DIGITALCOMMONS @WAYNESTATE

Journal of Modern Applied Statistical Methods

Volume 12 | Issue 2 Article 25

11-1-2013

Approximation Multivariate Distribution of Main Indices of Tehran Stock Exchange with Pair-Copula

G. Parham
Shahid Chamran University, Ahvaz, Iran, Parham_g@scu.ac.ir

A. Daneshkhah
Cranfield University, Cranfield, UK, ardaneshkhah@gmail.com

O. Chatrabgoun
Shahid Chamran University, Ahvaz, Iran, o-chatrabgoun@phdstu.scu.ac.ir

Recommended Citation

Parham, G.; Daneshkhah, A.; and Chatrabgoun, O. (2013) "Approximation Multivariate Distribution of Main Indices of Tehran Stock Exchange with Pair-Copula," *Journal of Modern Applied Statistical Methods*: Vol. 12 : Iss. 2 , Article 25.
DOI: 10.22237/jmasm/1383279840
Available at: http://digitalcommons.wayne.edu/jmasm/vol12/iss2/25

Journal of Modern Applied Statistical Methods
November 2013, Vol. 12, No. 2, 405-426.

ISSN 1538 − 9472

Approximation Multivariate Distribution of Main Indices of Tehran Stock Exchange with Pair-Copula

G. Parham
Shahid Chamran University
Ahvaz, Iran

A. Daneshkhah
Cranfield University
Cranfield, UK

O. Chatrabgoun
Shahid Chamran University
Ahvaz, Iran

The multivariate distribution of five main indices of Tehran stock exchange is approximated using a pair-copula model. A vine graphical model is used to produce an n-dimensional copula. This is accomplished using a flexible copula called a minimum information (MI) copula as a part of pair-copula construction. Obtained results show that the achieved model has a good level of approximation.

Keywords: Minimum information copula, pair-copula, vine.

Introduction

Sometimes in applied probability and statistics it is necessary to model multiple uncertainties or dependencies using multivariate distributions. To do it, it is common to use discrete model such as Bayesian networks but when modeling financial data, it is necessary to have model of continuous random variables. Copulas are quickly gaining popularity as modeling dependencies e.g. surveys by Nelsen (1999), Joe (1997). Copulas have found application in a number of areas of operations research including combining expert opinion and stochastic simulation, (e.g. Abbas et al. (2010) and references cited therein). A copula is a joint distribution on the unit square (or more generally on the unit n-cube) with uniform marginal distributions. Under reasonable conditions, a joint distribution for n-random variables can be found by specifying the univariate distribution for each variable, and in addition, specifying the copula. Following Sklar (1959) the joint distribution function of random vector $(X_1, \ldots, X_n)$ is

Dr. Parham is Faculty of Mathematical Sciences and Computer. Email at: Parham_g@scu.ac.ir. Dr. Daneshkhah is a Lecturer in Utility Asset Management. Email at: ardaneshkhah@gmail.com. Dr. Chatrabgoun is Faculty of Mathematical Sciences and Computer. Email at: o-chatrabgoun@phdstu.scu.ac.ir.

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

Where C is a copula distribution function, and $F_1, \dots, F_n$ are the univariate, or marginal, distribution functions. A special case is that of the 'Gaussian copula', obtained from Gaussian joint distribution and parameterized by the correlation matrix. Use of the Gaussian copula to construct joint distributions is equivalent to the NORTA method (normal to anything). 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 is 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.

As mentioned because of difficulty in specifying the copulas and restricted to the exact class, copula approximation is to some extend new topic in this case. The approach used herein allows a lot of flexibility in copula specification that was analyzed and some properties of it was said in Bedford et al. (2013) and developed by Daneshkhah et al. (2013), and for approximation multivariate distribution, a graphical model, called a vine, is used to systematically specify how two-dimensional copulas are stacked together to produce an n-dimensional copula.

The main objectives is to show that a vine structure can be used to approximate Tehran stock exchange multivariate copula to any required degree of approximation. The standing technical assumptions are that the multivariate copula density f under study is continuous and is non-zero. No other assumptions are needed. A constructive approach involves the use of minimum information (*MI*) copula that can be specified to any required degree of precision based on the data available. According to Bedford et al. (2013) good approximation locally guarantees good approximation globally.

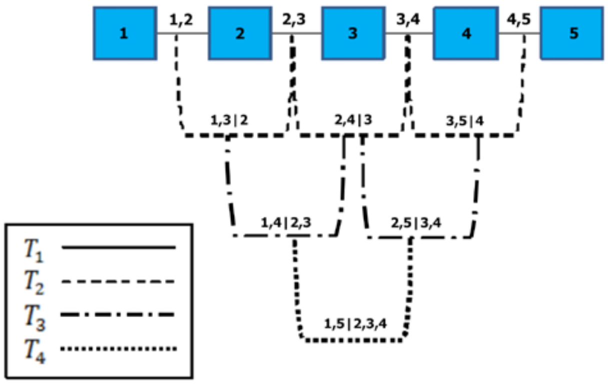

Figure 1. A Regular Vine with 5 Elements

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 to ignore the question about which vine structure is most appropriate, for some structures allow the use of less complex conditional copulas than others. Conversely, if only certain families of copulas are allowed then one vine structure might fit better than another.

