Probabilistic Modeling of Financial Uncertainties

Alireza Daneshkhah[1], Amin Hosseinian-Far[2], Omid Chatrabgoun[3], Tabassom Sedighi[4] and Maryam Farsi[5]

[1]Faculty of Engineering, Environment and Computing, Coventry University, CV1 2JH, UK
[2]Faculty of Business and Law, University of Northampton, NN2 7AL, UK,
[3]Faculty of Mathematical Sciences & Statistics, Malayer University, Iran
[4]Complex Systems Centre, Cranfield University, MK43 0AL, UK,
[6]Through-life Engineering Services Institute, Cranfield University, MK43 0AL, UK

ABSTRACT

Since the global financial crash, one of the main trends in the financial engineering discipline has been to enhance the efficiency and flexibility of financial probabilistic risk assessments. Creditors could immensely benefit from such improvements in analysis hoping to minimise potential monetary losses. Analysis of real world financial scenarios require modeling of multiple uncertain quantities with a view to present more accurate, near future probabilistic predictions. Such predictions are essential for an informed decision making. In this paper, we extend Bayesian Networks Pair-Copula Construction (BN-PCC) further using the minimum information vine model which results in a more flexible and efficient approach in modeling 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 Pair-Copula Construction (PCC) can approximate any non-Gaussian BN to any degree of approximation. Our proposed method has been applied to the portfolio data derived from a Brazilian case study. The results show that the fitting of the multivariate distribution approximated using the proposed model has been improved compared to other previously published approaches.

Keywords: Complex Dependencies, Financial Modeling, Heavy-tailed Densities, Non-Gaussian Bayesian Network, Vine Copula Model

INTRODUCTION

In recent years, there have been advancements in the efficiency improvement of computational algorithms for financial risk modeling (Ramachandran & Chang, 2014; Zhang et al., 2018), development of new methodologies for tackling financial prediction in the era of Big Data (Jeon et al., 2018) and the introduction of emerging technologies and platforms (Chang et al., 2012; Chang et al., 2017) such as cloud that can benefit financial modeling and prediction. There are numerous studies that have proved the assumption of normal distribution in financial asset returns wrong. In recent decades, there have been numerous research works and case studies that have verified the fact that financial applications entail heavy-tail distribution. This is where deviation from the mean is far greater than normal distribution. Moreover, understanding the implications of heavy-tail distributions is crucial, specifically when assessing financial risk. The key to financial risk assessment is to maximise the profit and/or return on investment whilst minimizing the potential realistic risks. Risks have impact, likelihood or probability which will assist us in calculating the risk value. However, catastrophes usually ensue upon convergence of the most extreme events and

therefore calculating the probability of risk manifestation is the key to an informed decision making. We should also note that heavy tailed and super heavy tailed distributions do not only appear in financial settings, and they also exist in a variety of other applications and domains, including but not limited to environment and weather data, electronic engineering for instance heavy tailed noise, hospital patients' stay statistics and others. In financial applications, Bayesian Networks (BNs) and copulas are two common approaches to modeling joint uncertainties with probability distributions. In particular, copulas have acclaimed more popularity; this is due to fact that with copula we can approximate the probability distribution of the data with heavy tail, which is in fact very common in financial applications (Ibragimov & Prokhorov, 2017).

