



A Weighted Entropic Copula from Preliminary Knowledge of Dependence

Ioana Panait

Abstract

This paper introduces a weighted entropic copula from preliminary knowledge of dependence. Considering a copula with common distribution we formulate the weighted entropy dependence model (WMEC). We give an approximator for the copula function of this problem. Also, we discuss some asymptotical properties regarding the unknown parameters of the model.

1 Introduction

”The copula function methodology has become the most significant new technique to handle the co-movement between markets and risk factors in a flexible way” said Cherubini et. al in the introduction of their book *Copula Methods in Finance* from 2004 [8].

Copula function has emerged as a useful tool for modeling stochastic dependence. These functions can model and explain the asymmetric dependence between random variables without looking at their marginal distributions- this being one of the main advantages of copula over density function. In essence, a copula is a multidimensional probability distribution with uniform marginals. For these reasons copula functions have been studied intensely over the years since their discovery: Embrechts, P., Lindskog, F., McNeil, A.J., [15]; Kolve,

Key Words: maximum entropy principle, copula function, weighted entropy.
2010 Mathematics Subject Classification: Primary 62H05; Secondary 62H20.
Received: 28.12.2016
Revised: 23.06.2017
Accepted: 29.06.2017

N., dos Anjos, U., Mendes, B. [22]; Mikosch [24]; Cherubini, U., Luciano, E., Vecchiato, W. [8] and not only.

The last decade has seen an explosion of manuscripts on the applications of copulas, especially to financial issues. These applications include modelling financial contagion [31], portofolio selection using a multivariate regime-switching copula to capture asymmetric dependence [12] or studies involving asymmetric pattern in volatility [26] and not only.

Among the many densities in statistical literature [5, 27], one special case of densities represent the maximum entropy densities first proposed by Jaynes (1957) [20]. These densities are obtained by maximization of an information criterion subject to mass and mean preserving constraints, denoted by *ME*. The principle of maximum entropy can be viewed as a criterion to select the best probability distributions compatible to some set of constraints. Applications of *ME* can be found in various fields of computer science, in statistical learning, especially natural language processing [13, 17]. The *ME* method was applied to obtain new classes of Lorenz curves by maximizing Tsallis entropy under mean and Gini equality and inequality constraints [28]. For more details on entropy optimization we suggest the reader to see [29].

The foundation of information theory was laid out by Shannon's 1948 paper where he introduced the well-known Shannon entropy which gives us the level of uncertainty of a random variable. Since then the field of information theory has exploded and many more entropies have been introduced: Tsallis entropy [34, 35, 36] which is along with Tsallis statistics used in various applications [33], weighted entropy introduced by Bellis and Guiaşu [6, 16] as a measure of useful information contained by a random variable, by taking into account both the probabilities and the values of the random variables, weighted cumulative paired interval entropy [4, 30] and many more. Some applications of the concept of entropy can be found in [1, 2, 3] where it has been used in the context of divergence rates for Markov chains, and in [27] in determining the level of uncertainty of statistical models used in reliability. For more details regarding information theory see [11].

In 2007, Dempster et. al [14] have discussed the problem of constructing empirical copulas from relative entropy. Some applications of maximum entropy copula can be found in [18] where it has been used to model multi-site streamflow dependence and in [21]. By applying the *ME* principle to copula functions, Chu and Satchell [10] have proposed a maximum entropy copula obtained by maximizing the bivariate Shannon entropy. Having limited information, they proposed a theoretical problem in order to obtain relative entropy measures of joint dependence. The constraints considered concerns to the need for uniform marginal distributions and to measures of association and rank correlations.

Considering the context described above we introduce a weighted entropy dependence model using copula functions.

The paper is organized as follows. In Section 2 we introduce the weighted entropy dependence model while in Section 3 we give an approximator for the copula function of this model. Section 4 is dedicated to the asymptotic properties of the unknown parameters the of *WMEC* model. Section 5 concludes this paper.

2 Weighted entropy dependence model

The study of copula functions and the role they play in probability theory and mathematical statistics is a subject still in its infancy. There are still many open issues and many things to discover.

In recent years, several studies have been conducted in the field of financial economics, as they have the potential to model and explain asymmetric dependence between random variables, separate from marginal distributions.

The copula is proposed by Sklar [32] as a method of constructing common distributions with given marginals. The advantage of copula functions is that the dependence of random variables can be parametrically specified, regardless of their marginals.

A bivariate copula $C(\cdot, \cdot)$ is defined over $[0, 1]^2$ with values in $[0, 1]$ and has the following properties [25]:

1. $C(x, 0) = C(0, y) = 0$, $C(x, 1) = x$ and $C(1, y) = y$ for every $x, y \in [0, 1]$
2. $C(x, y)$ is 2-increasing, i.e., $C(x_2, y_2) - C(x_2, y_1) - C(x_1, y_2) + C(x_1, y_1) \geq 0$ for every $x_1, x_2, y_1, y_2 \in [0, 1]$ such that $x_1 \leq x_2$ and $y_1 \leq y_2$

The Sklar's theorem links a $C(u, v)$ copula of a common distribution $F(x, y)$ by the relation $F(x, y) = C(F_1(x), F_2(y))$, where F_1 and F_2 are marginals. We have

$$f(x, y) = c(F_1(x), F_2(y))f_1(x)f_2(y); \quad c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v} \quad (1)$$

Using the ME method we introduce a new maximum entropy copula problem. In 2016, Chu and Satchell [10] constructed a maximum entropy copula (MEC) by maximizing the bivariate Shannon entropy under some set of constraints. The problem proposed by them is