Vine constructions for multivariate dependency

A copula is a multivariate distribution function with standard uniform marginal distributions. Using (1) it may be observed that a copula can be used, in conjunction with the marginal distributions, to model any multivariate distribution. However, apart from the multivariate Gaussian, Student, and the exchangeable multivariate Archimedean copulas, the set of higher-dimensional copulas proposed in the literature is limited and is not rich enough to model all possible mutual dependencies amongst the n variants (see Kurowicka & Cooke, 2006 for details of these copulas). 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) $n(n-1)/2$ copulas between n variables. (Note that $n(n-1)/2$ is the number of entries above the diagonal of an $n \times n$ correlation matrix - though these are algebraically related so not completely free variables). This structure was originally proposed by Joe (1997), and later reformulated and discussed in detail by Bedford and Cooke (2001, 2002), who considered simulation, information properties and the relationship to the multivariate normal distribution but who also considered a more general method called a Cantor tree construction. Kurowicka and Cooke (2006) considered simulation issues, and Aas et al. (2009) examined inference. The modeling scheme is based on a decomposition of a multivariate density into a set of bivariate copulas. The way these copulas are built up to give the overall joint distribution is determined through a structure called a vine, and can be easily visualized. A vine on n variables is a nested set of trees, where the edges of the tree j are the nodes of the tree $j+1$ (for $j=1,...,n-2$), and each tree has the maximum number of edges. For example, Figure 1 shows a vine with 5 variables which consists of four trees $\left(T_1,T_2,T_3,T_4\right)$ with 4, 3, 2 and 1 edges, respectively. 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, for $j=1,...,n-2$. There are $n\left(n-1\right)/2$ edges in a regular vine on n variables. The formal definition is as follows.

Definition: (Vine, regular vine) V is a vine on n elements if

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

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

V is a regular vine on n elements if additionally the proximity condition holds:

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

One of the simplest regular vines is shown in Figure 1 - this structure is called D-vine, see Kurowicka and Cooke, 2006, pp. 93. 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 two edges of T_1, for example 12 and 23, which are nodes joined by an edge in T_2 are taken, then of the numbers labeling these edges one is common to both (2), and they both have one unique number (1,3 respectively). The common number(s) will be called the conditioning set D_e for that edge e (in this example the conditioning set is simply {2}) and the other numbers will be called the conditioned set (in this example {1, 3}). For a regular vine the conditioned set always contains two elements.

A vine distribution is associated to a vine by specifying a copula to each edge of T_1 and a family of conditional copulas for the conditional variables given the conditioning variables, as shown by the following result of Bedford and Cooke (2001).

Theorem 1: Let $V=\left(T_1,...,T_{n-1}\right)$ be a regular vine on n elements. For each edge $e\left(j,k\right)\in T_i$, $i=1,...,n-1$ with conditioned set $\left\{j,k\right\}$ and conditioning set D_e, let the conditional copula and copula density be $C_{jk|D_e}$ and $c_{jk|D_e}$ respectively. Let the marginal distributions F_i with densities f_i, $i=1,...,n$ be given. Then the vine-dependent distribution is uniquely determined and has a density given by

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

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

The density decomposition associated with 5 random variables $X=\left(X_1,...,X_5\right)$ with a joint density function $f\left(x_1,\dagger\ldots,\dagger x_5\right)$ satisfying a copula-vine structure shown in Figure 1 with the marginal densities $f_1,\ldots,f_5$ is

$$\begin{aligned}
f_{12345}=\prod_{i=1}^{n}f\left(x_i\right)\times c_{12}\left(F\left(x_1\right),F\left(x_2\right)\right)c_{23}\left(F\left(x_2\right),F\left(x_3\right)\right)c_{34}\left(F\left(x_3\right),F\left(x_4\right)\right)c_{45}\left(F\left(x_4\right),F\left(x_5\right)\right)\\
\times c_{13}\left(F\left(x_1|x_2\right),F\left(x_3|x_2\right)\right)c_{24}\left(F\left(x_2|x_3\right),F\left(x_4|x_3\right)\right)c_{35}\left(F\left(x_3|x_4\right),F\left(x_5|x_4\right)\right)\\
\times c_{14}\left(F\left(x_1|x_2,x_3\right),F\left(x_4|x_2,x_3\right)\right)c_{25}\left(F\left(x_2|x_3,x_4\right),F\left(x_5|x_3,x_4\right)\right)\\
\times c_{15}\left(F\left(x_1|x_2,x_3,x_4\right),F\left(x_5|x_2,x_3,x_4\right)\right)
\end{aligned} \tag{3}$$

