On non-central squared copulas: Properties and applications

Bouchra R. Nasri (U. de Montréal)

October 21, 2020

Introduction and motivation I

- What are copulas?
- Consider a random vector $(X_1, X_2, \ldots, X_p)$. Suppose its margins are continuous, i.e. the marginal CDFs $F_i(x) = \Pr(X_i \le x)$ are continuous functions and F is their joint cumulative distribution.
- By applying the probability integral transform to each component, the random vector $(U_1, U_2, \ldots, U_p) = (F_1(X_1), F_2(X_2), \ldots, F_p(X_p))$ has uniformly distributed margins.
- The copula of $(X_1, X_2, \ldots, X_p)$ is defined as the joint cumulative distribution function of $(U_1, U_2, \ldots, U_p)$:

$$\begin{aligned} C(u_1, u_2, \ldots, u_p) &= \Pr(U_1 \le u_1, U_2 \le u_2, \ldots, U_p \le u_p) \\ &= F\left(F_1^{-1}(u_1), \ldots, F_p^{-1}(u_p)\right) \end{aligned} \tag{1}$$

Introduction and motivation II

- The Sklar representation (Sklar, 1959) express the joint distribution of $(X_1, \ldots, X_p)$ evaluated in $(x_1, \ldots, x_p)$ in terms of C and the margins $F_1(x_1) = P(X_1 \leq x_1), \ldots, F_1(x_p) = P(X_1 \leq x_p)$, as

$$F(x_1, \cdots, x_p) = C\left\{F_1(x_1), \cdots, F_p(x_p)\right\}. \tag{2}$$

- Copulas have been used for modelling several problems which required dependence such as:
  - the occurence of joint extreme events (Genest and Favre, 2007)
  - spatial interpolation (Bárdossy and Li, 2008; Quessy et al., 2015)
  - regression modelling (Noh et al., 2013, 2015; Nasri and Rémillard, 2019; Nasri et al., 2019).

Introduction and motivation III

- In the bivariate case, set $Y = X_2$ and $G = F_2$. Then the conditional distribution of Y giving $X_1 = x_1$ is

$$\Pr(Y \leq y | X_1 = x_1) = \partial_u C(u, v)|_{u=F_1(x_1),\ v=G(y)} . \qquad (3)$$

- Equation (3) makes it possible the use copula-based regression methods.
- More precisely, $\mathrm{E}(Y|X_1 = x_1)$ is the expectation of (3) and the conditional quantile $\mathcal{Q}(\alpha, x_1)$ is the inverse of this equation.
- For example, $\mathrm{E}(Y|X_1 = x_1) = \displaystyle\int yc\{F_1(x_1), G(y)\}g(y)dy$, if the density g of G exists, where $c(u, v) = \partial_u \partial_v C(u, v)$ is the copula's density.

Introduction and motivation IV

- More generally,

$$\mathrm{E}(Y|X_1 = x_1) = \int_0^{\infty} [1 - \partial_u C\{F_1(x_1), G(y)\}]\, dy \\ - \int_{-\infty}^{0} \partial_u C\{F_1(x_1), G(y)\} dy.$$

- Similarly, $\mathcal{Q}(\alpha, x_1) = G^{-1}[\mathcal{C}\{F(x_1), \alpha\}]$, where $\mathcal{C}(u,\alpha)$ is the quantile of order α of the cumulative function $\partial_u C(u,v)$ for a fixed u, i.e., $\partial_u C(u,v)|_{u,\ v=\mathcal{C}(u,\alpha)} = \alpha$.

Copula families I

In the literature, there are several families of copulas which display various dependence structures:

- The Archimedean copulas (Clayton, Gumbel, Frank, etc.) (Genest and MacKay, 1986). In this case, the copula can be written as

$$C(u_1, \ldots, u_p, v) = \phi\left\{\phi^{-1}(u_1) + \cdots + \phi^{-1}(v)\right\},$$

  where $\phi : [0, \infty] \mapsto [0, 1]$ satisfy $\phi(0) = 1$, $(-1)^k \frac{d^k \phi(s)}{ds^k} \geq 0$ for $k \in \{0, \ldots, p-1\}$, and $(-1)^{p-1} \frac{d^{p-1} \phi(s)}{ds^{p-1}}$ is non-increasing and convex (McNeil and Nešlehová, 2009).