BACKGROUND

Construction of multivariate distribution would assist us in appropriately examining dependencies between multivariate data in real world complexities. In recent years, copulas 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 obtained. A more flexible multivariate copula known as the vine copula model has been recently developed (Bedford et al., 2016) for modeling multivariate dependency. This hierarchical graphical model was firstly introduced in (Harry, 1997) and later was formulated in (Bedford & Cooke, 2001; Bedford & Cooke, 2002). Its structure is based on decomposition of a multivariate density into a cascade of bivariate copula. Pair copula construction solves the limitation in construction of multivariate copula, and furthermore considers the dependency between pair of variables. Bedford et al. (2016) enhanced the flexibility of the vine model by proposing a non-parametric bivariate copula. They proposed an alternative method which was based on using minimum information copulas. Their proposed approximation could offer any level of precision based on the model constraints. In contrast, our method offers greater flexibility in specifying copulas. It is straightforward to implement, with the only technical assumption of considering the copula density as continuous and non-zero. An appropriate approach to build an entropy copula or specifying dependency constraints is by moments (Daneshkhah et al., 2012). Moments can be specified either on the copula or on the underlying bivariate density. These moment constraints are considered as real-valued functions $\emptyset_1$,..., $\emptyset_k$ that are required to accept expected values e_1,…, e_k, respectively. The expected values are then computed based on the provided constraints i.e. provided data or experts' beliefs. In this paper, we improve the fitted multivariate density approximation proposed in (Bauer et al., 2012; Bauer & Czado, 2012) using a newly developed approximation method based on the entropy method. The conditional and joint probabilities of this BN can be derived by constructing an entropy copula between the nodes of interests given their parents' sets. An entropy 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 entropy copula. We verify our claim by comparing our approximation with the results illustrated in (Bauer et al., 2012) to model the global portfolio data from the perspective of an emerging market investor located in Brazil.

THE PROPOSED APPROXIMATION METHOD

The standard multivariate copula is not capable of modeling multivariate financial data. An alternative to the standard copula model is the PCC model which is more flexible and efficient for multivariate data. Nevertheless, PCC model optimization becomes a computationally challenging, when the number of variables increases. Bauer et al. (2012) attempted to resolve this by capturing conditional independences within data, and proposed a new model called Bayesian Network PCC (BN-PCC). This new model is structurally more flexible than PCC because of capturing conditional independences within a data structure. Additionally, the challenge of computing conditional distributions in graphical models for non-Gaussian distributions can be eased using bivariate copulas. Our proposed method extends this approach further through using minimum information vine model, with a view to computing multivariate dependencies of heavy-tailed distribution and tail dependence (as observed in the financial data) in a more efficient and flexible approach. Our proposed 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. (2012), the proposed model is not restricted to the use of parametric pair-copula models. In our approach pair-copulas can be approximated using the maximum entropy concept given the limited observed data by truncating the corresponding polynomials/bases after *k* terms, in order to meet the restrictions imposed by the data and problem under study. In the case study application, we examine three different bases: ordinary polynomial, orthonormal and Fourier series and propose the best fitting model based on a goodness-of-fit criterion.

APPLICATION OF THE CASE STUDY

OP, OPS and OFS basis families are used to approximate the multivariate distribution associated with the selected PCCDAG structure corresponding to case study portfolio data. In this section, we attempt to demonstrate the superior flexibility of our approach compared to the method proposed by Bauer et al. (2012).

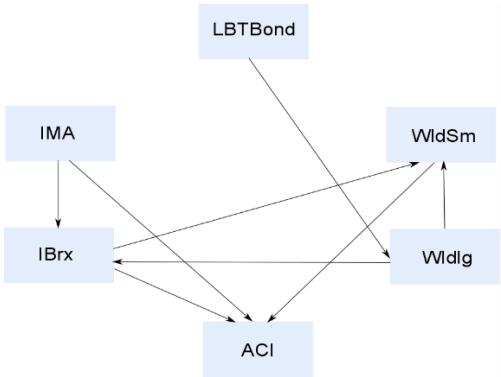

Figure 1: The DAG structure for Brazilian market investor's daily log returns; the structure has six dimensions and the log returns are occurring in the same period of time

Example: In this example, we have deployed the same data set used in Mendes et al. (2010) and in Chatrabgoun et al. (2018) with a view to clearly outline approximation method and for comparison purposes. The data consists of concurrent daily log-returns. The dimensions are Brazilian composite hedge fund index (ACI), a long-term inflation-indexed Brazilian treasury bonds index (IMA-C), 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). These data are gathered for the windows between January the 2nd, 2002 and October the 20th, 2008 in which 1629 data items are collected.

In the first step, the serial correlation within the six-time series are cleared away, to pursue the practice of independence of variable observations over time. Henceforth, the serial correlation in the conditional mean and the conditional variance are modelled by an AR (1) and a GARCH (1,1) model (Aas et al., 2009) respectively.