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

$$f_{12}(x,y) = f_1(x)\, f_2(y)\, c_{12}\left(F_1(x), F_2(y)\right)$$

or equivalently

$$f_{1|2}(x \mid y) = f_1(x)\, c_{12}\left(F_1(x), F_2(y)\right)$$

where c_{12} is the copula density and F_1, F_2 are the univariate distributions. Starting with

$$\begin{aligned} f_{12345}(x_1,\ldots,x_5) = {} & f_1(x_1)\, f_{2|1}(x_2 \mid x_1)\, f_{3|12}(x_3 \mid x_1, x_2) \\ & f_{4|123}(x_4 \mid x_1, x_2, x_3)\, f_{5|1234}(x_5 \mid x_1, x_2, x_3, x_4) \end{aligned}$$

inductively convert the latter expression in to that shown in (3). This results in

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

Next,

$$\begin{aligned} f_{3|12}(x_3|x_1, x_2) &= f_{3|2}(x_3|x_2)\, c_{13|2}\left(F_{1|2}(x_1 \mid x_2), F_{3|2}(x_3 \mid x_2)\right) \\ &= f_3(x_3)\, c_{23}\left(F_2(x_2), F_3(x_3)\right) c_{13|2}\left(F_1(x_1 \mid x_2), F_3(x_3 \mid x_2)\right) \end{aligned} \tag{4}$$

The calculation for the remaining term $f_{5|1234}(x_5 \mid x_1, x_2, x_3, x_4)$ is left to the reader.

Note that in the special case of a joint normal distribution, the normal copula would be used everywhere in the above expression and the conditional copulas would be constant (i.e. not depend on the conditioning variable). This means that the joint normal structure is specified by $n(n-1)/2$ (conditional) correlation values, which are algebraically free between -1 and +1 (unlike the values in a correlation matrix). See Bedford and Cooke (2002) for more details. The above theorem gives a constructive approach to build a multivariate distribution given a vine structure: If choices of marginal densities and copulas are made then the above formula will give a multivariate density. Hence, vines can be used to model general multivariate densities. However, in practice it is necessary to use copulas from a convenient class, and this class should ideally be one that allows any given copula to be approximated to an arbitrary degree. Having this class of copulas allows any multivariate distribution to be approximated 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, when specific families of copulas are used it seems that some vine structures do work better than others. That is, given a family of copulas, some vine structures may give a better degree of approximation than others. It is worth stressing the point that the flexibility of vines gives the potential to capture any fine grain structure within a multivariate distribution. A key aspect that cannot be modeled by Bayesian networks is that of conditional dependence. Bayesian networks are built around the concept of conditional independence -arrows from a parent node to two child nodes means that the child variables are conditionally independent given the parent variable. However, different models of conditional dependence are not available as building blocks in Bayesian networks. Multivariate Gaussian copulas do allow for a specification of conditional dependence, but do not allow that dependence to change - in a multivariate normal distribution, the conditional correlation of two variables given a third may be non-zero but is always constant. This approach, by contrast, allows the explicit modeling of non-constant conditional dependence.

The minimum information (MI) copula using the D_1AD_2 algorithm

Bedford et. al (2013) presented a way to approximate a copula using minimum information methods which demonstrate uniform approximation in the class of copula used. Bedford and Meeuwissen (1997) applied a so-called DAD algorithm to produce discretized MI copula with given rank correlation. This approach can be used whenever it is desirable to specify the expectation of any symmetric function of $U = F(x)$ and $V = F(y)$.

In order to have asymmetric specifications the D_1AD_2 algorithm must be used where A is a positive square matrix, thus, diagonal matrices D_1 and D_2 can be found such that the product of D_1AD_2 is doubly stochastic. It is possible to correlate the variables of interest X and Y by introducing constraints based on knowledge about functions of these variables. Suppose there are k of these functions, namely $h_1^{'}(X,Y), h_2^{'}(X,Y), \ldots, h_k^{'}(X,Y)$ and mean values $\alpha_1,...,\alpha_k$ are specified for all functions respectively from the data or the expert judgment.

Corresponding functions of the copula variables U and V, defined by $h_1(U,V) = h_1^{'}\left(F_1^{-1}(U), F_2^{-1}(V)\right)$, etc. can be defined and clearly these should also have the specified expectations $\alpha_1,...,\alpha_k$. The kernel

$$A(u,v)=exp(\lambda_1 h_1(u,v)+\ldots+\lambda_k h_k(u,v)) \tag{5}$$

is formed, where u denotes the realization of U and v the realization of V.

For practical implementations it is necessary to discretize the set of (u,v) values such that the whole domain of the copula is covered. This means that the kernel A described above becomes a 2-dimensional matrix A and that the matrices D_1 and D_2 are required to create a discretized copula density

$$P = D_1 A D_2 \tag{6}$$

Suppose that both U and V are discretized into n points, respectively u_i, and v_j, $i, j = 1,...,n$. Then $A = (a_{ij}), D_1 = diag\left(d_1^{(1)},\ldots,d_n^{(1)}\right)$ and $D_2 = diag\left(d_1^{(2)},\ldots,d_n^{(2)}\right)$ where $a_{ij} = A(u_i, v_j), d_i^{(1)} = D_1(u_i)$ and $d_i^{(2)} = D_2(u_i)$. The double stochastically of $D_1 A D_2$ with the extra assumption of uniform marginals means that

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

and

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

because for any given i and j the selected cell size in the unit square is $1/n$. Hence

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

and

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

The D_1AD_2 algorithm works by fixed point iteration and is closely related to iterative proportional fitting algorithms.