- $\phi(s) = (1+s)^{-1/\theta}$, $\theta > 0$ defines Clayton's copula while $\phi(s) = e^{-s^{1/\theta}}$, $\theta > 1$ defines Gumbel's copula.

Copula families II

- The elliptical copulas (Gaussian, Student, etc.) (Joe, 1997) are defined from multivariate elliptic distributions $Z_1, \ldots, Z_{p+1}$ such as the Gaussian or the Student. The copula can be written as

$$C(u_1, \ldots, u_p, v) = F_{p+1}\left\{F_1^{-1}(u_1), \ldots, F_1^{-1}(v)\right\},$$

  where F_{p+1} is the joint distribution function of $Z_1, \ldots, Z_{p+1}$ and F_1 is the common distribution function of Z_j.

- Archimedean and elliptical copulas generally give rise to monotonic dependence structures between Y and $X_1, \ldots, X_p$.
- In particular, in the bivariate case, the conditional expectation and the conditional quantiles are monotonic functions in x_1 (Nelsen, 2006; Dette et al., 2014).

Copula families III

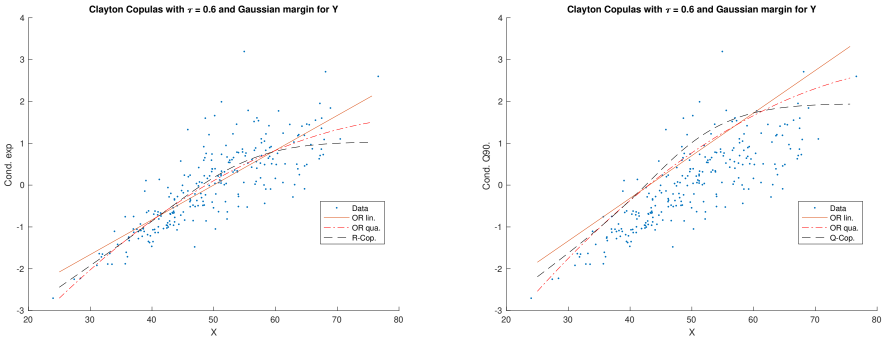

Figure 1: Estimation of the conditional expectation and the 90% conditional quantile using simulated data from Clayton's copula ($\tau = 0.6$) with normal margins for X and Y.

Dette et al. (2014) 's example I

- Problem: model the dependence between X and Y, where $Y = (X - 0.5)^2 + \epsilon_t$, with $X \sim \mathcal{U}(0,1)$ independent of $\epsilon \sim \mathcal{N}(0, \sigma^2)$, and $\sigma = 0.1$, using a copula-based mean regression model.
- Dette et al. (2014) looked for a copula family which could capture this kind of dependence. However, they limited their search to the basic families (elliptical and Archimedean), and their rotations.

Dette et al. (2014) 's example II

- However, it is shown in Nasri et al. (2019) that there are also many copula families for which $\mathrm{E}(Y|X=x)$ is not monotonic in x.
- Here is an example: consider the chi-square copula introduced by Bárdossy (2006); see also Quessy et al. (2016).
- A chi-square copula with parameters $a_1, a_2 \in [0, \infty)$, and $\rho \in [-1, 1]$ is the copula associated with the random variables $(Z_1 + a_1)^2$ and $(Z_2 + a_2)^2$, where Z_1 and Z_2 are joint standard Gaussian variables with correlation ρ.
- In Bárdossy (2006) and Quessy et al. (2016), the authors focused mainly on the case of $a_1 = a_2$ because they were interested in spatial interpolation.

Dette et al. (2014) 's example III

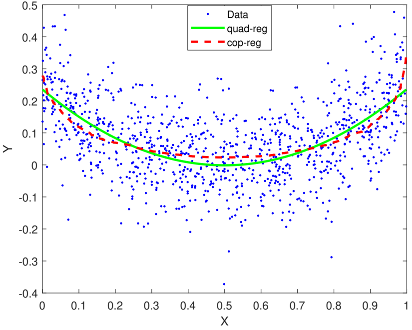

Figure 2: Quadratic regression curve (red) and conditional expectation estimator (green) for the simulated dataset. Here we used the chi-square copula with the parameters (rounded), we found $a_1 = 2.6$, $a_2 = 0$, and $\rho = 0.99$ Nasri et al. (2019).

