Accepted Manuscript

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

Author: Omid Chatrabgoun Amin Hosseinian-Far Victor Chang Nigel G. Stocks Alireza Daneshkhah

PII: S1877-7503(17)30988-2
DOI: http://dx.doi.org/doi:10.1016/j.jocs.2017.09.002
Reference: JOCS 755

To appear in:

Received date: 31-12-2016
Revised date: 8-8-2017
Accepted date: 3-9-2017

Please cite this article as: Omid Chatrabgoun, Amin Hosseinian-Far, Victor Chang, Nigel G. Stocks, Alireza Daneshkhah, Approximating Non-Gaussian Bayesian Networks using Minimum Information Vine Model with Applications in Financial Modelling, *<![CDATA[Journal of Computational Science]]>* (2017), http://dx.doi.org/10.1016/j.jocs.2017.09.002

Dr. Omid Chatrabgoun is an Associate Professor of Malayer University, Iran. He is a specialist of Copula and wavelet methods. He has active publication records available at https://scholar.google.com/citations?user=trIW0c8AAAAJ&hl=en

Dr. Alireza Daneshkhah is a Research Staff Member of Warwick Centre for Predictive Modelling., University of Warwick. He was a Lecturer at Cranfield University, UK. His areas of expertise include Bayesian Statistics, Bayesian network, Predictive modelling and Expert Judgement. His publication records can be seen at https://scholar.google.com/citations?user=Nd7OQx8AAAAJ&hl=en

Dr. Amin Hosseinian-Far is currently a Senior Lecturer and Course Leader in the Faculty of Arts, Environment and Technology at Leeds Beckett University. He received his BSc (Hons) in Business Information Systems from the University of East London, an MSc degree in Satellite Communication and Space Systems from the University of Sussex, a Postgraduate Certificate in Research and a PhD degree titled 'A Systemic Approach to an Enhanced Model for Sustainability' from the University of East London. He has held lecturing and research positions at the University of East London, and at a number of private HE institutions and strategy research firms in London. Prior to joining Leeds Beckett University, Dr Hosseinian-Far worked as Deputy Director of Studies at a large private higher education institute in the capital. Amin is a Member of the Institution of Engineering and Technology (IET), a Fellow of the Higher Education Academy (HEA), and a Fellow of the Royal Society of Arts (RSA).

Dr. Victor Chang is an Associate Professor in Information Management and Information Systems and also a Director of PhD Programme at International Business School Suzhou (IBSS), Xi'an Jiaotong Liverpool, China. Dr. Victor Chang was a Senior Lecturer in Computing at Leeds Beckett University, UK and a Visiting Researcher at the University of Southampton, UK. He has been a technical lead in web applications, web services, database, grid, cloud, storage/backup, bioinformatics, financial computing which subsequently have become his research interests. Victor has also successfully delivered many IT projects in Taiwan (place of birth), Singapore, Australia, and the UK since 1998. Victor is experienced in a number of different IT subjects and has 27 certifications with 97% on average. He completed PGCert (Higher Education, University Greenwich) and PhD (C.S, University of Southampton) within four years while working full-time, whereby the distance between his work and research is about hundreds of miles away. He has over 100 published peer-reviewed papers, including several high-quality journals up-to-date. Victor won £20,000 funding in 2001 (Singapore-Cambridge Trust) and £81,000 funding in 2009 (Department of Health). He was involved in part of the £6.5 million project in 2004, part of the £5.6 million project in 2006 and part of a £300,000 project in 2013. He also won another £200,000 to support his scholarly activities. He is a PI and Co-PI in some projects. Victor is a winner in 2011 European Identity Award in "On Premise to Cloud Migration". He has won a prestigious European award, Best Project in Research, at the European Cloud and Identity Award in 2016. He was selected to present his research at the United Nations in 2003 and at the House of Commons, UK, in 2011. He won the best papers in 2012 and 2015. Dr Victor Chang has taught numerous undergraduate and postgraduate modules since Year 2011. In some modules he taught, students like his teaching and enjoy his labs and lectures. He has given keynotes to international conferences and workshops extensively.

Dr. Omid Chatrabgoun

Dr. Alireza Daneshkhah

Dr. Amin Hosseinian-Far

Dr. Victor Chang

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

Omid Chatrabgoun[a], Amin Hosseinian-Far[b], Victor Chang[c], Nigel G. Stocks[d], Alireza Daneshkhah[e,*]

[a]*Department of Statistics, Faculty of Mathematical Sciences & Statistics, Malayer University, Malayer, Iran*
[b]*Faculty of Science, Technology, Engineering and Mathematics, The Open University, Milton Keynes, MK7 6AA, UK*
[c]*The International Business School Suzhou (IBSS), Xian Jiaotong-Liverpool University, China*
[d]*School of Engineering, University of Warwick, Coventry, CV4 7AL, UK*
[e]*Faculty of Engineering, Environment and Computing, Coventry University, Coventry, CV1 2JH, UK*

Abstract

Many financial modeling applications require to jointly model multiple uncertain quantities to present more accurate, near future probabilistic predictions. Informed decision making would certainly benefit from such predictions. Bayesian Networks (BNs) and copulas are widely used for modeling numerous uncertain scenarios. Copulas, in particular, have attracted more interest due to their nice property of approximating the probability distribution of the data with heavy tail. Heavy tail data is frequently observed in financial applications. The standard multivariate copula suffer from serious limitations which made them unsuitable for modeling the financial data. An alternative copula model called the Pair-Copula Construction (PCC) model is more flexible and efficient for modeling the complex dependence of financial data. The only restriction of PCC model is the challenge of selecting the best model structure. This issue can be tackled by capturing conditional independence using the Bayesian Network PCC (BN-PCC). The flexible structure of this model can be derived from conditional independences statements learned from data. Additionally, the difficulty of computing conditional distributions in graphical models for non-Gaussian distributions can be eased using pair-copulas. In this paper, we extend this approach further using the minimum information vine model which results in a more flexible and efficient approach in understanding the

*Corresponding author
Email address: Ali.Daneshkhah@coventry.ac.uk (Alireza Daneshkhah)

complex dependence between multiple variables with heavy tail dependence and asymmetric features which appear widely in the financial applications.

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

1. Introduction

In the recent years, the copula functions have gained popularity in constructing multivariate distributions and survey dependency structures. 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 evaluated. Building higher dimensional copula is generally a challenging task, and choosing a parametric family for a higher dimensional copula is rather more difficult and limited [16]. This drawback was tackled by applying a more flexible multivariate copula known as vine copula (or PCC) model which has recently developed for modeling multivariate dependency [4, 5, 12]. This modeling structure is based on a decomposition of a multivariate density into a cascade of bivariate copula. Since the vines demonstrate high flexibility and advantages in constructing multivariate distributions, they have recently been used to describe the inner-dependence structure and build the joint distribution of portfolio returns, and uncertain quantities in financial applications and risk analysis. One the main issues with the vines is that the bivariate copulas are restricted to a particular parametric class (Gaussian, multivariate t, etc.) [1]. As a result, the potential flexibility of the vine copula approach is not realized in practice.

