University of Strathclyde Glasgow

Bedford, T.J. and Daneshkhah, A. (2010) Approximating multivariate distributions with vines. Working paper. Management Science Working Series.

http://strathprints.strath.ac.uk/28364/

Strathprints is designed to allow users to access the research output of the University of Strathclyde. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (http://strathprints.strath.ac.uk) and the content of this paper for research or study, educational, or not-for-profit purposes without prior permission or charge. You may freely distribute the url (http://strathprints.strath.ac.uk) of the Strathprints website.

Any correspondence concerning this service should be sent to The Strathprints Administrator: eprints@cis.strath.ac.uk

Approximating Multivariate Distributions with Vines

Tim Bedford and Alireza Daneshkhah
Department of Management Science
University of Strathclyde
Glasgow, G1 1QE, UK

July 1, 2010

Abstract

In a series of papers, Bedford and Cooke used vine (or pair-copulae) as a graphical tool for representing complex high dimensional distributions in terms of bivariate and conditional bivariate distributions or copulae. In this paper, we show that how vines can be used to approximate any given multivariate distribution to any required degree of approximation. This paper is more about the approximation rather than optimal estimation methods. To maintain uniform approximation in the class of copulae used to build the corresponding vine we use minimum information approaches. We generalised the results found by Bedford and Cooke that if a minimal information copula satisfies each of the (local) constraints (on moments, rank correlation, etc.), then the resulting joint distribution will be also minimally informative given those constraints, to all regular vines. We then apply our results to modelling a dataset of Norwegian financial data that was previously analysed in Aas et al. (2009).

1 Introduction

Many areas of applied operations research require us to model multiple uncertainties using multivariate distributions. For many decision support settings it is possible to use discrete models such as Bayesian networks. In other areas, particularly when modelling financial data, it is necessary to have models of continuous multivariate random variables. Dependency modelling is therefore an area of great interest for a whole range of operations research applications.

There is a growing literature on the use of copulas to model dependencies. A copula is a joint distribution on the unit square (or more generally on the unit n-cube). Under reasonable conditions, we can uniquely determine a joint distribution for n random variables by specifying the univariate distribution for each variable, and in addition, specifying the copula. This holds because we can simply transform each variable by its own distribution function (sometimes called its quantile function) to ensure that the transformed variable has a uniform distribution, so that the joint distribution function F can be written

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)), \tag{1}$$

where C is a copula distribution function, and $F_1, \ldots, F_n$ are the univariate, or marginal, distribution functions. Hence we can use this formula constructively: Given a copula C and marginals $F_1, \ldots, F_n$ we can *define* F in this way. A special case is that of the 'Gaussian copula' which is equivalent to transforming each marginal to a normal distribution and then using a Gaussian joint distribution to model the dependency.

Clearly the use of a copula to model dependency is simply a translation of one difficult problem into another: instead of the difficulty of specifying the full joint distribution we have to the difficulty of specifying the copula. The main advantage is the technical one that copulas are normalized to have support on the unit square and uniform marginals. As many authors restrict the copulas to a particular parametric class (Gaussian, multivariate t, etc) the potential flexibility of the copula approach is not realized in practice. The approach used in this paper allows a lot of flexibility in copula specification. It utilizes a graphical model, called a vine, to systematically specify how two dimensional copulas are stacked together to produce an n-dimensional copula.

The main objective of this paper is to show that any vine structure can be used to approximate any given multivariate copula to any required degree of approximation. The technical assumptions we assume are that the multivariate density f under study has uniform marginals, is continuous and is non-zero. We illustrate this by modelling a dataset of Norwegian financial data that was previously analysed in [12].

Our constructive approach involves the use of minimum information copulas that can be specified to any required degree of precision based on the data available. We prove rigourously that good approximation 'locally' guarantees good approximation globally. Finally, we discuss rules of thumb that could be used to apply this in practice. In particular we discuss the problem of vine structure. A vine structure imposes no restrictions on the underlying joint probability distribution it represents (as opposed to the situation for Bayesian networks, for example). However this does not mean that we should ignore the question about which vine structure is most appropriate, for some structures allow the use of less complex conditional copulas than others. In particular, if we only allow certain families of copulas then one vine structure might fit better than another.

2 Vine constructions for multivariate dependence

A copula is a multivariate distribution function with standard uniform marginal distributions. Using Equation 1 we see that a copula can be used, in conjuction with the marginal distributions, to model any multivariate distribution. However, apart from the multivariate Gaussian, Student, and the exchangeable multivariate Archimedean copulae, the set of higher-dimensional copulae proposed in the literature is rather limited and are certainly not rich enough to model all possible mutual dependencies amongst the n variates (see Kurowicka and Cooke, 2006 for details of these copulae). Hence it is necessary to consider more flexible constructions.

A flexible structure, here denoted the *pair-copula construction* or *vine* allows for the free specification of (at least) $n(n-1)/2$ copulae. This structure was originally proposed by Joe (1996), and later reformulated and discussed in detail by Bedford and Cooke (2001, 2002), who considered simulation, information properities and the relationship to the multivariate normal distribution. Kurowicka and Cooke (2006) consider simulation issues and Aas et al. (2009) look at inference. Similar to the nested Archimedean constructions, the vine is hierarchical in nature. The modelling scheme is based on a decomposition of a multivariate density into a cascade of bivariate copulae. The way these copulae are built up to give the overall joint distribution is determined through a structure called a vine, and can be easily visualised. A vine on n variables is a nested set of trees, where the edges of the tree j are the nodes of the tree $j+1; j = 1, \ldots, n-2$, and each tree has the maximum number of edges. A *regular vine* on n variables is a vine in which two edges in tree j are joined by an edge in tree $j+1$ only if these edges share a common node, $j = 1, \ldots, n-2$. There are $n(n-1)/2$ edges in a regular vine on n variables. The formal definition is as follows.

Definition 1 ***(Vine, regular vine)*** *$\mathcal{V}$ is a vine on n elements if*

1. *$\mathcal{V} = (T_1, \ldots, T_{n-1})$.*

2. *T_1 is a connected tree with nodes $N_1 = \{1, \ldots, n\}$ and edges E_1; for $i = 2, \ldots, n-1$, T_i is a connected tree with nodes $N_i = E_{i-1}$.*

   $\mathcal{V}$ is a regular vine on n elements if additionally

3. *(proximity) For $i = 2, \ldots, n-1$, if a and b are nodes of T_i connected by an edge in T_i, where $a = \{a_1, a_2\}$, $b = \{b_1, b_2\}$, then exactly one of the a_i equals one of the b_i .*

One of the simplest regular vine is shown in Figure 1. Here, T_1 is the tree consisting of the straight edges between the numbered nodes. T_2 is the tree consisting of the curved edges that join the straight edges in T_1, and so on. For a regular vine each edge of T_1 is labelled by two numbers from $\{1, ..., n\}$. If we take two edges of T_1 which become linked nodes in T_2 then of the numbers labelling these edges one is common to both, and they both have one unique one. For example 12 and 23 are linked at the next level tree. The common number(s) will be called the *c*onditioning set D_e for that edge e (in this example the conditioning set is simply $\{2\}$) and the other numbers will be called the *c*onditioned set (in this example $\{1, 3\}$). For a regular vine the conditioned set always contains two elements.

With such a vine we associate conditional copulas to each edge, to couple the two variables in the conditioned set given the values in the conditioning set.

Bedford and Cooke (2001) express a regular vine distribution in terms of its density in the following theorem.

Theorem 1 *Let $\mathcal{V} = (T_1, \ldots, T_{n-1})$ be a regular vine on n elements. For each edge $e(j,k) \in T_i, i = 1, \ldots, n-1$ with conditioned set $\{j,k\}$ and conditioning set D_e, let the conditional*

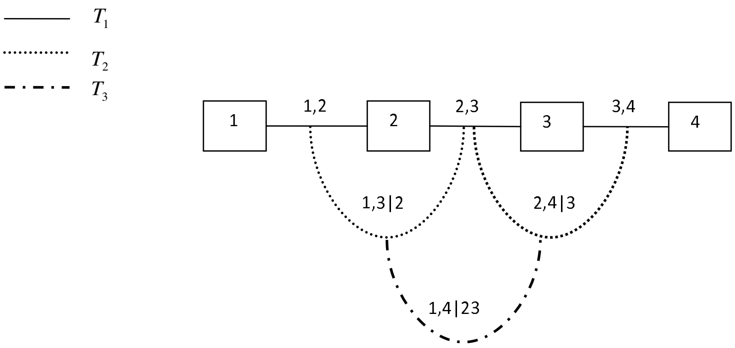

Figure 1: A regular vine with 4 elements

copula and copula density be $C_{jk|De}$ *and* $c_{jk|De}$ *respectively. Let the marginal distributions* F_i *with densities* $f_i, i = 1, \ldots, n$ *be given. Then the vine-dependent distribution is uniquely determined and has a density given by*

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i) \prod_{j=1}^{n-1} \prod_{e(j,k)\in E_i} c_{jk|D_e}(F_{j|D_e}, F_{k|D_e}) \tag{2}$$

The existence of regular vine distributions is discussed in detail by Bedford and Cooke (2002).