Non-central squared copulas I

- For a given copulas C, let $(U_1, \ldots, U_p) \sim C$, and apply the normal transformation $Z_1 = \Phi^{-1}(U_1), \ldots, Z_p = \Phi^{-1}(U_p)$.
- The non-central squared copula $\tilde{C}_{\mathrm{a}}$, $\mathrm{a} = (a_1, \ldots, a_p) \in [0, \infty)^p$, is the copula associated with the random vector $(Z_1 + a_1)^2, \ldots, (Z_p + a_p)^2$.
- $\tilde{C}_{\mathrm{a}}$ is also the the copula associated with the random vector $|Z_1 + a_1|, \ldots, |Z_p + a_p|$.
- For every $u \in (-1, 1)$, one can define $\tilde{h}_a(u) = \Phi\{h_a(u)\}$, with $h_a(u) = \mathrm{sign}(u) G_a^{-1}(|u|) - a$, and $G_a(x) = P\{|Z + a| \le x\} = \Phi(x - a) - \Phi(-x - a)$, $x \ge 0$.

Non-central squared copulas II

- The non-central squared copula $\tilde{C}_a$ is then given by

$$\begin{aligned}\tilde{C}_a(u_1,\ldots,u_p) &= P\left[\cap_{j=1}^{p}\{\tilde{h}_{a_j}(-u_j) < U_j \le \tilde{h}_{a_j}(u_j)\}\right] \\ &= \sum_{(\epsilon_1,\ldots,\epsilon_p)\in\{-1,1\}^p}\left(\prod_{j=1}^{p}\epsilon_j\right) C\left\{\tilde{h}_{a_1}(\epsilon_1 u_1),\ldots,\tilde{h}_{a_p}(\epsilon_p u_p)\right\}.\end{aligned}$$

- When $p=2$,

$$\begin{aligned}\tilde{C}_a(u_1,u_2) &= C\left\{\tilde{h}_{a_1}(u_1),\tilde{h}_{a_2}(u_2)\right\} - C\left\{\tilde{h}_{a_1}(-u_1),\tilde{h}_{a_2}(u_2)\right\} \\ &\quad - C\left\{\tilde{h}_{a_1}(u_1),\tilde{h}_{a_2}(-u_2)\right\} + C\left\{\tilde{h}_{a_1}(-u_1),\tilde{h}_{a_2}(-u_2)\right\}.\end{aligned}$$

Non-central squared copulas III

- Usually Archimedean copulas are not used for multivariate data since it imposes that all pairs have the same distribution. This is not the case here for $\tilde{C}_{\mathrm{a}}$ even if C is Archimedean, unless $a_1 = \cdots = a_p$.
- If C is the Gaussian copula, then $\tilde{C}$ is the chi-square copula (Quessy et al., 2016).
- If C is the Student copula, then $\tilde{C}_0$ is called the Fisher copula (Favre et al., 2018).
- If C is general and $a_1 = \cdots = a_p = 0$, then one gets the squared copula introduced in Quessy and Durocher (2019).

Non-central squared copulas IV

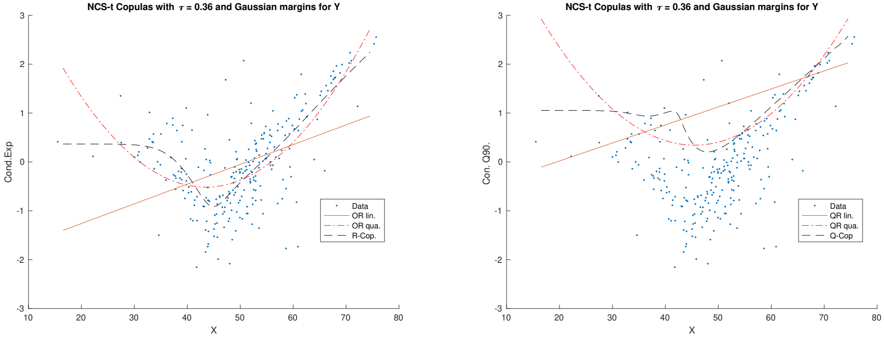

Figure 3: Estimation of the conditional expectation and the 90% conditional quantile using simulated data from NCS-t's copula ($\tau = 0.36$) with normal margins for X and Y.