There have been recently several attempts to tackle the drawback mentioned above for the multivariate copula including the vine model. The proposed methods are mainly focused on making the vine model more flexible and efficient using the non-parametric vine copula models. Kauermann et al. [13] proposed a non-parametric model using the spline to estimate multivariate copula density. However, the main purpose of this method was to tackle the curse of dimensionality, but it fails to do so. The methodology was improved by using penalized Bernstein polynomials and applied to the D-vine model, to estimate the bivariate copula density in each knot of the model [14]. However, the reported results are more promising, but no clear model selection algorithm is suggested and its performance for modeling weak dependency is still very poor. These methods were extended further in [21] by applying them on the simplified vine copula models. They exhibited that the kernel-based non-parametric estimators performed best, but its performance is worse than penalized B-spline estimators when there is weak dependence and no tail dependence.

Bedford et al. [6] enhanced the flexibility and efficiency of the vine model by proposing an alternative non-parametric method using the minimum information concept. A copula based on the minimum information concept can be constructed by specifying dependency constraints through the use of rank correlations/moments. It was demonstrated that a vine structure can be used to approximate any given multivariate copula to any required degree of approximation. They also illustrated that how this can be operationalized for use in practical situations involving uncertain risks.

Another challenge of the vine models is the selection of the best model based on the observed data. This issue has been recently addressed in [2, 3] by capturing conditional independences in the data which results in a new model called Bayesian network vine (or BN-PCC) model. This presentation provides more parsimonious model in different settings and is structurally more flexible than vine model. However, the BN-PCC suffers from the same drawbacks of the parametric vine models as discussed above. In this paper, we benefit from the simplification algorithm using BN proposed in [2, 3] and efficiency of the density approximation addressed in [6, 10] to approximate any non-Gaussian BN to any required degree of approximation. We illustrate the proposed BN-PCC in this paper is more flexible and efficient in modeling multivariate dependencies of heavy-tailed distribution and tail dependence as observed in the financial data and risk analysis domain, etc. The proposed model is not restricted to use the limited parametric pair-copula models, and can provide a precise approximation in the presence of the large/limited data and the restrictions imposed by the data and problem under study. We formulate these restrictions using various basis functions: Polynomial Series (PS); Orthonormal Polynomial Series (OPS); and Orthonormal Fourier Series (OFS).

The present paper is organised as follows. In Section 2, we present the vine construction associated with the non-Gaussian BN of multivariate data. In Section 3, we first briefly study the minimum information copula and show that how it can be used to approximate a bivariate copula density. We then develop it further to approximate the non-Gaussian BN. We improve this approximation in Section 4, using PS, OPS and OFS basis functions. In Section 5, we examine the performance of the proposed model and compare it with the alternative models given in [2, 3] by analysing the global portfolio data from the perspective of an emerging market investor located in Brazil [19]. A simulation study is illustrated in Section 6, and finally we conclude the paper in Section 7.

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

Considering the above-mentioned vine's drawbacks in modelling multivariate data, there have been several attempts to develop a method through using the nice properties of both graphical model and vine model, simultaneously. The main purpose is to benefit from the conditional independence in the graphs and then simplify the vine structure [11]. Simplified vine copula models give rise to very flexible models which are often found to be superior to other multivariate copula models [1]. Indeed, to make the model more tractable, one usually makes the simplifying assumption that the pair-copula densities do not change with conditional assumption [21].

A Bayesian network (BN) which is certainly the most common and applicable probabilistic graphical model represents a set of random variables (r.vs) and their conditional dependencies via a directed acyclic graph (DAG). The construction of a BN based on the assumption of a joint Gaussian distribution is quite straightforward, but this assumption is not a realistic for capturing the features of real world data such as tail behaviour and non-linear, asymmetric dependencies. This gap was filled in [2] by introducing non-Gaussian graphical model by combining useful properties of both pair-copula and DAG which was then called non-Gaussian BN-PCC. In this paper, we only briefly introduce the BNs' concepts required in this paper, and the preliminary notations of the BNs and their detailed theory can be found in [9].

As mentioned above, the decomposition of a multivariate distribution can be efficiently implemented by benefiting from the conditional independencies offered by a DAG. The density function $f(.)$ of n r.vs, $(X_1, \ldots, X_n)$ can be decomposed as a product of n conditional density functions as follows:

$$f(x_1,\ldots,x_n)=\prod_{i=1}^{n} f(x_i|pa(x_i)), \tag{1}$$

where $pa(x_i)$ represents the parent set of x_i. The density decomposition given in (1) illustrates that once the value of $pa(x_i)$ is learned, knowing the value of the other preceding variables is redundant.

Bauer et al. [2], in the following theorem, illustrate that how the multivariate density given in (1) can be represented in terms of the PCC model.

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 (CDF) 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 $c_{vw|pa(v,w)}, v\epsilon V, w\epsilon pa(v)$ and f can be decomposed as follows:*

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

Proof. See [2] and references cited therein.

The above theorem gives us a constructive approach to build a multivariate distribution given a DAG. In other words, by making suitable choices of marginal densities and pair-copula functions, the above presentation given in (2) provides us an approximation for the multivariate density. 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.

3. Approximating Multivariate Density: A minimum information copula approach

This section outlines a multivariate density approximation approach using the minimum information techniques in conjunction with the observed data or expert elicitation of observables [6]. This can be used construct a multivariate distribution using a Non-Gaussian BN-PCC 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

We apply an algorithm, named *DAD* [6] to generate discretized minimum information copula between two variables with given rank correlation. This method depends on this fact that the correlation is specified by means of the symmetric function U_1U_2. A similar method can be applied whenever the expectation of any symmetric function of U_1 and U_2 must be determined.

We suppose that there exist two r.vs X_1 and X_2, with CDFs $F_1(.)$ and $F_2(\cdot)$, respectively. The main purpose is then to correlate these r.vs according to some constraints which can be represented as the expected values of several functions. These functions should be chosen so that various types of dependency between the r.vs could be illustrated. Suppose there are l of these functions, i.e. $g'_1(X_1, X_2), \ldots, g'_l(X_1, X_2)$, and that we would like to compute their expected values based on the observed data, denoted by $\beta_1, \ldots, \beta_l$ for the considered functions, respectively. It should be noted that the mean values of the functions can be also determined in terms of expert opinions [6]. The associated functions of the copula variables, i.e. $U_1 \in [0, 1]$ and $U_2 \in [0, 1]$ can be simply specified and represented as follows:

$$g_i(U_1, U_2) = g'_i(F_1^{-1}(U_1); F_2^{-1}(U_2)), \quad i = 1, \ldots, l$$

where $g_i : [0, 1]^2 \to \mathbb{R}$, at which the mean values $\beta_1, \ldots, \beta_l$ can be specified that these functions should simultaneously take. In addition, suppose that g_i and g_j are linearly independent for any $i \neq j$. We then pursue a copula that possess these expected values. This optimisation problem could be either impractical or undetermined. Therefore, given tractability of the moment, a copula is considered to be minimum information (regarding the uniform distribution), which guarantees a unique and reasonable solution. The corresponding kernel is then given by

$$A(u_1, u_2) = \exp(\beta_1 g_1(u_1, u_2) + \ldots + \beta_l g_l(u_1, u_2)), \tag{3}$$

where u_1 and u_2 denote the realisations of U_1 and U_2, respectively.

