

#### **Original citation:**

Chatrabgoun, O., Hosseinian-Far, A., Chang, V., Stocks, Nigel G. and Daneshkhah, Alireza. (2017) Approximating non-gaussian bayesian networks using minimum information vine model with applications in financial modelling. Journal of Computational Science.

#### **Permanent WRAP URL:**

http://wrap.warwick.ac.uk/93677

#### Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

#### **Publisher's statement:**

© 2017, Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International http://creativecommons.org/licenses/by-nc-nd/4.0/

#### A note on versions:

The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP URL' above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

# Approximating Non-Gaussian Bayesian Networks using Minimum Information Vine Model with Applications in Financial Modelling

O. Chatrabgoun<sup>a</sup>, A. Hosseinian-Far<sup>b</sup>, V. Chang<sup>c</sup>, N. G. Stocks<sup>d</sup>, A. Daneshkhah<sup>e,\*</sup>

#### Abstract

Many applications of financial modelling require to jointly model multiple uncertain quantities to present more accurate, near future probabilistic predictions required in decision making. Bayesian networks (BNs) and copulas are two common approaches to modelling joint uncertainties with probability distributions in financial and bushiness professions. In particular, the copulas have attracted more attentions due to their nice property of approximating the probability distribution of the data with heavy tail which is very common in financial applications (e.g., financial asset returns, risk analysis of capital allocation within a financial ordination). The standard multivariate copulae suffer from some serious limitations which made then unsuitable for multivariate modelling of the financial data. An alternative copula model is the pair-copula construction (PCC) model which is more flexible and efficient for multivariate modelling of financial data. The only restriction of PCC model is that selecting the best model when the number of variables increases becomes a computationally a challenging problem. Bauer et al. [3] address this issue by capturing conditional independences in the data, and propose a new model called Bayesian network PCC (BN-PCC) which provides more parsimonious models in different settings. This new model is structurally more flexible than PCC due to the benefit of including conditional independences by the data structure. In addition, the difficulty of computing conditional distributions in graphical models for non-Gaussian distributions can be eased using bivariate copulas. In this paper, we extend this approach further using the minimum information vine model which results in a more flexible and efficient approach

Email address: A.Daneshkhah@warwick.ac.uk (A. Daneshkhah)

<sup>&</sup>lt;sup>a</sup>Department of Statistics, Faculty of Mathematical Sciences & Statistics, Malayer University, Malayer, Iran
<sup>b</sup>School of Computing, Creative Engineering & Engineering, Leeds Beckett University, Leeds, LS6 3QR, UK
<sup>c</sup>Xi'an Jiaotong-Liverpool University, Suzhou, China
<sup>d</sup>School of Engineering, University of Warwick, Covetry CV4 7AL, UK

<sup>&</sup>lt;sup>e</sup>Warwick Centre for Predictive Modelling, School of Engineering, University of Warwick, Coventry CV4 7AL, UK

<sup>\*</sup>Corresponding author

in modelling multivariate dependencies of heavy-tailed distribution and tail dependence as observed in the financial data. We demonstrate that the extended model based on minimum information PCC can approximate any given non-Gaussian BN to any required degree of approximation. Unlike the method developed by Bauer et al. [3], the proposed model is more flexible and is not restricted to use the parametric pair-copula models, but pair-copulas can be approximated using the maximum entropy (or minimum information) concept given the limited observed data by truncating the corresponding polynomials/bases after k terms to meet the restrictions imposed by the data and problem under study. We examine three different bases including ordinary polynomial, orthonormal and Fourier series and propose the best fitting model among them based on a goodness-of-fit criteria. Finally, we apply our method to modelling the global portfolio data from the perspective of an emerging market investor located in Brazil. The results show that the multivariate distribution approximated based on the proposed model in this paper is fitted far better other previously published methods.

#### Keywords:

Bayesian Network, Copula, Directed Acyclic Graph, Entropy, Orthonormal Series, Probabilistic Financial Modelling, Vine.

#### 1. Introduction

Soft computing is a collection of methodologies, which aim to exploit uncertainty. Modelling multiple uncertainties using multivariate distributions is required in real life problems as a soft computing methods. Construction of multivariate distribution would help us to appropriately examine dependencies between multivariate data in the real world complexities. A natural way to model multivariate data is to use the method proposed in [24] which is known as *Norta* method (normal to anything). Norta transforms the marginal distribution function of the variable to normal, induces a dependence structure and then transforms back. This method ignores the difference between product moment and rank correlation matrices of the joint normal and for higher dimensions, the set of rank correlation matrices may not be positive definite.

In recent years, copulas have gained popularity in constructing multivariate distributions and survey dependency structures. In the well known Sklar's theorem [41], univariate distributions link to each other to construct multivariate distribution i.e. copula function is a multivariate distribution

function which is marginally distributed on the interval [0, 1] uniformly. One of the main advantages of the copula function is to separate dependency structure from marginal distributions. Moreover, by using copula function, some quantities such as tail dependency which is the dependency between extreme values of the variables, can be obtained. This kind of dependency in particular is very important in real life problems.

Unfortunately, multivariate copulas are not as rich as bivariate copula. Building higher dimensional copula is generally a challenging task, and choosing a parametric family for the given higher dimensional copula is rather more difficult and limited (see [19]). This drawback in modelling of multivariate data by using copula, motivated statisticians to apply a flexible multivariate copulas known as pair copula for modelling multivariate dependency. This graphical model which can be seen as a classical hierarchical model was firstly introduced by Joe [25] and was later formulated by Bedford and Cooke [6, 7]. Aas et al. [1] developed and called it pair copula or vine and also decomposed a general multivariate distribution based on it and proposed a method to perform inferences. This modelling structure is based on a decomposition of a multivariate density into a cascade of bivariate copula. Pair copula construction solves the limitation in construction of multivariate copula and also considers the dependency between pair of variables. Two common forms of the pair copula are D-vine and C-vine, (for more information see [31]).

Using vine to model multivariate distribution apart from the advantages mentioned above is suffering from this drawback that when the number of variables increases, the number of bivariate copulas increases by quadratic rate (i.e., as for n variables we should use n(n-1)/2 bivariate copula), and this is against parsimonious axiom in modelling.

To overcome these vine modelling challenges, Bauer et al. [3] and later Bauer and Czado [4] proposed the pair-copula construction for modelling multivariate distribution represented by a Bayesian network (BN) for non-Gaussian distributions. Their method also permits to include the conditional independence assumptions induced by a BN. This approach is very useful to construct non-Gaussian distribution in order to capture features such as tail behaviour and non-linear, asymmetric dependency. In order to tackle these challenges and for modelling non-Gaussian multivariate distribution, it is more plausible to represent the pair-copula construction model in terms of a DAG [3, 4]. The method addressed in [3, 4] also permits to include the conditional independence assumptions induced by a BN. This approach is very useful to construct non-Gaussian distribution in order to

capture features such as tail behaviour and non-linear, asymmetric dependency.

Bedford et al. [8] stated that the use of a parametric copula to model dependency is simply a translation of one difficult problem into another: instead of the difficulty of specifying the full joint distribution, we have the difficulty of specifying the copula. The only and main advantage is the technical one that copulas are normalized to have support on the unit square and uniform marginal distributions. As a result, restricting copula functions to a particular parametric class (Gaussian, multivariate t, etc.) makes the potential flexibility of the copula approach not being realized in practice.

To settle this concern, there have been recently several studies proposing various non-parametric methods to tackle the issues mentioned above. For instance, Kauermann et al. [28] proposed a semiparametric method by using the spline to estimate multivariate copula density. In order to achieve a good and smooth fit, the spline coefficients are first penalized, and uniform marginals of the copula density is then approximated by placing linear constraints on the spline coefficients which quadratic programming is then required to derive the fitted model. However, the main purpose of the method proposed in [28, 38] was to tackle the curse of dimensionality using a semi-parametric approach, but it was not fully successful in achieving it. It was shown that the proposed method shifts the problem a little bit so that computation on 3, 4 (or 5) dimensions could be possible. In order to tackle the curse of dimensionality [38], the methodology was then applied to D-vines with penalized Bernstein polynomials or penalized B-splines, to estimate the bivariate copula density in each knot of the selected D-vine model [27]. However, the reported results are very promising, but there is no clear model selection algorithm. In addition, when there is weak dependency they do not perform well. Nagler et al. [39] extended the above approaches by applying them on the simplified vine copula models. Simplified vine copula models give rise to very flexible models which are often found to be superior to other multivariate copula models [1]. In order to make the model more tractable, one usually makes the simplifying assumption that the pair-copula densities do not change with conditional assumption [39]. There are several interesting factors driving the relative performance of the non-parametric estimators. The most important one is the strength of dependence. They concluded that the kernel estimators performed best, but do worse than penalized B-spline estimators when there is weak dependence or tail dependence.

Bedford et al. [8] proposed an alternative method based on using entropy copula (also known

as minimum information copula) that can be determined to any required degree of precision based on the available data (or expert judgements). The approach used in this paper, by contrast, allows a lot of flexibility in copula specification. It can be easily implemented in practice and it is only required to assume that the copula density of interest must be continuous and non-zero. Our constructive approach involves the use of entropy copulas that can be specified to any required degree of precision based on the data available. We illustrate properly that good approximation locally guarantees globally good approximation.

A natural way to build a minimum information copula or specifying dependency constraints is through the use of moments [8]. These can be specified either on the copula or on the underlying bivariate density. These moment constraints are considered as real-valued functions  $\phi_1, \ldots, \phi_k$  which are required to take expected values  $e_1, \ldots, e_k$ , respectively. The expected values are either computed from the available data or specified based on the experts' beliefs.

In this paper, we improve the fitted multivariate density approximation proposed in [3, 4] using a newly developed approximation method based on the entropy method (see also [22]). The conditional and joint probabilities of the selected DAG structure can be specified by constructing a minimum information copula between the nodes of interests given their parents' sets. In this study, we assume that the DAG structure is learned using either search and score methods (Sparse Candidate Algorithm; and Greedy Search methods) or constraint based methods, including PC, MMPC algorithms, Fast algorithm, etc (e.g., see [14, 29, 22]). A minimum information copula can be represented in terms of Polynomial Series (PS), and more flexible ones including Orthonormal Polynomial Series (OPS) and Orthonormal Fourier Series (OFS). We demonstrate that the approximation accuracy will be notably increased using the minimum information copula. We verify our claim by comparing our approximation with the results illustrated in [3, 4] to model the global portfolio data from the perspective of an emerging market investor located in Brazil.

The present paper is organised as follows. In Section 2, we introduce the pair-copula decomposition associated with the non-Gaussian BN of multivariate data. In Section 3, we first briefly study the entropy copula and show that how it can be used to approximate a bivariate copula density. We then develop it further to approximate the multivariate distribution associated with the given non-Gaussian BN. We improve this approximation in Section 4, using PS and OPS basis functions and OFS. In Section 5, we apply our method for modelling the global portfolio data from

the perspective of an emerging market investor located in Brazil. We then exhibit our approximation flexibility by comparing it with the method presented in [3, 4]. Section 6 is dedicated to a simulation study. We finally conclude our study in Section 7.