Non-central squared copulas: limiting behaviors of $\tilde{C}_{\mathrm{a}}$ I

- Recall that the non-central squared copula $\tilde{C}_{\mathrm{a}}$, $\mathrm{a} = (a_1, \ldots, a_p) \in [0, \infty)^p$, is the copula of the random vector $(Z_1 + a_1)^2, \ldots, (Z_p + a_p)^2$.
- By assuming $a_1, \ldots, a_p > 0$, $\tilde{C}_{\mathrm{a}}$ is the copula associated with the random variables $Z_1 + \frac{Z_1^2}{2a_1}$, ..., $Z_p + \frac{Z_p^2}{2a_p}$.
- One can see that by letting $a_1 \to \infty, \ldots, a_p \to \infty$, $\tilde{C}_{\mathrm{a}} \to C$.
- By assuming $a_1, \ldots, a_{p-1} = 0$, and letting $a_p \to \infty$, $\tilde{C}_{\mathrm{a}}$ is the copula associated with the random variables Z_1^2, ..., Z_{p-1}^2 and Z_p.

Non-central squared copulas: limiting behaviors of $\tilde{C}_a$ II

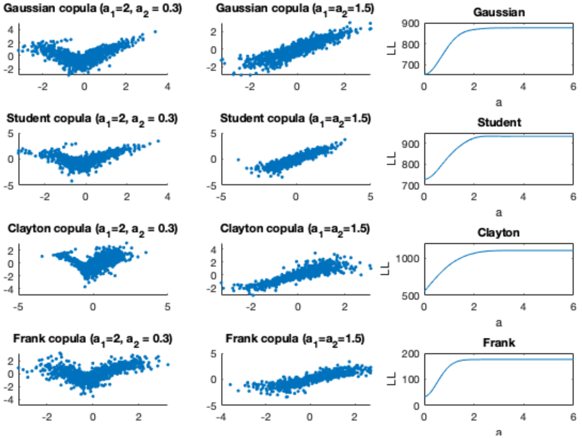

Figure 4: Scatterplots of random samples of size $n = 1000$ from $\tilde{C}_{a_1,a_2}$ with standard Gaussian margins ($\tau = 0.75$, $\nu = 12$). The graph of the pseudo log-likelihood for random samples of size $n = 1000$ from $\tilde{C}_{a,a}$ as a function of $a \in (0, 6]$ are displayed in column (c), where the true values of the non-centrality parameters are $a_1 = a_2 = 4$.

Non-central squared copulas: limiting behaviors of $\tilde{C}_{\mathrm{a}}$ III

- Set $\mathrm{a} = (a_1, \ldots, a_p) \in [0, \infty)^p$ and suppose that $a_j \geq b$ for all $j \in \{1, \ldots, k\}$, with $1 \leq k \leq p$. Then

$$\sup_{\mathrm{u} \in [0,1]^p} \left| \tilde{C}_{\mathrm{a}}(\mathrm{u}) - \tilde{C}_{\infty,\ldots,\infty,a_{k+1},\ldots,a_p}(\mathrm{u}) \right| \leq k\Phi(-b).$$

- In particular, if $a_j \geq b$, for all $j \in \{1, \ldots, p\}$, then

$$\sup_{\mathrm{u} \in [0,1]^p} \left| \tilde{C}_{a_1,\ldots,a_p}(\mathrm{u}) - C(\mathrm{u}) \right| \leq p\Phi(-b).$$