For practical performances, the same approach as given in [6, 10] is used to discretise the values of (u_1, u_2) such that the total space of the copula is covered. It is trivial to demonstrate that the kernel A exhibited in (3) is a two-dimensional matrix, and the main difficulty is then to specify the matrices D_1 and D_2. The following product then becomes a doubly stochastic matrix [6] which exhibits a discretised copula density

$$P = D_1 A D_2. \tag{4}$$

where $P \in [0,1]^2$.

We can use the D_1AD_2 method to uniquely approximate the joint density of the r.vs of interests with uniform marginal distributions and based on the computed Lagrange's coefficients, $(\lambda_1, \ldots, \lambda_l)$. It can be shown that the set of all possible expected values $(\beta_1, \ldots, \beta_l)$ satisfying in (5)

$$E[g_i(U_1, U_2)] = E[g_i'(F_1^{-1}(U_1); F_2^{-1}(U_2))] = \beta_i, \quad i = 1, \ldots, l \tag{5}$$

with respect to some probability distribution is convex. In addition, given all any values of $\{\beta_i\}_{i=1}^{l}$ lie in the interior of this convex set, there is a unique density function [6, 10] with parameters $(\lambda_1, \ldots, \lambda_l)$ computed based on the constraints given in (5).

In order to approximate the copula density based on the D_1AD_2 algorithm, an iterative algorithm is required which will be briefly explained here. We first discretise both (u_1, u_2) into n grid-points, represented as $\{(u_1^{(i)}, u_2^{(j)}), \ \ i, j = 1, \ldots, n\}$. The grid points can be uniformly selected over the copula domain, or chosen based on the purpose of study which will be discussed further in Section 5. We can then define

$$A = (a_{ij}), \quad D_1 = diag(d_1^{(1)}, \ldots, d_n^{(1)}), \quad D_2 = diag(d_1^{(2)}, \ldots, d_n^{(2)}),$$

where $a_{ij} = A(u_1^{(i)}, u_2^{(j)})$, $d_i^{(1)} = D_1(u_1^{(i)})$, $d_j^{(2)} = D_2(u_2^{(j)})$, and $diag(d_1^{(1)}, \ldots, d_n^{(1)})$ stands for a diagonal matrix with the diagonal entries, $(d_1^{(1)}, \ldots, d_n^{(1)})$. The doubly stochastic matrix presented in (4) will be then represented in the following forms

$$\forall i = 1, \ldots n, \quad \sum_j d_i^{(1)} d_j^{(2)} a_{ij} = 1/n, \quad \&$$
$$\forall j = 1, \ldots n, \quad \sum_i d_i^{(1)} d_j^{(2)} a_{ij} = 1/n.$$

The iterative numerical approach required for the D_1AD_2 algorithm is quite simple which begins with selecting arbitrary positive initial matrices for D_1 and D_2, and these matrices will be then successively updated by iterating the following maps

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

It is trivial to illustrate the above iteration scheme will eventually converge in the geometric rates to some matrices to achieve the approximation precision [6].

The next step is to find a suitable set of Lagrange's coefficients, $\{\lambda_i\}_{i=1}^{l}$'s associated with the expected values $\{\beta_i\}_{i=1}^{l}$ at which these values are calculated with respect to the copula density approximated using the D_1AD_2 given in (4). As a results, λ_i's, satisfying the constraints illustrated in (5), can be determined by solving the following set of nonlinear equations:

$$L_k(\lambda_1,\ldots,\lambda_l) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} P(u_1^{(i)}, u_2^{(j)})h_k(u_1^{(i)}, u_2^{(j)}) - \beta_k, \quad k = 1, 2, \ldots, l. \tag{6}$$

In this paper, we use the FMINSEARCH - MATLAB's optimization tool which is developed based on Nelder-Mead simplex method [6] to solve the above nonlinear system of equations. The function to be minimized is then given by

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

We can use a similar approach describe above to estimate a copula given the expected values evaluated based on the experts' opinions [6, 10].

3.2. Approximating Multivariate Density by Non-Gaussian BN-PCC

In the last subsection, we demonstrate how the bivariate copulas as building blocks of the given multivariate model can be approximated using bivariate minimum information copulas which results in a a family of bivariate copulas with some nice features. Since, our main aims is to build BN-PCC model in terms of multiple bivariate copulas, it is very crucial to illustrate the mentioned family of bivariate (conditional) copula densities encompassed in the multivariate distribution of interest must form a *compact set* in the space of continuous functions defined over $[0, 1]^2$. This would allow us to exhibit that the same family of copulas (with finite parameter) can be applied to provide an approximation to every conditional copulas with the same level of approximation.

In order to illustrate this, we need to accurately explain the method in which the densities are approximated. It is plausible to assume all densities are continuous and uniformly bounded away from zero. We denote the space of continuous real valued functions on $Z = [0,1]^p$ for some p with $\mathfrak{C}(Z)$. We define a norm on this space as follows

$$||f_{1\dots p}|| = \sup |f_{1\dots p}(x_1, \dots, x_p)|.$$

where $f_{1\dots p}(.)$'s are some real-valued functions on $Z = [0,1]^p$.

It is trivial to show that the above norm is finite, because Z is a compact set and the functions defined over Z are continuous. We now present the set of all possible two-dimensional copulas corresponding to f as

$$\mathfrak{C}(f) = \{c_{ij|i_1\dots i_p} : 1 \leq i, j, i_1, \dots, i_p \leq n, i, j \neq i_1, \dots, i_p\}. \tag{7}$$

Any $c_{ij|i_1\dots i_p} \in \mathfrak{C}(f)$ presents the copula of conditional density of (X_i, X_j) given $\{X_{i_1}, \dots, X_{i_p}\}$. It should be noted that the set $\mathfrak{C}(.)$ is not finite.

