Approximation Multivariate Distribution with Pair Copula Using the Orthonormal Polynomial and Legendre Multiwavelets Basis Functions

A. DANESHKHAH,1 G. PARHAM,2 O. CHATRABGOUN,3 AND M. JOKAR4

1Warwick Centre for Predictive Modelling, School of Engineering, University of Warwick, Coventry, UK
2Department of Statistics, Faculty of Mathematical Sciences and Computer, Shahid Chamran University, Ahvaz, Iran
3Department of Statistics, Faculty of Mathematical Sciences and Statistics, Malayar University, Malayar, Iran
4Faculty of Sciences, Department of Mathematics, Lorestan University, Lorestan, Iran

We concentrate on constructing higher dimensional distributions using a fast growing graphical model called Vine/ pair-copula model which has been introduced and developed by Joe, Cooke, Bedford, Kurowica, Daneshkhah, and others. They first construct a \( n \)-dimensional copula density by stacking together \( \frac{n(n-1)}{2} \) bivariate copula density, and they then approximate arbitrarily well these bivariate copulas and the corresponding multivariate distribution using a semi-parametric method. One constructive approach involves the use of minimum information copulas that can be specified to any required degree of precision based on the available data (or possibly based on the experts' judgments). By using this method, one is able to use a fixed finite dimensional family of copulas to be employed in terms of a vine construction, with the promise of a uniform level of approximation.

The basic idea behind this method is to use a two-dimensional ordinary polynomial series to approximate any log-density of a bivariate copula function by truncating the series at an appropriate point. We make this approximation method more accurate and computationally faster by using the orthonormal polynomial and Legendre multiwavelets (LMW) series as the basis functions. We show the derived approximations are more precise and computationally faster with better properties than the one proposed previous method in the literature. We then apply our method to modeling a dataset of Norwegian financial data that was previously analyzed in the series of articles, and finally compare our results by them. At the end, we present a method to simulate from the approximated models, and validate our approximation using the simulation results to recover the same dependency structure of the original data.

Received December 16, 2012; Accepted May 2, 2013
Address correspondence to A. Daneshkhah, Warwick Centre for Predictive Modelling, School of Engineering, University of Warwick, Coventry CV4 7AL; E-mail: a.daneshkhah@cranfield.ac.uk
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lssp.
Keywords  Density approximation; Legendre multiwavelets; Minimum information technique; Orthonormal polynomial series; Pair-copula construction; Vine.

Mathematics Subject Classification  62G05; 62G07.

1. Introduction

Bedford and Cooke (2001, 2002) introduced a probabilistic construction of multivariate distributions based on the simple graphical model called vine. This model represents an entirely new approach of building complicated multivariate and highly dependent models which can be seen as the classical hierarchical modeling. The principle behind the vine construction is to model dependency using simple local building blocs based on conditional independence (e.g., cliques in random fields). Aas et al (2009) called these building blocs, pair-copulas. They used the pair-copula decomposition of a general multivariate distribution and proposed a method to perform inference.

They investigated modeling complicated high-dimensional data by fitting different parametric bivariate copulas to construct the corresponding pair-copula model. However, there is a huge number of parametric bivariate copulas, but it is well known that building higher-dimensional copulas is generally a difficult problem, and choosing a parametric family for the given higher-dimensional copula is rather more difficult and limited (see Embrechts et al., 2003). As a result, the problem of choosing a parametric copula for a higher-dimensional copula is reduced to fitting a parametric bivariate copulas to data. Bedford et al. (2013a) stated that the use of a parametric copula to model dependency is simply a translation of one difficult problem into another: instead of the difficulty of specifying the full joint distribution we have the difficulty of specifying the copula. The only and main advantage is the technical one that copulas are normalized to have support on the unit square and uniform marginal distributions. Therefore, the potential flexibility of the copula, by restricting them to a particular parametric class (e.g., Gaussian, multivariate t-student, etc) is not realized in practice.

To overcome this difficulty, Bedford et al. (2013a) proposed an alternative approach at which a vine structure can be used to approximate any given multivariate copula to any required degree of approximation. This method can be easily implemented in practice. It is only required to assume that the multivariate copula density of interest must be continuous and non-zero.

This method is constructive and involves the use of minimum information copulas that can be determined to any required degree of precision based on the available data (or expert judgments). It can be shown that good approximation "locally" guarantees good approximation globally. It can be shown that a vine structure imposes no restrictions on the underlying joint probability distribution it represents (Bedford et al., 2013,). Furthermore, Kurowicka and Joe (2011) reported that this is essential to address this question that which vine structure is most appropriate where some structures allow the use of less complex conditional copulas than others. Conversely, if we only allow certain families of copulas then one vine structure might fit better than another. This question is still open and under study, and is beyond the scope of this article.

Thus, it is trivial to show that if there is any difficulty to fit a multivariate distribution by a pair-copulas model, then the problem is not related to the vine structure but the copulas/conditional copulas. As a result, the question “does a vine structure fit” only makes sense in the context of a given family of copulas. Therefore, we need to have a class of copulas with which we can approximate any given copula to an arbitrary degree.
A natural way to build a minimum information copula or specifying dependency constraints is through the use of moments (Bedford, 2006). These can be specified either on the copula or on the underlying bivariate density. We follow Bedford et al. (2013a) to consider the moment constraints in which real-valued functions \( \phi_1, \ldots, \phi_k \) are required to take expected values \( e_1, \ldots, e_k \), respectively. We then fit a minimum information copula that satisfies a set of constraints as above and which has minimum information (with respect to the uniform copula \( c(u, v) = uv \)) amongst the class of all copulas satisfying those constraints. It is trivial to show that this copula is the “most independent” bivariate density that satisfies these constraints. In addition, a specification of minimum information bivariate copulas naturally leads us to the minimum information vine distributions. Particularly, it can be shown that if a minimal information copula satisfied each of the (local) constraints (on moments, rank correlation, etc.), then the resulting joint distribution would also be minimally informative given those constraints (see Kurowicka and Cooke, 2006).

In order to calculate the minimum information copula associated with the constraints mentioned above, an iterative numerical method called \( D_1A D_2 \) algorithm is used by Bedford and Meeuwissen (1997). The number and type of the real-valued functions \( \phi_1, \ldots, \phi_k \) can control the accuracy of the approximation approach and the cost of computation. Bedford et al. (2013a) developed this method by using the ordinary polynomial bases to approximate a multivariate distribution of interest.

We improve this density approximation by using more flexible bases including orthonormal polynomial (hereafter OP) series and Legendre multiwavelets (hereafter LMW), and examine their properties and possible applications. We show that the accuracy of approximation will be increased and the computation cost will be considerably decreased using the OP basis functions. Furthermore, we will show that OP bases are more convenient than the other natural bases (e.g., polynomial series) for the purpose of computation.

The approximation of the copula density using minimum information copula can be improved even further using the wavelet bases which have been recently used for density estimation. The wavelets have become popular due to their ability to approximate a large class of functions, including those with localized, abrupt variations. However, a well-known attribute of wavelet bases is that they cannot be simultaneously symmetric, orthogonal, and compactly supported. Thus, a more general and vector valued construction of wavelets called Multiwavelets would overcome this drawback, and make them natural choices for estimating density functions, many of which exhibit local symmetries around features such as mode. In particular, using LMWs as basis functions will improve accuracy of the approximation incredibly and the computation cost will be also considerably decreased in comparison of the OP bases. We show the efficiency of this method using the introduced bases by comparing them with the model developed by Bedford et al. (2013a) and the one proposed by Aas et al. (2009) for modeling the Norwegian financial data.