It can be shown that a multivariate distribution can be arbitrarily well approximated by using a fixed family of bivariate copula. 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 $[0,1]^2$ (see Bedford et al. (2013) for proof). Based on this it can be shown that the same finite parameter family of copula can be used to give a given level of approximation to all conditional copula simultaneously.

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

$$u, v, uv, u^2, v^2, u^2, vu, v^2, \ldots.$$

Given an ordered basis $h_1, h_2, \ldots \in C\left([0,1]^2\right)$ and a required degree of approximation $\in > 0$ in the sup metric, Bedford et al. (2013) stated the following theorem.

Theorem 2: Given $\in > 0$, there is a k such that any member of $LNC(f)$ can be approximated to within error $\in > 0$ by a linear combination of $h_1, \ldots, h_k$.

First consider a practical guide to build a minimally informative copula structure briefly discussed to approximate any multivariate distribution. A multivariate distribution can be approximated as follows:

- Specify a basic family $B(k)$
- Specify a pair-copula structure
- For each part of pair-copula specify either

  1. mean $\alpha_1,...,\alpha_k$ for $h_1,...,h_k$ on each pairwise copula;
  2. functions $\alpha_m(ji|D_e)$ for the mean values as functions of the conditioning variables, for $m=1,...,k$, where D_e is the conditioning set for the edge e.

Data set

A data set of Tehran stock exchange is used that includes five time series of daily data: the overall index (*O*), the industry index (*I*), the free float index (*F*), the main board index (*M*) and the secondary index (*S*). All are for the period January 5^{th}2008 to October 30^{th} 2011. (number of observation equal to 668) These five variables are denoted by *O*, *I*, *F*, *M* and *S*, respectively.

First, remove serial correlation in the five time series i.e. the observation of each variable must be independent over time. Hence, the serial correlation in the conditional mean and the conditional variance are modeled by an AR(1) and a GARCH(1,1) model (Bollerslev, 1986), respectively. That is for time series i, the following model for log-return x_i;

$$x_{i,t} = c_i + \alpha_i x_{i,t-1} + \sigma_{i,t} z_{i,t}$$

$$E\left[z_{i,t}\right] = 0$$

and

$$Var\left[z_{i,t}\right] = 1$$

Where $\varepsilon_{i,t-1} = \sigma_{i,t} + z_{i,t}$. Aas et al. (2009)

The further analysis is performed on the standard residuals z_i. If the AR(1)-GARCH(1,1) models are successful at modeling the serial correlation in the conditional mean and the conditional variance, there should be no autocorrelation left in the standard residuals and squared standard residuals. The modified Qstatistic is used (Ljung and Box, 1979) and the Lagrange Multiplier Test (LM) Engle (1982), respectively, to check this. For all series and both tests, the null hypothesis that there is no autocorrelation left cannot be rejected at the 5% level. Because interest lies mainly in estimating the dependence structure of the risk factor, the standard residual vectors are converted to the uniform variables using the kernel method before further modeling.

It is necessary to generate a vine approximation fitted as in Figure 2 to this data set using minimum information distributions. It should be noted that the corresponding functions of the copula variables *X*, *Y*, *Z*, *U* and *V* associated with *O*, *I*, *F*, *M* and *S* can be found. These are defined by, for example, $h_i\left(X,Y\right)=h_i^{'}\left(F_1^{-1}\left(O\right),F_2^{-1}\left(I\right)\right)$ and should have the same specified expectation, in this case $E[h_i\left(X,Y\right)\Big]=E[h_i^{'}\left(O,I\right)\Big]$. The minimum information copulas calculated in this example are derived based on copula variables *X*, *Y*, *Z*, *U* and *V*.

It should be noticed that to generalize to other stock exchanges and other applications, a vine structure can be determined uniquely by specifying the order of variables in the first tree T_1. To specify this order, we can use correlation scatter plot, Kendall's τ or the tail dependence coefficient (see e.g., Aas et al., 2009) to measure the strongest bivariate dependencies among the variables in the first tree of the D-vine (or C-vine) of interest. Once the Kendall's τ or the tail dependence coefficients between any pair of the variables in the first tree calculated, then order these measures, and put the variables with the highest measures next to each other and place the ones with weak dependencies farther away. Skipping to present the numerical details of these measures, and following Aas et al. (2009), use the pair-copula construction given in Figure 2 as the selected D-vine structure. In the case, that there is no data to compute these measures to specify the vine structure (or variables order in the first tree), we can use the expert's judgement to elicit these measures or other relevant measures that are more convenient to express by the expert (see Bedford et al. (2013) for a relevant work).