#### 2. Pair-copula construction for non-Gaussian Bayesian Networks

Considering the above-mentioned vine's drawbacks in modelling multivariate data, it has been tried to develop a method through using the nice properties of graphical model and pair-copula, simultaneously. Hanea et al. [23] provided an opportunity to exploit the advantages of both worlds. Indeed, the purpose is to apply the conditional independence in graphs and the simplified vine structure. Simplified vine copula models give rise to very flexible models which are often found to be superior to other multivariate copula models [1]. In order to make the model more tractable, one usually makes the simplifying assumption that the pair-copula densities do not change with conditional assumption [39].

Graphical models [33] are probabilistic models in which conditional independence between variables can be shown using a simple graph, i.e. in a graphical model, vertex are the variables, and the conditional and the causal relationships between variables are shown by edges. Let's introduce the conditional independence concept as:

 $X_1$  is said to be conditionally independent of  $X_2$  given  $X_3$ , denoted by  $(X_1 \perp X_2 \mid X_3)$ , if for all configuration  $x_1, x_2, x_3$  of the variable in  $X_1, X_2, X_3$  satisfying  $p(X_3 = x_3)$ , it holds

$$p(X_1 = x_1 | X_2 = x_2, X_3 = x_3) = p(X_1 = x_1 | X_3 = x_3).$$

Equivalent definition for conditional independence can be stated as follow:

$$p(X_1 = x_1, X_2 = x_2 | X_3 = x_3) = p(X_1 = x_1 | X_3 = x_3)p(X_2 = x_2 | X_3 = x_3).$$

Bayesian networks models known as Directed acyclic graphical (hereafter DAG) and are certainly the most common and perhaps the most applicable version of a graphical model. The construction of the Bayesian networks was based on the assumption of a joint Gaussian distribution, however this approach lacks the necessary performance for capturing the features of real world data such as tail behaviour and non-linear, asymmetric dependencies. Bauer et al. [3] filled this gap and introduced non-Gaussian graphical model by combining useful properties of both pair-copula and

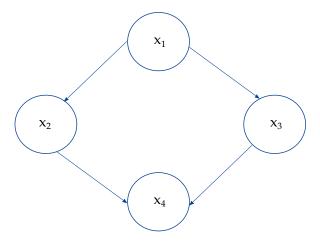


Figure 1: A DAG with 4 elements

DAG, and named it non-Gaussian DAG-PCC. Elidan [17, 18] gives another copula decomposition of distributions associated with a DAG that is based on generally higher variate copulas, and therefore lacks the flexibility of the pair-copula approach.

First, let us introduce some of the preliminary notations associated with graphical models, and also some of the basic concepts related to Bayesian networks. Further details on Bayesian network can be found in [14]. Applications of Bayesian networks range from artificial intelligence, decision support systems, and engineering to genetics, geology, medicine, and finance, see [40].

A graph is a pair G = (V, E) where V is the set of vertices (or nodes) and E is the set of edges. The set of edges E is a subset of the set  $V \times V$  of ordered pair of nodes. It is assumed that E contains only distinct pair of nodes so that there exists no loops, that is,  $(x,y) \in E \Longrightarrow x \neq y$ . Given two nodes x and y, the edge between them is said to be directed if  $(x,y) \in E$  but  $(x,y) \notin E$ , and written  $x \longrightarrow y$ . If the edges in each graph are all directed, such graph is called a directed one. Figure 1 illustrates a directed graph with vertices  $V = \{x_1, x_2, x_3, x_4\}$ . A cycle of length n is a path (i.e. a path of length n form x to y is a sequence  $x = x_0 = x_1 = \ldots = x_n = y$  of distinct vertices such that  $(x_{i-1}, x_i) \in E$  for all  $i = 1, 2, \ldots, n$ ) with the modification that the first and the last vertex are identical  $x_0 = x_n$ . A directed graph G = (V, E) is acyclic if it contains no directed cycle.

Given a DAG D, we define the set of descendants de(x) of x are the vertices y such that  $x \longrightarrow y$  but not  $y \longrightarrow x$  and the descendants of x are the nodes y such that there is a path from x to y, but not from y to x, and by similar definition non-descendants of x or nd(x) defined as

 $nd(x) = V \setminus (de(x) \bigcup x)$ . For  $x \longrightarrow y$ ; x is a parent of y and y is a child of x. The set of parents of a vertex y is denoted by pa(y), and the set of children of a vertex x represented by ch(x). For example in Figure 1, the set of parents of  $X_4$  is  $pa(X_4) = \{X_2, X_3\}$  and the set of children of  $X_3$  is given by  $ch(X_3) = \{X_4\}$ .

We only consider the density decomposition related to DAG. Factorization of the multivariate density can be done by using the conditional independence concept. The basic decomposition scheme offered by a DAG can be explained based on the conditional independence. Consider a joint density function f defined over n variables  $X_1, \ldots, X_n$ . The multivariate density function f can be decomposed as a product of n conditional density functions as follows:

If  $X_i$  is independent of all other predecessors given its parent then, we can write

$$f(x_1,...,x_3) = \prod_{i=1}^n f(x_i|pa(x_i)),$$

i.e. once we know the value of  $pa(x_i)$ , knowing the value of the other preceding variables is redundant. This property of f is known as Markovian or local Markov properties in the literature. We illustrate briefly that the density decomposition scheme associated with 4 random variables  $X = (X_1, \ldots, X_4)$  with a joint density function  $f(x_1, \ldots, x_4)$  are satisfying a DAG shown in Figure 1 and the marginal densities  $f(x_1), \ldots, f(x_4)$ . We make use of the expression

$$f(x_1,\ldots,x_4)=f(x_1)f(x_1|x_2)f(x_3|x_1,x_2)f(x_4|x_1,x_2,x_3)$$

The marginal distribution of  $X_1$  is known, therefore  $f_1$  is also known. The marginals of  $X_1$  and  $X_2$  are known, and the copula of  $X_1, X_2$  is also known; therefore by applying Sklar's theorem we can get  $f(x_1, x_2)$ , and hence

$$f(x_1|x_2) = c_{12}(F(x_1), F(x_2))f_2(x_2).$$

In order to get  $f(X_3|X_1,X_2)$ , based on the conditional independence  $X_2$  and  $X_3$  given  $X_1$ , we can write

$$f(x_3|x_1,x_2) = f(x_3|x_1) = c_{13}(F(x_1),F(x_3))f_3(x_3).$$

Similarly, by using conditional independence between  $X_1$  and  $X_4$  given  $X_2$  and  $X_3$ , holds that

$$f(x_4|x_1, x_2, x_3) = f(x_4|x_2, x_3).$$

Now, for  $v \in V$ , we order the elements of pa(v) increasingly (with respect to some strict total order  $<_v$  on pa(x)) and set

$$pa(v, w) = \{u \in pa(v) | u <_v w\}, \quad w \in pa(v).$$

These orders can be determined based on the Kendall's  $\tau$  rank correlation between variable v and pa(v). For example, for  $pa(4,3) = \{2\}$ , we can write

$$f(x_4|x_2,x_3) = c_{34|2}(F_{4|2}(x_4|x_2),F_{3|2}(x_3|x_2))f_{4|2}(x_4|x_2).$$

We can determine  $f_{4|2}(x_4|x_2)$  in a similar way as  $f_{2|1}(x_2|x_1)$  or  $f_{3|1}(x_3|x_1)$ . Therefore, f can be decomposed as:

$$f(x_1, \dots, x_4) = \prod_{i=1}^4 f(x_i) \times c_{12}(F(x_1), F(x_2)) \times c_{13}(F(x_1), F(x_3)) \times c_{24}(F(x_2), F(x_4)) \times c_{34|2}(F_{4|2}(x_4|x_2), F_{3|2}(x_3|x_2)).$$

$$(1)$$

As a result, we can state the following theorem.

**Theorem 1.** Let D=(V,E) be a DAG and let f be a multivariate density function on n variables with marginal density  $f_i$  and corresponding cumulative distribution function  $F_i$ ,  $i=1,2,\ldots,n$ . Then f is uniquely determined by its univariate margins  $f_i$ ,  $i=1,2,\ldots,n$  and its conditional pair-copula  $c_{vw|pa(v,w)}, v \in V, w \in pa(v)$  and f can be decomposed as follows:

$$f(x_1, \dots, x_n) = \prod_{v=1}^n f(x_v) \prod_{w \in pa(v)} c_{vw|pa(x,w)} (F_{v|pa(v,w)}, F_{w|pa(v,w)}).$$
 (2)

*Proof.* The proof of this theorem relies on graph theoretical considerations only (see [3] and references therein).

The above theorem gives us a constructive approach to build a multivariate distribution given a DAG: If we make choices of marginal densities and copula, then the above formula gives us a multivariate density. Hence, PCC-DAG can be used to model general multivariate densities. However, in practice, we have to use copula from a convenient class, and this class should ideally be the one that allows us to approximate any given copula to an arbitrary degree. In the following sections, we address this issue in more details. By having this class of copula, we can then approximate any multivariate distribution using a DAG.

When we apply Non Gaussian PCC-DAG for decomposing multivariate density f as seen in Theorem 1, some of the copula may not exist as  $C_{23}$  in Figure 1 and decomposition (1), i.e. since

the copula  $C_{23}$  is not available in the decomposition of the f, Bauer et al. (2012) exploited the conditional independence property  $(X_2 \perp X_3 | X_1)$  to get  $F_{3|2}(x_3 | x_2)$ . This method proposed in [3] requires quite complex numerical computation, and there is no general closed-form solution for it. To overcome this challenge, we divide the support of  $X_2$  to some bins and in each bin we calculate  $F_{3|2}(x_3|x_2)$ . Complete discussion are provided on the Brazilian case.

## 3. Approximating Multivariate Density: A minimum information copula approach

This section outlines an approach based on using the entropy techniques originated from [9] in conjunction with the observed data or expert elicitation of observables. This is used to define a copula that can help to build the joint distribution of two random variables. It can also be used to develop it further for constructing a multivariate distribution using a Non-Gaussian PCC-DAG model. The method that will be described below is based on using the  $D_1AD_2$  algorithm to determine the copula in terms of potentially asymmetric information about two variables of interests.

#### 3.1. The $D_1AD_2$ algorithm and minimum information copula

Bedford and Meeuwissen [9] applied a so-called DAD algorithm to produce discretized entropy copula between two variables with given rank correlation. This approach relies on the fact that the correlation is determined by means of the symmetric function UV. The same approach can be used whenever we wish to specify the expectation of any symmetric function of U and V ([5, 34]).