The density decomposition associated with 4 random variables $\mathbf{X} = (X_1, \ldots, X_4)$ with a joint density function $f(x_1, \ldots, x_4)$ satisfying a copula-vine structure (this structure is called D-vine, see Kurowicka and Cooke, 2006, pp. 93) shown in Figure 1 with the marginal densities $f_1, \ldots, f_4$ is

$$\begin{aligned} f_{1234}(x_1, \ldots, x_4) \quad = \quad & \prod_{i=1}^{4} f(x_i) \times c_{12}\{F(x_1), F(x_2)\} c_{23}\{F(x_2), F(x_3)\} c_{34}\{F(x_3), F(x_4)\} \\ & \times c_{13|2}\{F(x_1 \mid x_2), F(x_3 \mid x_2)\} c_{24|3}\{F(x_2 \mid x_3), F(x_4 \mid x_3)\} \\ & \times c_{14|23}\{F(x_1 \mid x_2, x_3), F(x_4 \mid x_2, x_3)\} \end{aligned} \tag{3}$$

This formula can be derived for this case using the general expression

$$f(x, y) = f_X(x) f_Y(y) c(F_X(x), F_Y(y)),$$

or equivalently

$$f(x|y) = f_X(x) c(F_X(x), F_Y(y)),$$

where c is the copula density and F_X, F_Y are the univariate distributions. Starting with

$$f_{1234}(x_1, \ldots, x_4) = f_1(x_1) f_2(x_2, \mid x_1) f_3(x_3 \mid x_1, x_2) f_4(x_4 \mid x_1, \ldots, x_3),$$

we inductively convert the latter expression into that shown in Equation 3. We have

$$f_{2|1}(x_2 \mid x_1) = f_2(x_2) c_{12}(F_1(x_1), F_2(x_2)).$$

Next,

$$\begin{aligned} f_{3|12}(x_3 \mid x_1, x_2) &= f_{3|2}(x_3 \mid x_2) c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \\ &= f_3(x_3) c_{23}(F_2(x_2), F_3(x_3)) c_{13|2}(F_1(x_1|x_2), F_3(x_3|x_2)). \end{aligned}$$

The calculation for $f_{4|123}(x_4 \mid x_1, \ldots, x_3)$ is left to the reader.

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

Unlike the situation with Bayesian networks, where not all structures can be used to model a given distribution, the theorem shows that - in principle - any vine structure may be used to model a given distribution. However, in practice it seems that some vine structures do work better than others, and so this must be a result of restricting to a particular family of copulas. That is, given a family of copulae, some vine structures may give a better degree of approximation than others. In fact, we could say that the question "does a vine structure fit?" only makes sense in the context of a given family of copulae.

3 Building bivariate minimum information copulae

The emphasis on this paper is on approximation rather than on optimal estimation techniques. We use minimum information methods to demonstrate uniform approximation in the class of copulae used.

This section discusses some preliminary ideas that will be needed, and in particular shows how the approaches work to determine a unique copula when there are just two variables of interest.

We note now though that for regular vines it is possible to compute a useful expression for the information of a distribution in terms of the information of the copulae. The results needed to do this are given in Bedford and Cooke (2002); Kurowicka and Cooke (2006) and references therein. Lemma 4.4 and Theorem 4.5 in Bedford and Cooke (2002) shows that if we take a minimal information copula satisfying each of the (local) constraints (on moments, rank correlation, etc.), then the resulting joint distribution is also minimally informative given those constraints. The similar expression is generalized to all regular vines.

3.1 Data: Expert judgment or random sample driven approaches

Quantitative models are typically parametrized either by expert judgement or estimation from data. Bedford and Cooke (2001), argue that expert judgement should be based on observable quantities. In our context, it is the uncertain quantities such as X that are observable. The quantile $F(X)$ is a quantity could be argued to be an observable quantity if the distribution function F is known. Now the use of copulas implies that we must in fact know the marginal disributions, so this might not seem an important point. However, experts could normally be expected to find it easier to consider the joint behaviour of the untransformed variables X_1, X_2, etc, than the joint behaviour of the transformed variables $F_1(X_1)$, $F_2(X_2)$. Hence when using experts to make assessments it is definitely preferable to use assessments on the untransformed variables. However, in the context of specifying joint distributions this causes more difficulties as there will generally be constraints. For example when two different marginal distributions are specified for X_1 X_2, then the product moment correlation might not be able to take values close to +/- 1. Fortunately, by working with minimum information distributions we can deal with this problem to some extent. This method allows interactive elicitation of expert opinions by giving guidance as to what values of uncertain quantities are compatible with the assessments already made (Bedford, 2006).

By contrast, when assessing distributions on the basis of data (large quantities of which may well be available for example in financial risk modelling problems), the data can be transformed to uniform after estimation of the marginals. This makes it possible to consider approximation, or encoding, of the data using a multivariate copula, and enables us to consider ways of judging how well that approximation can be made using given families of two-dimensional copulae. We shall consider two different approaches to do this.

3.2 The D_1AD_2 algorithm and minimum information copulae

Suppose there are k functions, $h_1, h_2, \ldots, h_k : [0,1]^2 \to \mathbb{R}$, for which we can specify the mean values $\alpha_1, \ldots, \alpha_k$ that these functions should take. We seek a copula that has these mean values, a problem which is usually either infeasible or underdetermined. Hence, assuming feasibility for the moment, we ask also that the copula be minimally informative (with respect to the uniform distribution), which guarantees a unique and reasonable solution: Define the kernel

$$A(u,v) = \exp(\lambda_1 h_1(u,v) + \ldots + \lambda_k h_k(u,v)). \tag{4}$$

According to the general theory of Borwein et al. (1994), Nussbaum (1989) there is a unique copula with minimum information satisfying the constraints that the mean value of h_i is α_i $(i = 1, ..., k)$, and this has density

$$d^{(1)}(u)d^{(2)}(v)A(u,v).$$

The parameters $(\lambda_1, \ldots, \lambda_k)$ depend on $(\alpha_1, \ldots, \alpha_k)$ in a nonlinear way. Fortunately there are numerical procedures to determine this relationship: Given $(\lambda_1, \ldots, \lambda_k)$ we can determine the functions $d^{(1)}(u)$ and $d^{(2)}(v)$ and then calculate the associated mean values for $h_1, h_2, \ldots, h_k$.

We numerically solve this function to obtain the unique $(\lambda_1, \ldots, \lambda_k)$ for which the mean values of $h_1, h_2, \ldots, h_k$ are $\alpha_1, \ldots, \alpha_k$.

The general theory says that the set of all possible expectation vectors $(\alpha_1, \ldots, \alpha_k)$ that could be taken by $(h_1, h_2, \ldots, h_k)$ under some probability distribution is convex, and that for every $(\alpha_1, \ldots, \alpha_k)$ in the interior of that convex set there is a density with parameters $(\lambda_1, \ldots, \lambda_k)$ for which $(h_1, h_2, \ldots, h_k)$ take these expectations.

This general approach to defining a copula was used by Bedford and Meeuwissen (1997) with a single function $h(u, v) = uv$, which essentially measures the Spearman rank correlation of the copula. Bedford (2006) and Lewandowski (2008) have considered larger groups of functions.