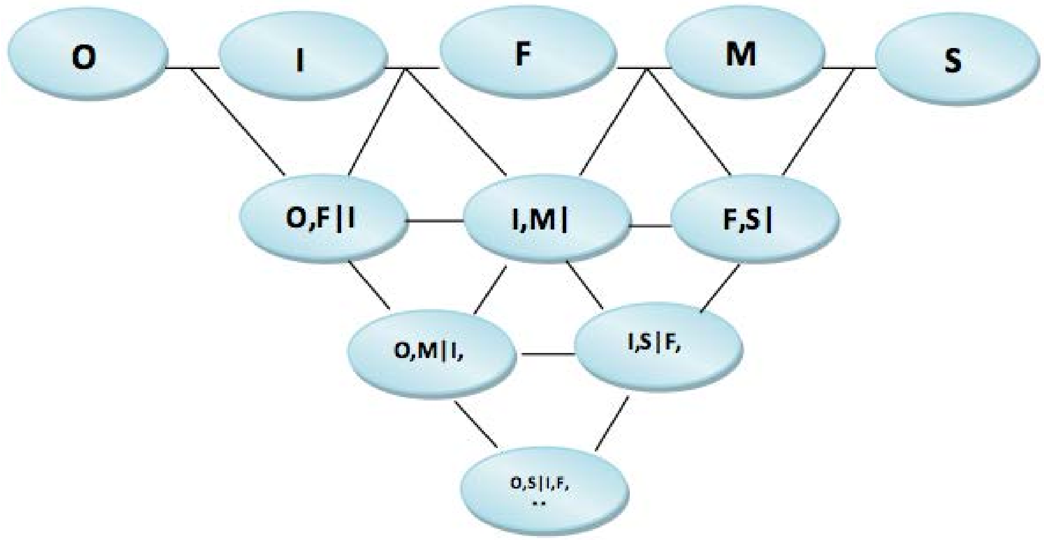

Figure 2. Selected vine structure for the Tehran stock exchange with 5 variables: overall index (O), industry index (I), free float index (F), main board index (M) and secondary index (S).

Initially minimally informative copulas are constructed between each set of two adjacent variables in the first tree, T_1. To do so it is necessary to decide upon which bases to take and how many discretization points to use in each case. The recommended procedure for first copula in T_1, between *O* and *I* is considered next.

Basis function

Which basis functions to include in the copula must first be decided. Basis functions could be chosen, starting with simple polynomials and moving to more complex ones, and including them until satisfied with our approximation. For example if the following basis functions in order is included,

$$OI, O^2I, OI^2, O^2I^2, O^3I, OI^3, O^2I^3, O^3I^2, O^4I, OI^4, OI^5, O^5I, O^3I^3, O^2I^4, O^4I^2$$

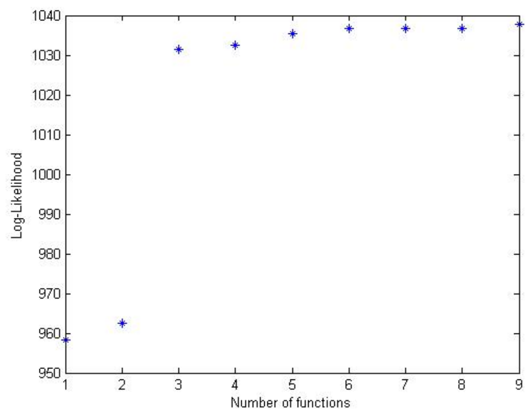

Figure 3. The log-likelihood of the minimally informative copula calculated based on different functions

then the log-likelihood for the copula changes as in Figure 3. There is a jump in the log-likelihood as the third basis function, OI^2 is added. This could imply that we are not adding the basis functions in an optimal manner. Instead at each stage, it is proposed to assess the log-likelihood of adding each additional basis function, then include the function which produces the largest increase in the log-likelihood. Thus the method is similar to a stepwise regression. Doing so for the initial copula yields the basis functions OI, O^2I, O^3I^2.

There is no longer a jump in the log-likelihood when adding the four basis function. The log-likelihood also increase more quickly and reaches its plateau value of around 1030 using fewer basis functions.

Fixing the values of the expectations of these functions by using the empirical data as follows

$$\alpha_1 = \frac{1}{667}\sum_{i=1}^{667} O_i I_i = 0.328,$$

$$\alpha_2 = \frac{1}{667}\sum_{i=1}^{667} O_i^2 I_i = 0.2428,$$

$$\alpha_3 = \frac{1}{667}\sum_{i=1}^{667} O_i^3 I_i^2 = 0.1578$$

The minimum information copula C_{OI} With respect to the uniform distribution given the three constraint above can then be constructed. In order to do so it is necessary to decide on the number of discretization points (or grid size). A larger grid size will provide a better approximation to the continuous copula but at the cost of more computation time. Similarly, the more iteration of the D_1AD_2 and the optimization algorithms that are run, the more accurate the approximation will become. This is again at the expense of speed. Comments on the convergence of the *DAD* algorithm are given in Bedford et al. (2013) and Daneshkhah et al. (2013). In terms of the optimization it is possible to specify how accurate the approximation should be and then judge the effect on the number of iterations required for convergence. In number of iterations needed will also depend on the grid size. In order to be consistent throughout the rest of the example, choose a grid size 50 � 50.