This method can be developed further using Borwein et al.'s idea [11] which enables us to have asymmetric specifications. In the revised method, we first determine a positive square matrix A, also called a *kernel*; two diagonal matrices  $D_1$  and  $D_2$  should be then found in such a way that the following product,  $D_1AD_2$  is doubly stochastic. The theory can be easily generalised for continuous functions [8, 15].

Now, suppose there are two random variables X and Y, with cumulative distribution functions  $F_X$  and  $F_Y$ , respectively. These are the variables of interest that we would like to correlate by introducing constraints, based on some knowledge about functions of these variables. Suppose there are k of these functions, namely  $h'_1(X,Y), h'_2(X,Y), \ldots, h'_k(X,Y)$ , and that we wish either to calculate their mean values in terms of the observed data, or the expert wishes to specify mean

values  $\alpha_1, \ldots, \alpha_k$  for all these functions, respectively. We can simply specify corresponding functions of the copula variables U and V, defined by  $h_i(U, V) = h'_i(F_1^{-1}(U); F_2^{-1}(V))$ ,  $i = 1, 2, \ldots, k$ , where  $h_i : [0, 1]^2 \to \mathbb{R}$ , at which we can specify the mean values  $\alpha_1, \ldots, \alpha_k$  that these functions should simultaneously take. Further suppose that  $h_i, h_j$  are linearly independent for  $i \neq j$ . We seek a copula that has these mean values, a problem which is usually either infeasible or undetermined. Hence, assuming feasibility for the moment, we also consider the copula to be maximum entropy (with respect to the uniform distribution), which guarantees a unique and reasonable solution. We form the kernel

$$A(u,v) = \exp(\lambda_1 h_1(u,v) + \ldots + \lambda_k h_k(u,v)), \tag{3}$$

where u denotes the realization of U and v the realization of V.

For practical implementations, we use the same method as proposed by [8], and later by [15] to discretize the set of (u, v) values such that the whole domain of the copula is covered. Thus, the aforementioned kernel A becomes a 2-dimensional matrix, and two matrices  $D_1$  and  $D_2$  should be then determined. As a result, the following product denoted by P over  $[0,1]^2$  becomes a doubly stochastic matrix which represents a discretized copula density

$$P = D_1 A D_2. (4)$$

The  $D_1AD_2$  algorithm can be used to generate a unique joint density with uniform marginals for each vector  $(\lambda_1, \ldots, \lambda_k)$ . The set of all possible expectation vectors  $(\alpha_1, \ldots, \alpha_k)$  that could be taken by  $(h_1, h_2, \ldots, h_k)$  under some probability distribution is convex, and that for every  $(\alpha_1, \ldots, \alpha_k)$  in the interior of that convex set, there is a density with parameters  $(\lambda_1, \ldots, \lambda_k)$  for which  $(h_1, h_2, \ldots, h_k)$  take these values [11, 8, 15].

We now explain the iterative algorithm required to approximate the mentioned copula density by this algorithm. Suppose that both (u, v) are discretized into n points, as  $u_i$ , and  $v_j$ , i, j = 1, ..., n respectively. Then, we write  $A = (a_{ij}), D_1 = diag(d_1^{(1)}, ..., d_n^{(1)}), D_2 = diag(d_1^{(2)}, ..., d_n^{(2)})$ , where  $a_{ij} = A(u_i, v_j), d_i^{(1)} = D_1(u_i), d_j^{(2)} = D_2(v_j)$ . We define the doubly stochastic matrix,  $D_1AD_2$  with the uniform marginals as follows

$$\forall i = 1, \dots n$$
  $\sum_{j} d_i^{(1)} d_j^{(2)} a_{ij} = 1/n$ , and  $\forall j = 1, \dots n$   $\sum_{j} d_i^{(1)} d_j^{(2)} a_{ij} = 1/n$ .

The idea behind  $D_1AD_2$  algorithm is simple. It starts with arbitrary positive initial matrices for  $D_1$  and  $D_2$ , and the new vectors will be then successively defined by iterating the following maps

$$d_i^{(1)} \mapsto \frac{1}{n \sum_j d_j^{(2)} a_{ij}} \ (i = 1, \dots, n), \quad d_j^{(2)} \mapsto \frac{1}{n \sum_i d_i^{(1)} a_{ij}}, \quad (j = 1, \dots, n).$$

It can be shown that this iteration scheme converges geometrically to the requested vectors [11].

Note that to compare different discretizations (for different n's) we should multiply each cell weight  $d_i(1)d_j(2)a_{ij}$  by  $n^2$  as this quantity approximates the continuous copula density with respect to the uniform distributions.

The mapping from the set of vectors of  $\lambda$ 's onto the set of vectors of the resulting expectations of functions  $(h_1, \ldots, h_k)$  has to be found numerically. Bedford et al. [8] proposed the optimization techniques for determining the  $\lambda_i$ 's and the corresponding copula. The expectations  $\alpha_i$  of k functions of variables X and Y are given by

$$E[h'_i(X,Y)] = E[h_i(U,V)] = \alpha_i, \quad i = 1, \dots, k.$$

We now wish to determine the appropriate set of  $\lambda$ 's for the given expectations  $\alpha_i$ , where the expectations have been calculated using the discrete copula density  $D_1AD_2$  given in (4). Hence, to determine  $\lambda_i$ 's satisfying the constraints, the following set of equations has to be solved

$$L_l(\lambda_1, \dots, \lambda_k) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n P(u_i, v_j) h_l(u_i, v_j) - \alpha_l, \quad l = 1, 2, \dots, k.$$
 (5)

The left hand sides of the above equations are just functions of  $\lambda$ 's and with optimization algorithms their roots can be found. One of the possible solvers for this task would be FSOLVE - MATLAB's optimization routine. An alternative method is to use another MATLAB's optimization procedure called FMINSEARCH, which implements the Nelder-Mead simplex method [32]. The minimized function is then

$$L_{sum}(\lambda_1,\ldots,\lambda_k) = \sum_{l=1}^k L_l^2(\lambda_1,\ldots,\lambda_k).$$

We refer the interested readers to [8, 34] to show how an expert could specify a copula through defining expected values.

### 3.2. Approximating Multivariate Density by Non-Gaussian PCC-DAG

In this section, we use techniques from approximation theory to show that any n-dimensional multivariate density which is  $C^2$  (that is, twice differentiable, with continuous second derivatives)

can be approximated arbitrarily well pointwise using a finite parameter set of 2-dimensional copulas in a PCC-DAG construction. The basic idea is that we can use a series expansion, like a two-dimensional Ordinary Polynomial Series (PS), Orthonormal Polynomial Series (OPS) and Fourier Series (OFS), to approximate any log-density function by truncating the series at an appropriate point. What is non-trivial, however, about this method, is that the same truncation can be used everywhere in a PCC-DAG construction giving an overall uniform pointwise approximation. Hence our method allows the use of a fixed finite dimensional family of copulas to be used in a PCC-DAG construction, with the promise of a uniform level of approximation. Since the approximations we make of copula densities might not be quite copula densities themselves, we need to transform them to make them copulas.

To demonstrate this, we first should show that the family of bivariate (conditional) copula densities contained in a given multivariate distribution forms a *compact set* in the space of continuous functions on  $[0,1]^2$ . Then, it can be shown that the same finite parameter family of copulae can be used to derive a given level of approximation to all conditional copulae simultaneously.

Here, we develop the approximation method used in [8, 15] to approximate any log-density function at a desired level of approximation which is more accurate and exhibits better properties. We first introduce the notations. The basic assumption is that all densities are continuous. We denote C(Z) as the space of continuous real valued functions on a space Z, where  $Z = [0,1]^r$  for some r, and the corresponding norm on C(Z) is given by

$$||f_{1...r}|| = \sup |f_{1...r}(x_1, \dots, x_r)|.$$

The set of all possible 2-dimensional (conditional) copulae is denoted by

$$C(f) = \{c_{ij|i_1...i_r} : 1 \le i, j, i_1, ..., i_r \le n, i, j \ne i_1, ..., i_r\},\$$

where  $c_{ij|i_1...i_r}$  is the copula of the conditional density of  $X_i, X_j$  given  $X_{i_1}, ..., X_{i_r}$ .

The famous Arzela-Ascoli theorem can be used to check the compactness of the following function space,  $K \subset C([0,1]^2)$ . This space is relatively compact, if the functions in K are equicontinuous and pointwise bounded. Further details about equicontinuous and pointwise bounded of the minimum information copula density can be found in [8].

It can be shown that the following two spaces are relatively compact (Theorem 3 in [8]).

$$\mathcal{M}(f) = \{ f_{i|i_1...i_r} : 1 \le i, i_1, ..., i_r \le n, i \ne i_1, ..., i_r \},$$

and

$$\mathcal{B}(f) = \{f_{ij|i_1...i_r} : 1 \le i, j, i_1, ..., i_r \le n, i, j \ne i_1, ..., i_r\},\$$

where  $f_{i|i_1...i_r}$  is the conditional density of  $X_i$  given  $X_{i_1},...,X_{i_r}$ , and  $f_{ij|i_1...i_r}$  is the conditional density of  $X_i, X_j$  given  $X_{i_1},...,X_{i_r}$ .

It is then straightforward to show that the set  $C(f) \subset C([0,1]^2)$  is relatively compact. In addition, since all the functions in C(f) are positive and uniformly bounded away from 0, the set  $\mathcal{LNC}(f) = \{\ln(g) : g \in C(f)\} \subset C([0,1]^2)$  is also relatively compact (see [8] for details and proofs).

As a result, the set  $C([0,1]^2)$  can be considered as a vector space, and in this context a base is simply a sequence of functions  $h_1, h_2, \dots \in C([0,1]^2)$  such that any function  $g \in C([0,1]^2)$  can be written as  $g = \sum_{i=1}^{\infty} \lambda_i h_i$ . In other words, it can be shown that given  $\epsilon > 0$ , there is a k such that any member of  $\mathcal{LNC}(f)$  (or  $\mathcal{C}(f)$ ) can be approximated to within error  $\epsilon > 0$  by a linear combination of  $h_1, h_2, \dots, h_k$ . There are lots of possible bases. We introduce three of these bases, PS, OPS and OFS in the next three subsections with some nice properties in their density approximation.

It should be noticed that the copula density approximated by the method described above might not be a copula density itself. Therefore, the resulting approximation needs to be transformed in such a way to obtain a copula. This can be done by weighting the approximated density. One of the most effective weighting schemes is the  $D_1AD_2$  algorithm mentioned in the previous section. If we have a continuous positive real valued function A(u,v) on  $[0,1]^2$ , then there are continuous positive functions  $d_1(u)$  and  $d_2(v)$ , such that  $d_1.d_2.A$  is a copula density, that is, it has uniform marginal distributions. This density is called C-Projection of A and denoted by C(A). Bedford et al. [8] present the following lemma in which it allows us to control the error made when approximating a copula by another function.