Non-central squared copulas: limiting behaviors of $\tilde{C}_{\mathrm{a}}$ IV

- For example, in the bivariate case, when $b = 3$, $\Phi(-3) = 0.0013$, so
  - $\sup_{\mathrm{u}\in[0,1]^2} \left| \tilde{C}_{\mathrm{a}}(\mathrm{u}) - \tilde{C}_{\infty,a_2}(\mathrm{u}) \right| \le .0013$, if $\mathrm{a} \in [3,\infty) \times [0,3]$,
  - $\sup_{\mathrm{u}\in[0,1]^2} \left| \tilde{C}_{\mathrm{a}}(\mathrm{u}) - \tilde{C}_{a_1,\infty}(\mathrm{u}) \right| \le .0013$, if $\mathrm{a} \in [0,3] \times [3,\infty)$,

    and
  - $\sup_{\mathrm{u}\in[0,1]^2} \left| \tilde{C}_{\mathrm{a}}(\mathrm{u}) - C(\mathrm{u}) \right| \le .0026, \quad$ if $\mathrm{a} \in [3,\infty)^2$.
- Therefore, in the bivariate case, the estimation of the parameters should be restricted to $[0,3]^2$.

Non-central squared copulas: dependence measures I

- In the bivariate case, one can investigate the behavior of two measures of dependence, namely Kendall's tau and Spearman's rho in terms of the non-centrality parameters a_1, a_2.
- For a given bivariate copula D, recall that

$$\tau(D) = -1 + 4 \int_{(0,1)^2} D(u_1, u_2) dD(u_1, u_2), \tag{4}$$

and

$$\rho_S(D) = 12 \int_{(0,1)^2} \{D(u_1, u_2) - u_1 u_2\} du_1 du_2, \tag{5}$$

where $(U_1, U_2) \sim D$.

Non-central squared copulas: dependence measures II

- Quessy et al. (2016) gives an explicit formula for $\tau(\tilde{C}_{a_1,a_2})$ where C is the Gaussian copula, but it is impossible to get an expression for $\tau(\tilde{C}_{a_1,a_2})$ in the general case.
- One can use numerical integration to compute values for both $\tau(\tilde{C}_{a_1,a_2})$ and $\rho_S(\tilde{C}_{a_1,a_2})$.

Non-central squared copulas: dependence measures III

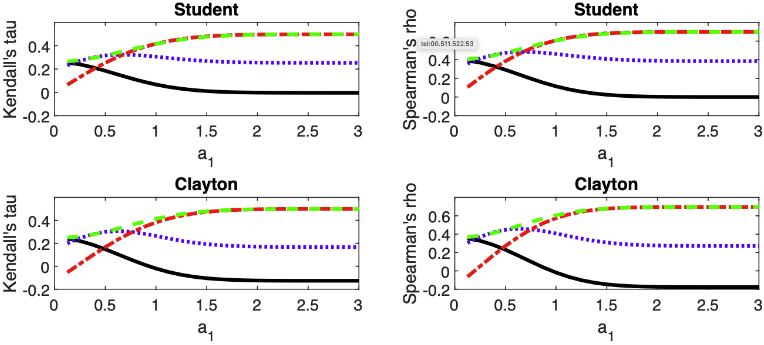

Figure 5: Graph of Kendall's tau and Spearman's rho for $\tilde{C}_a$ when the copula C is the Student copula and Clayton copula with Kendall's tau 0.5, as a function of $a_1 \in [0, 3]$, for $a_2 = 0$ (black line), $a_2 = 0.5$ (blue line), $a_2 = 2.5$ (red line), and $a_2 = a_1$ (green line).

Non-central squared copulas: dependence measures IV

- One can notice that the values of Kendall's tau and Spearman's rho seem to be zero for the Frank copula which is true also for the Student and the Gaussian copulas.
- This is indeed true for all copulas which are invariant with respect to the 180 rotation.
- If $C_{180} = C$, then $\tau\left(\tilde{C}_{\infty,0}\right) = \rho_S\left(\tilde{C}_{\infty,0}\right) = 0$.