The discrete version of this problem can be written in terms of matrices. Suppose that (u, v) are discretized into n points, respectively as u_i, and v_j, $i, j = 1, \ldots, n$. Then we write $A = (a_{ij}), D_1 = diag(d_1^{(1)}, \ldots, d_n^{(1)}), D_2 = diag(d_1^{(2)}, \ldots, d_n^{(2)})$, where $a_{ij} = A(u_i, v_j)$, $d_i^{(1)} = d^1(u_i)$, $d_j^{(2)} = d^2(v_j)$. The assumption of uniform marginals means that

$$\forall i = 1, \ldots n \quad \sum_j d_i^{(1)} d_j^{(2)} a_{ij} = 1/n, \quad \text{and}$$

$$\forall j = 1, \ldots n \quad \sum_1 d_i^{(1)} d_j^{(2)} a_{ij} = 1/n.$$

Hence

$$d_i^{(1)} = \frac{n}{\sum_j d_j^{(2)} a_{ij}} \quad and \quad d_j^{(2)} = \frac{n}{\sum_i d_i^{(1)} a_{ij}}$$

The problem of finding matrices D_1 and D_2 so that $D_1 A D_2$ is a stochastic matrix has been long studied. Sinkhorn and Knopp (1967) gave a simple algorthim, and the iterative proportional fitting algorithm (Cziszar, 1975) has been much used. IPF simply uses an iterative procedure to determine the entries of D_1 and D_2. The idea is very simple - start with arbitrary positive initial matrices for D_1 and D_2. Then successively define new vectors by iterating the maps

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

This iteration converges geometrically to give us the vectors required.

Nussbaum (1989) considered the problem in much greater generality, considering continuous densities and functions, and showed that the corresponding functional is a contraction mapping on a space of functions endowed with a Hilbert projective metric. We shall make use of this fact when considering the quality of approximations made to copulae below.

The methods described in this section will be used in two ways. Firstly to determine minimally informative approximations to copulae. Secondly to adjust densities that are near to, but not actually copulae, so as to make the adjusted density a copula.

3.3 Numerical construction of minimally informative copulae with the D_1AD_2 algorithm

As discussed above, for a given set of functions $(h_1, \ldots, h_k)$, the mapping from the set of vectors of λ's parameterizing the kernel A onto the expectations of the function $(\alpha_1, \ldots, \alpha_k)$ has to be found numerically. We employ optimization techniques for achieving the result. We wish to determine the appropriate set of λ's for given expectations α_i, where the expectations have been calculated using the discrete copula density D_1AD_2.

Define

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

We seek the roots of these functions. One of the possible solvers for this task would be FSOLVE - MATLAB's optimization routine. It implements various root finding techniques allowing for choosing the one suiting our problem best. However we also obtained good results by using another of Matlab's optimization procedures in the example below, namely FMINSEARCH, which implements the Nelder-Mead simplex method (Lagarias et al., 1998). The minimized function is

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

As an example we show how an expert could specify a copula though defining two expected values.

Example 1 *Suppose that we are given two uncertain quantities X and Y with distribution functions F_X and F_Y for which we want to specify a copula. Suppose that the marginal distributions (these distributions can also be specified by an expert) are also of X and Y are as follows*

$$X \sim N(0,1), \qquad Y \sim N(1,4)$$

In addition, suppose that an expert is willing to specify the expected values of XY and X^2Y, that is, the expected values of functions $h'_1(x,y) = xy$, $h'_2(x,y) = x^2y$. We cannot directly apply the methods described above as these functions are given in terms of X and Y rather than the copula variables. But, we can find corresponding functions of the copula variables U and V, defined by $h_i(u,v) = h'_i(F_X^{-1}(u), F_Y^{-1}(v))$ $(i = 1, 2)$. The expected values of (h_1, h_2) under the copula density equal the expected values of (h'_1, h'_2) under the required joint distribution.

To numerically implement the D_1AD_2 algorithm we discretize the copula space by fixing a size k of the square matrix kernel A and associate points in the unit square (u_i, v_j) with each element of the matrix, where $u_i = (i - 0.5)/k$ and $v_j = (j - 0.5)/k$. We then define $A_\lambda = \exp\{\lambda_1 h_1(u_i, v_j) + \lambda_2 h_2((u_i)v_j)\}$. The D_1AD_2 then gives us a discrete approximation to the copula with minimum information have a certain value of $E(XY)$ and $E(X^2Y)$ depending on the parameters λ_1, λ_2. Figure 2 shows the approximation of the copula for $\lambda_1 = 0.069806, \lambda_2 = -0.010728$.

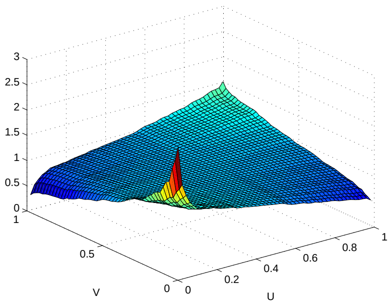

Figure 2: The minimally informative copula with $\lambda_1 = 0.069806, \lambda_2 = -0.010728$.

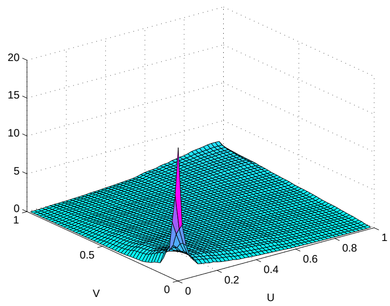

Figure 3: The minimally informative copula given the following constraints, $E(XY) = 1$ and $E(X^2Y) = 0.8$.

The functional relationship between the set of vectors of λ's and the set of vectors or resulting expectations of functions of (h_1, h_2) can be determined numerically. Given the expectations of the functions mentioned above fixed as follows, $E(XY) = 1$ and $E(X^2Y) = 0.8$, the optimal values for λ's are: $\lambda_1 = 0.38893$, $\lambda_2 = -0.11304$. Figure 3 shows the minimally informative (discretized) copula for these values. We also show both objective functions, $(h'_1(X,Y), h'_2(X,Y))$ in Figure 4 together with their counterparts in the copula space, $(h_1(U,V), h_2(U,V))$.

Figures 5, 6 show the expected values of $h'_1(X,Y)$ and $h'_2(X,Y)$ as a functions of λ_1 and λ_2. We may wish to specify more expectation values. Figure 7 shows the minimally informative copula given constraints on $E(XY)$, $E(X^2Y)$, $E(XY^2)$ and $E(X^2Y^3)$ when these expectations are fixed at $(0.1528, 0.9205, 0.1661, 13.5603)$, respectively.

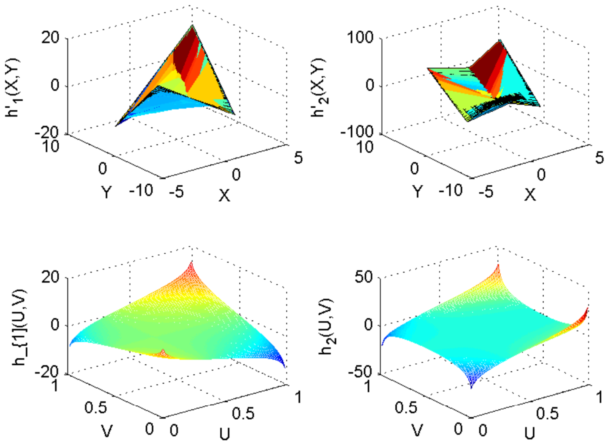

Figure 4: Plots of base functions and the corresponding functions on the copula domain

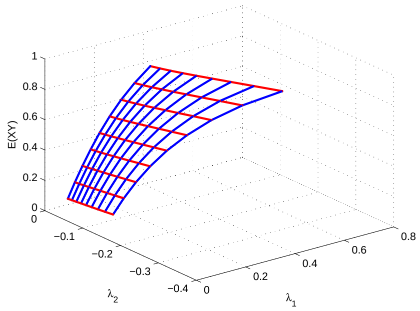

Figure 5: The presentation of $E(XY)$ as a function of λ_1 and λ_2

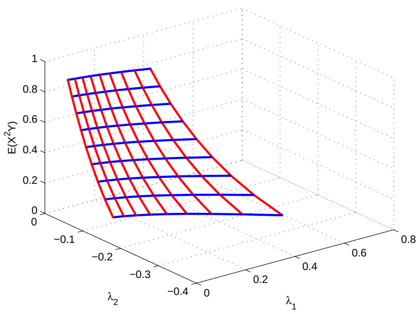

Figure 6: The presentation of $E(X^2Y)$ as a function of λ_1 and λ_2

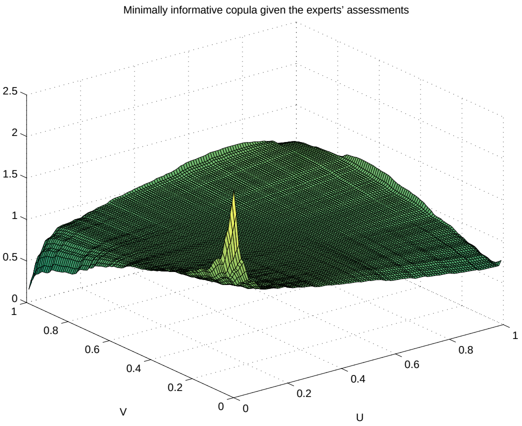

Figure 7: The minimally informative copula given the following constraints: $E(XY) = 0.1528$, $E(X^2Y) = 0.9205$, $E(XY^2) = 0.1661$ and $E(X^2Y^3) = 13.5603$.