The article is organized as follows. In Section 2, we introduce the pair-copula decomposition associated with the multivariate distribution of interest. In Section 3, we first briefly study the minimum information copula and how to use it to approximate a bivariate copula density. We then develop it to approximate the multivariate distribution of interest. We develop this method further using OP series (obtained based on Graham-Schmidt method) and LMWs as the basis functions in Section 4. In Section 5, we apply our method using the new bases for modeling Norwegian Financial returns data. We then exhibit its flexibility by comparing it with the other methods. At the end of this section, we demonstrate the validity of our approach using a developed version of the simulation method proposed by Kurowicka and Cooke (2006).
2. Vine Constructions of Multiple Dependence

Kurowicka and Cooke (2006) highlighted the point that however, the copula families, such as the exchangeable multivariate Archimedean copula or the nested Archimedean constructions, constitute a huge improvement, but they are still not rich enough to model all possible mutual dependencies amongst the \( n \) variables. This is also illustrated by Aas et al. (2009) and Bedford et al. (2013a). Therefore, a more flexible structure called \textit{pair-copula construction} or \textit{vine} proposed by them which allows for the free specification of \( n(n-1)/2 \) copulas and is hierarchical in nature. This modeling structure is based on a decomposition of a multivariate density into a cascade of bivariate copulas.

In other words, a vine associated with \( n \) 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 \textit{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 of vine and regular vine can be found in Kurowicka and Cooke (2006). The following theorem expresses a regular vine distribution in terms of its density.

**Theorem 2.1.** Let \( V = (T_1, \ldots, T_{n-1}) \) be a regular vine on \( n \) elements, where \( T_i \) is a connected tree with nodes \( N_i = \{1, \ldots, n\} \) and edges \( E_i \); for \( i = 2, \ldots, n - 1 \), \( T_i \) is a connected tree with nodes \( N_i = E_{i-1} \). 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 copula and copula density be \( C_{jk|D_e} \) and \( c_{jk|D_e} \) respectively. Let the marginal distributions \( F_i \) with densities \( f_i \), \( i = 1, \ldots, n \) be given. Then, the vine-dependent distribution is uniquely determined and has a density given by

\[
f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i) \prod_{j=1}^{n-1} \prod_{e(j, k) \in E_i} c_{jk|D_e}(F_j|D_e, F_k|D_e)
\]

**Proof.** See Bedford and Cooke (2001).
Multivariate Distribution Approximation

\[ \times c_{13}\{F(x_1 \mid x_2), F(x_3 \mid x_2)\} c_{24}\{F(x_2 \mid x_3), F(x_4 \mid x_3)\} \times c_{14}\{F(x_1 \mid x_2, x_3), F(x_4 \mid x_2, x_3)\} \]

(2)

It is trivial to show that if \( f(x_1, \ldots, x_n) \) is absolutely continuous to product \( f_1, \ldots, f_n \), it then can be represented by any vine-dependent distribution. The existence of regular vine distributions in details is discussed in Bedford and Cooke (2002). We illustrate briefly how such a distribution is determined using the regular vine in Fig. 1 as an example. We make use of the expression

\[ f(x_1, \ldots, x_4) = f(x_1) f(x_2 \mid x_1) f(x_3 \mid x_1, x_2) f(x_4 \mid x_1, \ldots, x_3) \]

The marginal distribution of \( X_1 \) is known, so we have \( f_1 \). The marginals of \( X_1 \) and \( X_2 \) are known, and the copula of \( X_1, X_2 \) is also known, so we can get \( f(x_1, x_2) \), and hence \( f(x_2 \mid x_1) \). In order to get \( f(x_3 \mid x_1, x_2) \) we can determine \( f(x_3 \mid x_2) \) in the similar way as \( f(x_2 \mid x_1) \). Next we calculate \( f(x_1 \mid x_2) \) from \( f(x_1, x_2) \). With \( f(x_1 \mid x_2) \), \( f(x_3 \mid x_2) \), and the conditional copula of \( X_1, X_3 \) given \( X_2 \) we can determine the conditional joint distribution \( f(x_1, x_3 \mid x_2) \), and hence the conditional marginal \( f(x_3 \mid x_1, x_2) \). Progressing in this way we obtain \( f(x_4 \mid x_1, \ldots, x_3) \). As a result, we can state the following theorem.

**Theorem 2.2.** Given a distribution with density function \( f(x_1, \ldots, x_n) \) and a vine \( V \) on \( n \) elements, there are copulas \( c_{jk}^{D_i} \) such that (1) is satisfied, that means

\[ f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i) \prod_{j=1}^{n-1} \prod_{e(j,k) \in E_i} c_{jk}^{D_i}(F_j^{D_i}, F_k^{D_i}) \]

**Proof.** It is trivial, one should follow the explanation given above to build a 4-dimensional multivariate distribution to prove this theorem. See also Bedford et al. (2013a) and references therein.

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 copulas 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 copulas 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 copulas, 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 copulas, 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 copulas.
3. Approximating Multivariate Density: A Minimum Information Approach

This section sets out to show that we can use the minimum information techniques originated from Bedford and Meeuwissen (1997) in conjunction with the observed data or expert elicitation of observables, to define a copula that can be used to build the joint distribution of two random variables, and then develop it further for constructing a multivariate distribution using a Vine model. The method that will be described below is based on using the $D_1 A D_2$ algorithm to determine the copula in terms of potentially asymmetric information about two variables of interests.

3.1. The $D_1 A D_2$ Algorithm and Minimum Information Copula

Bedford and Meeuwissen (1997) applied a so-called $DAD$ algorithm to produce discretized minimally informative copula between two variables with given rank correlation. This approach relies on the fact that the correlation is determined by the mean of the symmetric function $U V$. The same approach can be used whenever we wish to specify the expectation of any symmetric function of $U$ and $V$ (see Bedford, 2006; Lewandowski, 2008).

This method can be developed further using the idea stated in Borwein et al. (1994) which enables us to have asymmetric specifications. In the revised method, we first determine a positive square matrix $A$, also called a kernel, and two diagonal matrices $D_1$ and $D_2$ should be then found in such a way that the following product, $D_1 A D_2$ is doubly stochastic. The theory can be easily generalized for continuous functions (see Bedford et al., 2013a).

Now, suppose there are two random variables $X$ and $Y$, with cumulative distribution functions $F_X$ and $F_Y$, respectively. These are the variables of interest that we would like to correlate by introducing constraints based on some knowledge about functions of these variables. Suppose there are $k$ of these functions, namely $h_1'(X, Y), h_2'(X, Y), \ldots, h_k'(X, Y)$, and that we wish either to calculate their mean values in terms of the observed data, or the expert wishes to specify mean values $\alpha_1, \ldots, \alpha_k$ for all these functions, respectively.

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

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

where $u$ denote the realization of $U$ and $v$ the realization of $V$.

For practical implementations, we use the same method as proposed by Bedford et al. (2013a) to discretize the set of $(u, v)$ values such that the whole domain of the copula is covered. Thus, the aforementioned kernel $A$ becomes a 2-dimensional matrix, and two matrices $D_1$ and $D_2$ should be then determined. As a result, the following product denoted by $P$ over $[0, 1]^2$ becomes a doubly stochastic matrix which represents a discretized copula density.

$$P = D_1 A D_2$$ (4)
The $D_1 A D_2$ algorithm can be used to generate a unique joint density with uniform marginals for each vector $(\lambda_1, \ldots, \lambda_k)$. The set of all possible expectation vectors $(\alpha_1, \ldots, \alpha_k)$ that could be taken by $(h_1, h_2, \ldots, h_k)$ under some probability distribution is convex, and that for every $(\alpha_1, \ldots, \alpha_k)$ in the interior of that convex set there is a density with parameters $(\lambda_1, \ldots, \lambda_k)$ for which $(h_1, h_2, \ldots, h_k)$ take these values (see Borwein et al., 1994; and Bedford et al., 2013a).