**Lemma 1.** Let g be a non-negative continuous copula density. Given  $\epsilon > 0$ , there is a  $\delta$  such that  $||g - f|| < \delta$  then  $||g - C(f)|| < \epsilon$ .

Note that these reweighting functions have the same differentiability properties as the function f being reweighted. This can be seen from the integral equation that they satisfy:

$$d^{(1)}(u) = \frac{1}{\int d^{(2)}(v) f(u, v) dv} \quad and \quad d^{(2)}(v) = \frac{1}{\int d^{(1)}(u) f(u, v) du}.$$

Eventually, the term given in (2) can be used to show that good approximation of each conditional copula would result in a good approximation of the multivariate density of interest by using PCC-DAG.

#### 4. Building approximations using maximum entropy distributions

In this section, we provide a practical guide for building a maximum entropy PCC-DAG structure to approximate any multivariate distribution. In the previous section, we presented a method proposed in [8] that all conditional copulae can be approximated using linear combinations of basis functions. In this section, we are going to address the issue of how the appropriate parameter values can be chosen. We will also introduce a practical and efficient alternative, based on using the entropy criterion that lies very close to the approach described above. In other words, given the basis functions  $\{1, h_1, \ldots, h_k\} : [0, 1]^2 \to \mathbb{R}$ , we seek values  $\lambda_1, \ldots, \lambda_k$  so that  $\exp(\sum_1^k \lambda_i h_i)$  is close to the approximated copula density. This can be done by fitting the moments of  $h_i$  in the entropy framework. Therefore, if  $E_g[h_i(u, v)] = \alpha_i$ , we seek for the entropy copula density that also has these moments. This copula density can be uniquely determined, using the  $D_1AD_2$  algorithm, hence

$$d^{1}(u)d^{2}(v)\exp(\sum_{i=1}^{k}\lambda_{i}h_{i}(u,v)).$$

As mentioned above, a multivariate distribution can be modelled by a PCC-DAG structure where it can be defined as a decomposition of the given multivariate distribution into certain conditional copulae. The following algorithm has summarised the steps for approximating the given multivariate distribution associated with a PCC-DAG structure:

- 1. Specify a basis family, denoted by  $S(k) = \{h_1, h_2, \ldots\},\$
- 2. Specify a DAG structure,
- 3. For each part of the DAG, the bivariate copulae, specify either
  - mean  $\alpha_1, \ldots, \alpha_k$  for  $h_1, \ldots, h_k$  on each pairwise copula;
  - functions  $\alpha_m(ji \mid D_e)$  for the mean values as functions of the conditioning variables, for  $m = 1, \ldots, k$ .

One of the main aspect that would effect the aforementioned approximation is the basis family. Here, we examine the impact of three basis families, the *ordinary polynomial* series, the *orthonormal*  polynomial series and Fourier on approximating the entropy copulae and the multivariate distribution associated with the chosen DAG structure. We first briefly introduce these three basis functions.

#### 4.1. Ordinary Polynomial base

One of the simple basis that can be applied in entropy copula is ordinary polynomial basis. These basis were mainly used in [8] and can be defined simply as follows:

$$\psi_0(u) = 1, \psi_1(u) = u, \psi_2(u) = u^2, \psi_3(u) = u^3, \psi_4(u) = u^4, \dots$$

PS basis are so easy to determine and selecting it by expert judgement can be easier than other basis.

#### 4.2. Orthonormal polynomial base

In mathematics, particularly numerical analysis, a basis function is an element of the basis for a function space. The term is a degeneration of the term basis vector for a more general vector space; that is, each function in the function space can be represented as a linear combination of the basis functions. We say two polynomial functions  $g_1$  and  $g_2$  are orthonormal in the interval [0,1], if

$$\int_0^1 g_1(u)g_2(u)du = \begin{cases} 1 & \text{for } g_1(u) = g_2(u); \\ 0 & \text{for } g_1(u) \neq g_2(u). \end{cases}$$
 (6)

The OPS base can be calculated more conveniently than some natural basis. In fact, if the basis is an OPS basis, adding a new item to the expansion does not change the coefficient of the already found shorter expansion [21]. But, if the basis is not orthonormal, any new item in general has a nonzero projection on previous items. It means that the already found coefficients of the expansion would have to be changed. That is one of the reasons for using OPS basis functions as the basis family, S(k). It is reasonable to consider Gram-Schmidt OPS basis which is one of the famous OPS basis functions on [0, 1].

To construct this OP basis over the interval [0,1], we use the *Gram-Schmidt process* as follows

$$\varphi_{0}(u) = 1$$

$$\varphi_{n}(u) = \frac{u^{n} - \sum_{j=0}^{n-1} \frac{\int_{0}^{1} u^{n} \varphi_{j}(u) du}{\int_{0}^{1} \varphi_{j}^{2}(u) du} \varphi_{j}(u)}{||u^{n} - \sum_{j=0}^{n-1} \frac{\int_{0}^{1} u^{n} \varphi_{j}(u) du}{\int_{0}^{1} \varphi_{j}^{2}(u) du} \varphi_{j}(u)||} \qquad n \ge 1.$$

The first few functions are

$$\varphi_0(u) = 1,$$

$$\varphi_1(u) = \sqrt{3}(-1+u),$$

$$\varphi_2(u) = \sqrt{5}(1-6u+6u^2),$$

$$\varphi_3(u) = \sqrt{7}(-1+12u-30u^2+20u^3),$$

$$\varphi_4(u) = \sqrt{9}(1-20u+90u^2-140u^3+70u^4),$$

$$\varphi_5(u) = \sqrt{11}(-1+30u-210u^2+560u^3-630u^4+252u^5).$$

#### 4.3. Fourier base

Trigonometric or Fourier basis is the other type of orthonormal basis. Computational speed of these basis for some data is considerable. Especially, these basis function present appropriate fit in the peridic data. The first functions are

$$\begin{split} \phi_0(u) &= 1, \quad \phi_1(u) = \sqrt{2}cos(2\pi u), \quad \phi_2(u) = \sqrt{2}sin(2\pi u), \\ \phi_3(u) &= \sqrt{2}cos(4\pi u), \quad \phi_4(u) = \sqrt{2}sin(4\pi u), \\ \phi_5(u) &= \sqrt{2}cos(6\pi u), \quad \phi_6(u) = \sqrt{2}sin(6\pi u). \end{split}$$

## 5. Application: Global portfolio data from the perspective of an emerging market investor located in Brazil

In this section, we apply the approximation method presented in this paper using OP, OPS and OFS basis families, S(k) (as mentioned in the first step in the algorithm above) to approximate the multivariate distribution associated with the selected PCC-DAG structure corresponding to the global portfolio data from the perspective of an emerging market investor located in Brazil. We then exhibit the potential flexibility of our approach by comparing it with the method cited in [3, 4].

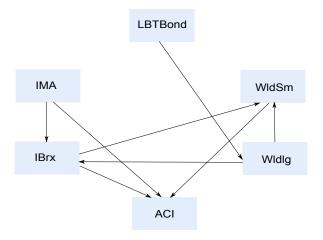


Figure 2: Selected DAG structure for six dimensional contemporaneous daily log-returns of the global portfolio data from the perspective of an emerging market investor located in Brazil.

Example: In this example, we use the same data set as previously studied in [37] to illustrate the approximation method introduced in this paper. The data consists of six dimensional contemporaneous daily log-returns: Brazilian composite hedge fund index (the ACI, Arsenal Composite Index), a long-term inflation-indexed Brazilian treasury bonds index (the IMA-C index, computed by the Brazilian Association of Financial Institutions, Andima), Brazilian stock index with the 100 largest capitalization companies (IBRX), Index of large world stocks computed by MSCI (WLDLg), Index of small capitalization world companies computed by MSCI (WLDSm), and index of total returns on US treasury bonds computed by Lehman Brothers Barra (LBTBond). They are recorded over the period January 2, 2002 to October 20, 2008 in which 1629 data are collected. We denote these six variables ACI, IMA, IBrX, Wldlg, WLdSm and LBIBond, respectively.

We shall first remove serial correlation in these six time series, that is, the observation of each variable must be independent over time. Hence, the serial correlation in the conditional mean and the conditional variance are modelled by an AR(1) and a GARCH(1,1) model (see [10]), respectively. Thus, the following model for log-return  $x_i$  is considered for the  $i^{th}$  time series

$$x_{i,t} = c_i + \alpha_i x_{i,t-1} + \sigma_{i,t} z_{i,t},$$
  
 $E[z_{i,t}] = 0$  and  $Var[z_{i,t}] = 1,$   
 $\sigma_{i,t}^2 = \alpha_{i,0} + a_i \epsilon_{i,t-1}^2 + b_i a \sigma_{i,t-1}^2,$ 

where  $\epsilon_{i,t-1} = \sigma_{i,t} + z_{i,t}$  [1].

The further analysis is performed on the standardized residuals  $z_i$ . If the AR(1)-GARCH(1,1) models are successful at modelling the serial correlation in the conditional mean and the conditional variance, there should be no autocorrelation left in the standardized residuals and squared standardized residuals. We can use the modified Q-statistic and the Lagrange multiplier test, respectively, to confirm this [1]. For all series, the null hypothesis, 'no autocorrelation left for the both tests', cannot be rejected with %5 significance. Since, we are mainly interested in estimating the dependence structure of the risk factor, the standardized residual vectors are converted to the uniform variables using the kernel method before further modeling. We denote the converted time series of ACI, IMA, IBrX, Wldlg, WLdSm and LBIBond by 1,2,3,4,5 and 6, respectively.

Here, we want to generate a PCC-DAG approximation fitted to this data set using entropy distributions based on the different basis. Indeed, the real challenge is in connecting DAG models to vines. We try to specify DAG structure in our data. One approach is to apply structure learning algorithms such as the PC algorithm (see [42], Section 5.4.2) to the data  $\Phi^{-1}(.)$ , where  $\Phi$  denotes the standard normal cdf. This transformation is needed, since the tests for conditional independence performed by the PC algorithm(at the %5 significance level) are based on the assumption of normality. As an alternative approach, expert knowledge is frequently exploited to define the DAG (see [30], Chapter 5). Moreover, there are structure selection algorithms for Non Gaussian DAG's available in Bauer and Czado (2016) which are similarly based on the PC algorithm. We adopt the DAG structure presented in Figure 2 by applying the PC algorithm. Furthermore, the presented structure for non Gaussian DAG available in Bauer and Czado (2016) produces the same results. According to the presented DAG, we decompose the multivariate density of our data by applying Theorem 1 in order to derive PCC-DAG structure i.e. given the presented DAG, Theorem 1 prescribes which pair copulas are required to be specified in the definition of our model. Note that variable 1(ACI) has three parents (2(IMA), 3(IBrX), and 5(WldSm)) as the order of the parents based on the heuristic rule of modelling strong bivariate dependences prior to weak dependences. Our decision was based on estimates  $\hat{\tau}$  of kendall's  $\tau$  variable 1,5 ( $\hat{\tau}=0.209$ ), variable 1,3 ( $\hat{\tau} = 0.197$ ), and variable 1,2 ( $\hat{\tau} = 0.127$ ), respectively. Similar rule can be applied for variables 3(IBrX) and its parents (2(IMA) and 4(WLdLg) based on  $\hat{\tau}$  as  $\hat{\tau}_{32} = 0.0858$ , and  $\hat{\tau}_{34} = 0.424$ . Moreover, variable 5 has two parents (3(IBrX) and 4(WIdIg)) which are  $\hat{\tau}_{53} = 0.402$  and  $\hat{\tau}_{54} = 0.75$ .