4 Copula compactness

One of the main aims of this paper is to show that we can arbitrarily well approximate a multivariate distribution by using a fixed family of bivariate copulae. A key step to demonstrating this is to show that the family of bivariate (conditional) copula densities contained in a given multivariate distribution forms a compact set in the space of continuous functions on $[0,1]^2$. Based on this we can then show that the same finite parameter family of copulae can be used to give a given level of approximation to all conditional copulae simultaneously.

It is worth defining more precisely the way in which we approximate densities. We assume that all densities are continuous. Write $C(Z)$ for the space of continuous real valued functions on a space Z, where we shall always take $Z = [0,1]^r$ for some r. A norm on the space $C(Z)$ is given by

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

Since our functions are assumed continuous on Z, and since Z is compact, the norm of any such function is finite. We shall be particularly interested in the set

$$\mathcal{C}(f) = \{c_{ij|i_1\ldots i_r} : 1 \leq i, j, i_1, \ldots, i_r \leq n, i, j \neq i_1, \ldots, i_r\}$$

where $c_{ij|i_1\ldots i_r}$ is the copula of the conditional density of X_i, X_j given $X_{i_1}, \ldots, X_{i_r}$. It will be important to show that this set is relatively compact in the space of all continuous real valued functions $C([0,1]^2)$, because then we can show that the copula densities can be uniformly approximated. We consider compactness relative to the topology induced by the sup norm.

Compactness of a set K can be defined equivalently through one of two properties, each of which we shall use:

1. Any open cover of K has a finite subcover. In other words if K is a subset of an infinite union of open sets, then it is in fact also a subset of a finite union of those open sets.

2. Any sequence of points (which in our case are functions) of K has a convergent subsequence.

The famous Arzela-Ascoli Theorem gives another way of checking compactness when dealing with function spaces. It says that a subset $K \subset C([0,1]^2)$ is relatively compact if the functions of K are equicontinuous and pointwise bounded. We recall that a set of functions is equicontinuous if for all $\epsilon > 0$ and (u,v) there is a $\delta > 0$ such that if the Euclidean distance $|(u,v) - (u',v')| < \delta$ then

$$|g(u,v) - g(u',v')| < \epsilon \qquad \forall g \in K,$$

and that K is pointwise bounded if

$$\sup\{||g|| : g \in K\} < \infty.$$

As a first step to showing the relative compactness of $\mathcal{C}(f)$ we first give our attention to two other spaces: The set of conditional marginal densities

$$\mathcal{M}(f) = \{f_{i|i_1\ldots i_r} : 1 \leq i, i_1, \ldots, i_r \leq n, i \neq i_1, \ldots, i_r\},$$

where $f_{i|i_1\ldots i_r}$ is the conditional density of X_i given $X_{i_1}, \ldots, X_{i_r}$, and the set of conditional bivariate densities

$$\mathcal{B}(f) = \{f_{ij|i_1\ldots i_r} : 1 \leq i, j, i_1, \ldots, i_r \leq n, i, j \neq i_1, \ldots, i_r\}$$

where $f_{ij|i_1\ldots i_r}$ is the conditional density of X_i, X_j given $X_{i_1}, \ldots, X_{i_r}$. Note that as we have defined it, a member of $\mathcal{M}(f)$ is a function of one variable - in other words, all the different marginals that we get for different conditions are individually members of $\mathcal{M}(f)$. Similarly for $\mathcal{B}(f)$. Hence $\mathcal{M}(f) \subset C([0,1])$ and $\mathcal{B}(f) \subset C([0,1]^2)$.

Theorem 2 *The sets $\mathcal{M}(f) \subset C([0,1])$ and $\mathcal{B}(f) \subset C([0,1]^2)$ are relatively compact.*

Proof:

We have assumed that our multivariate density $f_{1\ldots n}$ is a continuous function defined on $[0,1]^n$. Since all marginal densities $f_{i_1\ldots i_r}$ are obtained by integrating out variables from $f_{1\ldots n}$, it is clear that

$$|f_{i_1\ldots i_r}(x_{i_1}, \ldots, x_{i_r})| \leq \sup |f_{1\ldots n}(x_1, \ldots, x_n)|,$$

where the sup is taken over the variables x_i ($i \neq i_1, \ldots, i_r$). Hence

$$|f_{i|i_1\ldots i_r}(x_i|x_{i_1}, \ldots, x_{i_r})| = |\frac{f_{ii_1\ldots i_r}(x_i x_{i_1}, \ldots, x_{i_r})}{f_{i_1\ldots i_r}(x_{i_1}, \ldots, x_{i_r})}| \leq ||f||/\alpha$$

where $\alpha > 0$ is a lower bound on the values taken by f. This shows that there is a pointwise bound for all the functions in $\mathcal{M}(f)$.

In order to show equicontinuity we first note that each function $f_{i_1 \dots i_r}$ is uniformly continuous. Since there are only a finite number of such functions, we can always ensure that given $\epsilon > 0$ there is a $\delta > 0$ such that for any $i_1 \dots i_r$ if

$$|(x_{i_1}, \dots, x_{i_r}) - (y_{i_1}, \dots, y_{i_r})| < \delta$$

then

$$|f_{i_1 \dots i_r}(x_{i_1}, \dots, x_{i_r}) - f_{i_1 \dots i_r}(y_{i_1}, \dots, y_{i_r})| < \epsilon.\alpha.$$

Hence if $|x_i - y_i| < \delta$ then

$$|f_{i|i_1 \dots i_r}(x_i | x_{i_1}, \dots, x_{i_r}) - f_{i|i_1 \dots i_r}(y_i | x_{i_1}, \dots, x_{i_r})| \leq$$

$$|f_{ii_1 \dots i_r}(x_i, x_{i_1}, \dots, x_{i_r}) - f_{ii_1 \dots i_r}(y_i, x_{i_1}, \dots, x_{i_r})|/\alpha \leq \epsilon$$

so that $\mathcal{M}(f)$ must also be an equicontinous family.

A similar argument shows that $\mathcal{B}(f)$ is an equicontinous family.

□

We can now show

Theorem 3 *The set $\mathcal{C}(f) \subset C([0,1]^2)$ is relatively compact.*

Proof

For any element $c_{ij|i_1 \dots i_r}$ of $\mathcal{C}(f)$, we have

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

Hence if we take a sequence of elements in $\mathcal{C}(f)$ then there are corresponding sequences of elements of $\mathcal{M}(f)$ and $\mathcal{B}(f)$. Since $\mathcal{M}(f)$ is relatively compact there must be a convergent subsequence, and looking along that same subsequence there must be a subsequence of that for which the corresponding functions in $\mathcal{B}(f)$ converge. Now, along this subsequence the right hand side of the above expression converges, so the elements of $\mathcal{C}(f)$ on this same sequence must converge (and to the same thing). In particular there is a convergent subsequence. Hence $\mathcal{C}(f)$ is relatively compact.