The next step is to exhibit that $\mathfrak{C}(f)$ given in (7) is relatively compact in $\mathfrak{C}([0,1]^2)$ consists of all continuous real valued functions defined over $[0,1]^2$. This would help us to demonstrate that the copula densities can be uniformly approximated. The best way to prove $\mathfrak{C}(f)$ is relatively compact would be to prove the compactness of the following two spaces: $\mathfrak{M}(f)$ and $\mathfrak{B}(f)$. The former one, $\mathfrak{M}(f)$ defines the set of conditional marginal densities, and $\mathfrak{B}(f)$ shows the set of conditional bi-variate densities. We illustrate these space as follows

$$\mathfrak{M}(f) = \{f_{i|i_1\dots i_p} : 1 \leq i, i_1, \dots, i_p \leq n, i \neq i_1, \dots, i_p\},$$
$$\mathfrak{B}(f) = \{f_{ij|i_1\dots i_p} : 1 \leq i, j, i_1, \dots, i_p \leq n, i, j \neq i_1, \dots, i_p\},$$

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

The compactness of these spaces are illustrated in [6]. To prove the compactness of $\mathfrak{C}(f)$, any member $c_{ij|i_1,\dots,i_p} \in \mathfrak{C}(f)$ can be written as

$$c_{ij|i_1,\dots,i_p}(u_i, u_j|x_{i_1}, \dots, x_{i_p}) = \frac{f_{ij|i_1\dots i_p}(x_i, x_j|x_{i_1}, \dots, x_{i_p})}{f_{i|i_1\dots i_p}(x_i|x_{i_1}, \dots, x_{i_p}) f_{j|i_1\dots i_p}(x_j|x_{i_1}, \dots, x_{i_p})} \tag{8}$$

Now, if we consider a sequence of component in elements in $\mathfrak{C}(f)$, we can then find comparable sequences of components in $\mathfrak{M}(f)$ and $\mathfrak{B}(f)$. By knowing that $\mathfrak{M}(f)$ and $\mathfrak{B}(f)$ are relatively compact [6], a convergent subsequence in $\mathfrak{M}(f)$ would results in corresponding convergent functions in $\mathfrak{B}(f)$. This would result in convergence of the right-hand side of (8). That means the components of $\mathfrak{C}(f)$ associated with the considered sequence above have to be converge to the same expression. This proves the compactness of $\mathfrak{C}(f)$ (see also [6] for further details). We can immediately conclude that $\mathfrak{LC}(f) = \{\log(h) : h \in \mathfrak{C}(f)\} \subset \mathfrak{C}([0,1]^2)$ is a compact set. This is evident from the compactness of $\mathfrak{C}(f)$ and this fact that all elements in $\mathfrak{C}(f)$ are positive and uniformly bounded away from zero.

We now combine the results derived above and introduced in the previous section to approximate the copulas based on the sequences of functions in $\mathfrak{C}([0,1]^2)$. Suppose $g_1, g_2, \ldots,$ is any arbitrary and and countable sequence in $\mathfrak{C}([0,1]^2)$ with the following property which any function $h \in \mathfrak{C}([0,1]^2)$ can be illustrated in the following form

$$h = \sum_{i=1}^{\infty} \lambda_i g_i$$

where $\lambda_i \in \mathbb{R}$.

It is trivial to demonstrate that any finite set of basis components, $g_1, \ldots, g_n$ is linearly independent. As a result, given a sorted basis $g_1, g_2, \ldots \in \mathfrak{C}([0,1]^2)$ and a desired approximation level, $\epsilon > 0$, any component of $\mathfrak{LC}(f)$ can be estimated to within the required rate of approximation by a linear combination of $g_1, \ldots, g_l$, where l is appropriately selected to attain the required degree of approximation. The value of l is also dependent on the basis functions which are used to approximate the copula density. In this paper, we only use PS, OPS and OFS basis function to approximate the copula densities of the uncertain quantities due to their nice properties.

However, we can get similar results for $\mathfrak{LC}(f)$, but the number of basis functions used to attain the requested order of approximation could be different. It should be also noted that the proposed approximation for the copula of interest based on the linear combination, $\sum_{i=1}^{l} \lambda_i g_i$ is not totally approved to be a copula density itself. We now discuss how this approximation can be at it can be slightly modified to achieve a copula which produces plausible approximation. The adjustment can be done by weighting the derived density using the D_1AD_2 algorithm as discussed in Subsection 3.1. Using this algorithm the approximated copula based on the linear combination of the basis functions can be considered as a continuous positive real valued function denoted by $A(u_1, u_2)$ on $[0,1]^2$ which could be not a density. In order to make this a density, two continuous positive functions $d_1(u_1)$ and $d_2(u_2)$ exist, such that $d_1(u_1).d_2(u_2).A(u_1, u_2)$ becomes a copula density with uniform marginal distributions. We denote this product by $\mathfrak{C}(A) = d_1(u_1).d_2(u_{2)}.A(u_1, u_2)$ which is also called a C-projection of A. We can summarise the process of ensuring that approximating densities are copula densities in the following lemma which also enables us to manage the precision level of approximating a copula [6].

Lemma 1. *Let h be a positive continuous copula density. Given the order of approximation $\epsilon > 0$, there exists a positive real value $\gamma > 0$ such that if $||h - f|| < \gamma$, then $||h - \mathfrak{C}(f)|| < \epsilon$.*

It should be noted that the following relationships between the re-weighting functions can be presented

$$d^{(1)}(u_1) = \frac{1}{\int d^{(2)}(u_2) f(u_1, u_2) du_2} \quad \& \quad d^{(2)}(u_2) = \frac{1}{\int d^{(1)}(u_1) f(u_1, u_2) du_1}.$$

This also approves that these re-weighting functions possess the same differentiability properties as the function f being re-weighted.

Eventually, the equation presented in (2) can be used to demonstrate that good approximation of each conditional copula would result in a good approximation of the multivariate density represented by the BN-PCC.

4. Building approximations using minimum information distributions

In Section 3, we present a method that all conditional copulas in the BN-PCC model can be approximated using linear combinations of basis functions. In this section, we provide a practical guide for approximating the multivariate density of the observed data using a minimum information BN-PCC. In order to approximate the joint density of several variables connected through a DAG, the densities between any pair of variables are approximated using the minimum information copulas as described in Section 3 based on the expected values of the selected basis functions and the required approximation precision. Each copula appeared in the representation of the multivariate density given in (2) is approximated, in terms of the linear combination of the selected basis functions, $\{1, g_1, \ldots, g_l\} : [0,1]^2 \to \mathbb{R}$, by

$A(u_1, u_2) = \exp(\sum_1^l \lambda_i g_i(u_1, u_2))$. The Lagrange coefficients $\{\lambda_i\}_{i=1}^l$ are determined by solving the set of nonlinear equations given in (6) as explained in Subsection 3.1. The final copula density can be uniquely determined, by adjusting $A(u_1, u_2)$ using the $D_1 A D_2$ algorithm as follows

$$d^1(u_1) d^2(u_2) \exp(\sum_{i=1}^{l} \lambda_i g_i(u_1, u_2)).$$

We summarise the steps required for approximating the multivariate data connected through a BN with the density factorisation given in (2) using the proposed method in Algorithm 1.

Algorithm 1 Approximating a non-Gaussian BN using minimum information vine copula model

1: **Input:**

1. Observed data $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_p) \in \mathbb{R}^{n \times p}$, where $\mathbf{x}_i = (x_{i1}, \ldots, x_{in})^T$, for $i = 1, \ldots, p$;
2. The approximation precision, ϵ.

2: Learn a DAG structure from the observed data;

3: Specify a basis family, as $\{g_1, g_2, \ldots\}$;

4: For each copula in (2) associated with the DAG, determine either

1. the expected values of the basis function, $\beta_1, \beta_2, \ldots$ or
2. the mean values, $\beta_m(ji \mid D_e)$ as the functions of the conditioning variables, for $m = 1, \ldots, l$.