Based on these ordering, and according to the measure of dependencies kendall's  $\tau$ , the resulting multivariate density decomposition is:

$$f_{1,\dots,6}(x_1,\dots,x_6) = \prod_{i=1}^6 f_i(x_i) \times c_{15}(F_1(x_1), F_5(x_5)) \times c_{45}(F_4(x_4), F_5(x_5)) \times c_{46}(F_4(x_4), F_6(x_6)) \times c_{34}(F_3(x_3), F_4(x_4)) \times c_{13|5}(F_{1|5}(x_1|x_5), F_{3|5}(x_3|x_5)) \times c_{23|4}(F_{2|4}(x_2|x_4), F_{3|4}(x_3|x_4))$$

$$\times c_{35|4}(F_{3|4}(x_3|x_4), F_{5|4}(x_5|x_4)) \times c_{12|35}(F_{1|35}(x_1|x_3, x_5), F_{2|35}(x_2|x_3, x_5)). \tag{7}$$

We now derive the entropy copulae in association with some moment constraints between copula variables 1,2,3,4,5,6 in the density decomposition (7). We initially construct maximum entropy copulas for unconditional copula  $c_{15}$ ,  $c_{46}$ ,  $c_{34}$ ,  $c_{45}$ . Now, is essential to decide which bases should be taken and how many discretization points should be used in each case. We start to outline our procedure for the unconditional copula  $c_{15}$ . Other unconditional copula  $c_{46}$ ,  $c_{34}$ ,  $c_{45}$  can be followed in a similar way.

We could simply choose basis functions based on the method described in [15] i.e. starting with simple bases, and moving to more complex ones, and including them until we are satisfied with our approximation. Our OP basis functions are as follows,

$$\psi_{1}(.)\psi_{1}(.), \psi_{1}(.)\psi_{2}(.), \psi_{2}(.)\psi_{1}(.), \psi_{1}(.)\psi_{3}(.), \psi_{3}(.)\psi_{1}(.),$$

$$\psi_{2}(.)\psi_{2}(.), \psi_{2}(.)\psi_{3}(.), \psi_{3}(.)\psi_{2}(.), \psi_{1}(.)\psi_{4}(.), \psi_{4}(.)\psi_{1}(.),$$

$$\psi_{1}(.)\psi_{5}(.), \psi_{5}(.)\psi_{1}(.), \psi_{2}(.)\psi_{4}(.), \psi_{4}(.)\psi_{2}(.), \psi_{3}(.)\psi_{3}(.), \dots$$

OPS basis function constructed using Gram-Schmidt process

$$\varphi_{1}(.)\varphi_{1}(.), \varphi_{1}(.)\varphi_{2}(.), \varphi_{2}(.)\varphi_{1}(.), \varphi_{1}(.)\varphi_{3}(.), \varphi_{3}(.)\varphi_{1}(.),$$

$$\varphi_{2}(.)\varphi_{2}(.), \varphi_{2}(.)\varphi_{3}(.), \varphi_{3}(.)\varphi_{2}(.), \varphi_{1}(.)\varphi_{4}(.), \varphi_{4}(.)\varphi_{1}(.),$$

$$\varphi_{1}(.)\varphi_{5}(.), \varphi_{5}(.)\varphi_{1}(.), \varphi_{2}(.)\varphi_{4}(.), \varphi_{4}(.)\varphi_{2}(.), \varphi_{3}(.)\varphi_{3}(.), \dots$$

and then considered OFS basis functions are:

$$\phi_{1}(.)\phi_{1}(.), \phi_{1}(.)\phi_{2}(.), \phi_{2}(.)\phi_{1}(.), \phi_{1}(.)\phi_{3}(.), \phi_{3}(.)\phi_{1}(.),$$

$$\phi_{2}(.)\phi_{2}(.), \phi_{2}(.)\phi_{3}(.), \phi_{3}(.)\phi_{2}(.), \phi_{1}(.)\phi_{4}(.), \phi_{4}(.)\phi_{1}(.),$$

$$\phi_{1}(.)\phi_{5}(.), \phi_{5}(.)\phi_{1}(.), \phi_{2}(.)\phi_{4}(.), \phi_{4}(.)\phi_{2}(.), \phi_{3}(.)\phi_{3}(.), \dots$$

Following the explanations to select basis function in an optimal manner, we add the basis functions by using stepwise method in [15]. In this method, at each stage, we propose to assess the log-likelihood of adding each additional basis function. We then include the function which produces the largest increase in the log-likelihood. Also, according to [15], in order to get optimal results, first four bases have been considered.

We are now able to construct the entropy copula density  $C_{15}$  with respect to the uniform distributions given the corresponding OP, OPS and OFS constraints above, using the method described in this paper. We are initially required to determine the number of discretization points (grid size). Clearly, a larger grid size will provide a better approximation to the continuous copula, however more computation time will be required. Similarly, the approximation will become more precise, if we run the  $D_1AD_2$  algorithm with more iterations, and therefore, this would cost us more computation time. It can be concluded that the number of iterations will depend on the grid size. We consider the approximation errors in the range  $1 \times 10^{-1}$  to  $1 \times 10^{-24}$ . Thus, the larger the number of grid points used, the larger the number of iterations required for convergence; which is true at all error levels. For all grid sizes, a higher of number of iterations are required initially for improving the accuracy of computation; once the error is reduced, the number of iterations can be decreased. In this example, we choose a grid size of  $200 \times 200$  throughout.

Based on the information given above regarding the grid size, number of iterations and error size, we can derive the entropy copula  $C_{15}$  associated with the chosen constraints. Expectations  $\alpha$  of the selected basis, Lagrange multiplies values (parameter values)  $\lambda$  and Log-Likelihood are summarized in Table 3. Log-Likelihood (L) for PS, OPS, and OFS basis are 93.49, 98.59, and 38.76, respectively. The corresponding copulas in terms of the OP, OPS and OFS bases are plotted in Panels (a), (b), and (c) in Figure 3 respectively.

Method	Base	$(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$	$(\lambda_1,\lambda_2,\lambda_3,\lambda_4)$	L
PS	$(\psi_1\psi_1, \psi_2\psi_1, \psi_5\psi_5, \psi_1\psi_2)$	( 0.27,0.18,0.04,0.19)	( 14.2,-7.9,3.5,-4.1)	93.49
OPS	$(\varphi_1\varphi_1,\varphi_2\varphi_2,\varphi_4\varphi_2,\varphi_2\varphi_4)$	( 0.29,0.13,0.08,0.07)	( 0.31,0.09,0.08,0.04)	95.59
OFS	$(\phi_2\phi_2, \phi_1\phi_1, \phi_3\phi_2, \phi_3\phi_4)$	( 0.16,0.08,0.07,0.07)	( 0.16,0.08,0.07,0.04)	37.76

Table 1: The minimally informative copula given moment constraints for OP, OPS, and OFS bases between 1 and 5

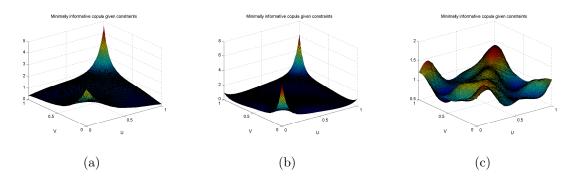


Figure 3: The minimally informative copula given moment constraints between variable 1 and 5; Panel (a): PS basis, Panel (b): OPS basis, and Panel (c): OFS basis

Consider the minimum information copula computation in PCC-DAG structure such that instead of choosing the grid equidistant (or uniform grid), we could choose points where more points are included in the tail of the distribution. This could result in outperforming of Gaussian models by non-Gaussian models approximated based on the method described in this paper. However, we have used Chebyshev points for copula approximation in our grid using minimum information method instead of uniform grid, since they allow for more points in the tail or boundaries of our approximation. This is very important especially in financial applications. Chebyshev points are roots of Chebyshev Polynomial; the discussion and some details are presented in [36]. In order to compare uniform grid and Chebyshev, we consider the above discussion with respect to the uniform grid size deriving the minimum information copula  $C_{15}$  which is associated with the previous chosen constraints. Figure (4) illustrates the entropy copula  $C_{15}$  Chebyshev grid which allow for more points in the tail.

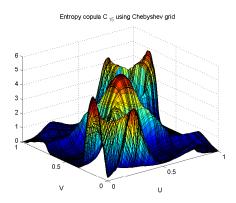


Figure 4: Entropy copula  $C_{15}$  using Chebyshev grid.

One of the main advantages of using OPS and OFS bases over the ordinary polynomial series (considered in Bedford et al., 2015) is that the  $D_1AD_2$  algorithm converges much faster using these bases. This is because of the following nice property of these two bases that adding a new basis to the kernel defined in (3) and used to construct the entropy copula, does not change the Lagrange multipliers of the already used in the kernel. But, this is not the case when one is applying the PS bases (as proposed in [8]) to calculate the entropy copula. In this situation, we need to run the  $D_1AD_2$  algorithm each time a new basis is added to the already chosen bases, and the parameter values are changing accordingly. Therefore, more iterations are required for the  $D_1AD_2$  algorithm using the OPS bases is 9.83 seconds and for the OFS bases is 8.89, while this time for the PS bases is 29.87 seconds which is almost twofold of the former one and almost two and half times more than the latter one.

The other unconditional copula in the decomposition (7) i.e.  $C_{46}$ ,  $C_{34}$ , and  $C_{45}$  could be calculated in the similar way. Using the step-wise method, we select the four PS, OPS and OFS bases that along with their corresponding constraints, resulting Lagrange multipliers, and Log-Likelihood (L) are given in Table 2. The approximated maximum entropy copula for these unconditional copula in terms of the PS, OPS and OFS bases is shown in Panels of Figure 5.