To generate a PCC-DAG approximation fitted to this data set using entropy distributions and based on the different basis, there is the real challenge of connecting DAG models to vines. To specify DAG structure in our data, we apply structure learning algorithms such as the PC algorithm to the data $\emptyset^{-1}(.)$, where $\emptyset$ denotes the standard normal cdf. Such a transformation is required, as the assessment of conditional independence at %5 significance level by the PC algorithm assumes normality. As an alternative approach, expert knowledge is frequently exploited to define DAG (Kurowicka & Cooke, 2006). There are also structure selection algorithms for non-Gaussian DAG's available in Bauer et al. (2012) which are similarly based on the PC algorithm. We adopt the DAG structure presented in Figure 1 by applying the PC algorithm. Moreover, the presented structure for non-Gaussian available DAG in (Bauer et al., 2012) produces the same results. Considering the DAG, we decompose the multivariate density of the data by applying Theorem 1 in order to derive PCC-DAG structure i.e. given the presented DAG, Bauer et al.'s theorem Bauer et al. (2012) 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), and 5 (WldSm)) as the order of the parents are 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 τ variable 1,5 (τ = 0.209), variable 1,3 (τ = 0.197), and variable 1,2 (τ = 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. Also, variable 5 have two parents (3(IBrX) and 4(WIdIg)) which are $\hat{\tau}_{53}$ = 0.402 and $\hat{\tau}_{54}$ = 0.75. Based on these ordering, according to the measure of dependencies Kendall's, the resulting multivariate density decomposition is:

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

We now derive the entropy copulae associated with some of the moment constraints between copula variables 1, 2, 3, 4, 5, 6 in the density decomposition above. We initially construct maximum entropy copulas for unconditional copula $c_{15}, c_{46}, c_{34}, c_{45}$. Subsequently, it is imperative to establish the bases that should be selected and the number of discretization points in each case. We have outlined the approach for the unconditional copula c_{15}. The approach for unconditional copulas c_{46}, c_{34}, c_{45} is similar. Basis functions could be chosen based on the method described in Daneshkhah et al. (2012) 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

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

Here, we define $\varphi_i(.)$ over [0; 1] as follows (Daneshkhah et al., 2012)

$$\varphi_0(u) = 1$$

$$\varphi_0(u) = \frac{u^n - \sum_{j=0}^{n-1} \frac{\int_0^1 u^n \varphi_j(u)du}{\int_0^1 \varphi_j^2\,(u)du} \varphi_j(u)}{\left\| u^n - \sum_{j=0}^{n-1} \frac{\int_0^1 u^n \varphi_j(u)du}{\int_0^1 \varphi_j^2\,(u)du} \varphi_j(u) \right\|} \quad n \geq 1.$$

The OFS basis functions are similarly presented as:

$$\emptyset_1(.)\emptyset_1(.), \emptyset_1(.)\emptyset_2(.), \emptyset_2(.)\emptyset_1(.), \emptyset_1(.)\emptyset_3(.), \emptyset_3(.)\emptyset_1(.),$$
$$\emptyset_2(.)\emptyset_2(.), \emptyset_2(.)\emptyset_3(.), \emptyset_3(.)\emptyset_2(.), \emptyset_1(.)\emptyset_4(.), \emptyset_4(.)\emptyset_1(.),$$
$$\emptyset_1(.)\emptyset_5(.), \emptyset_5(.)\emptyset_1(.), \emptyset_2(.)\emptyset_4(.), \emptyset_4(.)\emptyset_2(.), \emptyset_3(.)\emptyset_3(.), \ldots$$

where the general form of $\emptyset_1(.)$ is given by

$\emptyset_0(u) = 1$, $\quad \emptyset_1(u) = \sqrt{2\cos(2\pi u)}$, $\emptyset_2(u) = \sqrt{2\sin(2\pi u)}$,

$$\emptyset_3(u) = \sqrt{2\cos(4\pi u)}, \emptyset_4(u) = \sqrt{2\sin(4\pi u)}\ ,$$
$$\emptyset_5(u) = \sqrt{2\cos(6\pi u)}, \emptyset_6(u) = \sqrt{2\sin(6\pi u)}\ ,$$

Following the explanations to select basis function in an optimal manner, we add the basis functions by using stepwise method outlined in Daneshkhah et al. (2012). In this method, at each phase, log-likelihood of adding each additional basis function is assessed. Subsequently, function that develops the largest increase in the log-likelihood is included. Additionally, to obtain optimal results, first four bases should be considered (Daneshkhah et al., 2012).