Having done this, the *MI* copula C_{OI} can now be found. This gives parameter value of

$$\lambda_1 = 907.8,$$
$$\lambda_2 = -1025.1,$$
$$\lambda_3 = 389.41$$

The result has been summarized in table 1 and copula plotted in Figure 4. Note that the Log-likelihood for this copula is 1031.4.

Table 1. Minimum information copula between O and I

Base	**Expectation**	**Parameter Value**
OI	0.3280	907.8
O^2I	0.2428	-1025.1
O^3I^2	0.1578	389.41

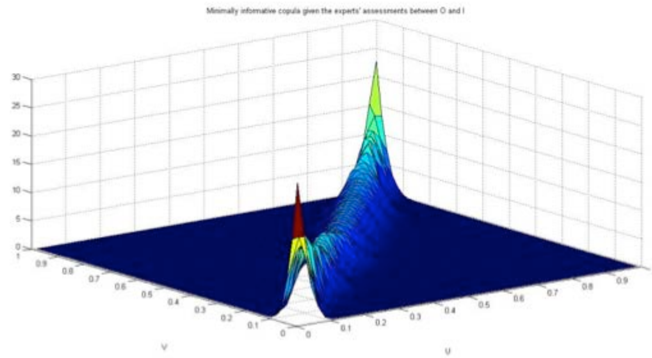

Figure 4. Minimum information copula between *O* and *I*

The second copula in T_1 is C_{IF}. Using the stepwise method as illustrated the following results obtained and the log-likelihood is $l_{IF} = 521.8$. The summarized result are given in Table 2, and Figure 5 shows the fitted copula.

Table 2. Minimum information copula between *I* and *F*.

Base	Expectation	Parameter Value
IF	0.3209	81.2
I^3F^3	0.1254	38.6
I^3F	0.1851	-75.7

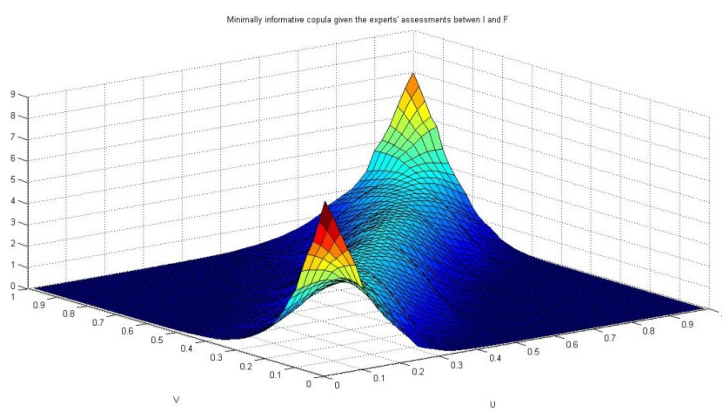

Figure 5. Minimum information copula between *I* and *F*

The third marginal copula is between F and M. Given a 50×50 grid and a required error of no more than 1×10^{-12} the three bases chosen using the stepwise procedure, the constraint for each base and the resulting parameter values are given in Table 3 and Figure 6 . The log-likelihood for this copula is 462.31.

Table 3. Minimum information copula between *F* and *M*

Base	**Expectation**	**Parameter Value**
FM	0.3195	60.97
F^4M^2	0.1252	26.42
FM^3	0.1839	-45.46

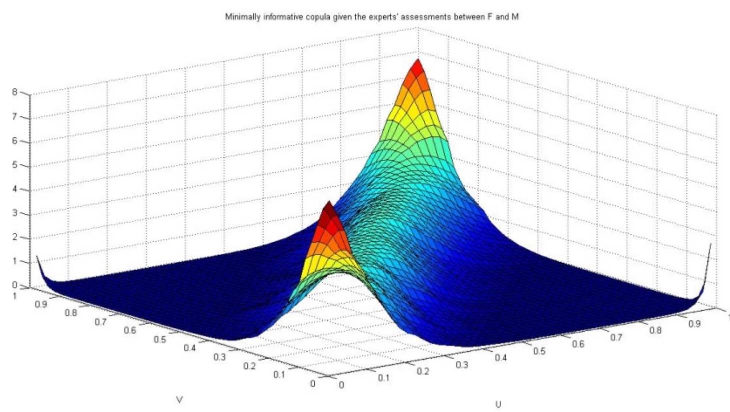

Figure 6. Minimum information copula between *F* and *M*

and the last copula in first tree, T_1, between M and S is C_{MS} . The result are summarized in Table 4 and Figure 7.