Now, the conditional copulas  $C_{13|5}$ ,  $C_{23|4}$  and  $C_{35|4}$  can similarly be approximated using the entropy approach. We only illustrate construction of the conditional maximum entropy copula between  $C_{13|5}$ .  $C_{23|4}$  and  $C_{35|4}$  can be similarly approximated in a similar way. In order to calculate this copula, we divide the support of 5 into some arbitrary sub- intervals or bins and then construct

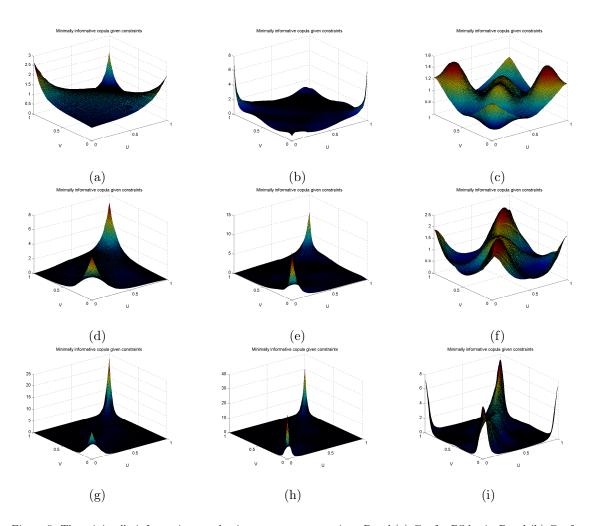


Figure 5: The minimally informative copula given moment constraints, Panel (a): $C_{46}$  for PS basis, Panel (b): $C_{46}$  for OPS basis, Panel (c): $C_{46}$  for OFS basis, Panel (d): $C_{34}$  for PS basis, Panel (e): $C_{34}$  for OPS basis, Panel (f): $C_{34}$  for OPS basis, Panel (g): $C_{45}$  for PS basis, Panel (h): $C_{45}$  for OPS basis, Panel (i): $C_{45}$  for OPS basis.

Copula	Base	$(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$	$(\lambda_1,\lambda_2,\lambda_3,\lambda_4)$	L
C46	PS: $(\psi_1 \psi_1, \psi_5 \psi_5, \psi_5 \psi_1, \psi_1 \psi_4)$	( 0.23,0.02,0.06,0.08)	( 1.4,6.5,-4.7,-4.6)	44.19
	OPS: $(\varphi_1\varphi_1, \varphi_2\varphi_2, \varphi_4\varphi_2, \varphi_5\varphi_5)$	(-0.18,0.13,-0.06,0.06)	(-0.18,0.12,-0.06,0.06)	51.03
	OFS: $(\phi_2\phi_2, \phi_1\phi_1, \phi_2\phi_4, \phi_4\phi_2)$	(-0.11,0.1,-0.08,-0.07)	(-0.11,0.1,-0.08,0.02)	30.37
C34	PS: $(\psi_1 \psi_1, \psi_1 \psi_2, \psi_2 \psi_5, \psi_2 \psi_1)$	( 0.29,0.21,0.08,0.21)	( 36,27.5,10.4,-5.3)	379.02
	OPS: $(\varphi_1\varphi_1, \varphi_2\varphi_2, \varphi_5\varphi_3, \varphi_1\varphi_2)$	( 0.57,0.35,0.1,-0.07)	(0.73, 0.23, 0.09, 0.01)	392.4
	OFS: $(\phi_2\phi_2, \phi_1\phi_1, \phi_4\phi_2, \phi_2\phi_4)$	(0.35, 0.3, 0.19, 0.01)	( 0.4,0.3,0.2,-0.003)	245.49
C45	$PS:(\psi_1\psi_1, \psi_5\psi_5, \psi_1\psi_2, \psi_1\psi_4)$	(0.32, 0.07, 0.23, 0.15)	( 144,-18.4,-96.3,42.3)	1479.6
	OPS: $(\varphi_1\varphi_1, \varphi_2\varphi_2, \varphi_3\varphi_3, \varphi_3\varphi_1)$	( 0.88,0.78,0.67,-0.01)	(2.8,0.73,0.67,-0.01)	1506.3
	OFS: $(\phi_2\phi_2, \phi_1\phi_1, \phi_2\phi_4, \phi_3\phi_1)$	( 0.8,0.7,0.1,0.09)	( 1.6,1.2,0.52,-0.001)	1366.1

Table 2: The minimally informative copula given moment constraints for  $C_{46}$ ,  $C_{34}$ , and  $C_{45}$ 

the conditional copula within each bin. To do so, we select bases in the same way as for the unconditional copulas and fit the copula to the calculated mean values or constraints. In this case, we use four bins so that the first copula is for  $13|5 \in (0,0.25)$ . The other bins are  $13|5 \in (0.25,0.5)$ ,  $13|5 \in (0.5,0.75)$ , and  $13|5 \in (0.75,1)$ . We can follow this process again for the remaining bins. Table 3 shows the mean values or constraints (denoted by  $\alpha_i$ ) and corresponding Lagrange multipliers ( $\lambda_i$ ) required to build the conditional entropy copula between 1|5 and 3|5 for PS, OPS and OFS bases, respectively. The log-likelihood of the approximated copula in each bin is also reported in these tables. The Log-Likelihood over all bins for  $C_{23|4}$  and  $C_{35|4}$  for (PS, OPS, OFS) basis are (16.13, 39.1, 38.63) and (223.69, 345.15, 246.99), respectively.

We can obtain the conditional maximum entropy copula,  $C_{12|35}$ , similarly by dividing each of the conditioning variables' supports into four bins. Then the entropy copulas for 1|35 and 2|35 are calculated on each combination of bins for 3 and 5 which makes 16 bins altogether for it. The bins, bases and log-likelihoods associated with each copula based on the PS, OPS and OFS basis are given in Table 4.

The log-likelihood of the overall non-Gaussian PCC-DAG model using the PS, OPS and OFS bases are 2390.44, 2669.69 and 2093.75, respectively. The use of log-likelihood as a goodness-of-fit criterion is not inconsistent with minimum information modeling. Jaynes [26] uses the parameter

Interval	Bases	$(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$	$(\lambda_1,\lambda_2,\lambda_3,\lambda_4)$	L
	PS: $(\psi_1 \psi_1, \psi_1 \psi_2, \psi_1 \psi_3, \psi_1 \psi_4)$	(0.12,0.06,0.03,0.02)	(38.1,-115,129.7,-44.3)	35.9
0 < M < 0.25	OPS: $(\varphi_1\varphi_1, \varphi_4\varphi_3, \varphi_1\varphi_4, \varphi_5\varphi_1)$	(0.51,-0.15,-0.2,0.12)	(0.4,-0.08,-0.1,0.02)	52.99
	OFS: $(\phi_5\phi_5, \phi_1\phi_1, \phi_2\phi_2, \phi_2\phi_3)$	(0.09, 0.12, 0.2, 0.08)	(0.15,0.09,0.18,-0.05)	18.21
	PS: $(\psi_2\psi_1, \psi_1\psi_3, \psi_1\psi_5, \psi_1\psi_4)$	(0.13, 0.08, 0.04, 0.05)	(2.5,40.7,57.9,-98.4)	5.4
0.25 < M < 0.5	OPS: $(\varphi_1\varphi_1, \varphi_2\varphi_3, \varphi_5\varphi_4, \varphi_1\varphi_4)$	(0.12,-0.06,0.08,0.04)	(0.12,-0.08,0.1,-0.05)	9.2
	OFS: $(\phi_2\phi_5, \phi_5\phi_5, \phi_1\phi_1, \phi_2\phi_2)$	(-0.12,-0.01,0.06,0.06)	(-0.1,-0.01,0.05,-0.01)	6.7
	PS: $(\psi_5\psi_5, \psi_4\psi_5, \psi_1\psi_1, \psi_3\psi_4)$	(0.04,0.05,0.32,0.07)	(7.4,4.3,2.1,-10.3)	7.19
0.5 < M < 0.75	OPS: $(\varphi_3\varphi_3, \varphi_1\varphi_1, \varphi_3\varphi_1, \varphi_2\varphi_3)$	(0.07, 0.1, 0.1, 0.07)	(0.06, 0.13, 0.16, -0.05)	10.3
	OFS: $(\phi_4\phi_2, \phi_5\phi_3, \phi_1\phi_3, \phi_2\phi_4)$	(0.14, 0.03, 0.05, 0.07)	(0.13,0.04,0.06,-0.04)	6.3
	PS: $(\psi_1 \psi_1, \psi_5 \psi_5, \psi_2 \psi_1, \psi_5 \psi_2)$	(0.4,0.09,0.3,0.14)	(11.7,0.65,-10.3,2.7)	30.5
0.75 < M < 1	OPS: $(\varphi_1\varphi_1, \varphi_2\varphi_5, \varphi_4\varphi_1, \varphi_1\varphi_5)$	(0.4,0.12,0.06,0.06)	(0.4,0.09,0.06,0.07)	40.5
	OFS: $(\phi_2\phi_2, \phi_4\phi_2, \phi_3\phi_1, \phi_2\phi_4)$	(0.2, 0.14, 0.09, 0.09)	(0.14,0.13,0.08,-0.01)	17.58

Table 3: Minimaly informative copula given moment constraints between 1 and 3 given 5

maximum likelihood estimates associated with the form of the minimum information distribution to justify the connection in the constraint rule of expectations and frequencies. The use of log-likelihood is also recommended as a plausible model selection tool within minimum information framework [8].

However, the best model should be typically selected by trading-off between the goodness of fit of the candidate models and their model complexity. When there are several competing copulas, and one wishes to know which copula fits the data best. The selected copula model should be the one that minimizes the Kullback-Leibler information between the copula model and the true unknown copula. The Akaike information criteria (AIC) can be thus considered as a plausible tool, which also minimizes the information between the candidate models, to select the best model from the parametric families and non-parametric candidate models [20]. In this regard, the AIC is first computed for each model with the same data, then the "best" model is the one with the least AIC value. It is claimed that AIC is also more computationally efficient than other copula selection methods [20]. Since, the comparison of the log-likelihood of the proposed non-parametric model and the parametric model [3, 4] is not conclusive, as the model complexity measured by the number of parameters is not considered. As suggested above, we compare these models based on the AIC