We now explain the iterative algorithm required to approximate the mentioned copula density by this algorithm. Suppose that both $(u, v)$ are discretized into $n$ points, respectively as $u_i$ and $v_j$, $i, j = 1, \ldots, n$. Then, we write $A = (a_{ij})$, $D_1 = \text{diag}(d_1^{(1)}, \ldots, d_n^{(1)})$, $D_2 = \text{diag}(d_1^{(2)}, \ldots, d_n^{(2)})$, where $a_{ij} = A(u_i, v_j)$, $d_1^{(1)} = D_1(u_i)$, $d_2^{(2)} = D_2(v_j)$. We define the doubly stochastic matrix, $D_1 A D_2$ with the uniform marginals as follows

$$\forall i = 1, \ldots, n \sum_j d_1^{(1)} d_2^{(2)} a_{ij} = 1/n, \text{ and}$$

$$\forall j = 1, \ldots, n \sum_i d_1^{(1)} d_2^{(2)} a_{ij} = 1/n,$$

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

$$d_1^{(1)} \mapsto \frac{1}{n \sum_j d_2^{(2)} a_{ij}} (i = 1, \ldots, n), \quad d_2^{(2)} \mapsto \frac{1}{n \sum_i d_1^{(1)} a_{ij}}, \quad (j = 1, \ldots, n)$$

It can be shown that this iteration scheme converges geometrically to the requested vectors (see Borwein et al., 1994).

Note that to compare different discretizations (for different $n$) we should multiply each cell weight $d_1^{(1)} d_2^{(2)} a_{ij}$ by $n^2$ as this quantity approximates the continuous copula density with respect to the uniform distributions.

The mapping from the set of vectors of $\lambda$’s onto the set of vectors of resulting expectations of functions $(h_1, \ldots, h_k)$ has to be found numerically. Bedford et al. (2013a) proposed the optimization techniques to determine the $\lambda_i$’s and corresponding copula. The expectations $\alpha_i$ of $k$ functions of variables $X$ and $Y$ are given by

$$E[h_i(X, Y)] = E[h_i(U, V)] = \alpha_i, \quad i = 1, \ldots, k.$$
MATLAB’s optimization procedure called FMINSEARCH, which implements the Nelder-Mead simplex method (see Lagarias et al., 1998). The minimized function is then

\[
L_{\text{sum}}(\lambda_1, \ldots, \lambda_k) = \sum_{i=1}^{k} L_i^2(\lambda_1, \ldots, \lambda_k).
\]

We refer the interested reader to Lewandowski (2008) and Bedford et al. (2013a) to show how an expert could specify a copula through defining expected values.

### 3.2. Approximating Multivariate Density by Vine

In this section, we use techniques from approximation theory to show that any \(n\)-dimensional multivariate density which is \(C^2\) (that is, twice differentiable, with continuous second derivatives) can be approximated arbitrarily well pointwise using a finite parameter set of 2-dimensional copulas in a vine construction. The basic idea is that we can use a series expansion, like a two-dimensional Polynomial series, OP series or LMW, to approximate any log-density function by truncating the series at an appropriate point. What is non-trivial, however, about this method, is that the same truncation can be used everywhere in a vine construction and gives overall uniform pointwise approximation. Hence, our method 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. Since the approximations we make of copula densities might not be quite copula densities themselves, we need to transform them to make them copulas.

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

Here, we develop the approximation method used by Bedford et al. (2013a) to approximate any log-density function at the desired level of approximation which is more accurate and exhibits better properties. We first introduce some notations. The basic assumption is that all densities are continuous. We denote \(C(Z)\) as the space of continuous real valued functions on a space \(Z\), where \(Z = [0, 1]^r\) for some \(r\), and the corresponding norm on \(C(Z)\) is given by

\[
||f_{i_1 \ldots i_r}|| = \sup |f_{i_1 \ldots i_r}(x_1, \ldots, x_r)|.
\]

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

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

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

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

It can be shown that the following two spaces are relatively compact (Bedford et al., 2013a, Theorem 3).

\[
\mathcal{M}(f) = \{f_{i_1 \ldots i_r} : 1 \leq i, i_1, \ldots, i_r \leq n, i \neq i_1, \ldots, i_r\}.
\]
Multivariate Distribution Approximation

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

where \( f_{ij|i_1,\ldots,i_r} \) is the conditional density of \( X_i \) given \( X_{i_1}, \ldots, X_{i_r} \), and \( f_{ij|i_1,\ldots,i_r} \) is the conditional density of \( X_i, X_j \) given \( X_{i_1}, \ldots, X_{i_r} \).

It is then straightforward to show that the set \( \mathcal{C}(f) \subset \mathcal{C}([0, 1]^2) \) is relatively compact. In addition, since all the functions in \( \mathcal{C}(f) \) are positive and uniformly bounded away from 0, the set \( \mathcal{LN}(\mathcal{C})(f) = \{ \ln(g) : g \in \mathcal{C}(f) \} \subset \mathcal{C}([0, 1]^2) \) is also relatively compact (see Bedford et al., 2013a, for details and proofs).

As a result, the set \( \mathcal{C}([0, 1]^2) \) can be considered as a vector space, and in this context a base is simply a sequence of functions \( h_1, h_2, \ldots \in \mathcal{C}([0, 1]^2) \) such that any function \( g \in \mathcal{C}([0, 1]^2) \) can be written as \( g = \sum_{i=1}^{\infty} \lambda_i h_i \). In other words, it can be shown that given \( \epsilon > 0 \), there is a \( k \) such that any member of \( \mathcal{LN}(\mathcal{C})(f) \) (or \( \mathcal{C}(f) \)) can be approximated to within error \( \epsilon > 0 \) by a linear combination of \( h_1, h_2, \ldots, h_k \). There are lots of possible bases, for example, the following polynomial series

\[ u, v, uv, u^2, v^2, u^2v, uv^2, \ldots \]

which was mainly used in Bedford et al. (2013a).

In the next section, we will improve this density approximation based on the minimum information techniques considerably using the OP series and LMWs instead the ordinary polynomial series as the basis functions. We also exhibit other nice properties of these approximations.

It should be noticed that the approximated copula density by the method described above might not be a copula density itself. Therefore, the resulting approximating needs to be transformed in such a way to obtain a copula. This can be done by weighting the approximated density. One of the most effective weighting schemes is the \( D_1 A D_2 \) algorithm mentioned in the previous section. If we have a continuous positive real valued function \( A(u, v) \) on \([0, 1]^2\) then there are continuous positive functions \( d_1(u) \) and \( d_2(v) \), such that \( d_1 d_2 A \) is a copula density, that is, it has uniform marginal distributions. This density is called \( C\)-Projection of \( A \) and denoted by \( C(A) \). Bedford et al. (2013a) presented the following lemma at which it allows us to control the error made when approximating a copula by another function.

**Lemma 3.1.** Let \( g \) be a non-negative continuous copula density. Given \( \epsilon > 0 \) there is a \( \delta \) such that if \( ||g - f|| < \delta \) then \( ||g - C(f)|| < \epsilon \).