Table 4: Minimum information copula between M and S

Base	**Expectation**	**Parameter Value**
MS	0.2928	25.52
MS^2	0.2064	-23.22
M^2S^4	0.0989	8.44

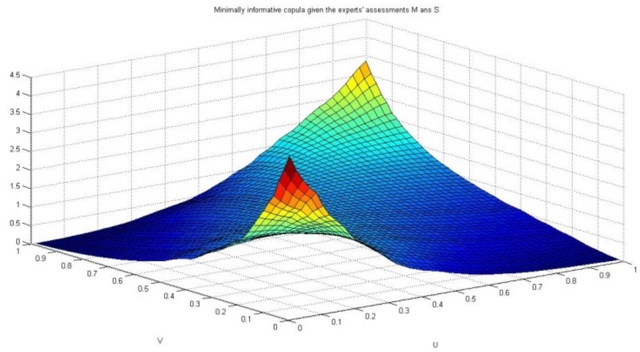

Figure 7: Minimum information copula between *M* and *S*

The conditional copulas in the second tree, T_2, can similarly be approximated using the minimum information approach. Initially the conditional *MI* copula between *O*|*I* and *F*|*I* is constructed. In order to calculate this copula, divide the support of *I* into some arbitrary sub-intervals or bins and then construct the conditional copula within each bin. To do so, find bases in the same way as for the marginal copulas and fit the copulas to the expectations calculated for these. Two bins are used so that the first copula is for $O,F|I \in (0,0.5)$. The bases for this copula are

$$h_1^{'}\left(O,F \mid I \in (0,0.5)\right) = OF, h_2^{'}\left(O,F \mid I \in (0,0.5)\right) = OF^2,$$
$$h_3^{'} = \left(O,F \mid I \in (0,0.5)\right) = OF^3$$

The expectations given these basis functions which will constrain the *MI* copula are

$$\alpha_1 = 0.0902, \quad \alpha_2 = 0.0368, \quad \alpha_3 = 0.168$$

This process can be followed again for the remaining bins. Table 5 shows the constraints and corresponding Lagrange multipliers required to build the conditional *MI* copula between $O|I \in (0.5,1)$ and $F|I \in (0.5,1)$. It also gives the log-likelihood in each case.

Table 5. Minimum information copula between *O* and *F* given *I*

Condition	Base	Expectation	Parameter Value	Log-likelihood
0< I <= 0. 5	(OF, OF^2, OF^3)	(.0902,.0368,.0168)	(274.76,-627.3 , 482.6)	195.94
0. 5< I <= 1	(O^2F^4,O^5F, O^4F^2)	(0.247,0.254,0.247)	(-18.2,-69.98 , 81.93)	162.92

Similarly, the *MI* copula can be constructed between remaining nodes in T_2, one of them *I*|*F* and *M*|*F* and another between *F*|*M* and *S*|*M* based on 2 bins and 3 constraints found in the usual manner. The resulting *MI* copula are summarized in Table 6 and Table 7.

Table 6. Minimum information copula between *I* and *M* given *F*

Condition	Base	Expectation	Parameter Value	Log-likelihood
0< F <= 0.5	(IM, IM2, I4M2)	(0.1193,0.06,0.017)	(982.3 , -881.7 , 298.2)	551.3
0.5<F<= 1	(I3M3, I4M2, I4M)	(0.258,0.259,0.302)	(704.4 , -242.1 , -216.8)	555.4

Table 7. Minimum information copula between *F* and *S* given *M*

Condition	Base	Expectation	Parameter Value	Log-likelihood
0<M<= 0.5	(FS, F3S, FS2)	(0.1314,0.0258,0.078)	(51.3 , -51.9 , -11.3)	87
0.5<M<= 1	(F5S, FS5, F3S3)	(0.222,0.197,0.193)	(5.3 , -5.3 , 6.5)	73.7

O|(*I*,*F*) and *M*|(*I*,*F*) are calculated on each combination of bins for *I*,*F*. Thus in T_3 there are 4 bins altogether. The bins, bases and log-likelihoods (l) associated with each copula are given in Table 8.

Similarly the *MI* copulas for *I*|(*F*,*M*) and *S*|(*F*,*M*) are calculated on each combination of bins for *F*,*M*. Table 9 shows the result in this case.

Table 8. Minimum information copula between *O* and *M* given *I* and *F*

Condition	Base	Expectation	Parameter Value	*I*
I <=0.5& F <=0.5	(OM, OM^2, OM^3)	(.082,.031,.0124)	(2685.4 , -7892.9 , 783)	405.95
I <=0.5 & F >0.5	(OM, O^5M, O^4M^2)	(0.164,0.006,0.007)	(2046.3 , -27710 , 1263)	41.9
I >0.5 & F <=0.5	(OM^5, OM^3, O^2M^4)	(0.153,0.245,0.152)	(1481 , -206 , -556)	37.9
I >0.5 & F >0.5	(O^5M, OM, OM^3)	(0.282, 0.582, 0.39)	(728.4, -2054.7, 1025.2)	243.4

Table 9. Minimum information copula between *I* and *S* given *F* and *M*

Condition	Base	Expectation	Parameter Value	*I*
F <=0.5 & M <=0.5	(IS, IS2, IS3)	(.0994,.0542,.0345)	(108.9 , -190.2 , 3243.1)	92.3
F <=0.5 & M >0.5	(I2S, IS2, I5S)	(0.202,0.17,0.061)	(32.2 , -5.6 , -16.5)	6
F >0.5 & M <=0.5	(IS2, I2S3, I2S4)	(0.218,0.082,0.067)	70.9 , -72.9 , 26.7)	4.7
F >0.5 & M >0.5	(I4S2, I3S, I2S)	(0.233, 0.344, 0.42)	(7.5, 2.8,-0.4)	79.9