Interval	Bases (PS,OPS,OFS)	L(PS,OPS,OFS)
0 < 3 < 0.25 &		
0 < 5 < 0.25	$(\psi_1\psi_1, \psi_3\psi_1, \psi_2\psi_4, \psi_5\psi_2), (\varphi_1\varphi_1, \varphi_3\varphi_3, \varphi_3\varphi_1, \varphi_3\varphi_5), (\phi_2\phi_2, \phi_1\phi_1, \phi_4\phi_5, \phi_5\phi_2)$	(19.2,16,8)
0 < 3 < 0.25 &		
0.25 < 5 < 0.5	$(\psi_3\psi_5, \psi_2\psi_3, \psi_4\psi_4, \psi_5\psi_4), (\varphi_5\varphi_5, \varphi_3\varphi_5, \varphi_2\varphi_3, \varphi_2\varphi_4), (\phi_4\phi_4, \phi_1\phi_2, \phi_5\phi_4, \phi_2\phi_1)$	(0.9, 8.5, 3.6)
0 < 3 < 0.25 &		
0.5 < 5 < 0.75	$(\psi_4\psi_1, \psi_1\psi_5, \psi_2\psi_3, \psi_5\psi_5), (\varphi_1\varphi_5, \varphi_2\varphi_4, \varphi_2\varphi_1, \varphi_2\varphi_3), (\phi_2\phi_4, \phi_5\phi_1, \phi_1\phi_3, \phi_1\phi_4)$	(2.6, 16.4, 14.9)
0 < 3 < 0.25 &		
0.75 < 5 < 1	$(\psi_1\psi_1, \psi_1\psi_2, \psi_1\psi_4, \psi_5\psi_1), (\varphi_4\varphi_5, \varphi_4\varphi_3, \varphi_5\varphi_2, \varphi_2\varphi_2), (\phi_3\phi_3, \phi_3\phi_5, \phi_3\phi_4, \phi_4\phi_1)$	(0.53, 4.4, 5.3)
0.25 < 3 < 0.5 &		
0 < 5 < 0.25	$(\psi_1\psi_3, \psi_2\psi_2, \psi_5\psi_5, \psi_1\psi_5), (\varphi_1\varphi_1, \varphi_4\varphi_2, \varphi_3\varphi_4, \varphi_2\varphi_5), (\phi_2\phi_2, \phi_4\phi_2, \phi_3\phi_1, \phi_3\phi_3)$	(9.1,8.8,6)
0.25 < 3 < 0.5 &		
0.25 < 5 < 0.5	$(\psi_1\psi_1, \psi_2\psi_1, \psi_3\psi_1, \psi_5\psi_2), (\varphi_4\varphi_5, \varphi_5\varphi_3, \varphi_1\varphi_1, \varphi_3\varphi_4), (\phi_4\phi_2, \phi_1\phi_2, \phi_1\phi_3, \phi_3\phi_1)$	(4.4,10.5,4.9)
0.25 < 3 < 0.5 &		
0.5 < 5 < 0.75	$(\psi_3\psi_5, \psi_1\psi_1, \psi_2\psi_3, \psi_1\psi_2), (\varphi_4\varphi_2, \varphi_3\varphi_5, \varphi_1\varphi_2, \varphi_5\varphi_1), (\phi_4\phi_2, \phi_4\phi_1, \phi_5\phi_2, \phi_1\phi_2)$	(2.4,5.5,3.8)
0.25 < 3 < 0.5 &		
0.75 < 5 < 1	$(\psi_5\psi_1, \psi_1\psi_2, \psi_2\psi_2, \psi_4\psi_1), (\varphi_2\varphi_1, \varphi_3\varphi_5, \varphi_5\varphi_2, \varphi_1\varphi_1), (\phi_2\phi_4, \phi_1\phi_2, \phi_4\phi_2, \phi_1\phi_1)$	(4.9,7.9,2.9)
0.5 < 3 < 0.75 &		
0 < 5 < 0.25	$(\psi_5\psi_5, \psi_4\psi_3, \psi_2\psi_1, \psi_1\psi_1), (\varphi_1\varphi_5, \varphi_4\varphi_3, \varphi_5\varphi_5, \varphi_5\varphi_2), (\phi_2\phi_4, \phi_1\phi_2, \phi_5\phi_4, \phi_5\phi_1)$	(3.7,7.5,3.7)
0.5 < 3 < 0.75 &		
0.25 < 5 < 0.5	$(\psi_3\psi_5, \psi_1\psi_3, \psi_2\psi_4, \psi_1\psi_1), (\varphi_2\varphi_3, \varphi_3\varphi_2, \varphi_5\varphi_1, \varphi_5\varphi_5), (\phi_4\phi_2, \phi_4\phi_1, \phi_5\phi_2, \phi_1\phi_2)$	(2.8, 7.1, 3.8)
0.5 < 3 < 0.75 &		
0.5 < 5 < 0.75	$(\psi_1\psi_2, \psi_5\psi_4, \psi_4\psi_4, \psi_5\psi_3), (\varphi_1\varphi_1, \varphi_2\varphi_4, \varphi_3\varphi_5, \varphi_5\varphi_5), (\phi_2\phi_2, \phi_3\phi_3, \phi_1\phi_4, \phi_3\phi_1)$	(4.5,6,4)
0.5 < 3 < 0.75 &		
0.75 < 5 < 1	$(\psi_1\psi_1, \psi_1\psi_5, \psi_1\psi_4, \psi_1\psi_3), (\varphi_1\varphi_5, \varphi_1\varphi_1, \varphi_2\varphi_2, \varphi_2\varphi_3), (\phi_2\phi_4, \phi_1\phi_2, \phi_4\phi_2, \phi_1\phi_1)$	(2.6,3,2.9)
0.75 < 3 < 1 &		
0 < 5 < 0.25	$(\psi_5\psi_1, \psi_1\psi_1, \psi_3\psi_1, \psi_1\psi_2), (\varphi_2\varphi_4, \varphi_4\varphi_3, \varphi_5\varphi_2, \varphi_2\varphi_1), (\phi_1\phi_2, \phi_2\phi_5, \phi_4\phi_3, \phi_4\phi_2)$	(1.2,7.2,1.7)
0.75 < 3 < 1 &		
0.25 < 5 < 0.5	$(\psi_5\psi_3, \psi_3\psi_1, \psi_5\psi_5, \psi_2\psi_3), (\varphi_4\varphi_5, \varphi_2\varphi_4, \varphi_3\varphi_3, \varphi_1\varphi_5), (\phi_5\phi_5, \phi_1\phi_4, \phi_4\phi_4, \phi_5\phi_2)$	(0.99,2.3,3)
0.75 < 3 < 1 &		
0.5 < 5 < 0.75	$(\psi_5\psi_5, \psi_2\psi_5, \psi_1\psi_4, \psi_5\psi_1), (\varphi_4\varphi_2, \varphi_1\varphi_5, \varphi_5\varphi_1, \varphi_2\varphi_1), (\phi_4\phi_2, \phi_3\phi_1, \phi_5\phi_2, \phi_3\phi_4)$	(2.2,6.5,6.9)
0.75 < 3 < 1 &		
0.75 < 5 < 1	$(\psi_3\psi_2, \psi_1\psi_5, \psi_2\psi_3, \psi_1\psi_1), (\varphi_2\varphi_1, \varphi_1\varphi_2, \varphi_5\varphi_1, \varphi_1\varphi_5), (\phi_1\phi_1, \phi_3\phi_2, \phi_4\phi_3, \phi_2\phi_3)$	(6.7,5.7,2.2)

Table 4: entropy copula for given moment constraints between 1 and 2 given 3 and 5  $\,$ 

Type of copula	AIC
Bauer et al. (2012) method	-3078.62
entropy copula based	-4187.24
on OFS basis	
entropy copula	-4780.88
based on PS basis	
entropy copula	-5339.38
based on OPS basis	

Table 5: Comparison between different models.

which includes the model complexity for the parametric method. The AIC of the overall Non-Gaussian PCC-DAG model using the PS, OPS and OFS bases are -4780.88, -5339.38 and -4187.24, respectively. These values are considerably less than the AIC of the parametric non-Gaussian PCC-DAG models using Bauer et al. (2012) method, (as the AIC equals to -3078.62). We illustrate the corresponding results in Table 5.

The size of observed data could be considered as a source of potential error when the minimum information vine model is applied for modeling a high-dimensional problem. As the dimensionality (or number of uncertain variables) increases, the number of trees representing the structure of pair-copula model will also increase (see also [12]). The conditional distributions/expectations at lower levels of a deeper vine model must then be estimated based on fewer data points which can be then less accurate and noisier (see also [16] for further details with an application in modelling flood events with the limited data). This problem could be resolved by ignoring some unnecessary conditional dependencies (the so-called simplifying assumption) in the sense discussed in [2, 43]. An alternative method is to approximate fully conditional pair-copula models using Gaussian processes [35]. This model shows promising results with better predictive performance than the method that ignores conditional dependencies. This simplifying pair-copula model is more appropriate for high-dimensional problems.

In this paper, the conditional independence statements play a crucial role in simplifying the model structure [22]. Thus, the computation of the conditional probabilities in practice could not

be complex unless the parent set is considerably large. In the case that the parent set of a variable of interest is very large, the corresponding conditional probability could be estimated using the Gaussian process emulators as suggested in [35].

#### 6. Simulation study

We can now discuss the data simulations derived from the presented minimum information PCC-DAG in order to provide comparisons between correlations in the simulated data and in the observed data. This is based on 2000 simulations. We follow the simulation method proposed by [30] subjected to sampling from the cumulative distributions. This simulation method has been followed by Daneshkhah et al. [15] references cited therein. Moreover, this simulation has been updated using provided simulation method from PCC-DAG in parametric status in [4, 3]. Their sampling strategy is as follows: sample two independent variables distributed uniformly on intervals [0, 1], denoted by  $U_1, U_2$ , and calculate values of the original variables using the following equations:

$$x_1 = u_1,$$
  $x_2 = F_{2|1}^{-1}(u_2|x_1),$ 

where  $x_i$  is realization values of  $X_i$ , and  $u_i$  is realization values of  $U_i$ . Finally, this has been applied to all variables in PCC-DAG. Please note that the order of variables for childs and parents in this simulation is important.

It can be observed that the simulated data and the original data have similar dependency patterns. Table 6 shows the rank correlations between the variables of interest calculated from the original observed data, that are based on the simulated data taken from the fitted PCC-DAG through entropy copula on OPS basis. Other bases can be similarly simulated, however they are not reported and considered here. By comparing these correlations, we can conclude that the results show strong consistency and the estimated correlations based on the entropy PCC-DAG are closer to the ones that are originated from the observed data. Furthermore, we can compare entropy PCC-DAG and parametric PCC-DAG which estimated correlations based on the entropy PCC-DAG using OPS basis are closer to the ones that are derived from the observed data rather than the parametric PCC-DAG.

Table 6: Correlation coefficients of the original data and the simulated data

			Original			
	LBTBond	IMA	IBrX	WldSm	Wldlg	ACI
LBTBond		-0.069	-0.373	0.596	0.571	-0.465
IMA			0.112	0.024	0.022	0.211
IBrX				0.197	0.197	0.435
WldSm					0.938	-0.080
Wldlg						-0.093

## Parametric PCC-DAG

	LBTBond	IMA	IBrX	WldSm	Wldlg	ACI
LBTBond		-0.054	-0.360	0.458	0.479	-0.453
IMA			0.110	0.020	0.019	0.185
IBrX				0.240	0.223	0.342
WldSm					0.924	-0.088
Wldlg						-0.117

## Entropy PCC-DAG

	LBTBond	IMA	IBrX	WldSm	Wldlg	ACI
LBTBond		-0.066	-0.366	0.526	0.513	-0.458
IMA			0.112	0.022	0.020	0.197
IBrX				0.207	0.212	0.417
WldSm					0.933	-0.083
Wldlg						-0.102

#### 7. Discussion and Conclusions

Gaussian distributions are generally used for modelling and computing financial asset returns, risk assessment of capital allocation by banks, and estimating risks associated with financial portfolios in actuarial science. However, the existing internal Gaussian models are limited when it comes to inference from tails. As opposed to normal Gaussian distributions, copulae are known to be a suitable and powerful means for overcoming the flaws in the existing techniques. An example for the application of copulae in the above-mentioned areas, would be the claim allocations and fees' assignments for investigators, experts, etc. as part of allocated loss adjustment expense processes. An additional case for the application of copulae, would be risk assessments conducted by banks and credit institutions for credit and market evaluations and judgements; an existing flaw with many of the existing techniques, known to be internal bottom-up approaches, for such risks assessments, as such techniques are incapable of modelling joint distribution of non-identical risks.