5: Approximate each pair-wise copula in (2) using the minimum information method based on the selected basis and computed β's values.

6: Approximate the multivariate density associated with the BN by replacing the approximated copulas in (2).

4.1. The Basis Family

In approximating multivariate density presented above, the log-density of each pair-copula appeared in (2) density is approximated by truncating the selected basis family at a point determined in accordance to the required approximation error. The proposed approximation could be computed must faster with better fit to data by selecting more efficient basis functions [10]. We demonstrate that the *orthonormal polynomial* series and *Fourier* basis function will outperform the approximation derived by using the ordinary polynomial series using the proposed method and the alternative methods studied in [2, 3]. In the following subsections, we introduce these basis families and briefly discuss their advantages and drawbacks in approximating a multivariate density using the proposed method.

4.1.1. Ordinary Polynomial base

One of the simple basis that can be applied in minimum information copula is ordinary polynomial basis. These basis was mainly used in Bedford et al. (2016) and can be defined simply as follows:

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

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

4.1.2. Orthonormal polynomial base

Two polynomial functions g_1 and g_2 are called orthonormal on $[0, 1]$, if

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

The orthonormal polynomial base (OPS) can be then constructed more conveniently than some other natural basis using this definition. The main benefit of these basis function over the OP basis is that the D_1AD_2 algorithm converges in a swifter manner. This is mainly due to property of orthogonal basis family at which adding a new bases does not change the already used Lagrange coefficients in $A(u_1, u_2) = \exp(\sum_1^l \lambda_i g_i(u_1, u_2))$. This is not the case for the OP bases where any new item in general has a non-zero projection on previous items. It means that the already derived coefficients of the series expansion could be altered.

The most common orthonormal polynomial basis function is the Gram-Schmidt OPS which can be defined over $[0, 1]$ as follows

$$\varphi_0(u) = 1$$

$$\varphi_n(u) = \frac{u^n - \sum_{j=0}^{n-1} \frac{\int_0^1 u^n \varphi_j(u)du}{\int_0^1 \varphi_j^2(u)du} \varphi_j(u)}{||u^n - \sum_{j=0}^{n-1} \frac{\int_0^1 u^n \varphi_j(u)du}{\int_0^1 \varphi_j^2(u)du} \varphi_j(u)||} \quad n \geq 1.$$

4.1.3. Fourier base

Trigonometric or Fourier basis is the other type of orthonormal basis. The computational speed of these basis function for the periodic data is much faster. The first six Fourier bassis functions are defined as

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

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 to approximate the multivariate distribution associated with the selected BN-PCC 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 [2, 3].

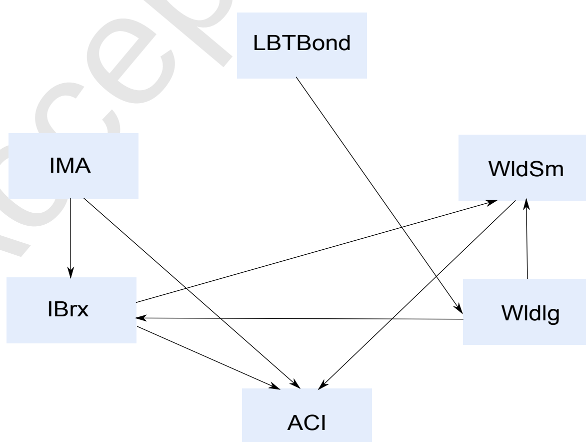


Figure 1: Selected DAG structure for six dimensional contemporaneous daily log-returns of the global portfolio data from the perspective of an emerging market investor located in Brazil.

We use the same data set as originally studied in [19] to illustrate the approximation method introduced in this paper. The data consists of six dimensional contemporaneous daily log-returns: 1) Arsenal composite index (ACI); 2) IMA-C index which represents the Brazilian treasury bonds inflation; 3) IBRX, a stock index related to 100 biggest capitalization companies; 4) WLDLg is an index of large world stocks; 5) WLDSm is an index of small capitalization world companies; 6) LBTBond is an index of total returns on US treasury bonds. The total of 1629 data are collected over the period 02-01-2002 to 20-10-2008.

The serial correlation in these six time series must be first removed, i.e. the observed data of each variable must be independent over time. Thus, we respectively model the serial correlation in the conditional mean and variance the AR(1) and GARCH(1,1) models [7]. The following model for log-return of x_i is then proposed:

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

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

The further analysis is performed on the standardized residuals z_i. If 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 (see [1] for the details of these statistics). 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 into the uniform variables using the kernel method before any further modelling. 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 BN-PCC approximation fitted to this data set using the minimum information distributions based on the different basis functions described above. The main challenge for this approximation is in linking DAG models to the vines. We first need to learn DAG structure from the observed data. One approach is applying the structure learning algorithms, such as the PC algorithm (see [22] Section 5.4.2) to Φ^{-1}(data), 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 elicit the DAG structure (see [15], Chapter 5). There are also model structure selection algorithms for the non-Gaussian data [3] which is based on the PC algorithm again. We adopt the DAG structure presented in Figure 1 derived by applying the PC algorithm introduced in [3] for non-Gaussian distributions. Given the derived DAG, we can decompose the multivariate density of our data by applying Theorem 1 in order to derive BN-PCC model. In other words, given the DAG structure, Theorem 1 prescribes which pair copulas need to be specified in the definition of our model. Note that variable 1(ACI) has three parents (2(IMA), 3(IBrX), 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 $\hat{\tau}$ of Kendall's estimates between the variables: $\hat{\tau}_{15} = 0.209$, $\hat{\tau}_{13} = 0.197$, and $\hat{\tau}_{12} = 0.127$. Similar rule can be applied for variables 3(IBrX) and its parents (2(IMA) and 4(WLdLg)) based on $\hat{\tau}_{32} = 0.0858$, and $\hat{\tau}_{34} = 0.424$. The $\hat{\tau}$'s Kendall estimates between 5(WldSm) and its parents (3(IBrX) and 4(WIdIg)) are: $\hat{\tau}_{53} = 0.402$ and $\hat{\tau}_{54} = 0.75$. Based on these ordering, the resulting multivariate density decomposition is:

$$
\begin{aligned}
f_{1,\ldots,6}(x_1,\ldots,x_6) = \prod_{i=1}^{6} f_i(x_i) \times c_{15}(F_1(x_1), F_5(x_5)) \times c_{45}(F_4(x_4), F_5(x_5)) \times c_{46}(F_4(x_4), F_6(x_6)) \\
\times c_{34}(F_3(x_3), F_4(x_4)) \times c_{13|5}(F_{1|5}(x_1|x_5), F_{3|5}(x_3|x_5)) \times c_{23|4}(F_{2|4}(x_2|x_4), F_{3|4}(x_3|x_4)) \\
\times c_{35|4}(F_{3|4}(x_3|x_4), F_{5|4}(x_5|x_4)) \times c_{12|35}(F_{1|35}(x_1|x_3,x_5), F_{2|35}(x_2|x_3,x_5))
\end{aligned}
\tag{9}
$$