ProblemEM : maximize $W(c) = - \int_0^1 \int_0^1 c(u, v) \log c(u, v) du dv$ (2)

subject to

$$\int_0^1 \int_0^1 c(u, v) du dv = 1 \tag{3}$$

$$\int_0^u \int_0^1 c(x, v) dx dv = u, \quad \forall u \in [0, 1] \tag{4}$$

$$\int_0^1 \int_0^v c(u, y) du dy = v, \quad \forall v \in [0, 1] \tag{5}$$

$$\int_0^1 \int_0^1 h(u, v; \hat{\theta}_N) c(u, v) du dv = 0 \tag{6}$$

where $h(u, v; \hat{\theta}_N)$ is an arbitrary function used to impose constraints concerning the copula function with respect to measures of association and rank correlations [10].

Considering the problem above and how it was constructed, we introduce a new dependence model defined by the optimization problem:

Problem WME

$$\min_f E_{(X,Y)} \left\{ w(F_1(X), F_2(Y)) \left[\frac{1}{f_1(X)f_2(Y)} \log \frac{f(X, Y)}{f_1(X)f_2(Y)} - \frac{1}{f_1(X)f_2(Y)} + \frac{1}{f(X, Y)} \right] \right\} \tag{7}$$

s.t.

$$E_{(X,Y)}[a(X, Y)] = \mu_0$$

where f is a two-dimensional density, f_1 and f_2 are unidimensional densities corresponding to the random variables X and Y , respectively, and \log is the natural logarithm, $supp(f_1) = \{x \in R : f_1(x) \neq 0\}$, $supp(f_2) = \{x \in R : f_2(x) \neq 0\}$, and a is an arbitrary function so $\mu_0 < \infty$. Function $w(u, v)$ is a positive and bounded weight function. Also, we suppose that the expectations exist.

The main goal of this paper is solving problem (7). The solution is a weighed maximum entropy copula, denoted by *WMEC*. We assume for the rest of the paper that *WMEC* is a differentiable function so that its copula density exists.

From Sklar’s theorem, the optimized function reduces to:

$$CH^w(c) = -E_{(U,V)} \{ w(U, V) [\log c(U, V) - 1 + c(U, V)^{-1}] \} \tag{8}$$

where (U, V) is a random vector on $[0, 1]^2$ with $c(u, v)$ probability density function.

Taking into account the optimized reduced function, the *WME* problem can be reformulated in the following manner:

$$Max_c CH^w(c) = -E_{(U,V)} \{w(U, V)[\log c(U, V) - 1 + c(U, V)^{-1}]\}, \quad (9)$$

with restrictions

$$E_{(U,V)}[I(U \in [0, 1], V \in [0, 1])] = 1 \quad (10)$$

$$E_{(U,V)}[I(U \in [0, u], V \in [0, 1])] = u, \quad (\forall) u \in [0, 1], \quad (11)$$

$$E_{(U,V)}[I(U \in [0, 1], V \in [0, v])] = v, \quad (\forall) v \in [0, 1], \quad (12)$$

$$E_{(U,V)}[a(U, V; \hat{\lambda}_s)] = 0. \quad (13)$$

where $E_{(U,V)}$ is calculated in relation to the density of the random vector (U, V) , that is to say with respect to $c(u, v)$, here I being the indicator function.

Relation (10) implies that $c(u, v)$ is a joint density on the unit circle. Relations (11) and (12) imply that the marginals of $c(u, v)$ are $U_{[0,1]}$ distributions, i.e. uniformly distributed on the interval $[0, 1]$. Relation (13) imposes a constraint on the joint behaviour of U and V . This constraint concerns measures of association and rank correlations. To be observed that we can have more than one constraint like (13).

A measure of association is defined as $\rho = \int_0^1 \int_0^1 a(u, v)dC(u, v)$, where a is a bivariate function such that $|\rho| < \infty$. This measure is also referred as *copula-based measure of dependence*. This measure of association can be

estimated by the rank statistic $\hat{\rho} = \frac{1}{N} \sum_{i=1}^N a(\frac{R_i}{N}, \frac{S_i}{N})$, where (R_i, S_i) represents the ranks of (X_i, Y_i) in a sample of size N . For example, if we have constraints concerning Spearman correlation, we take $a(u, v; \hat{\lambda}_s) = a(u, v) = uv$, $u, v \in [0, 1]$. Throughout this paper, we sometimes omit s for brevity - we write $a(u, v; \lambda)$.

The following lemmas we use in the next section where we give an approximator for *WMEC*.

Lemma 1. [9] *We consider*

$$\begin{aligned} E_{(U,V)} \{I(U \in [0, 1], V \in [a, b])\} &= \beta - \alpha \\ E_{(U,V)} \{I(U \in [a, b], V \in [0, 1])\} &= \beta - \alpha, \end{aligned} \quad (14)$$

where α and β are arbitrary numbers from $[0, 1]$, $\alpha < \beta$ and $a < b$.

Using a dyadic sequence from $[0, 1]$, (14) is equivalent to:

$$\begin{aligned} E_{(U,V)} \{I(U \in [j \cdot 2^{-m}, (j + 1) \cdot 2^{-m}], V \in [0, 1])\} &= \\ E_{(U,V)} \{I(U \in [0, 1], V \in [j \cdot 2^{-m}, (j + 1) \cdot 2^{-m}])\} &= \frac{1}{2^m} \end{aligned} \quad (15)$$

(\forall) $j = 0, 1, \dots, (2^m - 1)$ and m is large enough.

