}s + \frac{2}{3}s^2]^{-2} ds.$$

Then, we are able to compute :

$$\begin{aligned} \int_0^1 [\frac{13}{6} - \frac{2}{3}s + \frac{2}{3}s^2]^{-2} ds = 36 \int_0^1 (13 - 4s + 4s^2) ds &\overset{u=(2s-1)}{=} 36 \int_0^1 ((2s-1)^2 + 12)^2 ds, \\ = 18 \int_{-1}^{1} (u^2+12)^{-2} du &\overset{\text{Lemma}}{=} \frac{3}{4}\left(\frac{2}{13} + \int_{-1}^{1} \frac{1}{u^2+12} du\right), \\ \overset{v=u/(2\sqrt{3})}{=} \frac{6}{52} + \frac{3}{8\sqrt{3}} \int_{\frac{-1}{2\sqrt{3}}}^{\frac{1}{2\sqrt{3}}} \frac{1}{v^2+1} dv. \end{aligned}$$

$$\int_0^1 [\frac{13}{6} - \frac{2}{3}s + \frac{2}{3}s^2]^{-2} ds = \frac{\sqrt{3}}{8}\left(atan(\frac{1}{2\sqrt{3}}) - atan(-\frac{1}{2\sqrt{3}})\right) + \frac{6}{52} \approx 0.23707.$$

Summing all the components of the variance gives $Var\left(\int_{[0,1]} N_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})du\right) \approx 0.00713 > 1/150$, which gives our counterexample.