We can now compose the entropy copula density C15 with respect to the uniform distributions given the corresponding OP, OPS and OFS constraints above. We would initially require identifying the number of discretization points (grid size). A larger grid size provides an improved approximation to continuous copula with the expense of more computation time. Similarly, in our scenario, the approximation becomes more precise, if we

run the D_1AD_2 algorithm with further iterations, nevertheless, this would consume more computational resources including time. Therefore, it could 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 required for convergence; this is the case over any error level. Any grid size follows the same pattern where initially a large increase in the number of iterations is required to enhance accuracy and a smaller increase is needed when the error is smaller. We have selected a grid size of 200×200 throughout this example. Considering grid size, number of iterations and error size rules (outlined above), we can derive the entropy copula c_{15} associated with the chosen constraints. Expectations α 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.

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

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	$(\emptyset_2\emptyset_2, \emptyset_1\emptyset_1, \emptyset_3\emptyset_2, \emptyset_3\emptyset_4)$	(0.16, 0.08, 0.07, 0.07)	(0.16, 0.08, 0.07, 0.04)	37.76

One of the main advantages of using OPS and OFS bases over the ordinary polynomial series (Bedford et al., 2016) is that the D_1AD_2 algorithm converges in a swifter manner. This is due to property of orthogonal bases (OPS & OFS) where adding a new bases does not alter the already used Lagrange coefficients in the kernel. On the other hand, this is not the case when PS bases is applied [6] to calculate the entropy copula. Therefore, when OP is used, there is a need to run D_1AD_2 algorithm each time a new bases is added to the already chosen bases. This is due to changes in parameter values each time new terms are added. 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 and OFS bases are 9.83 & 8.89 seconds respectively, whereas this time for the PS bases is 29.87 seconds.

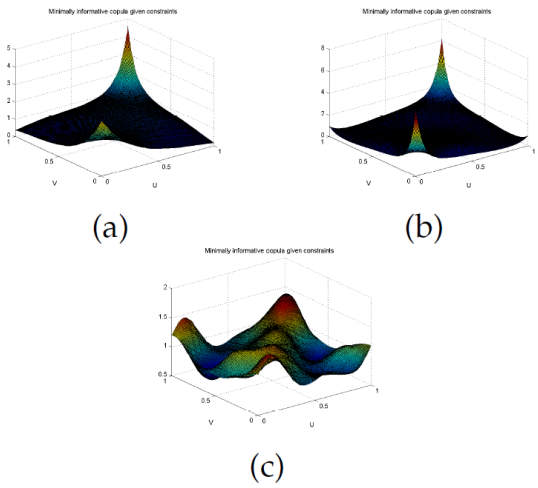

Figure 2: The minimally informative copula given moment constraints between variable 1 and 5; Panel (a): PS basis, Panel (b): OPS basis, and Panel (c): OFS basis

The other unconditional copula in the decomposition (7) i.e. c_{46},c_{34} and c_{45} could be similarly calculated. Following a stepwise method, PS, OPS and OFS bases along with their corresponding constraints, resulting Lagrange multipliers, and Log-Likelihood (L) are given in Table 2.

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

Method	Base	$(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$	$(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$	L
PS	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: $(\emptyset_2\emptyset_2, \emptyset_1\emptyset_1, \emptyset_2\emptyset_4, \emptyset_4\emptyset_2)$	(-0.11,0.1,-0.08,-0.07)	(-0.11,0.1,-0.08,0.02)	30.37
OPS	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: $(\emptyset_2\emptyset_2, \emptyset_1\emptyset_1, \emptyset_4\emptyset_2, \emptyset_2\emptyset_4)$	(0.35,0.3,0.19,0.01)	(0.4,0.3,0.2,-0.003)	245.49
OFS	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: $(\emptyset_2\emptyset_2, \emptyset_1\emptyset_1, \emptyset_2\emptyset_4, \emptyset_3\emptyset_1)$	(0.8,0.7,0.1,0.09)	(1.6,1.2,0.52,-0.001)	1366.1

The approximated maximum entropy copula for these unconditional copula, in terms of the PS, OPS and OFS bases, are illustrated in Panels of Figure 3.