Lemma 2. [19] [Du Bois- Reymond] *Let $b(t)$ be a continuous function on the interval $[t_0, t_1]$. Suppose that the following equality holds for any continuous function $v(t)$ with mean zero value (i.e. $\int_{t_0}^{t_1} v(t)dt = 0$):*

$$\int_{t_0}^{t_1} b(t)v(t)dt = 0.$$

Then, $b(t) = b = \text{const}$. Conversely, if $b(t) = \text{const}$, then $\int_{t_0}^{t_1} b(t)v(t)dt = 0$.

Lemma 3. [23] *The indicator function $\mathbf{I}_{(x, \infty)}(y)$ can be approximated by a continuous function $\Phi_s(y, x)$, defined as:*

$$\Phi_s(y, x) = \frac{s}{\sqrt{2\pi}} \int_{-\infty}^y \exp\{-(v - x)^2 s^2 / 2\} dv.$$

$\Phi_s(y, x)$ has the following properties:

$$\begin{aligned} \lim_{s \rightarrow \infty} \Phi_s(y, x) &\implies \mathbf{I}_{(x, \infty)}(y), \\ \lim_{s \rightarrow \infty} \frac{\partial \Phi_s(y, x)}{\partial y} &\implies \delta(y - x), \end{aligned}$$

where $\delta(\cdot)$ is Dirac's delta function.

3 An approximator for WMEC model

In this section we give an approximator for the copula function of the weighted maximum entropy problem (*WME*). In order to obtain the approximator of *WMEC* we use definite integrals.

Theorem 4. *WMEC*, $\widehat{c}(u, v)$ can be approximated by an approximator, $\widehat{c}_{m,n}^w(u, v)$, for m, n large enough, with

$$\widehat{c}_{m,n}^w(u, v) = B_{m,n}(u, v) \{E_{(Z_1, Z_2)}[B_{m,n}(Z_1, Z_2)]\}^{-1}, \quad (16)$$

where

$$(Z_1, Z_2) \sim U_{[0,1]^2}, \text{ i.e. } (Z_1, Z_2) \text{ is bivariate uniform distributed on } [0, 1]^2$$

and

$$B_{m,n} = \exp\left\{-\sum_{j=0}^{2^m-1} [\widehat{\mu}_{1,j}(\phi(n(j \cdot 2^{-m} - u)) + \phi(-n((j+1)2^{-m} - u)))+\right.$$

$$\left. + \widehat{\mu}_{2,j}(\phi(n(j \cdot 2^{-m} - v)) + \phi(-n((j+1) \cdot 2^{-m} - v)))] - \widehat{\mu}_{1,2^m} a(u, v; \widehat{\lambda}) - \alpha_0 \widehat{c}(u, v)\right\},$$

with

$$\widehat{\theta}_m = \{\widehat{\mu}_{1,0}, \dots, \widehat{\mu}_{1,2^m-1}, \widehat{\mu}_{2,0}, \dots, \widehat{\mu}_{2,2^m-1}\}$$

which contains the minimum values of the following potential function:

$$D_{m,n}(\widehat{\theta}_m, \widehat{\lambda}) = E_{(Z_1, Z_2)} \exp \left\{ -\sum_{j=0}^{2^m-1} (w(Z_1, Z_2))^{-1} \left[\widehat{\mu}_{1,j}(\phi(n(j \cdot 2^{-m} - Z_1)) \right. \right. \\ \left. \left. + \phi(-n((j+1)2^{-m} - Z_1)) - 1 + 2^{-n} \right) \right. \\ \left. + \widehat{\mu}_{2,j}(\phi(n(j \cdot 2^{-m} - Z_2)) + \phi(-n((j+1) \cdot 2^{-m} - Z_2)) - 1 + 2^{-m}) \right] \\ \left. - \widehat{\mu}_{1,2^m} a(Z_1, Z_2; \widehat{\lambda}) - \alpha_0 \widehat{c}(Z_1, Z_2) \right\}$$

for given α_0 and $\widehat{c}(u, v)$, where $\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\{-\frac{1}{2}y^2\} dy$.

Proof. Using Lemma 1, the Lagrangian function associated with the *WME* problem can be expressed as follows:

$$\begin{aligned}
 L(c, \theta_m; \widehat{\lambda}) &= \\
 &= -E_{(U,V)} \{w(U, V)[\log c(U, V) - 1 + c(U, V)^{-1}] - \mu_{1,-1} \\
 &\times [E_{(U,V)} \{I(U \in [0, 1], V \in [0, 1])\} - 1] \\
 &- \sum_{j=0}^{2^m-1} \{\mu_{1,j} E_{(Z_1, Z_2)} \{I(Z_1 \in [j \cdot 2^{-m}, (j+1) \cdot 2^{-m}], Z_2 \in [0, 1]) \cdot [c(Z_1, Z_2) - 2^{-m}]\} \\
 &+ \mu_{2,j} E_{(Z_1, Z_2)} \{I(Z_1 \in [0, 1], Z_2 \in [j \cdot 2^{-m}, (j+1) \cdot 2^{-m}]) \cdot [c(Z_1, Z_2) - 2^{-m}]\} \\
 &- \mu_{1,2^m} \cdot E_{(Z_1, Z_2)} a(Z_1, Z_2; \widehat{\lambda}) \\
 &= E_{(Z_1, Z_2)} \{w(Z_1, Z_2)[\log c(Z_1, Z_2) - 1 + c(Z_1, Z_2)^{-1}] + \mu_{1,-1}[c(Z_1, Z_2) - 1] \\
 &+ \sum_{j=0}^{2^m-1} \{\mu_{1,j} \cdot I(Z_1 \in [j \cdot 2^{-m}, (j+1) \cdot 2^{-m}]) \\
 &+ \mu_{2,j} \cdot I(Z_2 \in [j \cdot 2^{-m}, (j+1) \cdot 2^{-m}])\} \cdot [c(Z_1, Z_2) - 2^{-m}] \\
 &+ \mu_{1,2^m} a(Z_1, Z_2; \widehat{\lambda}) \cdot c(Z_1, Z_2)\}.
 \end{aligned}$$