Table 10. Minimum information copula between O and S given I,F and M

Condition	Base	Expectation	Parameter Value	*I*
I <=0.5&F <=0.5& M <=0.5	(OS, OS2, OS3)	(.0976,.0503,.03)	(111.7 , -173.6 , 80.7)	81.1
I <=0.5 & F <=0.5& M >0.5	(O5S, OS5, OS4)	(0.005,0.002,0.001)	(229.5 , -476.5 , -640)	3.7
I <=0.5 & F >0.5& M <=0.5	(OS, O4S, O4S2)	(0.207,0.014,0.008)	(44.7, -211.8 , 16.6)	1.9
I <=0.5 & F >0.5& M >0.5	(OS5, OS, O4S2)	(0.001, 0.12, 0.003)	(22.9, -985.7, 253.5)	0.1
I >0.5 & F <=0.5& M <=0.5	(O4S, O3S, O2S)	(0.131, 0.203, 0.321)	(17.5, 23.8,- 36.4)	0.2
I >0.5 & F <=0.5& M >0.5	(O3S, OS, O4S2)	(0.194, 0.36, 0.095)	(11.5, -8.5,- 1.1)	0.72
I >0.5 & F >0.5& M <=0.5	(O2S4, O2S2, O2S3)	(0.136, 0.185, 0.158)	(722.7, -211.9,- 633)	0.84
I >0.5 & F >0.5& M >0.5	(O3S3, OS4, OS)	(0.226, 0.266, 0.519)	(11.8, -8.4,-3.14)	77.1

The conditionally *MI* copula in the fourth tree, T_4, can be obtained. In this situation, first divide each of the conditioning variables' supports into 2 bins as in T_2 and T_3, then the *MI* copulas for $O|(I,F,M)$ and $S|(I,F,M)$ are calculated on each combination of bins for I,F,M. Thus in T_4 there are 8 bins altogether. The bins, bases and log-likelihoods associated with each copula are given in Table 10.

Comparison to the other approaches

Table 11. Comparison to the other approaches

Type of Copula	**Variables (X,Y)**	**Parameters**	***l***
Gaussian copula	(O,I)-(I,F)-(F,M)-(M,S) (O|I,F|I)-(I|F,M|F)-(F|M,S|M) (O|I,F,M|I,F) (I|F,M,S|F,M) (O|I,F,M,S|I,F,M)	Gaussian copula are used as building blocks	3721.04
t-copula	(O,I)-(I,F)-(F,M)-(M,S) (O|I,F|I)-(I|F,M|F)-(F|M,S|M) (O|I,F,M|I,F) (I|F,M,S|F,M) (O|I,F,M,S|I,F,M)	t- copula are used as building blocks	3987.1
MI copula	(O,I)-(I,F)-(F,M)-(M,S) (O|I,F|I)-(I|F,M|F)-(F|M,S|M) (O|I,F,M|I,F) (I|F,M,S|F,M) (O|I,F,M,S|I,F,M)	Details are provided in this article	4845.12

As mentioned, multivariate copula function are limited and weak to modeling multivariate dependency, the proposed method was compared with two different multivariate copula function. When the multivariate Gaussian copula was fit to this data the Log-likelihood is 3458.7 and by multivariate t-copula is 3468.4. In order to make a comparison the log-likelihood of the data sample was computed for three different copula models used on the same vine structure: The Gaussian copula, the t-copula used by Aas (2009), and our minimum information copula. The results are shown in Table 11.

Conclusion

If choices of marginal densities are made for any indexes of Tehran stock exchange and copulas between them then the above formula will give a multivariate density for each proposed level of variables.

Acknowledgements

The authors are grateful to Professor Tim Bedford, Dr. Kevin Wilson and Dr. Kjersti Aas for helpful comments.

References

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

Abbas, A. E., Budescu, D. V., & Gu, Y. H. (2010). Assessing joint distributions with isoprobability contours. *Management Science, 56*, 997–1011.

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

Bedford, T., Daneshkhah, A., & Wilson, K. (2013). Approximate Uncertainty Modeling with Vine copulas. To appear in *European Journal of Operation Research.*

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

Daneshkhah, A., Parham, G., Chatrabgoun, O., & Jokar, M. (2013). Approximation Multivariate Distribution with pair copula Using the Orthonormal Polynomial and Legendre Multiwavelets basis functions. Submitted to the *Journal of Communications in Statistics - Simulation and Computation.*

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom. *Econometrica, 50*(4), 987–1007.

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

Kurowicka, D., & Cooke, R. (2006). *Uncertainty analysis with high dimensional dependence modelling*. New York: John Wiley.

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika, 65*(2), 297–303.

Nelsen, R. B. (1999). *An introduction to copulas*. New York: Springer-Verlag, Inc.

Sklar, A. (1959). Fonctions de repartition a n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Universite de Paris, 8*, 229–231.