□

Since all the functions in $\mathcal{C}(f)$ are positive and uniformly bounded away from 0 it follows that

Corollary 1 *The set $\mathcal{LNC}(f) = \{\ln(g) : g \in \mathcal{C}(f)\} \subset C([0,1]^2)$ is relatively compact.*

4.1 Linear bases and approximate copulae

The set $C([0,1]^2)$ can be considered a vector space, and in this context a basis is simply sequence of functions $h_1, h_2, \dots \in C([0,1]^2)$ for which any function $g \in C([0,1]^2)$ can be written as $g = \sum_{i=1}^{\infty} \lambda_i h_i$. There are lots of possible bases, for example

$$u, v, uv, u^2, v^2, u^2 vuv^2, \dots.$$

Given an ordered basis $h_1, h_2, \ldots \in C([0,1]^2)$ and a required degree of approximation $\epsilon > 0$ in the sup metric, we can consider the collection of open sets

$$U_{k,\epsilon} = \{g \in C([0,1]^2) : \inf ||g - \sum_{i=1}^{k} \lambda_i h_i|| < \epsilon\}$$

where the inf in the above definition is to be taken over all possible values of the λ_i. Now, $U_{k,\epsilon}$ is clearly open and furthermore

$$U_{k,\epsilon} \subset U_{k+1,\epsilon}, \qquad \bigcup_{k=1}^{\infty} = C([0,1]^2).$$

So the $U_{k,\epsilon}$ form an open cover of $\mathcal{LNC}(f)$ and hence by definition of compactness there is a k such that $U_{k,\epsilon}$ covers $\mathcal{LNC}(f)$. We can state this as a result

Theorem 4 *Given $\epsilon > 0$, there is a k such that any member of $\mathcal{LNC}(f)$ can be approximated to within error $\epsilon > 0$ by a linear combination of $h_1, h_2, \ldots, h_k$.*

The same result holds for $\mathcal{C}(f)$ (though not necessarily with the same k).

Finally, we remark that though we have been looking only at approximation in the sense of the sup norm, one could easily look at higher order approximation. For example, if we assume that the density $f_{1\ldots n}$ is continuously differentiable then all the derivatives are continuous functions and the same arguments as used above show that they form an equicontinuous and pointwise bounded family. Following through we find that the copulae generated from $f_{1\ldots n}$ are also continuously differentiable. By using a slightly different norm on the continuously differentiable functions $C^1([0,1]^2) \subset C([0,1]^2)$,

$$||g||_1 = ||g|| + ||\frac{d}{du}g|| + ||\frac{d}{dv}g||,$$

we can guarantee that a similar approximation result to the above holds with pointwise approximation of the derivatives as well.

4.2 Ensuring that approximating densities are copula densities

Since the approximations we make of a copula density might not be quite a copula density itself, we need to transform it to obtain a copula. This is done by weighting the density as described above in Section 3.2. If we have a continuous positive real valued function $A(u,v)$ on $[0,1]^2$ then there are continuous positive functions $d_1(u)$ and $d_2(v)$ such that $d_1.d_2.A$ is a copula density, that is, it has uniform marginals. We call this density the *C-Projection* of A and denote it $C(A)$. It will also be convenient to denote by $N(h)$ the normalization of a non-negative function h with finite integral.

The next lemma allows us to control the error made when approximating a copula by another function.

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

Proof: We show that by taking f sufficiently close to g one can ensure that the reweighting functions for f are as close to 1. This then implies that $C(f)$ is close to g.

Without loss of generality we can assume that f is normalized. The proof uses the fact that we can use the Borwein-Lewis-Nussbaum approach to find functions $d_{1f}(u)$ and $d_{2f}(v)$ such that $d_{1f}.d_{2f}.f$ has uniform marginals. Such functions d_{1g} and d_{2g} exist also for g but are constant , $d_{1g}(u) = d_{2g}(v) = 1$. As discussed above, these reweighting functions are fixed points of a functional that is a contraction mapping when using the Hilbert metric $\mathfrak{D}$ on the appropriate space of pairs of functions (d_1, d_2).

We denote by L_f the functional associated to f. Since this is a contraction mapping there exists a $\lambda_f \in (0, 1)$ such that

$$\mathfrak{D}(L_f(a, b), L_f(c, d)) < \lambda_f \mathfrak{D}((a, b), (c, d)).$$

If we set $a_0 = 1$, $b_0 = 1$ and $(a_{n+1}, b_{n+1}) = L_f(a_n, b_n)$, then we have convergence to the required pair of functions (d_{1f}, d_{2f}) that reweight f to become uniform.

Now, by choosing f close enough to g we can ensure two things. First that the contraction rate associated to L_f is close to that of L_g, in particular less than some chosen $\lambda < 1$. Second we can ensure that

$$\mathfrak{D}(L_g(1, 1), L_f(1, 1)) = \mathfrak{D}(1, 1), L_f(1, 1))$$

is as small as required. This implies that

$$\begin{aligned}
\mathfrak{D}((1, 1), (a, b)) &\leq \sum_{n=0}^{\infty} \mathfrak{D}((a_n, b_n), (a_{n+1}, b_{n+1})) \\
&\leq \mathfrak{D}((a_0, b_0), (a_1, b_1)) \sum_{n=0}^{\infty} \lambda^n \\
&= \frac{\mathfrak{D}((1, 1), L_f(1, 1))}{1 - \lambda}.
\end{aligned}$$

Hence the reweighting functions for f are close to the identity, and so $C(f)$ is close to g. □

Tim, according to Contraction *Mapping Principle Theorem*, if we let $X = (d_{1n}, d_{2n})$ be the appropriate space of pairs of functions (d_1, d_2), and $L_f(L_g) : X \to X$ be a contraction mapping with contractivity coefficient λ_f. Let $x_0 = (1, 1) \in X$ and inductively define $(d_{1(n+1)}, d_{2(n+1)}) = L_f(d_{1(n)}, d_{2(n)}), n \geq 0$. The L_f has a unique fixed point (d_{1f}, d_{2f}), the sequence $(d_{1(n)}, d_{2(n)})$ converges to (d_{1f}, d_{2f}) and

$$\mathfrak{D}((d_{1f}, d_{2f}), (d_{1(n)}, d_{2(n)})) \leq \lambda_f^n \mathfrak{D}((d_{1f}, d_{2f}), (1, 1)).$$

I think that proves what we have mentioned on the Lemma and is similar to your proof. The proof of the theorem is based on showing that $(d_{1(n)}, d_{2(n)})$ is a Cauchy

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

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

5 Constructing approximations using minimally informative distributions

The above discussion has shown that we can approximate all conditional copulae using linear combinations of basis functions. We did not address the question of how you choose the appropriate parameter values, and indeed finding the parameters that would minimize the sup norm for a given copula is not of itself an appealing procedure. A pragmatic alternative that lies very close to the approach described above is to use the minimum information criterion. In other words given $\{1, h_1, \ldots, h_k\} : [0,1]^2 \to \mathbb{R}$ we seek values $\lambda_1, \ldots, \lambda_k$ so that $\exp(\sum_1^k \lambda_i h_i)$ is close to the copula density we are approximating.

In the minimum information framework we do this by fitting the moments of h_i. So if $\int\int h_i g du dv = \alpha_i$ then we search for the copula density with minimum information (with respect to the independent distribution) that also has those moments. It can be shown that this copula density is unique and has the form

$$d^1(u)d^2(v)\exp(\sum_1^k \lambda_i h_i(u,v)).$$

When we use a vine structure to model a multivariate distribution, the vine defines a decomposition of the multivariate distribution into certain conditional copulae, associated to the conditioned and conditioning sets of the vine. For example, if $\{i,j\}$ is the conditioned set and D_e is the conditioning set in one part of a vine, then the family of conditional copulae for x_i, x_j given D_e has to be specified. Using the minimum information approach means that we should specify mean values for the functions h_r given the variables in D_e, that is, we have to specify the conditional means $\alpha_m(ij \mid D_e)$.

A multivariate distributions can be then approximated as follows:

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

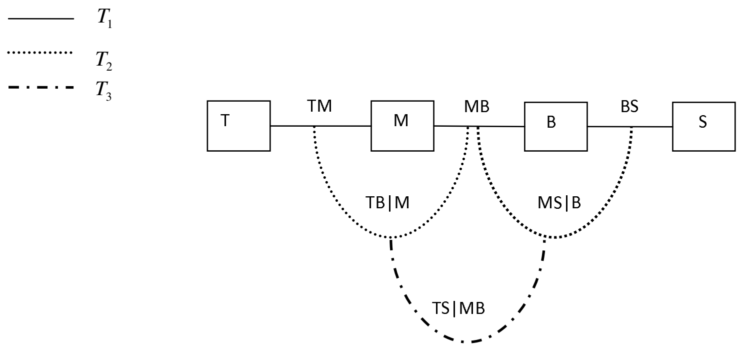

Figure 8: Selected vine structure for the Norwegian stock data set with 4 variables: Norwegian stock index (T), MSCI world stock index (M), Norwegian bond index (B) and SSBWG hedged bond index (S).

We illustrate the procedure by applying it to a financial data set.