We now derive the minimum information copulae in association with some moment constraints between copula variables 1, 2, 3, 4, 5, 6 in the density decomposition (9) . We initially construct minimum information 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 [10] 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,

$$\psi_1(.)\psi_1(.), \psi_1(.)\psi_2(.), \psi_2(.)\psi_1(.), \psi_1(.)\psi_3(.), \psi_3(.)\psi_1(.),$$
$$\psi_2(.)\psi_2(.), \psi_2(.)\psi_3(.), \psi_3(.)\psi_2(.), \psi_1(.)\psi_4(.), \psi_4(.)\psi_1(.),$$
$$\psi_1(.)\psi_5(.), \psi_5(.)\psi_1(.), \psi_2(.)\psi_4(.), \psi_4(.)\psi_2(.), \psi_3(.)\psi_3(.), \ldots$$

OPS basis function constructed using Gram-Schmidt process

$$\varphi_1(.)\varphi_1(.), \varphi_1(.)\varphi_2(.), \varphi_2(.)\varphi_1(.), \varphi_1(.)\varphi_3(.), \varphi_3(.)\varphi_1(.),$$
$$\varphi_2(.)\varphi_2(.), \varphi_2(.)\varphi_3(.), \varphi_3(.)\varphi_2(.), \varphi_1(.)\varphi_4(.), \varphi_4(.)\varphi_1(.),$$
$$\varphi_1(.)\varphi_5(.), \varphi_5(.)\varphi_1(.), \varphi_2(.)\varphi_4(.), \varphi_4(.)\varphi_2(.), \varphi_3(.)\varphi_3(.), \ldots$$

and then considered OFS basis functions are:

$$\phi_1(.)\phi_1(.), \phi_1(.)\phi_2(.), \phi_2(.)\phi_1(.), \phi_1(.)\phi_3(.), \phi_3(.)\phi_1(.),$$
$$\phi_2(.)\phi_2(.), \phi_2(.)\phi_3(.), \phi_3(.)\phi_2(.), \phi_1(.)\phi_4(.), \phi_4(.)\phi_1(.),$$
$$\phi_1(.)\phi_5(.), \phi_5(.)\phi_1(.), \phi_2(.)\phi_4(.), \phi_4(.)\phi_2(.), \phi_3(.)\phi_3(.), \ldots$$

Following the explanations to select basis function in an optimal manner, we add the basis functions by using stepwise method in [10]. 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 [10], in order to get optimal results, first four bases have been considered.

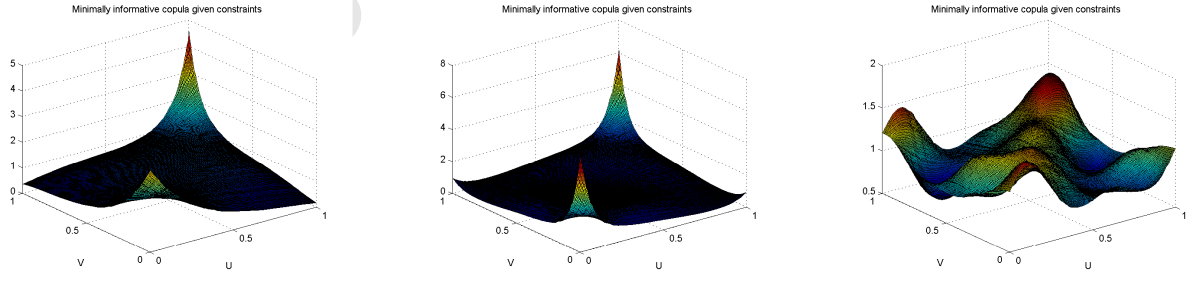
We are now able to construct the minimum information copula density C_{15} with respect to the uniform distributions given the corresponding OP, OPS and OFS constraints above, using the method described in this paper. We first need to determine the number of discretization points (grid size). Simply, a larger grid size will provide a better approximation to the continuous copula, but at the cost of more computation time. Similarly, the approximation will become more precise, if we run the D_1AD_2 algorithm in more iterations. Indeed, 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 1×10^{-1} to

1×10^{-24}. Thus, the larger the number of grid points used, the larger the number of iterations are needed for convergence which is true over all error levels. The grid sizes all follow the same pattern with large increases in the number of iterations needed for improved accuracy initially and smaller increases when the error is smaller. We choose a grid size of 200×200 throughout of this example.

Based on the information given above regarding the grid size, number of iterations and error size, we can derive the minimum information copula C_{15} associated with the chosen constraints. Expectations β of the selected basis, Lagrange multiplies values (parameter values) λ and Log-Likelihood are summarized in Table 1. 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 2, respectively.

Method	Base	$(\beta_1, \beta_2, \beta_3, \beta_4)$	$(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$	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



(a) (b) (c)

Figure 2: 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

Note that, the minimum information copula in BN-PCC structure can be also computed by choosing the grid points such that more points are included in the tail of the distribution instead of the uniform grid points. This could result in outperforming the Gaussian models by the non-Gaussian models approximated based on the proposed method. To verify this claim, we have used Chebyshev points for our grid in copula approximation using minimum information method instead of uniform grid. The main reason behind choosing the Chebyshev points is that they allow for more points in the tail or boundaries of our approximation which are very important, particularly in the Financial applications. Chebyshev points are roots of Chebyshev polynomial which full discussion with some details are presented in [18]. In order to compare uniform grid points with Chebyshev ones, we use the same information given to compute the minimum information copula of interest based on the Chebyshev grid points. Figure (3) shows the minimum information copula C_{15} illustrated over the Chebyshev grid. As mentioned above, by including more points in the tails, the copula density with the heavy tails can be more accurately approximated as illustrated in Figure 3.

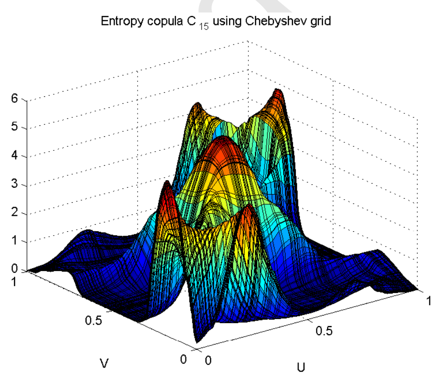


Figure 3: The minimum information copula C_{15} using Chebyshev grid.

One of the main advantages of using OPS and OFS bases over the ordinary polynomial series (studied in details in [6]) 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 minimum information 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 [6] to calculate the minimum information 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 to converge. The optimisation time 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 stepwise 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 4. The approximated minimum information copula for these unconditional copula in terms of the PS, OPS and OFS bases is shown in Panels of Figure 4.