where $(Z_1, Z_2) \sim U_{[0,1]^2}$, i.e. (Z_1, Z_2) is bivariate uniform distributed.

Taking the first derivative of Lagrange function $L(c, \theta_m, \widehat{\lambda})$ with respect to c we get:

$$\begin{aligned}
 E_{(Z_1, Z_2)} \left\{ w(Z_1, Z_2) \log c(Z_1, Z_2) + \mu_{1,-1} + \sum_{j=0}^{2^m-1} [\mu_{1,j} I(Z_1 \in [j \cdot 2^{-m}, (j+1) \cdot 2^{-m}]) + \right. \\
 \left. + \mu_{2,j} I(Z_2 \in [j \cdot 2^{-m}, (j+1) \cdot 2^{-m}])] + \mu_{1,2^m} a(Z_1, Z_2; \widehat{\lambda}) \right\},
 \end{aligned}$$

Define

$$\begin{aligned}
 d_n(u, v) &= w(u, v) \log c(u, v) + \mu_{1,-1} + \sum_{j=0}^{2^m-1} \{\mu_{1,j} I(u \in [j \cdot 2^{-m}, (j+1) \cdot 2^{-m}]) + \\
 &+ \mu_{2,j} I(v \in [j \cdot 2^{-m}, (j+1) \cdot 2^{-m}])\} + \mu_{1,2^m} a(u, v; \widehat{\lambda}),
 \end{aligned}$$

then applying Lemma 2 to the function

$$d(u, v) = \frac{d_n(u, v)}{\widetilde{c}(u, v) - 1},$$

where $\tilde{c}(u, v)$ is an arbitrary copula density so that,

$$E_{(Z_1, Z_2)}[\tilde{c}(Z_1, Z_2) - 1] = 0$$

we obtain the following representation:

$$\begin{aligned} \hat{c}_{m,n}(u, v) = & \exp\{-(\mu_{1,-1} - \alpha_0) - \sum_{j=0}^{2^m-1} \{\mu_{1,j} I(u \in [j \cdot 2^{-m}, (j+1) \cdot 2^{-m}]) + \\ & + \mu_{2,j} I(v \in [j \cdot 2^{-m}, (j+1) \cdot 2^{-m}])\} - \mu_{1,2^m} a(u, v; \hat{\lambda}) - \alpha_0 \tilde{c}(u, v)\}. \end{aligned} \quad (17)$$

where α_0 is a generic constant.

By replacing (17) in relation (10), the leading term $1 + \mu_{1,-1} - \alpha_0$ is canceled out and we get:

$$\hat{c}_{m,n}(u, v) = B_{m,n}^*(u, v) \cdot B_{m,n}(u, v), \quad (18)$$

where

$$\begin{aligned} B_{m,n}(u, v) = & \exp \left\{ - (w(u, v))^{-1} (\mu_{1,-1} - \alpha_0) \right. \\ & - (w(u, v))^{-1} \sum_{j=0}^{2^m-1} \{ \hat{\mu}_{1,j} I(u \in [j \cdot 2^{-m}, (j+1) \cdot 2^{-m}]) \\ & + \hat{\mu}_{2,j} I(v \in [j \cdot 2^{-m}, (j+1) \cdot 2^{-m}]) \} \\ & \left. - \hat{\mu}_{1,2^m} (w(u, v))^{-1} a(u, v; \hat{\mu}) - \alpha_0 (w(u, v))^{-1} \tilde{c}(u, v) \right\}. \end{aligned}$$

(because $\int_{[0,1]^2} \hat{c}_{m,n}(u, v) dudv = 1$, according to (10))

and

$$B_{m,n}^*(u, v) = (E_{(Z_1, Z_2)} (B_{m,n}(Z_1, Z_2)))^{-1}$$

By (15) we obtain that

$$B_{m,n}^*(u, v) \cdot E_{(Z_1, Z_2)}(I(Z_1 \in [j \cdot 2^{-m}, (j+1) \cdot 2^{-m}], Z_2 \in [0, 1])B_{m,n}(Z_1, Z_2)) = 2^{-n}.$$

$$B_{m,n}^*(u, v) \cdot E_{(Z_1, Z_2)}(I(Z_2 \in [j \cdot 2^{-m}, (j+1) \cdot 2^{-m}], Z_1 \in [0, 1])B_{m,n}(Z_1, Z_2)) = 2^{-n}.$$

$$B_{m,n}^*(u, v) \cdot E_{(Z_1, Z_2)}(a(Z_1, Z_2, \hat{\lambda})B_{m,n}(Z_1, Z_2)) = 0. \quad \text{for all } j = 0, \dots, (2^m-1). \quad (19)$$