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

\[ d^{(1)}(u) = \frac{1}{\int d^{(2)}(v)f(u, v)dv} \text{ and } d^{(2)}(v) = \frac{1}{\int d^{(1)}(u)f(u, v)du} \]

Eventually, the term given in (1) can be used to show that good approximation of each conditional copula would result in a good approximation of the multivariate density of interest.
4. Building Approximations Using Minimally Informative Distributions

In this section, we give practical guide to build a minimally—informative vine structure to approximate any multivariate distribution. In the previous section, we present a method proposed by Bedford et al. (2013b) that all conditional copulas can be approximated using linear combinations of basis functions. In this section, we are going to address the issue of how the appropriate parameter values can be chosen. We also introduce a practical and efficient alternative based on using the minimum information criterion that lies very close to the approach described above. In other words, given the basis functions \( \{1, h_1, \ldots, h_k\} : [0, 1]^2 \rightarrow \mathbb{R} \), we seek values \( \lambda_1, \ldots, \lambda_k \) so that \( \exp(\sum_1^k \lambda_i h_i) \) is close to the approximated copula density. This can be done by fitting the moments of \( h_i \) in the minimum information framework. Therefore, if \( E_g[h_i(u, v)] = \alpha_i \), we seek for the minimum information copula density that also has these moments. This copula density can uniquely be determined, using the \( D_1 A D_2 \) algorithm, as follows

\[
d^1(u)d^2(v)\exp\left(\sum_1^k \lambda_i h_i(u, v)\right).
\]

As mentioned above, a multivariate distribution can be modeled by a vine structure where it can be defined as a decomposition of the given multivariate distribution into certain conditional copulas, associated with the conditioned and conditioning sets of the vine. The following algorithm is summarized the steps to approximate the given multivariate distribution associated with a vine structure:

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

One of the main aspect that would effect the aforementioned approximation is the basis family. Here, we examine the impact of two basis families, the orthonormal polynomial series and Legendre multiwavelets on approximating the minimum information copulas and the multivariate distribution associated with the chosen vine structure. We first briefly introduce these two basis functions.

4.1. Constructing Orthonormal Polynomial Base

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

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

The OP base can be more convenient than some natural basis for the purpose of calculation. In fact, if the basis is an OP basis, adding a new item to the expansion does not change
coefficient of the already found shorter expansion (Gui, 2009). But if the basis is not orthonormal, any new item has in general nonzero projection on previous items. It means that the already found coefficients of the expansion would have to be changed. That is one of the reason we use OP basis functions as the basis family, $S(k)$. It is reasonable to consider Gram-Schmidt OP basis which is one of the famous OP basis functions on $[0, 1]$.

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

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

The first few functions are

$$
\varphi_0(u) = 1, \quad \varphi_1(u) = \sqrt{3}(-1 + u), \quad \varphi_2(u) = \sqrt{5}(1 - 6u + 6u^2), \\
\varphi_3(u) = \sqrt{7}(-1 + 12u - 30u^2 + 20u^3), \quad \varphi_4(u) = \sqrt{9}(1 - 20u + 90u^2 - 140u^3 + 70u^4) \\
\varphi_5(u) = \sqrt{11}(-1 + 30u - 210u^2 + 560u^3 - 630u^4 + 252u^5), \ldots
$$

4.2. Constructing Legendre Multiwavelets Base

The use of wavelets for density estimation has recently gained in popularity due to their ability to approximate a large class of functions, including those with localized, abrupt variations. However, a well-known attribute of wavelet bases is that they can not be simultaneously symmetric, orthogonal, and compactly supported. Therefore, a more general, vector-valued, construction of wavelets is proposed by Locke and Peter (2013) to overcome this disadvantage, and making them natural choices for estimating density functions, many of which exhibit local symmetries around features such as a mode. Locke and Peter (2013) introduced the methodology of wavelet density estimation using multiwavelet bases and illustrate several empirical results where multiwavelet estimators outperform their wavelet counterparts at coarser resolution levels.

In this section, we use the multiwavelet bases to approximate the minimum information copula. The main advantage of using these bases over the polynomial bases introduced in the previous subsection is that the wavelets (and in particular, multiwavelets) are better choices where the functions of interest contain discontinuities and sharp spikes. In addition, in order to preserve the orthonormality property among the multiwavelet bases, we use LMW bases.

In order to construct these bases, we need to introduce some notions and definitions which are briefly described in the following subsections.

4.2.1. Multiresolution Analysis. Wavelet theory is based on the idea of multiresolution analysis (MRA). Usually it is assumed that an MRA is generated by one scaling function, and dilates and translates of only one wavelet $\phi \in L^2(\mathbb{R})$ form a stable basis of $L^2(\mathbb{R})$.

We can generate a reference subspace or sample space $V_0$ as $L^2$-closure of the linear span of the integer translation of the following functions $\phi^m \in L^2(\mathbb{R}), m = 0, \ldots, r,$
\[ V_0 = \text{clos}_{L^2} < \phi^m(\cdot - k) : k \in Z >, \quad m = 0, \ldots, r, \]

and consider subspace
\[ V_j = \text{clos}_{L^2} < \phi^m_{j,k} : k \in Z >, \quad j \in Z \quad \text{and} \quad m = 0, \ldots, r, \]

where \( \phi^m_{j,k} = \phi^m(2^j x - k) : j, k \in Z, \quad m = 0, \ldots, r. \)

Now, we are able to present a proper definition of multiresolution analysis as follows.

**Definition 4.1.** Functions \( \phi^m \in L^2(R) \), are said to generate a multiresolution analysis (MRA) if they generate a nested sequence of closed subspaces \( V_j \) that satisfies

\begin{align}
(i) & \quad \ldots \subset V_{-1} \subset V_0 \subset V_1 \subset \ldots \\
(ii) & \quad \text{clos}_{L^2}(\bigcup_{j \in Z} V_j) = L^2(R) \\
(iii) & \quad \bigcap_{j \in Z} V_j = 0 \\
(iv) & \quad \phi^m(x) \in V_j \iff \phi^m(x + 2^{-j}) \in V_j \iff \phi^m(2x) \in V_{j+1} \\
(v) & \quad \{ \phi^m(\cdot - k) \}_{k \in Z} \; \text{form a Riesz basis of } V_0 
\end{align}

If \( \phi^m \) generates an MRA, then \( \phi^m \) are called scaling functions. In case that the different integer translate of \( \phi^m \) are orthogonal (with respect to the standard linear product \( \langle f, g \rangle = \int_{-\infty}^{+\infty} f(x)g(x)dx \) for two functions in \( L^2(R) \), denoted by \( \phi^m(\cdot - k) \perp \phi^m(\cdot - \bar{k}) \) for \( m \neq \bar{m}, k \neq \bar{k} \), the scaling functions are called an orthogonal scaling functions.

As the subspaces \( V_j \) are nested, there exist complementary orthogonal subspaces \( W_j \) such that
\[ V_{j+1} = V_j \bigoplus W_j, \quad j \in Z \]

here and in the following \( \bigoplus \) denotes orthogonal sums.

This yields an orthogonal decomposition of \( L^2(R) \), namely;
\[ L^2(R) = \bigoplus_{j \in Z} W_j, \]

**Definition 4.2.** Functions \( \psi^m \in L^2(R) \) are called wavelets, if they generate the complementary orthogonal subspaces \( W_j \) of a MRA, i.e.,
\[ W_j = \text{clos}_{L^2} < \psi^m_{j,k} : k \in Z >, \quad j \in Z, \quad \text{and} \quad m = 0, \ldots, r, \]

where \( \psi^m_{j,k} = \psi^m(2^j x - k) \), \( j, k \in Z \).