Copula	Base	$(\beta_1, \beta_2, \beta_3, \beta_4)$	$(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$	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}

Now, the conditional copulas $C_{13|5}$, $C_{23|4}$ and $C_{35|4}$ can similarly be approximated using the proposed approach. We only illustrate construction of the conditional minimum information copula, $C_{13|5}$, and the other two copulas, $C_{23|4}$ and $C_{35|4}$ can be similarly approximated. In order to calculate this copula, we divide the support of 5 into some arbitrary sub-intervals or bins and then construct 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

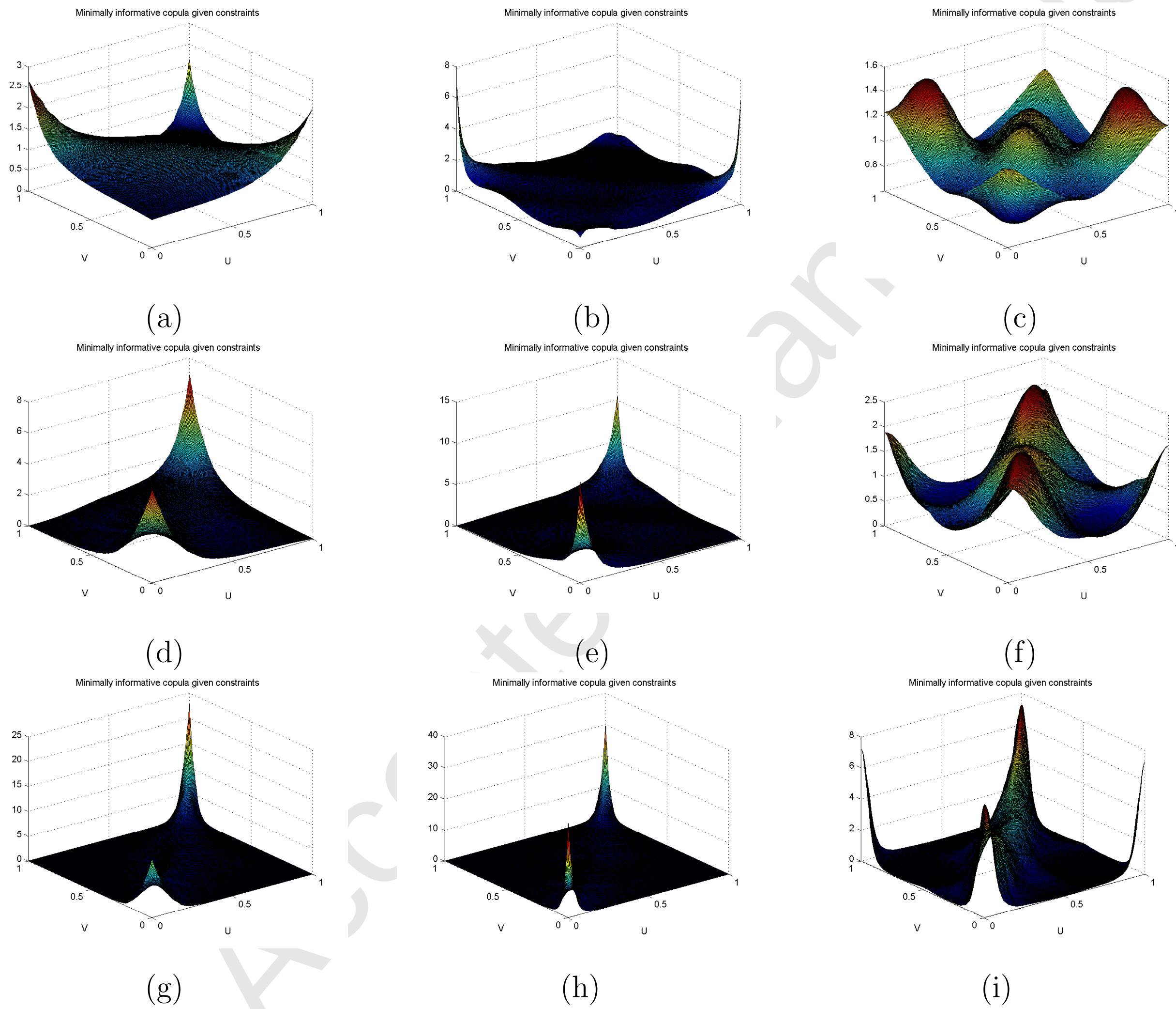


Figure 4: The minimally informative copula given moment constraints, Panel (a):C_{46} for PS basis, Panel (b):C_{46} for OPS basis, Panel (c):C_{46} for OFS basis, Panel (d):C_{34} for PS basis, Panel (e):C_{34} for OPS basis, Panel (f):C_{34} for OFS basis, Panel (g):C_{45} for PS basis, Panel (h):C_{45} for OPS basis, and Panel (i):C_{45} for OFS basis.

values or constraints. Here, 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. Tables 3 show the mean values or constraints (denoted by β_i) and corresponding Lagrange multipliers (λ_i) required to build the conditional minimum information 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.

Interval	Bases	$(\beta_1, \beta_2, \beta_3, \beta_4)$	$(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$	L
$0 < M < 0.25$	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
	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
$0.25 < M < 0.5$	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
	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
$0.5 < M < 0.75$	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
	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
$0.75 < M < 1$	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
	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: Minimally informative copula, $C_{13|5}$, given the moment constraints between $(1, 3)$ given 5

We can obtain the conditional minimum information copula, $C_{12|35}$, similarly by dividing each of the conditioning variables' supports into four bins. Then the minimum information 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 BN-PCC model using the PS, OPS and OFS bases, derived by adding the log-likelihoods of the copulas constructed above, are 2390.44, 2669.69 and 2093.75, respectively. Since the comparison based on comparing the log-likelihood of presented non-parametric model in this paper and the parametric model given in [2] is not

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: The minimum information copula, $C_{12|35}$ for the given moment constraints between $(1, 2)$ given $(3, 5)$

Type of copula	AIC
Bauer et al [2] method	-3078.62
Minimum information copula using OFS base	-4187.24
Minimum information copula using PS base	-4780.88
Minimum information copula using OPS base	-5339.38

Table 5: Comparison between the models proposed in this paper and the ones given in [2, 3].

sufficient, and the model complexity measured by the number of parameters is left without consideration. Therefore, we compare these methods based on the Akaike information criteria (AIC) which includes the model complexity. The AIC of the overall Non-Gaussian BN-PCC 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 fitted Non-Gaussian BN-PCC models to the data using Bauer et al [2] method (with AIC equals to -3078.62). We illustrate the corresponding results in Table 5.

6. Simulation study

We now discuss the simulation of data from the presented minimum information BN-PCC and make comparisons between correlations in the simulated and observed data in terms of 2000 simulations. The simulation method is simple and is based on sampling from the CDFs [15]. The same methodology has been used in [3] to draw sample from the parametric BN-PCC. This simulation strategy is explained in the following steps:

1. Draw n samples from two independent r.vs distributed uniformly on $[0, 1]$, the samples are shown by $\{(u_1^{(i)}, u_2^{(i)}),\ i = 1, \ldots, n\}$;
2. Compute the values of the original variables using the following equations:

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

   where $x_j^{(i)}$ is realization of X_j.
3. Repeat this to any copula in BN-PCC model given in (2) by appropriately taking into account the order of the child and parents variable in this simulation.