So (18) can be rewritten as follows:

$$\begin{aligned} \widehat{c}_{m,n}(u, v) &= -B_{m,n}^* (u, v) \sum_{j=0}^{2^m-1} (\widehat{\mu}_{1,j} 2^{-m} + \widehat{\mu}_{2,j} 2^{-m}) \\ &\quad \times \exp \left\{ \sum_{j=0}^{2^m-1} \left\{ \widehat{\mu}_{1,j} \cdot ((w(u, v))^{-1} \cdot I(u \in [j \cdot 2^{-m}, (j+1) \cdot 2^{-m}]) - 2^{-m}) \right. \right. \\ &\quad \left. \left. + \widehat{\mu}_{2,j} ((w(u, v))^{-1} \cdot I(v \in [j \cdot 2^{-m}, (j+1) \cdot 2^{-m}]) - 2^{-m}) \right. \right. \\ &\quad \left. \left. - \widehat{\mu}_{1,2^m} a(u, v; \widehat{\lambda}) - \alpha_0 \widetilde{c}(u, v) (w(u, v))^{-1} \right\} \right\}. \end{aligned}$$

The potential function can be written as follows:

$$D_{m,n}(\widehat{\theta}_m, \widehat{\lambda}) = \int_0^1 \int_0^1 \exp\{A_m + B_m - \widehat{\mu}_{1,2^m} a(u, v, \widehat{\lambda}) - \alpha_0 (w(u, v))^{-1} \widetilde{c}(u, v)\} du dv$$

where

$$A_m = - \sum_{j=0}^{2^m-1} \widehat{\mu}_{1,j} (w(u, v))^{-1} \cdot I(u \in [j \cdot 2^{-m}, (j+1) \cdot 2^{-m}]) - 2^{-m}$$

and

$$B_m = - \sum_{j=0}^{2^m-1} \widehat{\mu}_{2,j} (w(u, v))^{-1} \cdot I(v \in [j \cdot 2^{-m}, (j+1) \cdot 2^{-m}]) - 2^{-m}$$

Then, (19) is equivalent to the following system of equations:

$$\begin{aligned} \frac{\partial D_{m,n}(\widehat{\theta}_m, \widehat{\lambda})}{\partial \mu_{1,j}} &= 0 \\ \frac{\partial D_{m,n}(\widehat{\theta}_m, \widehat{\lambda})}{\partial \mu_{2,j}} &= 0 \\ \frac{\partial D_{m,n}(\widehat{\theta}_m, \widehat{\lambda})}{\partial \mu_{1,2^m}} &= 0 \end{aligned} \tag{20}$$

for all $j = 0, \dots, (2^m - 1)$.

We remark that since the second order derivative of $D_{m,n}(\widehat{\theta}_m, \widehat{\lambda})$ is the covariant matrix of

$$\left\{ \left(w(Z_1, Z_2) \right)^{-1} \cdot I(Z_1 \in [j \cdot 2^{-m}, (j+1) \cdot 2^{-m}]), \right. \\ \left. \left(w(Z_1, Z_2) \right)^{-1} \cdot I(Z_2 \in [j \cdot 2^{-m}, (j+1) \cdot 2^{-m}]), \left(w(Z_1, Z_2) \right)^{-1} a(Z_1, Z_2, \hat{\lambda}) \right\}_{j=0}^{2^m-1},$$

$D_{m,n}(\hat{\theta}_m, \hat{\lambda})$ is positively defined.

We obtain that the solutions of the equation (20) are its minimum values $D_{m,n}(\hat{\theta}_m, \hat{\lambda})$, which depends on $\hat{\lambda}$, α_0 and $\tilde{c}(u, v)$.

Since the potential function $D_{m,n}(\hat{\theta}_m, \hat{\lambda})$ and $WMEC$ (18) are not smooth, according to common practice, they must be smooth.

From Lemma 2, for n large enough, we have an integral representation for

$$\sum_{j=0}^{2^m-1} \mu_{1,j} \cdot I(u \in [j \cdot 2^{-m}, (j+1) \cdot 2^{-m}]),$$

Thus, following the results from [10], Lemma 2 leads us to the following:

$$\begin{aligned} I(u \in [j \cdot 2^{-m}, (j+1) \cdot 2^{-m}]) &= 1 - I(u < j \cdot 2^{-m}) - I(u > (j+1) \cdot 2^{-m}) \\ &= 1 - \lim_{n \rightarrow \infty} \frac{n}{\sqrt{2\pi}} \int_{-\infty}^{j \cdot 2^{-m}} \exp\{-(x-u)^2 n^2 / 2\} dx \\ &\quad - \lim_{n \rightarrow \infty} \frac{n}{\sqrt{2\pi}} \int_{-\infty}^{-(j+1) \cdot 2^{-m}} \exp\{-(x+u)^2 n^2 / 2\} dx \\ &= 1 - \lim_{n \rightarrow \infty} \phi(n(j \cdot 2^{-m} - u)) - \lim_{n \rightarrow \infty} \phi(-n((j+1) \cdot 2^{-m} - u)). \end{aligned}$$

So,

$$\begin{aligned} \sum_{j=0}^{2^m-1} \mu_{1,j} \cdot I(u \in [j \cdot 2^{-m}, (j+1) \cdot 2^{-m}]) &= \\ 1 - \lim_{n \rightarrow \infty} \phi(n(j \cdot 2^{-m} - u)) - \lim_{n \rightarrow \infty} \phi(-n((j+1) \cdot 2^{-m} - u)). \end{aligned}$$