Obviously, \( \psi^m_{j,k} \perp \psi^m_{j',k'} \) for \( j \neq j', \quad m \neq \bar{m}, \quad \text{and} \quad k \neq \bar{k}, \) if \( \langle 2^{j/2}\psi^m_{j,k}, 2^{j'/2}\psi^m_{j',k'} \rangle = \delta_{j,j'}\delta_{k,k'}\delta_{m,m} \), then, \( \psi^m \) are called orthogonal wavelets, where
\[ \delta_{i,k} = \begin{cases} 
1 & \text{for } i = k; \\
0 & \text{for } i \neq k.
\end{cases} \]
Now, we are able to define Legendre scaling functions and its corresponding multi-wavelets according to MRA definition give above.

4.2.2. Construction of Scaling Functions. The LMWs system with multiplicity \( r \) consists of \( r \) scaling functions and \( r \) wavelets. The \( r \)th order Legendre scaling functions are the set of \( r + 1 \) functions \( \phi^0(x), \ldots, \phi^r(x) \) where \( \phi^i(x) \) is a polynomial of \( i \)th order and all \( \phi^i(x) \)'s form orthogonal basis (Shamsi and Razzaghi, 2005), that is, for \( i = 0, 1, \ldots, r \),

\[
\phi^i(x) = \sum_{k=0}^{i} a_{ik} x^k, \quad \text{for } i = 0, 1, \ldots, r \tag{8}
\]

The coefficient \( a_{ik} \) are chosen so that \( a_{ik} \geq 0 \), and

\[
\int_0^1 \phi^i(x)\phi^k(x)dx = \delta_{i,k}, \quad \text{for } i, k = 0, 1, \ldots, r \tag{9}
\]

The scaling functions \( \phi^i(x) \) have symmetry, anti-symmetry properties for odd or even \( i \), respectively. The two-scale relations for Legendre scaling functions of order \( r \), are in the form (Albert et al., 2002);

\[
\phi^i(x) = \sum_{j=0}^{r} p_{i,j} \phi^j(2x) + \sum_{j=0}^{r} p_{i,r+j+1} \phi^j(2x - 1), \quad \text{for } i = 0, 1, \ldots, r \tag{10}
\]

The coefficients \( p_{i,j} \) determined uniquely by substituting equation (8) to (10). Now we would like to mention two remarks on the two scale relations.

1. Since \( \phi^i(x) \) is a \( i \)-th order polynomial, the right hand side of (10) has at most \( i \)th order scaling functions. Therefore, \( p_{i,j} = p_{i,r+j+1} = 0 \) for \( i < j \).
2. The two scale relations for the Legendre scaling function of order \( n \) which is lower than \( r \) is a subset of first \( n \) two-scale relations for \( \phi^i \) for \( i = 0, 1, \ldots, n \) form \( r \)th order two scale relations.

4.2.3. Construction of Wavelets. The two-scale relation for the \( r \)th order LMWs is given in the following form (Albert et al., 2002):

\[
\psi^i(x) = \sum_{j=0}^{r} q_{i,j} \phi^j(2x) + \sum_{j=0}^{r} q_{i,r+j+1} \phi^j(2x - 1), \quad \text{for } i = 0, 1, \ldots, r \tag{11}
\]

The \( 2(r+1)^2 \) unknown coefficients \( \{q_{i,j}\} \) in (11) can be determined in terms of the following \( 2r(r+1) \) vanishing moment conditions (12) and \( 2(r+1) \) orthogononal conditions (13).

Vanishing moments

\[
\int_0^1 \psi^i(x)x^jdx = 0, \quad \text{for } i = 0, 1, \ldots, r; j = 0, 1, \ldots, i + r. \tag{12}
\]

Orthogonality

\[
\int_0^1 \psi^i(x)\psi^j(x)dx = \delta_{i,j}, \quad \text{for } i, j = 0, 1, \ldots, r. \tag{13}
\]
For example, the Legendre scaling functions of order 5 consist of six functions as follows:

\[
\phi^0(x) = 1 \quad \text{for } 0 \leq x \leq 1 \\
\phi^1(x) = \sqrt{3}(-1 + 2x) \quad \text{for } 0 \leq x \leq 1 \\
\phi^2(x) = \sqrt{5}(1 - 6x + 6x^2) \quad \text{for } 0 \leq x \leq 1 \\
\phi^3(x) = \sqrt{7}(-1 + 12x - 30x^2 + 20x^3) \quad \text{for } 0 \leq x \leq 1 \\
\phi^4(x) = \sqrt{9}(1 - 20x + 90x^2 - 140x^3 + 70x^4) \quad \text{for } 0 \leq x \leq 1 \\
\phi^5(x) = \sqrt{11}(-1 + 30x - 210x^2 + 560x^3 - 630x^4 + 252x^5) \quad \text{for } 0 \leq x \leq 1
\]

The closed form solution to the LMWs of order 5, \(\psi^0(x), \psi^1(x)\psi^2(x), \psi^3(x), \psi^4(x)\) and \(\psi^5(x)\) are given below which are determined using the conditions (12) and (13).

\[
\psi^0(x) = \begin{cases} 
3.55 - 146.72x + 1, 419.86x^2 - 5, 300.81x^3 & \text{for } 0 \leq x \leq \frac{1}{2}; \\
+8, 519.15x^4 - 4, 997.9x^5 \\
-502.87 + 4, 122.32x - 13, 346.68x^2 + 21, 203.23x^3 & \text{for } \frac{1}{2} \leq x \leq 1 \\
-16, 470.37x^4 + 4, 997.907x^5
\end{cases}
\]

\[
\psi^1(x) = \begin{cases} 
-3.47 + 181.55x - 2, 188.78x^2 + 10, 023.38x^3 & \text{for } 0 \leq x \leq \frac{1}{2} \\
-19, 433.09x^4 + 13, 500.89x^5 \\
-2, 080.47 + 15, 646.19x - 46, 291.67x^2 & \text{for } \frac{1}{2} \leq x \leq 1 \\
+67, 299.87x^3 - 48, 071.93x^4 + 13, 500.89x^5
\end{cases}
\]

\[
\psi^2(x) = \begin{cases} 
2.81 - 174.03x + 2, 438.52x^2 - 12, 760.78x^3 & \text{for } 0 \leq x \leq \frac{1}{2} \\
+27, 823.96x^4 - 21, 415.36x^5 \\
-4, 084.87 + 29, 360.26x - 83, 053.61x^2 & \text{for } \frac{1}{2} \leq x \leq 1 \\
+1.16 \times 10^5x^3 - 79, 252.82x^4 + 21, 415.36x^5
\end{cases}
\]

\[
\psi^3(x) = \begin{cases} 
1.71 - 121.14x + 1, 911.69x^2 - 11, 113.58x^3 & \text{for } 0 \leq x \leq \frac{1}{2} \\
+26, 588.59x^4 - 22, 203.27x^5 \\
4, 935.99 - 34, 300.49x + 93, 930.24x^2 - 1.27 \times 10^5x^3 & \text{for } \frac{1}{2} \leq x \leq 1 \\
+84, 427.78x^4 - 22, 203.27x^5
\end{cases}
\]

\[
\psi^4(x) = \begin{cases} 
-0.71 + 56.63x - 998.10x^2 + 6, 413.33x^3 & \text{for } 0 \leq x \leq \frac{1}{2} \\
-16, 797.83x^4 + 15, 222.11x^5 \\
3, 895.43 - 26, 219.63x + 69, 675.97x^2 - 91, 443.07x^3 & \text{for } \frac{1}{2} \leq x \leq 1 \\
+59, 312.70x^4 - 15, 222.11x^5
\end{cases}
\]
5. Application: Norwegian Financial Returns

In this section, we apply the approximation method presented in this article using OP and LMW basis families, $S(k)$ (as mentioned in the first step in the algorithm above) to approximate the multivariate distribution associated with the selected vine structure corresponding to the Norwegian financial returns. We then exhibit the potential flexibility of our approach by comparing it with the other methods cited in Bedford et al. (2013a) and Aas et al. (2009).