Non-central squared copulas: parameters' estimation (Genest et al., 1995) I

Table 1: Relative RMSE and bias (in parenthesis) in percentage for the estimation of the parameters (a_1, a_2, τ) for $\tilde{C}_{a_1,a_2}$ when the copula C is Gaussian and Clayton family with $\tau = 0.5$

(a_1, a_2)	$n = 250$			$n = 500$		
	a_1	a_2	τ	a_1	a_2	τ
	Gaussian					
(0.5, 1.5)	48.0(−4.3)	35.5(−0.1)	12.8(4.2)	23.6(−4.9)	26.2(0.0)	8.8(2.3)
(0.5, 2.5)	35.7(−2.1)	24.8(−9.7)	9.7(3.1)	13.0(−1.5)	17.9(−2.3)	5.7(1.7)
(1.0, 2.0)	40.4(1.0)	35.2(−21.9)	11.9(4.2)	24.8(−0.3)	30.0(−16.4)	7.5(2.1)
(1.5, 2.0)	45.2(21.4)	22.7(−7.2)	7.2(1.7)	36.7(16.7)	19.0(−6.3)	4.8(0.6)
(1.5, 2.5)	42.6(18.6)	31.8(−26.8)	7.0(1.6)	37.2(16.7)	29.6(−26.4)	5.0(0.7)
(2.0, 2.5)	37.4(23.0)	23.0(−11.7)	6.8(2.0)	36.6(26.1)	19.8(−9.4)	4.4(0.7)
	Clayton					
(0.5, 1.5)	19.5(−0.4)	12.6(1.0)	8.9(0.5)	13.2(0.0)	7.2(0.2)	6.1(0.3)
(0.5, 2.5)	22.2(−2.1)	13.9(−1.2)	7.0(1.2)	10.7(−1.1)	8.1(0.2)	4.8(0.4)
(1.0, 2.0)	12.8(−2.7)	16.7(−4.9)	7.8(2.4)	7.2(−0.9)	9.6(−0.9)	5.3(0.7)
(1.5, 2.0)	18.7(1.6)	21.8(−17.0)	8.2(1.8)	13.6(1.0)	20.5(−15.9)	5.7(0.9)
(1.5, 2.5)	17.0(−1.8)	37.2(−32.6)	8.1(2.4)	12.5(−1.4)	38.4(−34.3)	5.5(1.5)
(2.0, 2.5)	18.7(6.5)	22.8(−19.2)	6.8(0.9)	14.8(5.0)	21.6(−18.8)	4.8(0.5)

Non-central squared copulas: parameters' estimation (Genest et al., 1995) II

Table 2: Relative RMSE and bias (in parenthesis) in percentage for the estimation of the parameters (a_1, a_2, τ) for $\tilde{C}_{a_1,a_2}$ when the copula C is Frank, and Gumbel family with $\tau = 0.5$.

(a_1, a_2)	$n = 250$			$n = 500$		
	a_1	a_2	τ	a_1	a_2	τ
	Frank					
(0.5, 1.5)	47.4(−2.2)	47.0(9.4)	12.9(3.8)	19.9(−5.0)	45.2(10.7)	9.2(2.9)
(0.5, 2.5)	39.8(−2.5)	38.2(−24.4)	11.2(3.3)	15.8(−3.2)	34.0(−20.2)	7.6(2.4)
(1.0, 2.0)	69.3(23.3)	41.3(−13.7)	13.0(4.5)	65.1(24.0)	38.7(−13.6)	8.6(2.0)
(1.5, 2.0)	61.9(32.8)	35.7(2.2)	9.2(2.9)	58.8(32.8)	32.6(2.4)	6.1(1.5)
(1.5, 2.5)	61.4(32.6)	33.7(−19.5)	8.8(2.6)	61.6(37.2)	30.8(−16.4)	5.6(1.1)
(2.0, 2.5)	41.0(15.0)	29.8(−14.0)	8.2(2.6)	40.8(21.6)	26.7(−10.3)	5.1(1.3)
	Gumbel					
(0.5, 1.5)	35.0(−4.8)	38.1(1.6)	10.9(3.7)	20.2(−2.9)	34.6(4.0)	7.8(1.7)
(0.5, 2.5)	20.9(−3.2)	30.4(−15.9)	9.5(3.2)	14.9(−1.0)	24.1(−9.6)	6.3(2.2)
(1.0, 2.0)	65.6(17.1)	40.5(−21.7)	12.6(4.3)	53.1(14.0)	36.8(−19.9)	8.8(2.3)
(1.5, 2.0)	65.9(38.6)	30.2(−4.7)	8.5(2.0)	57.9(33.3)	28.0(−2.7)	5.5(0.8)
(1.5, 2.5)	65.4(37.3)	33.5(−22.9)	8.4(2.4)	57.9(33.6)	30.9(−21.8)	5.4(0.6)
(2.0, 2.5)	43.0(27.2)	27.3(−14.1)	7.7(2.1)	43.0(32.1)	23.7(−11.3)	4.8(0.9)