Example 2 In this example, we use the same data set studied in Aas et al (2009). These are four time series of daily data: the Norwegian stock index (TOTX), the MSCI world stock index, the Norwegian bond index (BRIX) and the SSBWG hedged bond index, for the period from 0.4.01.1999 to 0.8.07.2003. We denote these four variables by T, M, B and S, respectively.

We want to generate vine approximation fitted to this data set to any given multivariate density using minimum information distribution. We select a similar vine structure with 4 elements shown in Figure 1 for this data presented in Figure 8. It should be noticed that, we can find the corresponding functions of the copula variables X, Y, Z and W associated with T, M, B, S, respectively, defined by $h_i(X,Y) = h'_i(F_1^{-1}(X), F_2^{-1}(Y))$, etc., and clearly these should also have the same specified expectation, that is, $E(h'_i(T,M)) = E(h_i(X,Y))$, etc. The minimum information copulae calculated in this example are derived based on the copula variables, X, Y, Z, W.

We first can construct a minimally informative copula between any two variables joining together in the first tree, T_1. As an example, we show the construction of a minimally informative copula between two variables M and T denoted by C_{TM} under the following constraints: $h'_1(M,T) = MT$, $h'_2(M,T) = TM^2$, $h'_3(M,T) = T^2M$ and $h'_4(M,T) = MT^3$. In other words, we use the Fourier copula of order 4 or a base with 4 elements to approximate this copula. We fix the values of the expectations of these functions as follows

$$\alpha_1 = \frac{1}{1094}\sum_{i=1}^{1094} T_i M_i = 0.2314, \alpha_2 = \frac{1}{1094}\sum_{i=1}^{1094} T_i M_i^2 = 0.1497$$

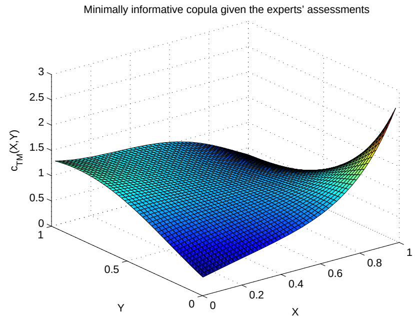

Figure 9: The minimally informative copula between T and M variables of Norwegian Stock data.

$$\alpha_3 = \frac{1}{1094}\sum_{i=1}^{1094} T_i^2 M_i = 0.1465, \alpha_4 = \frac{1}{1094}\sum_{i=1}^{1094} T_i^3 M_i = 0.1058$$

The minimum information copula C_{TM} with respect to the uniform distribution given the four constraints mentioned above has been constructed on the same grid of 50 by 50 equally spaced points and presented in Figure 9.

We now want to study the influence of adding more constraints on the approximation of the copula density. As we discussed above and as expected a minimum information copula should fit better to the data based on more constraints. We verify this point by fitting and comparing two minimum information copulae based on 3 and 12 constraints between variables M and B.

We first use the Fourier copula of order 12 or a base with 12 elements to approximate the copula between variables M and B, denoted by C_{MB}. The selected objective functions for this base are:

$$h'_1(M,B) = MB, \quad h'_2(M,B) = B^2M, \quad h'_3(M,B) = M^2B, \quad h'_4(M,B) = M^3B$$

$$h'_5(M,B) = M^2B^2, \quad h'_6(M,B) = MB^3, \quad h'_7(M,B) = B^2M^3, \quad h'_8(M,B) = M^2B^3$$

$$h'_9(M,B) = M^3B^3, \quad h'_{10}(M,B) = MB^4 \quad h'_{11}(M,B) = M^4B, \quad h'_{12}(M,B) = M^5B$$

The minimum information copula C_{MB} with respect to the uniform distribution given the constraints above has been constructed on the same grid of 50 by 50 points. The constraints presented as the expectations of the objective functions and their Lagrange multipliers required to construct the minimally informative copula between M and B are reported in Table 1.

The minimally informative copula density C_{MB} given the constraints reported in Table 1 is presented in Figure 10.

The minimum information copula C_{MB} with respect to the uniform distribution and given the first three constraints reported in the first column of Table 1, that is, $E[h'_1(M,Z)] =$

The Constraints	The constraints approximated based on C_{MB}	Lagrange multipliers
$E[h'_1(M,B)] = 0.2905$	0.29026	26.2459
$E[h'_2(M,B)] = 0.2075$	0.20747	-27.4369
$E[h'_3(M,B)] = 0.2066$	0.20683	-3.3349
$E[h'_4(M,B)1] = 0.1611$	0.16122	-9.3389
$E[h'_5(M,B)] = 0.1527$	0.153	-0.3289
$E[h'_6(M,B)] = 0.1624$	0.16238	1.6079
$E[h'_7(M,B)] = 0.1217$	0.12181	25.9629
$E[h'_8(M,B)] = 0.1223$	0.1224	8.3688
$E[h'_9(M,B)] = 0.0989$	0.0988	-14.0917
$E[h'_{10}(M,B)] = 0.1340$	0.1338	6.5137
$E[h'_{11}(M,B)] = 0.1324$	0.1323	-17.0045
$E[h'_{12}(M,B)] = 0.1126$	0.11228	10.1875

Table 1: The constraints, the approximated expectations, and their Lagrange multipliers to construct the minimal informative copula between M and B.

0.2905, $E[h'_2(M,Z)] = 0.2075$ and $E[h'_3(M,Z)] = 0.2066$ has been constructed on the same grid of 50 by 50 points, presented in Figure 11. Their Lagrange multipliers are $\lambda_1 = 7.4845$, $\lambda_2 = 2.3297$ and $\lambda_3 = -2.6631$.

The log likelihoods of these two copulae based on 12 constraints and 3 constraints (and constructed on 200 by 200 grid points) are respectively: $logL^{12pt}_{c_{MB}} = 92.1645$ and $logL^{3pt}_{c_{MB}} = 87.1966$. This is in agreement with the point made above.

The conditional copulae in the second tree, T_2 can be similarly approximated based on the minimally informative copula described above. We first construct the conditional minimum information copula between $T \mid M$ and $B \mid M$ given the following constraints represented as the conditional expectations of some objective functions:

$$h'_1(T,B) = TB, \quad h'_2(T,B) = TB^2, \quad h'_3(T,B) = T^2B$$

To calculate this conditional copula, we divide the support of M into some arbitrary subintervals or bins (here, we use 10 bins), and we then compute the expectations of the aforementioned functions on each bin as the constraints. As a result, the minimum information copula $C_{(T,B)|M}$ with respect to the uniform distribution and the following constraints on the first bin, where $0 < M < 0.1$

$$E[h'_1(T,B) \mid M \in (0,0.1)] = 0.1594, \; E[h'_2(T,B) \mid M \in (0,0.1)] = 0.0678,$$

$$E[h'_3(T,B) \mid M \in (0,0.1)] = 0.1224$$

has been constructed on the same grid of 50 by 50 points and shown in Figure 12.

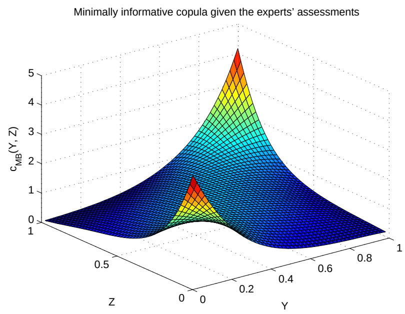

Figure 10: The minimally informative copula between M and B variables of Norwegian Stock data based on 12 constraints presented in the first column of Table 1.

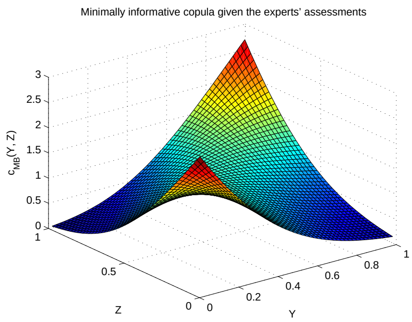

Figure 11: The minimally informative copula between M and B variables of Norwegian Stock data based on the first 3 constraints presented in the first column of Table 1.