Example: In this example, we use the same dataset as considered by the aforementioned authors to illustrate the approximation method introduced in this article. The data consists of 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. They are recorded over the period 04 January 1999 to 08 July 2003 at which 1094 data are collected. We denote these four variables $T$, $B$, $M$, and $S$, respectively.

We first shall remove serial correlation in these four time series, that is, the observation of each variable must be independent over time. Hence, the serial correlation in the conditional mean and the conditional variance are modeled by an AR(1) and a GARCH(1,1) model (Bollerslev, 1986), respectively. Thus, the following model for log-return $x_i$ is considered for the $i$th time series

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

$$E[z_{i,t}] = 0 \quad \text{and} \quad Var[z_{i,t}] = 1$$

$$\sigma^2_{i,t} = \alpha_{i,0} + a_i \epsilon^2_{i,t-1} + b_i \sigma^2_{i,t-1}$$

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

The further analysis is performed on the standardized 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 standardized residuals and squared standardized residuals. We can use the modified Q-statistic and the Lagrange multiplier test, respectively, to check this (Aas et al., 2009). For all series, the null hypothesis that there is no autocorrelation left for the both tests cannot be rejected at the 5% level. Since, we are mainly interested in estimating the dependence structure of the risk factor, the standardized residual vectors are converted to the uniform variables using the kernel method before further modeling. We denote the converted time series of $T$, $M$, $B$, and $S$ by $X$, $Y$, $Z$, and $W$, respectively.

Here, we want to generate a vine approximation fitted to this dataset using minimum information distributions. We adopt a similar vine structure used in the previous studies and presented in Fig. 2. It should be noticed that 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 $\tau$ 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 $\tau$ or the tail dependence coefficients between any pair of the variables in the first tree calculated, we should 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. We skip to present the numerical details of these measures, and follow Aas et al. (2009), we use the pair-copula construction given in Fig. 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 judgment to elicit these measures or other relevant measures that are more convenient to express by the expert (see Bedford et al. (2013b) for a relevant work).

We now derive the minimum information copulas in association with some moment constraints between copula variables, $X, Y, Z, W$. We initially construct minimally informative copulas between each set of two adjacent variables in the first tree, $T_1$. It is also essential to decide which bases should be taken and how many discretization points should be used in each case. We start illustrate our procedure for the first copula in the first tree between $T, M$.

We could simply choose basis functions, starting with simple OPs or LMWs bases, and moving to more complex ones, and include them until we are satisfied with our approximation. We first started with the following OP basis functions, constructed using Gram-Schmidt process

\[
\begin{align*}
\varphi_1(x)\varphi_1(y), \varphi_1(x)\varphi_2(y), \varphi_2(x)\varphi_1(y), \varphi_1(x)\varphi_3(y), \varphi_3(x)\varphi_1(y), \\
\varphi_2(x)\varphi_2(y), \varphi_2(x)\varphi_3(y), \varphi_3(x)\varphi_2(y), \varphi_1(x)\varphi_4(y), \varphi_4(x)\varphi_1(y), \\
\varphi_1(x)\varphi_5(y), \varphi_5(x)\varphi_1(y), \varphi_2(x)\varphi_4(y), \varphi_4(x)\varphi_2(y), \varphi_3(x)\varphi_3(y), \ldots
\end{align*}
\]

and then considered the following LMWs basis functions which is constructed based on the presented method given in subsection 5.2

\[
\begin{align*}
\phi^1(x)\psi^0(y), \phi^1(x)\psi^1(y), \phi^2(x)\psi^0(y), \phi^1(x)\psi^2(y), \phi^2(x)\psi^1(y), \\
\phi^3(x)\psi^0(y), \phi^1(x)\phi^1(y), \psi^1(x)\psi^1(y), \psi^0(x)\psi^2(y), \psi^2(x)\psi^0(y), \\
\phi^1(x)\phi^2(y), \phi^2(x)\phi^1(y), \phi^0(x)\psi^3(y), \phi^3(x)\psi^0(y), \phi^4(x)\psi^3(y), \ldots
\end{align*}
\]
We realized that adding the basis functions in this way is not optimal, and propose an alternative method which is in nature similar to a stepwise regression. In this method, at each stage, we propose to assess the log-likelihood of adding each additional basis function. We then include the function which produces the largest increase in the log-likelihood.

Figure 3 shows the changes of log-likelihood in terms of adding OP (*) and LMWs (△) bases. We see that there is no longer a jump in the log-likelihood when adding the sixth basis function (i.e., by adding additional basis function does not make difference and changes in log-likelihood are negligible). The log-likelihood also increases more quickly and reaches its plateau value of 60.66 using 6 OP basis functions, and 63.36 using 6 LMWs bases. Furthermore, Bedford et al. (2013a) show that the use of log-likelihood in this way is not inconsistent with minimum information modeling. However, we examine that our approximation presented in this article using even fewer bases still superior than the one reported in Bedford et al. (2013a), but to be consistent with the algorithm explained above and able to compare our results with Bedford et al. (2013a) with the same number of bases, we use the following six OP bases chosen by using the so-called stepwise method.

$$\varphi_1(T)\varphi_1(M), \varphi_2(T)\varphi_2(M), \varphi_1(T)\varphi_2(M), \varphi_1(T)\varphi_3(M), \varphi_3(T)\varphi_3(M), \varphi_4(T)\varphi_1(M)$$

and the following LMW basis functions are similarly selected

$$\phi_1^1(T)\phi_1^1(M), \phi_2^2(T)\phi_2^2(M), \phi_4^4(T)\phi_5^5(M), \phi_1^4(T)\phi_2^2(M), \phi_3^3(T)\phi_3^3(M), \psi^2_2(T)\phi_4^4(M).$$

**Figure 3.** The log-likelihood of the minimally informative copula between $T$ and $M$, calculated based on OP (*) and LMWs (△) bases.
The minimally informative copula given the experts' assessment

Figure 4. The minimally informative copula between $T$ and $M$ using the OP bases. The corresponding expectations of the selected OP bases using the Norwegian financial returns data are calculated as

$$\alpha_1 = -0.229, \alpha_2 = 0.210, \alpha_3 = 0.080, \alpha_4 = -0.102,$$

$$\alpha_5 = -0.112, \alpha_6 = 0.046,$$

and also for the selected LMW bases are given by

$$\alpha_1 = 0.480, \alpha_2 = 0.229, \alpha_3 = -0.002, \alpha_4 = 0.019,$$

$$\alpha_5 = 0.086, \alpha_6 = 0.019,$$

We now able to construct the minimum information copula density $C_{T,M}$ with respect to the uniform distributions given the corresponding OP and LMW constraints above using the method described in this article. We first need to determine the number of discretization points (grid size). It is trivial to conclude that a larger grid size will provide a better approximation to the continuous copula but at the cost of more computation time. Similarly, the approximation will become more precise if we run the $D_1 A D_2$ algorithm in more iterations. Indeed, this would cost us more computation time. It can be concluded that the number of iterations will depend on the grid size. We consider the approximation errors in the range $1 \times 10^{-1}$ to $1 \times 10^{-24}$. Thus, the larger the number of grid points used, the larger the number of iterations are needed for convergence which is true over all error levels. The grid sizes all follow the same pattern with large increases in the number of iterations needed for improved accuracy initially and smaller increases when the error is smaller. We choose a grid size of $200 \times 200$ throughout of this example.
Based on the information given above regarding the grid size, number of iterations and error size, we can derive the minimum information copula $C_{TM}$ associated with the chosen constraints. The corresponding copulas in terms of the OP and LMWs bases are plotted in Figs. 4 and 5, respectively. We present Lagrange multiplies values (or parameter values) for the approximated copula density, using the OP bases as follows