Immediately, we get:

$$\begin{aligned}
 D_{m,n}(\widehat{\theta}_m, \widehat{\lambda}) &= \int_{[0,1]^2} \exp \left\{ \sum_{j=0}^{2^m-1} (w(u,v))^{-1} \left[\mu_{1,j}(w^{-1}(u,v))(\phi(n(j \cdot 2^{-m} - u)) \right. \right. \\
 &\quad \left. \left. + \phi(-n((j+1)2^{-m} - u)) - 1) + 2^{-m} \right] \right. \\
 &\quad \left. + \mu_{2,j}(w(u,v))^{-1} (\phi(n(j \cdot 2^{-m} - v)) \right. \\
 &\quad \left. + \phi(-n((j+1)2^{-m} - v)) - 1) + 2^{-m} \right] \\
 &\quad \left. - \mu_{1,2^m} a(u,v; \widehat{\lambda}) - \alpha_0 \widetilde{c}(u,v) \right\} dudv.
 \end{aligned}$$

and then we have:

$$\begin{aligned}
 B_n &\approx \exp \left\{ \sum_{j=0}^{2^m-1} \left[\widehat{\mu}_{1,j}(\phi(n(j \cdot 2^{-m} - u)) + \phi(-n((j+1)2^{-m} - u))) (w(u,v))^{-1} \right. \right. \\
 &\quad \left. \left. + \widehat{\mu}_{2,j}(\phi(n(j \cdot 2^{-m} - v)) + \phi(-n((j+1)2^{-m} - v))) \right] (w(u,v))^{-1} \right. \\
 &\quad \left. - \widehat{\mu}_{1,2^m} a(u,v; \widehat{\lambda}) (w(u,v))^{-1} - \alpha_0 \widetilde{c}(u,v) (w(u,v))^{-1} \right\}
 \end{aligned}$$

Now, we can write

$$\widehat{c}_{m,n}^w(u,v) = B_{m,n}^*(u,v) \cdot B_{m,n}(u,v)$$

which can be symmetrized by taking $\mu_{1,j} = \mu_{2,j}$ for all $j = 0, \dots, (2^m - 1)$ and by considering $a(u,v; \widehat{\lambda})$ to be a symmetric function.

To complete the demonstration, we have to proof that the *WMEC* approximator,

$$\widehat{C}_{m,n}^w(u,v) = \int_0^u \int_0^v \widehat{c}_{m,n}^w(u,v) dudv,$$

is the second order increasing. Let $[u_1, u_1 + \Delta] \times [v_1, v_1 + \Delta]$ be a closed range included in $[0, 1]^2$.

We can observe that, because $\widehat{c}_{m,n}(u,v)$ is a positive function, then:

$$\widehat{C}_{m,n}^w(u_1 + \Delta, v_1 + \Delta) - \widehat{C}_{m,n}^w(u_1 + \Delta, v_1) - \widehat{C}_{m,n}^w(u_1, v_1 + \Delta) + \widehat{C}_{m,n}^w(u_1, v_1) \tag{21}$$

$$= \int_{u_1}^{u_1 + \Delta} \int_{v_1}^{v_1 + \Delta} \widehat{c}_{m,n}^w(u,v) dudv \tag{22}$$

which is also non-negative.

Now, for m and n large enough, we get the *WMECs*. □

4 Some asymptotical properties

From constraint (13), we have some unknown parameters that we need to estimate. These unknown parameters are represented by $\hat{\theta}_s$. In order to estimate $\hat{\theta}_s$ we use a random sample of size s .

Let $D_m(\mu, \theta)$ be the approximate potential function with dependence parameters θ , according to Section 3, where $\hat{\mu}_s$ and μ^0 denote the minimal values of $D_m(\mu, \theta)$ for $\theta = \hat{\theta}_s$ and $\theta = \theta^0$, respectively. The Hessian matrices of $D_m(\mu, \theta)$ are $H_{1,m}(\mu, \theta) = \nabla_{\mu\mu'} D_m(\mu, \theta)$ and $H_{2,m}(\mu, \theta) = \nabla_{\mu\theta'} D_m(\mu, \theta)$.

Following the line from [10] and [9] for our model, we will consider the following hypotheses:

C_1^w . $\hat{\theta}_s \xrightarrow{p} \theta^0 \in \text{Int}(A)$, where A is some non-empty compact set and $\text{Int}(A)$ is the set of interior points of A . The dimension of the set A is equal to the number of dependent constraints.

The set

$$B = \{\mu \in R^{\dim(\mu)} / \nabla_{\mu} D_m(\mu, \theta) = 0, \forall \theta \in A\} \quad (23)$$

is also non-empty and compact, where $\dim(B)$ is equal to the number of Lagrange multipliers in the expression $D_m(\mu, \theta)$.

So, the number of marginal constraints is $\dim(\mu) - \dim(A)$.

C_2^w . The map from A to B is a diffeomorphism (i.e., there is a continuous bijective correspondence).

C_3^w . $D_m(\mu, \theta)$ is a strictly convex function of μ for all θ and uniformly continuous (in probability) in θ ,

i.e.

$$\sup_{\mu \in B} \left| D_m(\mu, \hat{\theta}_s) - D_m(\mu, \theta^0) \right| \xrightarrow{p} 0, \text{ as } \left| \hat{\theta}_s - \theta^0 \right| \xrightarrow{p} 0.$$

C_4^w . The vector of dependence parameter estimates is asymptotically normal such that

$$s^{1/2}(\hat{\lambda}_s - \lambda^0) \xrightarrow{d} s(0, \Psi),$$

where Ψ is an asymptotic variance-covariance matrix of $\hat{\theta}_s$.