Bin	The constraints $(E[h'_1 \mid M], E[h'_2 \mid M], E[h'_3 \mid M])$	Lagrange multipliers $(\lambda_1, \lambda_2, \lambda_3)$
$0 < M < 0.1$	(0.1594, 0.0678, 0.1224)	(10.9383, -8.4123, -6.9916)
$0.1 < M < 0.2$	(0.1785, 0.0857, 0.1252)	(-4.2491, 8.1402, -5.2132)
$0.2 < M < 0.3$	(0.207, 0.1181, 0.1357)	(2.1269,-1.7432,-1.4931)
$0.3 < M < 0.4$	(0.1891,0.1032,0.1171)	(-7.2137,2.0704,1.9255)
$0.4 < M < 0.5$	(0.2587,0.1748,0.1653)	(-8.7337,5.2627,3.8922)
$0.5 < M < 0.6$	(0.2377,0.1538,0.1526)	(-12.5348,-1.6083,14.9014)
$0.6 < M < 0.7$	(0.2712,0.1802,0.1673)	(3.9591,-7.3273,3.5925)
$0.7 < M < 0.8$	(0.2595,0.1736,0.1618)	(7.3803,-15.482,8.9438)
$0.8 < M < 0.9$	(0.3156,0.2386,0.1945)	(9.2597,-9.2321,-0.4385)
$0.9 < M < 1$	(0.2626,0.2087,0.1618)	(-0.7429,-1.1895,0.6117)

Table 2: The constraints and corresponding Lagrange multipliers associated with the conditional minimal informative copula between $T \mid M \in (0,1)$ and $B \mid M \in (0,1)$ for each bin

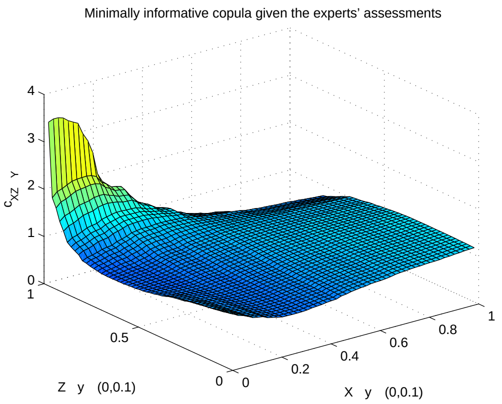

Figure 12: The minimally informative copula between $T \mid M \in (0, 0.1)$ and $B \mid M \in (0, 0.1)$ variables of Norwegian Stock data given $E[h'_1(T,B) \mid 0 < M < 0.1], E[h'_2(T,B) \mid 0 < M < 0.1], E[h'_3(T,B) \mid 0 < M < 0.1]$ constraints .

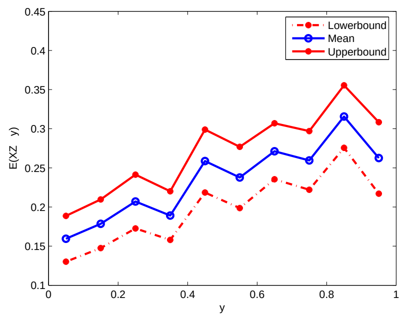

Figure 13: The changes of $E[h'_1(T,B) \mid 0 < M < 1]$ over the bins.

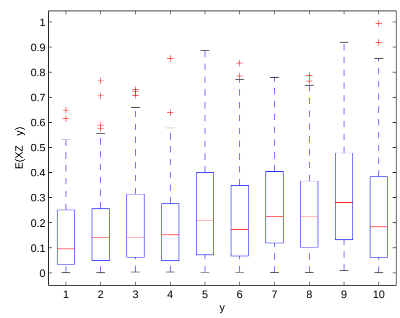

Figure 14: Box-plot demonstration of the $E[h'_1(T,B) \mid 0 < M < 1]$.

Table 2 shows the constraints and the corresponding Lagrange multipliers required to build conditional minimum information copula between $T \mid M \in (0,1)$ and $B \mid M \in (0,1)$ for 10 bins.

It is important to study the changes of the conditional expectation, $E[h'_1(T,B) \mid M]$ ($E[XZ \mid y]$) for different values of M or over the bins. Figure 13 shows this conditional expectation, $E[h'_1(T,B) \mid M]$, calculated from the minimum information copula $C(T \mid M, B \mid M)$ where M varies on $(0,1)$ along with the 95% confidence interval around the mean. As we can observe form this figure the changes of this measure is not...

The Box-plot demonstration of this conditional expectation, $E[h'_1(T,B) \mid 0 < M < 1]$ is illustrated in Figure 14. Similarly, we construct the conditional minimum information copula between $M \mid B$ and $S \mid B$ given the following constraints represented as the conditional expectations of some objective functions:

$$h'_1(M,S) = MS, \quad h'_2(M,S) = MS^2, \quad h'_3(M,S) = M^2S$$

Table 3 shows the constraints and the corresponding Lagrange multipliers required to build conditional minimum information copula between $M \mid B \in (0,1)$ and $S \mid B \in (0,1)$ for 10

Bin	The constraints $(E[h'_1 \mid B], E[h'_2 \mid B], E[h'_3 \mid B])$	Lagrange multipliers $(\lambda_1, \lambda_2, \lambda_3)$
$0 < B < 0.1$	(0.1472,0.1098,0.0702)	(0.4552,-0.3733,-0.4281)
$0.1 < B < 0.2$	(0.2105,0.1467,0.1177)	(-11.3758,-5.3255,19.6037)
$0.2 < B < 0.3$	(0.1941,0.1274,0.1087)	(-3.9076,0.3817,2.5244)
$0.3 < B < 0.4$	(0.2217,0.1493,0.1348)	(-2.1896,-3.4284,6.6761)
$0.4 < B < 0.5$	(0.237,0.1502,0.1561)	(-2.9583,1.0778,2.3114)
$0.5 < B < 0.6$	(0.2727,0.178,0.1824)	(3.1245,5.2801,-6.2032)
$0.6 < B < 0.7$	(0.271,0.1857,0.1707)	(8.0195,2.6283,-9.3362)
$0.7 < B < 0.8$	(0.2807,0.178,0.2)	(12.1643,-3.657,-6.1391)
$0.8 < B < 0.9$	(0.3023,0.1929,0.2274)	(-14.6401,7.9416,6.9864)
$0.9 < B < 1$	(0.3207,0.2056,0.2632)	(-7.1654,5.4488,0.8459)

Table 3: The constraints and corresponding Lagrange multipliers associated with the conditional minimal informative copula between $M \mid B \in (0,1)$ and $S \mid B \in (0,1)$ for each bin.

bins (the bins are obtained by dividing the support of B in 10 equal length sub-interval).

Figure 15 shows this conditional expectation, $E[h'_1(M,S) \mid B]$, calculated from the minimum information copula $C(M \mid B, S \mid B)$ where B varies on $(0,1)$ along with the 95% confidence interval around the mean. As we can observe form this figure the changes of this measure is not...

The Box-plot demonstration of this conditional expectation, $E[h'_1(T,B) \mid 0 < M < 1]$ is illustrated in Figure 16.

The conditional minimally informative copula in the third tree, T_3 can be similarly obtained as described above. In this situation, we first divide the conditioning variables supports into some sub-intervals, and we then construct the minimum information copula $T \mid (M,B)$ and $S \mid (M,B)$ given the some constraints represented as the conditional expectations of some objective functions where the conditioning variables varies over the specified bins. Figure 17 shows a minimally informative copula between $T \mid \{M \in (0.33), B \in (0,0.33)\}$ and $S \mid \{M \in (0.33), B \in (0,0.33)\}$ with respect to the uniform distribution and given three constraints $e_1 = E[h'_1(T,S) \mid y \in \{M \in (0.33), B \in (0,0.33)\}] = 0.394, e_2 = E[h'_2(T,S) \mid \{M \in (0.33), B \in (0,0.33)\}] = 0.295, e_3 = E[h'_3(T,S) \mid \{M \in (0.33), B \in (0,0.33)\}] = 0.3115$ which is constructed on the same grid of 50 by 50 data points. The objective functions used as the constraints are:

$$h'_1(T,S) = TS, \quad h'_2(T,S) = TS^2, \quad h'_3(T,S) = T^2S$$

Table 4 shows the constraints and corresponding Lagrange multipliers which enable us to construct the minimum information copula over the corresponding bin.

Figure 18 illustrates the changes of the conditional expectation $E[h_1(T,S) \mid 0 < M <$

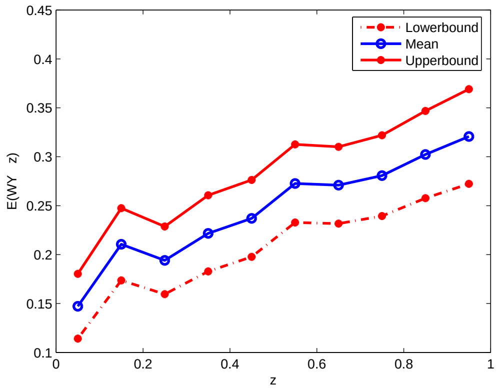

Figure 15: The conditional expectation $E[h_1'(M,S) \mid 0 < B < 1]$ derived from the minimally informative copula between $M \mid B \in (0,1)$ and $S \mid B \in (0,1)$ obtained above.