$$
\lambda_1 = -0.199, 5, \quad \lambda_2 = 0.165, 1, \quad \lambda_3 = 0.091, 2, \quad \lambda_4 = -0.077, 4,
$$

$$
\lambda_5 = -0.077, 2, \quad \lambda_6 = 0.052, 7
$$

and in the similar way these parameter values for the minimum information copula based on the LMWs bases are given by

$$
\lambda_1 = 1.984, 5, \quad \lambda_2 = 1.615, 8, \quad \lambda_3 = 0.002, 3, \quad \lambda_4 = -0.026, 3,
$$

$$
\lambda_5 = -7.416, 7, \quad \lambda_6 = 3.681, 9
$$

One of the main advantages of using OP and LMWs bases over the ordinary polynomial series (considered in Bedford et al., 2013a) is that the $D_1 AD_2$ algorithm converges much faster using these bases. This is because of the following nice property of these two bases that adding a new basis to the kernel defined in (3) and used to construct the minimum information copula, does not change the Lagrange multipliers of the already used in the kernel. This is shown in Table 1 for the OP bases. But, this is not the case when one is applying the ordinary polynomial bases to calculate the minimum information copula. In this situation, we need to run the $D_1 AD_2$ algorithm each time a new basis is added to the already chosen bases, and the parameter values are changing accordingly. Therefore,
Lagrange multipliers associated with the new added OP basis and their corresponding log-likelihood

| Base                                      | Parameter values                          | Log-likelihood |
|-------------------------------------------|-------------------------------------------|----------------|
| $\varphi_1(T)\varphi_1(M)$               | $-0.199,5$                                | 29.36          |
| Previous one, $\varphi_2(T)\varphi_2(M)$ | $-0.199,5, 0.165,1$                      | 49.2           |
| Previous one, $\varphi_1(T)\varphi_2(M)$ | $-0.199,5, 0.165,1, 0.091,2$             | 52.8           |
| Previous basis, $\varphi_1(T)\varphi_3(M)$ | $-0.199,5, 0.165,1, 0.091,2,$             | 56.16          |
|                                          | $-0.077,4$                                |                |
| Previous basis, $\varphi_3(T)\varphi_3(M)$ | $-0.199,5, 0.165,1, 0.091,2,$             | 59.04          |
|                                          | $-0.077,4, -0.077,2$                     |                |
| Previous basis, $\varphi_4(T)\varphi_1(M)$ | $-0.199,5, 0.165,1, 0.091,2,$             | 60.66          |
|                                          | $-0.077,4, -0.077,2,$                    |                |
|                                          | $0.052,7$                                 |                |

more iterations are required for the $D_1AD_2$ algorithm to converge. The optimization time required for the $D_1AD_2$ algorithm using the OP bases is 35.864 seconds and for the LMWs bases is 29.359, while this time for the ordinary polynomial bases is 72.93 seconds which is almost twofold of the former one and almost two and half times more than the latter one.

Furthermore, by comparing log-likelihoods of the minimum information copulas associated with the ordinary polynomial, OP and LMWs bases reported in Table 2, it can be concluded that the latter one produce more reliable copula density approximation in the sense that its log-likelihood is the largest.

It should be noticed that log-likelihood of the approximated copula using only five bases of OP or LMWs is still larger than the fitted copula based on the six ordinary polynomial bases. In addition, the approximated copulas in term of the proposed bases in this article unlike the one represented in Bedford et al. (2013a) are not sensitive to the initial values chosen for the parameter values (Lagrange multipliers) in the $D_1AD_2$ algorithm.

The second copula in the first tree ($T_1$) is $C_{MB}$. Using the stepwise method, we select the six OP and LMWs bases that along with their corresponding constraints and resulting Lagrange multipliers are given in Table 5 and Table 6, respectively. The approximated minimally informative copula in terms of the OP bases is shown in Fig. 6. Note that the minimum information copula associated with the LMWs bases is very similar to the one given Fig. 6, but to some extent is slightly smoother.

The third marginal copula is between $B$ and $S$. Similarly, the six bases are selected using the stepwise procedure, and the corresponding constraints and resulting Lagrange

| Type of Bases                                      | Number of bases | Log-likelihood |
|---------------------------------------------------|-----------------|----------------|
| Ordinary Polynomial (Bedford et al., 2013a)       | 6               | 58.125.6       |
| Orthonormal polynomial (Subsection 5.1)           | 6               | 60.66          |
| Lagrange multiwavelets (Subsection 5.2)           | 6               | 63.36          |
Table 3
The minimally informative copula for OP bases between M and B

| Base                | $\alpha_i$ | $\lambda_i$ | Log-likelihood |
|---------------------|------------|-------------|----------------|
| $\varphi_1(M)\varphi_1(B)$ | 0.480,3    | 0.570,1     |                |
| $\varphi_2(M)\varphi_2(B)$ | 0.229,8    | 0.084,7     |                |
| $\varphi_1(M)\varphi_3(B)$ | 0.084,1    | 0.043,3     | 158.001,3      |
| $\varphi_2(M)\varphi_4(B)$ | 0.098,9    | 0.100,0     |                |
| $\varphi_4(M)\varphi_1(B)$ | 0.075,7    | 0.083,0     |                |
| $\varphi_1(M)\varphi_5(B)$ | −0.011,2   | −0.053,1    |                |

multipliers are given in Table 5 and Table 6 for orthonormal and LMWs, respectively. The approximated minimally informative copula in terms of the OP bases is shown in Fig. 7.

The conditional copulas in the second tree, $T_2$ can similarly be approximated using the minimum information approach. We only illustrate construction of the conditional minimum informative copula between $T|M$ and $B|M$, and the other conditional copulas in this tree can be similarly approximated. In order to calculate this copula, we divide the support of $M$ into some arbitrary sub-intervals or bins and then construct the conditional copula within each bin. To do so, we select bases in the same way as for the marginal copulas and fit the copula to the calculated mean values or constraints. Here, we use four bins so that the first copula is for $T, B|M \in (0, 0.25)$. The selected OP bases for this copula are

$$h_1(T, B|M \in (0, 0.25)) = \varphi_2(T)\varphi_1(B), h_2(T, B|M \in (0, 0.25)) = \varphi_3(T)\varphi_1(B)$$

$$h_3(T, B|M \in (0, 0.25)) = \varphi_3(T)\varphi_1(B), h_4(T, B|M \in (0, 0.25)) = \varphi_4(T)\varphi_1(B)$$

$$h_5(T, B|M \in (0, 0.25)) = \varphi_1(T)\varphi_3(B), h_6(T, B|M \in (0, 0.25)) = \varphi_2(T)\varphi_3(B)$$

and the LMWs bases are also given by

$$h_1'(T, B|M \in (0, 0.25)) = \phi_1(T)\phi_1(B), h_2'(T, B|M \in (0, 0.25)) = \phi_2(T)\phi_2(B)$$

$$h_3'(T, B|M \in (0, 0.25)) = \psi_2(T)\phi_2(B), h_4'(T, B|M \in (0, 0.25)) = \psi_1(T)\phi_1(B)$$