It can then be recognised that the dependence pattern of the simulated and original data are similar. Table 6 displays the rank correlations between the interested variables calculated from the original observed data, and based on the simulated data taken from the fitted BN-PCC through minimum information copula based on OPS basis. Other base to shorthand and to prevent a repeat procedure in the simulation has been removed. By comparing the computed correlations, it can be accomplished that there is a strong consistency between the mentioned correlations. A similar comparisoon can be implemented between the minimum information BN-PCC derived based in the OPS basis functions and the parametric counterpart. The presented results shows stronger consistency between the estimated correlations based on the proposed method and the observed ones.

7. Discussion and Conclusions

One of the applications of Gaussian distributions are in modeling 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 copula in the above-mentioned areas, would be the claim allocations and fees' assignments for investigators, experts, etc. as part of Allocated Loss Adjustment Expense (ALAE) processes. An additional case for the application of copula, 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, is that those techniques are not capable of modeling 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 modeling such probability distributions. In the applications where tail properties are important for predictive probabilistic modeling, many of the existing techniques are limited and inadequate. One of the well-known techniques that can appropriately infer from tail properties is the multivariate Gaussian copula. As stated above, many of the current techniques used for financial application modeling, as-

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 BN-PCC

	$LBTBond$	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

Minimum information BN-PCC

	$LBTBond$	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

Table 6: The rank correlation coefficients calculated from the observed data & simulated from the proposed method

sume a normal Gaussian distribution of events for simplifying the complex nature of financial scenarios [6, 8, 23]. The proposed methodology for utilising vine structure for approximation, would enable the modeller to establish non-constant conditional correlations, and minimise the chance of risk underestimation.

In this paper, we have developed a novel method to approximate the complex dependence between several variables using the BN-PCC model. 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 of the constraints on each pairwise copula. We have considered a wide range of computationally efficient basis functions including PS, OPS and OFS bases. Using either OPS and OFS bases, the density approximation can be implemented much faster due to the suitable properties explained above. The functions used in our method can be altered to other suitable functions fitted for other applications. For instance, frequent runs of complex codes/function for specifying the minimum information distribution could be computationally very expensive, one could use Gaussian process emulators or Kriging models as a way to speed up the computations. Moreover, the Gaussian process can be used for estimating fully conditional vines and making the computation of the density approximation more tractable [17].

The existing methods, such as the Bayesian logic program, relational dependency networks, relational Markov networks, Bayesian networks build a graph to represent the conditional dependence structure between random variables. However, they tend to force the local quantitative part of the model to take a simple form, but the complex dependencies between high-dimensional variables are difficult to capture. We have already illustrated that the proposed method would be very efficient in understanding the complex dependence between the multiple variables, with moderately large sizes as discussed in Section 5, with tail dependence and asymmetric characteristics which appear widely in the financial applications. In future research, we wish to explore and illustrate the potential of the proposed method for modeling complex dependence between a set of high-dimensional variables which is a critical but challenging problem in many financial applications including financial markets, driving complex market movements, portfolio return and risk, etc. One approach to tackle the curse of dimensionality for the BN-PCC model is to construct a BN structure from the simplified vine model. However, this issue is not explicitly addressed in [2, 3], but there have recently been some interests in connecting BN's with the simplified vine for the similar purposes of this paper. The simplified vine could be developed based on limited number of parents and truncated vine copulas under reasonable conditions [20] or through truncated partial regular vine copula model [23].

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) is acknowledged.

References

[1] 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] Bauer, A., Czad,. C., and Klein, T. (2012). Pair-copula constructions for non-Gaussian DAG models. *The Canadian Journal of Statistics* **40(1)**, 86-109.

[3] Bauer, A., Czad,. C. (2016). Pair-copula Bayesian networks. *Journal of Computational and Graphical Statistics,* **25(4)**, 1248-1271.

[4] Bedford, T. and Cooke. R. M. (2001). Probability density decomposition for conditionally dependent random variables modelled by vines. *Ann. of Math and AI*, **32**, 245–268.

[5] Bedford, T., and Cooke. R. M. (2002). Vines - a new graphical model for dependent random variables. *Annals of Statistics*, **30(4)**: 1031–1068.

[6] Bedford, T., Daneshkhah, A., and Wilson, K. (2016). Approximate Uncertainty Modelling with Vine copulas, *Risk Analysis*, **36(4)**, 792-815.

[7] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* , **31**, 307–327.

[8] Chang, V. (2014). The Business Intelligence as a Service in the Cloud. *Future Generation Computer Systems* **37**, 512-534.

[9] Cowell, R. G., Dawid, A. P., Lauritzen, S. L., and Spiegelhalter, D. J. (2003). Probabilistic Networks and Expert Systems, 2nd ed., Springer, New York.

[10] Daneshkhah, A., Parham, G., Chatrabgoun, O., and Jokar, M. (2016). Approximation Multivariate Distribution with pair copula Using the Orthonormal Polynomial and Legendre Multiwavelets basis functions. *Comm. in Stat. - Simu. & Comput.*, **45(2)**, 389-419.

[11] Hanea, A.M., Kurowicka, D., and Cooke, R.M. (2006). Hybrid method for quantifying and analyzing Bayesian belief nets. *Quality & Reli. Eng. Int.*, **22**, 709–729.

[12] Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.

[13] Kauermann G, Schellhase C, Ruppert D (2013). Flexible Copula Density Estimation with Penalized Hierarchical B-splines, *Scandinavian Journal of Statistics*, **40(4)**: 685-705.

[14] Kauermann G, Schellhase C (2014). Flexible Pair-Copula Estimation in D-vines with Penalized Splines, *Statistics and Computing*, **24(6)**, 1081-1100.

[15] Kurowicka, D., and Cooke. R. (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*. John Wiley.

[16] Kurowicka, D. and Joe, H. (2011). *Dependence Modeling: Vine Copula Handbook*. World Scientific, Singapore.

[17] Lopez-Paz D, Hernandez-Lobato J. M, 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.*

[18] Mason, J. C., and Handscomb, D. (2003).*Chebyshev polynomials*. by CRC Press LLC.

[19] 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.

[20] Muller, D, and Czad, C. (2016). Representing sparse Gaussian DAGs as sparse R-vines allowing for non-Gaussian dependence, *arXiv preprint arXiv: 1604.04202*.

[21] Nagler, T., Schellhase, C., Czado. C. (2017). Nonparametric estimation of simplified vine copula models: comparison of methods, *arXiv preprint arXiv:1701.00845*.

[22] Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search, 2nd ed.*, MIT Press, Cambridge, Massachusetts.

[23] Wei, W., Yin, J., Li, J., and Cao, L. (2014), Modelling Asymmetry and Tail Dependence among Multiple Variables by Using Partial Regular Vine, in *Proceedings of the SIAM International Conference on Data Mining Conference (SDM14)*, 776-784.

- We demonstrate a new way of Approximating Non-Gaussian Bayesian Networks.
- We develop and use Minimum Information Vine Model with Applications in Financial Modelling to illustrate our research work.
- We explain both theoretical work and how to develop step-by-step, and eventually validate our research work.
- Our proposed work has research contributions for computational and algorithmic finance.