There are non-identical approaches to inference in multivariate distributions. Bayesian networks and copulae are generally very suitable for modelling such probability distributions. In the applications where tail properties are important for predictive probabilistic modelling, many of the existing techniques are limited and inadequate. One of the well-known techniques that can conveniently infer from tail properties is the multivariate Gaussian copula. As stated above, many of the current techniques used for financial application modelling, assume a normal Gaussian distribution of events for simplifying the complex nature of the financial scenarios (as discussed in [8, 13]). The proposed methodology for utilising vine structure for approximation, would enable the modeller to simply establish non-constant conditional correlations, and minimise the chance of risk underestimation.

In this paper, we extended the novel method originally presented by [3, 4] to approximate a multivariate distribution by any Non-Gaussian PCC-DAG structure. The main idea to implement this approximation method is to use the entropy copulae that can be determined to any required degree of precision, based on the available data. The approximation method used in this paper is flexible and easy to implement. The standing technical assumptions we require is that the multivariate density of DAG under study is continuous and is non-zero. In order to approximate a multivariate distribution for the observed data, one only needs to specify a DAG structure, a basis family, and the expected values for the certain functions associated with some constraints on each pairwise copula. Our focus in this paper was to introduce pair-copula structure that can

be approximated by any given non-Gaussian DAG to any required degree of approximation with different bases family. We concentrated on the PS, OPS and OFS bases. The OPS and OFS bases exhibit an appropriate property, which makes the distribution approximation faster in the sense that adding a new element to their expansion does not change coefficient of the already found shorter expansion, where any new item has in general non-zero projection on previous items.

Any functions can be used to create the minimum information copulas used here, and in some applications it may be natural to use functions that are themselves computed in computer codes. Because of the frequent evaluation calls needed to determine the minimum information model, it then makes sense to use emulators (particularly, Gaussian process as proposed in [43]) or Kriging models as a way to speed up the computations.

As a future work, we wish to extend the methodologies presented in [39] to estimate the non-Gaussian DAG model and compare it with the proposed method in this paper. It would be compelling to investigate how the model can be simplified further using the approaches presented in [39] and considering the conditional independence statements. Furthermore, it would be desired to extend the modified AIC given in Nagler et al. [39] to a more comprehensive model selection tool so that the non-parametric model presented in this paper could be compared with other non-parametric/parametric model candidates.

#### ACKNOWLEDGMENTS

This research was supported by funding from the UK Engineering & Physical Sciences Research Council (Strategic Package: Centre for Predictive Modelling in Science and Engineering - Grant No. EP/L027682/1).

#### References

- Aas, K., Czado, K. C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance, Mathematics and Economics*, 44, 182–198.
- [2] Acar E. F., Genest C., Neslehova J. (2012). Beyond simplified pair-copula constructions. *Journal of Multivariate Analysis*, **110**, 74–90.
- [3] Bauer, A., Czado,. C., and Klein, T. (2012). Pair-copula constructions for non-Gaussian DAG models. The Canadian Journal of Statistics 40(1), 86-109.
- [4] Bauer, A., Czado, C. (2016). Pair-copula Bayesian networks. Journal of Computational and Graphical Statistics, 25(4), 1248-1271.
- [5] Bedford, T. (2006). Interactive expert assignment of minimally-informative copulae, Management Science Working Paper No. 5.
- [6] Bedford, T. and Cooke. R. M. (2001). Probability density decomposition for conditionally dependent random variables modelled by vines. Annals of Mathematics and Artificial Intelligence 32, 245–268.
- [7] Bedford, T., and Cooke. R. M. (2002). Vines a new graphical model for dependent random variables. *Annals of Statistics*, **30(4)**: 1031–1068.
- [8] Bedford, T., Daneshkhah, A., and Wilson, K. (2016). Approximate Uncertainty Modelling with Vine copulas, Risk Analysis, 36(4), 792-815, doi: 10.1111/risa.12471.
- [9] Bedford, T., and Meeuwissen, A. (1997). Minimally informative distributions with given rank correlation for use in uncertainty analysis. *Journal of Statistical Computation and Simulation*, 57(1 - 4): 143 - 174.

- [10] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics, 31, 307–327.
- [11] Borwein, J., Lewis, A., and Nussbaum, R. (1994). Entropy minimization, DAD problems, and doubly stochastic kernels. *Journal of Functional Analysis*, 123, 264-307.
- [12] Brechmann, E. C., Czado, C., Aas, K. (2012). Truncated regular vines in high dimensions with application to financial data. *Canadian Journal of Statistics*, 40(1), 68-85.
- [13] Chang, V. (2014). The Business Intelligence as a Service in the Cloud. Future Generation Computer Systems 37, 512-534.
- [14] Cowell, R. G., Dawid, A. P., Lauritzen, S. L., and Spiegelhalter, D. J. (2003). Probabilistic Networks and Expert Systems, 2nd ed., Springer, New York.
- [15] Daneshkhah, A., Parham, G., Chatrabgoun, O., and Jokar, M. (2016a). Approximation Multivariate Distribution with pair copula Using the Orthonormal Polynomial and Legendre Multiwavelets basis functions. Communications in Statistics - Simulation and Computation, 45(2), 389-419.
- [16] Daneshkhah, A., Remesan, R., Chatrabgoun, O., and Holman, I. (2016b). Modelling flood characterizations with parametric and minimum information Pair-Copula, *Journal of Hydrology*, 540, 469–487.
- [17] Elidan, G. (2010a). Copula Bayesian networks. In Advances in Neural Information Processing Systems 23, Lafferty, J., Williams, C. K. I., Shawe-Taylor, J., Zemel, R. S., and Culotta, A., editors. NIPS Foundation, La Jolla, California, 559567.
- [18] Elidan, G. (2010b). Inference-less density estimation using copula networks. In Proceedings of the 26th Conference on Uncertainty in Articial Intelligence, Grunwald, P. and Spirtes, P., editors. AUAI Press, Corvallis, Oregon, 151159.
- [19] Embrechts, P., F. Lindskog, and A. J. McNeil. (2003). Modelling Dependence with Copulas and Applications to Risk Management. In Handbook of Heavy Tailed Distributions in Finance. Amsterdam: Elsevier/North-Holland, 31, 307–327.

- [20] Fang, Y., Madsen, L., and Liu, L. (2014). Comparison of Two Methods to Check Copula Fitting, IAENG International Journal of Applied Mathematics, 44(1), 53-61.
- [21] Gui, W. (2009). Adaptive Series Estimators for Copula Densities. PhD thesis, Florida State University College of Arte and Sciences.
- [22] Haff, I. H., Aas, K., Frigessi, A., Lacal, V., (2016). Structure learning in Bayesian Networks using regular vines, Computational Statistics and Data Analysis, 101: 186-208.
- [23] Hanea, A.M., Kurowicka, D., and Cooke, R.M. (2006). Hybridmethod for quantifying and analyzing Bayesian belief nets. Quality and Reliability Engineering International, 22, 709–729.
- [24] Iman, R. and Helton, J. (1985). A comparison of uncertainty and sensitivity analysis techniques for computer models. Technical Report, NUREG/CR-3904 SAND84-1461 RG, Sandia National Laboratories, Albuquerque, New Mexico, USA.
- [25] Joe, H. (1997). Multivariate Models and Dependence Concepts. Chapman & Hall, London.
- [26] Jaynes E. T. (2003). Probability Theory: The Logic of Science. Cambridge: Cambridge University Press.
- [27] Kauermann G., Schellhase C. (2014). Flexible Pair-Copula Estimation in D-vines with Penalized Splines, *Statistics and Computing*, **24(6)**, 1081-1100.
- [28] Kauermann G., Schellhase C., Ruppert, D. (2013). Flexible Copula Density Estimation with Penalized Hierarchical B-splines, Scandinavian Journal of Statistics, 40(4): 685-705.
- [29] Koski, T. J. T, and Noble, J. M. (2012) A Review of Bayesian Networks and Structure Learning, MATHEMATICA APPLICANDA, 40(1), 53103.
- [30] Kurowicka, D., and Cooke. R. (2006). Uncertainty Analysis with High Dimensional Dependence Modelling. John Wiley.
- [31] Kurowicka, D. and Joe, H. (2011). Dependence Modeling: Vine Copula Handbook. World Scientific, Singapore.

- [32] Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright. P. E. (1998). Convergence properties of the Nelder-Mead simplex method in low dimensions. SIAM Journal of Optimization, 9(1): 112–147.
- [33] Lauritzen, S. L. (1996). Graphical Models. Oxford University Press, Oxford.
- [34] Lewandowski, D. (2008). High Dimensional Dependence: Copulae, Sensitivity, Sampling. PhD thesis, Delft University.
- [35] Lopez-Paz D, Hernandez-Lobato JM, Ghahramani Z. 2013. Gaussian process vine copulas for multivariate dependence. Pp. 10–18 in Proceedings of the 30th International Conference on Machine Learning, Vol. 28.
- [36] Mason, J. C., and Handscomb, D. (2003). Chebyshev polynomials. by CRC Press LLC.
- [37] Mendes, B. V. M., Semeraro, M. M., and Leal, R. P. C. (2010). Pair-Copulas Modeling in Finance, Financial Markets and Portfolio Management, 24(2), 193-214.
- [38] Nagler, T., Czado. C. (2016). Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas, *Journal of Multivariate Analysis*, **151**, 69-89.
- [39] Nagler, T., Schellhase, C., Czado. C. (2017). Nonparametric estimation of simplified vine copula models: comparison of methods, arXiv preprint arXiv:1701.00845.
- [40] Pourret, O., Naim, P., and Marcot, B., editors. (2008). Bayesian Networks: A Practical Guide to Applications. John Wiley & Sons, Chichester.
- [41] Sklar, A. (1959). Fonctions de r'epartition 'a n dimensions et leurs marges. Publications de l'Institut de Statistique de l'Universit e de Paris, 8, 229–231.
- [42] Spirtes, P., Glymour, C., and Scheines, R. (2000). Causation, Prediction, and Search, 2nd ed., MIT Press, Cambridge, Massachusetts.
- [43] Stöeber J, Joe H, Czado C. (2013). Simplified pair copula constructionsLimitations and extensions. Journal of Multivariate Analysis, 119:101–118.