The assumption C_2^w states that the relationship between A and B is a bijective correspondence relation (i.e., for a given set of dependence parameter estimates $\widehat{\theta}_s$ in A there exists uniquely a set of Lagrange multipliers $\widehat{\mu}_s$ in B which contains a unique subset of the Lagrange multipliers that determining the dependence constraints). The assumption ensures that the potential function has uniquely minimal values for a given set of parameters. Conversely, these minimal values are uniquely determined by a set of parameters.

Theorem 5. *Considering $C_1^w - C_4^w$, we have*

$$\widehat{\mu}_s \xrightarrow{p} \mu^0, \tag{24}$$

$$s^{1/2}(\widehat{\mu}_s - \mu^0) \xrightarrow{d} s(0, H_{1,m}^{-1}(\mu^0, \theta^0)H_{2,m}(\mu^0, \theta^0)\Psi H_{2,m}'(\mu^0, \theta^0)H_{1,m}^{-1'}(\mu^0, \theta^0)).$$

We note that \xrightarrow{p} and \xrightarrow{d} means the convergence in probability and law (distribution), respectively.

In particular, if dependent restrictions are linear in parameters, i.e. $a(u, v; \theta) = a(u, v) - \theta$, the potentially associated function of the problem WEM is:

$$\begin{aligned} D_m(\mu, \widehat{\theta}) = & \int_{[0,1]^2} \exp\{\mu_{1,-1} + \sum_{k=0}^{2^m-1} (w(u, v))^{-1}[\mu_{1,k}(\Phi(k - 2^m u) \\ & + \Phi(2^m u - k - 1) - 1 + 2^{-m}) \\ & + \mu_{2,k}(\Phi(k - 2^m v) + \Phi(2^m v - k - 1) - 1 + 2^{-m})] \\ & - k'(a(u, v) - \widehat{\theta})\} dudv - \mu_{1,-1}, \end{aligned} \tag{25}$$

where $\boldsymbol{\lambda} = \{\mu_{1,-1}, \mu_{1,0}, \mu_{2,0}, \dots, \mu_{1,k}, \mu_{2,k}, \dots, \mu_{1,2^n-1}, \mu_{2,2^n-1}, k'\}$ and $\mu_{1,-1}$ corresponds to the restriction

$$\int_{[0,1]^2} c(u, v) dudv = 1.$$

Considering $H_{2,n}(\theta^0, \boldsymbol{\lambda}) = \mathbf{I}$, we obtain

Theorem 6. *If (25) satisfies $C_1^w - C_4^w$, then we have:*

$$\widehat{\mu}_s \xrightarrow{p} \mu^0.$$

$$s^{1/2}(\widehat{\mu}_s - \mu^0) \xrightarrow{d} s(0, H_{1,n}^{-1}(\mu^0, \theta^0)\mathbf{I}\Psi\mathbf{I}'H_{1,m}^{-1'}(\mu^0, \theta^0)),$$

where \mathbf{I} is a diagonal matrix of size $\dim(\mu^0) \times \dim(\theta^0)$.

The previous theorem suggests that the efficiency of estimators $\widehat{\mu}_s$ increases as we use more marginal restrictions. However, if we add too many marginal restrictions, the efficiency of the estimators may decrease as this may increase the probability of the covariance $\{u, v, a(u, v; \widehat{\theta}_s)\}$ in $D_n(\mu, \widehat{\theta}_s)$ to be negative. Therefore, the Hessian matrix $H_{1,n}(\mu^0, \theta^0)$ contains some negative elements that can cause the asymptotic dispersion $s^{1/2}(\widehat{\mu}_s - \mu^0)$ to decrease excessively.

5 Conclusions

In this paper we propose a maximum entropy copula by maximizing a weighted entropy. Based on the work of Chu and Satchell [10] we introduce a new dependence model, namely the weighted maximum entropy model, denoted by *WME*. This problem has as solution a copula function- the weighted maximum entropy copula, denoted by *WMEC*. We approximate this solution using sets of definite integrals. Also, we discuss some asymptotical properties of the unknown parameters concerning measures of association.

References

- [1] Barbu V.Ş., Karagrigoriou A., Preda V. (2017), Entropy and divergence rates for Markov chains: I. The Alpha-Gamma and Beta-Gamma case, Proceedings of the Romanian Academy-series A, vol 4- to appear
- [2] Barbu V.Ş., Karagrigoriou A., Preda V. (2018), Entropy and divergence rates for Markov chains: II. The weighted case, Proceedings of the Romanian Academy-series A, vol 1- to appear
- [3] Barbu V.Ş., Karagrigoriou A., Preda V. (2018), Entropy and divergence rates for Markov chains: III. The Cressie and Read case and applications, Proceedings of the Romanian Academy-series A, vol 2- to appear
- [4] Bălcău C., Constantin D., Panait I. Ioana, Some bounds for the weighted cumulative paired interval entropy, International Journal of Risk Theory, Vol 7 (no.2), 2017
- [5] Băncescu I., Comparing the expected system lifetimes of k-out-of-m systems using transmuted-G distributions, Proceedings of the Romanian Academy- to appear (2017).
- [6] Belis M., Guiaşu S., 1968. A quantitative-qualitative measure of information in cybernetic systems. IEEE Transactions on Information Theory 14, 593-594