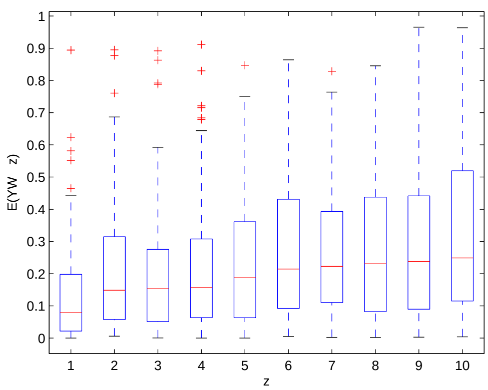

Figure 16: Box-plot demonstration of $E[h_1(M,S) \mid 0 < B < 1]$.

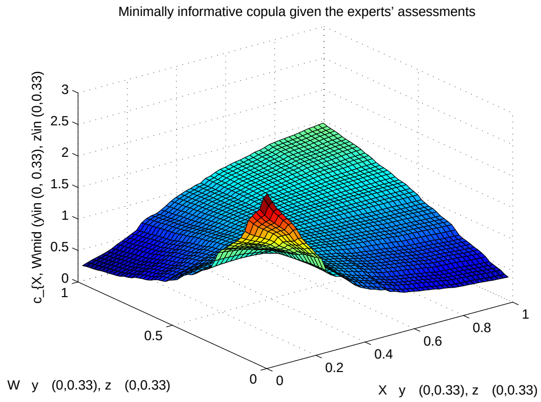

Figure 17: The minimally informative copula between $T \mid \{M \in (0.33), B \in (0, 0.33)\}$ and $S \mid \{M \in (0.33), B \in (0, 0.33)\}$ variables of Norwegian Stock data given $e_1 = E[h_1'(T,S) \mid \{M \in (0.33), B \in (0, 0.33)\}] = 0.394, e_2 = E[h_2'(T,S) \mid \{M \in (0.33), B \in (0, 0.33)\}] = 0.295, e_3 = E[h_3'(T,S) \mid \{M \in (0.33), B \in (0, 0.33)\}] = 0.3115$ constraints .

Bins	$(E[h_1'(T,S) \mid M, B]$ $E[h_1'(T,S) \mid M, B]$, $E[h_1'(T,S) \mid M, B])$	Lagrange multipliers $(\lambda_1, \lambda_2, \lambda_3)$
$0 < M < 0.33, 0 < B < 0.33$	(0.394,0.295,0.3115)	(5.0563,0.0806,-1.1935)
$0 < M < 0.33, 0.33 < B < 0.66$	(0.2995,0.1975,0.2192)	(5.8976,2.1862,-3.8)
$0 < M < 0.33, 0.66 < B < 1$	(0.2089,0.1346,0.1381)	(-0.5927,5.6017,0.1003)
$0.33 < M < 0.66, 0 < B < 0.33$	(0.2548,0.1731,0.1541)	(6.8429,2.8645,-6.5986
$0.33 < M < 0.66, 0.33 < B < 0.66$	(0.2459,0.1661,0.1623)	(10.2143,-3.9284,-1.6448)
$0.33 < M < 0.66, 0.66 < B < 1$	(0.2414,0.1643,0.1657)	(-5.0439,5.6452,3.3403)
$0.66 < M < 1, 0 < B < 0.33$	(0.2992,0.222,0.1976)	(0.3324,2.6365,1.6332)
$0.66 < M < 1, 0.33 < B < 0.66$	(0.2766,0.1942,0.1895)	(-0.0135,3.8203,-1.2379)
$0.66 < M < 1, 0.66 < B < 1$	(0.2163,0.1473,0.1334)	(16.729,-0.3679,-12.5079)

Table 4: The constraints and corresponding Lagrange multipliers associated with the conditional minimally informative copula between $T \mid (M, B)$ and $S \mid (M, B)$ for each bin.

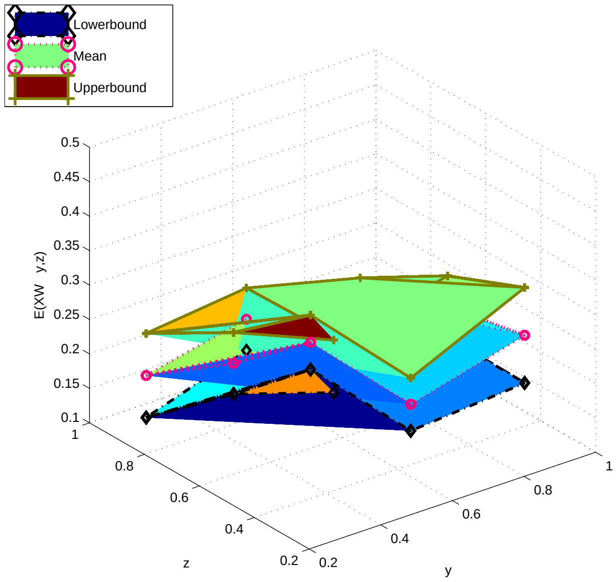

Figure 18: The conditional expectation $E[h_1(T,S) \mid 0 < M < 1, 0 < B < 1]$ derived from the minimally informative copula between $T \mid \{M \in (0,1), B \in (0,1)\}$ and $S \mid \{M \in (0,1), B \in (0,1)\}$ obtained above.

$1, 0 < B < 1]$ (the middle plane and recognised by "O" in the figure) and 95% confidence bound (we use "+" to display the upperbound, and "$\diamondsuit$" denotes the lowerbound) over the bins specified in Table 4.

6 Conclusion

In this paper, we present a novel method to approximate a multivariate distribution by any vine structure to any degree of approximation. Our approach uses the minimum information copulas that can be specified to any required degree of precision based on the data available. We prove rigourously that good approximation 'locally' guarantees good approximation globally. This approximation allows the use of a fixed finite dimensional family of copulas to be used in a vine construction, with the promise of a uniform level of approximation.In other words, we can use the same bases to approximate each copula in each tree of the corresponding vine.

However, a vine structure imposes no restrictions on the underlying joint probability distribution it represents, but this is crucial to investigate which vine structure is most appropriate. The choice of vine structure becomes more significant when we truncate class of copulae to make search strategy simpler. Therefore, the approximation of a multivariate distribution using a vine structure for a given multivariate copula depends on the bases represent the truncated class of copula and approximate level ϵ. This approximation can be made more accurate by adding more bases to achieve the desired level of approximation ϵ.

We wish to extend our method by using a series expansion, like a two-dimensional Fourier series or generalized Fourier series, to approximate any log-density function by truncating the series at an appropriate point.

We are also considering the possibility of extending this work by using the emulators to estimate the expensive bases and therefore to approximate the resulting couples and vine.

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] Bedford, T. (2006). Interactive expert assignment of minimally-informative copulae.

[3] Bedford, T. and Cooke. R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence* **32**, 245268.

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

[5] Bedford, T., and Meeuwissen, A. (1997). Minimally informative distributions with given rank correlation for use in uncertainty analysis. *Journal of Statistical Computation and Simulation*, **57(1 - 4)**: 143 - 174.

[6] Borwein, J., Lewis, A., and Nussbaum, R. (1994). Entropy minimization, DAD problems, and doubly stochastic kernels. *Journal of Functional Analysis*, **123**: 264–307.

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

[8] Knight, P. H. (2007). The Sinkhorn-Knopp Algorithm: Convergence and Applications. CERFACS Technical Report TR/PA/06/42, Strathclyde University Mathematics Department Technical Report 8-2006.

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

[10] Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright. P. E. (1998). Convergence pro- perties of the Nelder-Mead simplex method in low dimensions. S*IAM Journal of Optimization*, **9(1)**:112–147.

[11] Lewandowski, D. (2008).*High Dimensional Dependence: Copulae, Sensitivity, Sampling.* PhD thesis, Delft University.

[12] Sinkhorn, R., and Knopp, P. (1967). Concerning nonnegative matrices and doubly stochastic matrices. *Pacific J. Math.*, **21**, 343348.