Conclusion I

- In this presentation, new families of multivariate copulas were introduced, extending the chi-square copula and the Fisher copula.
- By varying the non-centrality parameters, one can model non-monotonic dependence.
- The limiting behavior of these copulas was shown as well as the behavior of some dependence measures.
- The estimation of parameters was discussed.

Conclusion II

- This work is already published in *Statistics and Probability Letters* (Nasri, 2020). All numerical computations can be done by using the R package NCSCopula available on CRAN (4666 download) `https://cran.r-project.org/web/packages/NCSCopula`.
- In the paper you can also find some extra calculations, more precisely the tail indexes for these families of copulas and the solution of a conjecture formulated by Quessy et al. (2016) for the chi-square copulas when $a_1 = a_2 > 0$

References I

Bárdossy, A. (2006). Copula-based geostatistical models for groundwater quality parameters. *Water Resources Research*, 42(11).

Bárdossy, A. and Li, J. (2008). Geostatistical interpolation using copulas. *Water Resources Research*, 44(7).

Dette, H., Van Hecke, R., and Volgushev, S. (2014). Some comments on copula-based regression. *Journal of the American Statistical Association*, 109(507):1319–1324.

Favre, A.-C., Quessy, J.-F., and Toupin, M.-H. (2018). The new family of Fisher copulas to model upper tail dependence and radial asymmetry: Properties and application to high-dimensional rainfall data. *Environmetrics*, 29(3):e2494.

Genest, C. and Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12:in press.

References II

Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82:543–552.

Genest, C. and MacKay, R. J. (1986). Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *The Canadian Journal of Statistics*, 14(2):145–159.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*, volume 73 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

McNeil, A. J. and Nešlehová, J. (2009). Multivariate Archimedean copulas, d-monotone functions and l_1-norm symmetric distributions. *Ann. Statist.*, 37(5B):3059–3097.

Nasri, B. R. (2020). On non-central squared copulas. *Statistics & Probability Letters*, 161:108704.

References III

Nasri, B. R. and Rémillard, B. N. (2019). Copula-based dynamic models for multivariate time series. *Journal of Multivariate Analysis*, 172:107–121.

Nasri, B. R., Rémillard, B. N., and Bouezmarni, T. (2019). Semi-parametric copula-based models under non-stationarity. *Journal of Multivariate Analysis*, 173:347–365.

Nelsen, R. B. (2006). *An Introduction to Copulas*, volume 139 of *Lecture Notes in Statistics*. Springer-Verlag, New York, second edition.

Noh, H., Ghouch, A. E., and Bouezmarni, T. (2013). Copula-based regression estimation and inference. *Journal of the American Statistical Association*, 108(502):676–688.

Noh, H., Ghouch, A. E., and Keilegom, I. V. (2015). Semiparametric conditional quantile estimation through copula-based multivariate models. *Journal of Business & Economic Statistics*, 33(2):167–178.

References IV

Quessy, J.-F. and Durocher, M. (2019). The class of copulas arising from squared distributions: properties and inference. *Econometrics and Statistics*, 12:148–166.

Quessy, J.-F., Rivest, L.-P., and Toupin, M.-H. (2015). Semi-parametric pairwise inference methods in spatial models based on copulas. *Spatial Statistics*, 14:472 – 490.

Quessy, J.-F., Rivest, L.-P., and Toupin, M.-H. (2016). On the family of multivariate chi-square copulas. *Journal of Multivariate Analysis*, 152:40–60.

Sklar, M. (1959). Fonctions de répartition à N dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231.