- [7] Berger A. L., Della Pietra V. J., Della Pietra S. A., "A maximum entropy approach to natural language processing", *Journal Computational Linguistics*, Volume 22 Issue 1, 1996, Pages 39-71
- [8] Cherubini, U., Luciano, E., Vecchiato, W., 2004. *Copula Methods in Finance*. John Wiley & Sons, New York
- [9] B. Chu, 2011. Recovering copulas from limited information and an application to asset allocation. *J. Bank. Financ.*, 35, 1824-1842.
- [10] B. Chu, S. Satchell, 2016, Recovering the Most Entropic Copulas from Preliminary Knowledge of Dependence, *Econometrics*, 4, 20, doi:10.3390/econometrics4020020
- [11] T. Cover, J. Thomas, 1991. *Elementary of Information Theory*; Wiley: New York, NY, USA.
- [12] Chollete L., Heinen A., Valdesogo A., Modelling international financial returns with a multivariate regime-switching copula, *J. Financ. Econom.*, 2009, 7, 437-480
- [13] D. Dănciulescu, 2015. Formal Languages Generation in Systems of Knowledge Representation based on Stratified Graphs, *Informatica*, vol. 26, no. 3, pp. 407-417.
- [14] M.A.H. Dempster, E.A. Medova, S.W. Yang, 2007. Empirical copulas for cdo tranche pricing using relative entropy. *Int. J. Theor. Appl. Financ*, 10, 679-701.
- [15] Embrechts, P., Lindskog, F., McNeil, A.J., 2003. Modelling dependence with copulas and applications to risk management, in: Rachev, S.T. (Ed.), *Handbook of Heavy Tailed Distribution in Finance*. Elsevier, North-Holland, Amsterdam. chapter 8, pp. 329-384.
- [16] S. Guiaşu, 1971. Weighted entropy. *Reports on Mathematical Physics* 2 (3), 165-179
- [17] G. Grigoraş, D. Dănciulescu, A. Bandoi, 2011. Hierarchically identification. *Recent Researches in Tourism and Economic Development* 511-513. In *Proceedings of the 1st International Conference on Tourism and Economic Development (TED 2011)*, Drobeta Turnu Severin, Romania, October 2011.
- [18] Hao, Z. Singh, V.P., 2013. Modeling multi-site streamflow dependence with maximum entropy copula. *Water Resour. Res.*, 49, doi:10.1002/wrcr.20523.

- [19] Ioffe, A.D., Tihomirov, V.M. Theory of Extremal Problems; Lions, J.L., Papanocolalaou, G., Rockafellar, R.T., Eds.; Studies in mathematics and its applications; North Holland Publishing Company: Amsterdam, The Netherlands; New York, NY, USA; Oxford, UK, 1979; Volume 6.
- [20] Jaynes, E.T. (1957) "Information theory and statistical mechanics", I. Physical Review 106, 620–630
- [21] J. Kapur, 1989. Maximum-Entropy Models in Science and Engineering; Wiley: New York, NY, USA.
- [22] Kolve, N., dos Anjos, U., Mendes, B., 2006. Copulas: a review and recent developments. Stochastic Models 22, 617-660.
- [23] Kutoyants, Y.A, 2004. Statistical Inference for Ergodic Diffusion Processes; Springer series in statistics; Springer-Verlag: London, UK; Berlin, Heidelberg, Germany.
- [24] Mikosch, T., 2006. Copulas: tales and facts. Extremes 9, 3-20.
- [25] R.B. Nelsen, 2006. An Introduction to Copulas; Springer: New York, NY, USA.
- [26] Ning C., Xu D., Wirjanto T.S., Is volatility clustering of asset returns asymmetric? J. Bank. Financ. 2015, 52, 62-76
- [27] Preda V., Băncescu I. (2016), A new family of distributions with a general generic distribution for reliability studies. Log-concavity and Application, International Journal of Risk Theory, Alexandru Myller Publishing Iasi, 1(6), 13-38
- [28] V. Preda, S.Dedu, C. Gheorghe, New classes of Lorenz curves by maximizing Tsallis entropy under mean and Gini equality and inequality constraints, Physica A, 436, 925-932, (2015)
- [29] Vasile Preda, Costel Bălcău, Entropy optimization with applications, Editura Academiei Române, București, 2010
- [30] V. Preda, C. Bălcău, I.I. Panait, 2017. A weighted cumulative paired interval entropy. Submitted.
- [31] Rodriguez J.C., Measuring financial contagion: A copula approach. J. Empir. Financ. 2007, 14, 401-423
- [32] A. Sklar 1959. Fonctions de repartition a n dimensions et leurs marges, Publ. Inst. Statist. Univ. Paris, 8: 229-231

- [33] Soares, Abner D., Newton J. Moura Jr, and Marcelo B. Ribeiro. "Tsallis statistics in the income distribution of Brazil." *Chaos, Solitons & Fractals* 88 (2016): 158-171.
- [34] Tsallis, C. (1988), Possible generalizations of Boltzmann-Gibbs statistics, *Journal of Statistical Physics* 52, 479–487
- [35] Tsallis, C. (1998), On the fractal dimension of orbits compatible with Tsallis statistics, *Physical Review E* 58, 1442–1445
- [36] Tsallis, C. (2002), Entropic nonextensivity: A possible measure of complexity, *Chaos, Solitons and Fractals* 12, 371–391

Ioana PANAIT,
PhD School of Mathematics,
University of Bucharest,
Str. Academiei, nr 14, București, Romania.
Email: yoana.panait@yahoo.com