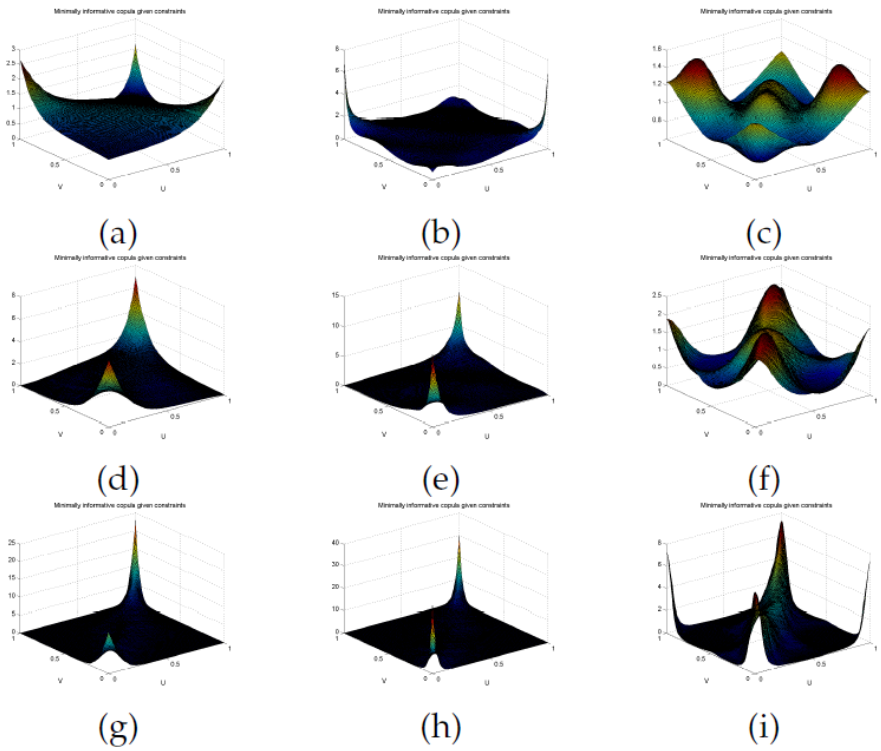

Figure 3: The minimally informative copula given moment constraints, Panel (a):C_{46} for PS basis, Panel (b):C_{46} for OPS basis, Panel (c):C_{46} for OFS basis, Panel (d):C_{34} for PS basis, Panel (e):C_{34} for OPS basis, Panel (f):C_{34} for OFS basis, Panel (g):C_{45} for PS basis, Panel (h):C_{45} for OPS basis, and Panel (i):C_{45} for OFS basis.

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. To calculate this copula, we divide the support of 5 into some arbitrary subintervals or bins and then construct the conditional copula within each bin. To do this, we select bases in the same way as for the unconditional copulas; we then fit the copula to the calculated mean values or constraints. We have used four bins so that the first copula is for 13|5 ∈ (0, 0.25). The other bins are 13|5 ∈ (0.25, 0.5), 13|5 ∈ (0.5, 0.75), and 13|5 ∈ (0.75, 1). This process can be similarly followed for the remaining bins.

Table 3: Comparison between different models.

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

The log-likelihood of the overall Non-Gaussian PCCDAG 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. Now, since the comparison based on comparing the log-likelihood of presented nonparametric model in this paper and the parametric model of Bauer et al. (2012) is not 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 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 fitted Non-Gaussian PCC-DAG models to the data using Bauer et al. (2012) method, (with AIC equals to -3078.62). The corresponding results are outlined in Table 3.

CONCLUSIONS

Gaussian distributions are generally used for 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, copulas are known to be a suitable and powerful means for overcoming the flaws in the existing techniques. An example for the application of copulas in the abovementioned 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 copulas, 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. BN's and copulas are generally very suitable for modelling 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 nicely 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 (Bedford et al., 2016; Chang, 2014; Chang et al., 2015)). 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 Bauer et al. (2012) to approximate a multivariate distribution by any Non-Gaussian PCC-DAG structure. The novelty of our approach lies within the use of entropy copulas for approximation for any level of precision. Furthermore, this approximation offers flexibility and manageability. The flexibility of the approach is where any function could be chosen for entropy copula constructions. Nevertheless, two conditions should be met within the DAG; the multivariate density is continuous and is non-zero. In our method, for approximating a multivariate distribution, a DAG structure, a basis family, and the expected values for certain functions are required to be identified. Finally, where there is a need to perform a computerized analysis of the basis functions, our method offers the ability for approximating and modeling of relatively more complex scenarios.

Acknowledgements

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

Aas, K., Czado, K. C., Frigessi, A., & Bakken, H. (2009). Paircopula constructions of multiple dependence. *Insurance, Mathematics and Economics, 44*, 182-198.

Bauer, A., & Czado, C. (2012). Pair-copula Bayesian networks. *Journal of of Computational and Graphical Statistics, 25*(4), 1248-1271.

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

Bedford, T., & Cooke, R. M. (2001). Probability Density Decomposition for Conditionally Dependent Random Variables Modeled by Vines. *Annals of Mathematics and Artificial Intelligence, 32*(1-4), 245–268.

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

Bedford, T., Daneshkhah, A., & Wilson, K. J. (2016). Approximate Uncertainty Modeling in Risk Analysis with Vine Copulas. *Risk Analysis, 36*(4), 792-815.

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

Chang, V., & Ramachandran, M. (2017). Financial Modeling and Prediction as a Service. *Journal of Grid Computing, 15*(2), 177–195.

Chang, V., Walters, R. J., & Wills, G. (2015). Monte Carlo risk assessment as a service in the cloud. *International Journal of Business Integration and Management, 7*(3).

Chang, V., Wills, G., Li, C.-S., Walters, R. J., De Roure, D., & Chee, C. (2012). The financial clouds review. In S. Aljawarneh, *Cloud Computing Advancements in Design, Implementation, and Technologies* (pp. 125-145). IGI Global.

Chatrabgoun, O., Hosseinian-Far, A., Chang, V., Stocks, N. G., & Daneshkhah, A. (2018). Approximating non-Gaussian Bayesian networks using minimum information vine model with applications in financial modelling. *Journal of Computational Science, 24*, 266-276.

Daneshkhah, A., Parham, G., Chatrabgoun, O., & Jokar, M. (2012). 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.

Harry, J. (1997). *Multivariate Models and Dependence Concepts.* London: Chapman & Hall.

Ibragimov, R., & Prokhorov, A. (2017). *Heavy Tails and Copulas: Topics in Dependence Modelling in Economics and Finance.* Singapore: World Scientific.

Jeon, S., Hong, B., & Chang, V. (2018). Pattern graph tracking-based stock price prediction using big data. *Future Generation Computer Systems, 80*, 171-187.

Kurowicka, D., & Cooke, R. (2006). Uncertainty Analysis with High Dimensional Dependence Modelling. In W. A. Shewhart, & S. S. Wilks, *Uncertainty Analysis.* John Wiley.

Mendes, B. V., Semeraro, M. M., & Leal, R. P. (2010). Pair-Copulas Modeling in Finance. *Financial Markets and Portfolio Management*, *24*(2), 193-214.

Ramachandran, M., & Chang, V. (2014). Financial Software as a Service: A Paradigm for Risk Modelling and Analytics. *International Journal of Organizational and Collective Intelligence (IJOCI)*, 65-89.

Zhang, Q., Wang, J., Lu, A., Wang, S., & Ma, J. (2018). An improved SMO algorithm for financial credit risk assessment – Evidence from China's banking. *Neurocomputing*, 314-325.