Table 4
The minimally informative copula for LMWs bases between M and B

| Base                | $\alpha_i$ | $\lambda_i$ | Log-likelihood |
|---------------------|------------|-------------|----------------|
| $\phi_1(M)\phi_1(B)$ | 0.480,3    | 377.364,2   |                |
| $\phi_2(M)\phi_2(B)$ | 0.229,8    | 193.925,4   |                |
| $\phi_3(M)\phi_3(B)$ | 0.098,9    | 253.235,8   | 159.72         |
| $\phi_4(M)\phi_1(B)$ | 0.075,7    | 281.705,7   |                |
| $\psi_1(M)\phi_1(B)$ | 0.053,1    | −622.023,4  |                |
| $\psi_2(M)\phi_1(B)$ | 0.046,3    | −12.280,2   |                |
Table 5
The minimally informative copula for OP bases between B and S

| Base | \(\alpha_i\) | \(\lambda_i\) | Log-likelihood |
|------|--------------|--------------|----------------|
| \(\varphi_1(B)\varphi_1(S)\) | -0.155,7     | -0.146,7     |                |
| \(\varphi_2(B)\varphi_2(S)\) | 0.101,0       | 0.083,6       |                |
| \(\varphi_3(B)\varphi_1(S)\) | -0.051,0      | -0.042,6      | 20.13          |
| \(\varphi_1(B)\varphi_4(S)\) | -0.037,8      | -0.036,5      |                |
| \(\varphi_2(B)\varphi_1(S)\) | 0.025,3       | 0.025,7       |                |
| \(\varphi_5(B)\varphi_1(S)\) | 0.022,2       | 0.024,0       |                |

Table 6
The minimally informative copula for LMWs bases between B and S

| Base | \(\alpha_i\) | \(\lambda_i\) | Log-likelihood |
|------|--------------|--------------|----------------|
| \(\phi^1(B)\phi^1(S)\) | 0.480,3       | -70.35       |                |
| \(\phi^2(B)\phi^2(S)\) | 0.229,8       | 91.72        |                |
| \(\psi^2(B)\phi^2(S)\) | 0.053,9       | 16.61        | 25.07          |
| \(\psi^1(B)\phi^1(S)\) | 0.053,1       | -22.29       |                |
| \(\psi^5(B)\phi^5(S)\) | 0.001,1       | 2.39         |                |
| \(\psi^4(B)\phi^4(S)\) | -0.009,8      | -3.49        |                |

Figure 6. The minimally informative copula between \(M\) and \(B\) using the OP bases.
The corresponding constraints of OP and LMWs bases to approximate $C_{T,B|M \in (0,0.25)}$

| Bases  | $\alpha_1$  | $\alpha_2$  | $\alpha_3$  | $\alpha_4$  | $\alpha_5$  | $\alpha_6$  |
|--------|--------------|--------------|--------------|--------------|--------------|--------------|
| OP     | $-0.299,5$   | $-0.124,0$   | $-0.163,4$   | $-0.031,7$   | $-0.058,5$   | $-0.063,0$   |
| LMWs   | $0.480,3$    | $0.229,8$    | $0.053,9,$   | $0.053,1$    | $0.001,1$    | $-0.009,8$   |

$h'_5(T,B|M \in (0,0.25)) = \psi^5(T)\phi^4(B)$, $h'_6(T,B|M \in (0,0.25)) = \psi^4(T)\phi^2(B)$

The mean values for OP and LMWs bases which will constrain the minimum information copula are given in Table 7.

We can follow this process again for the remaining bins. Tables 8 and 9 show the mean values or constraints (denoted by $\alpha_i$) and corresponding Lagrange multipliers ($\lambda_i$) required to build the conditional minimum information copula between $T|M$ and $B|M$ for OP and LMWs bases, respectively. The log-likelihood of the approximated copula in each bin is also reported in these tables.

Note that the resulting minimum information copula over all bins for OP bases is 47.54 and for LMWs is 58.41 while this amount for the ordinary polynomial bases is only 29.242 which indicates superiority of the former bases.

We can obtain the conditional minimum informative copula in the third tree, $T_3$, similarly by dividing each of the conditioning variables’ supports into four bins. Then, the minimum information copulas for $T|(B,M)$ and $S|(B,M)$ are calculated on each combination of bins for $M$ and $B$ which makes 16 bins altogether for this tree. The bins,
### Table 8

Minimal informative copula for orthonormal basis between $T$ and $B$ given $M$

| Interval     | Bases                                      | $\alpha_i$     | $\lambda_i$     | Log-likelihood |
|--------------|--------------------------------------------|-----------------|-----------------|----------------|
| $0 < M < 0.25$ | $\varphi_2(T)\varphi_1(B), \varphi_5(T)\varphi_1(B)$ | $-0.2995, -0.1240$ | $-0.3511, -0.135$ | 18.22          |
|              | $\varphi_3(T)\varphi_1(B), \varphi_4(T)\varphi_1(B)$ | $-0.1634, -0.0317$ | $-0.057, 0.0776$  |                |
|              | $\varphi_1(T)\varphi_3(B), \varphi_2(T)\varphi_3(B)$ | $-0.0585, -0.0630$ | $-0.0705, 0.001$  |                |
|              | $\varphi_3(T)\varphi_1(B), \varphi_2(T)\varphi_1(B)$ | $0.1504, 0.0562$  | $0.1902, 0.1051$  |                |
| $0.25 < M < 0.5$ | $\varphi_4(T)\varphi_1(B), \varphi_1(T)\varphi_4(B)$ | $0.1030, 0.0836$  | $0.1363, 0.0944$  | 9.05           |
|              | $\varphi_1(T)\varphi_2(B), \varphi_4(T)\varphi_2(B)$ | $0.0823, -0.0621$ | $0.0804, -0.0094$ |                |
|              | $\varphi_2(T)\varphi_1(B), \varphi_1(T)\varphi_3(B)$ | $0.1184, -0.1080$ | $0.1679, -0.2311$ |                |
| $0.5 < M < 0.75$ | $\varphi_2(T)\varphi_4(B), \varphi_1(T)\varphi_5(B)$ | $0.0956, -0.0815$ | $0.1459, -0.2047$ | 9.74           |
|              | $\varphi_1(T)\varphi_2(B), \varphi_3(T)\varphi_1(B)$ | $-0.0627, 0.0245$ | $-0.1869, 0.1253$ |                |
|              | $\varphi_1(T)\varphi_1(B), \varphi_2(T)\varphi_4(B)$ | $-0.2659, 0.1568$ | $-0.3177, 0.1135$ |                |
| $0.75 < M < 1$ | $\varphi_4(T)\varphi_1(B), \varphi_1(T)\varphi_5(B)$ | $0.1025, -0.0079$ | $0.1290, 0.0526$  | 10.53          |
|              | $\varphi_1(T)\varphi_3(B), \varphi_3(T)\varphi_3(B)$ | $-0.1737, -0.0376$ | $-0.1007, 0.0456$ |                |
| Interval | Bases | $\alpha_i$ | $\lambda_i$ | Log-likelihood |
|---------|-------|------------|------------|---------------|
| $0 < M < 0.25$ | $\phi^2(T)\phi^1(B), \phi^5(T)\phi^1(B)$ | $-0.020, -0.021$ | $-2.53, 2.17$ | |
| $0.25 < M < 0.5$ | $\psi^4(T)\phi^3(B), \phi^4(T)\phi^5(B)$ | $-0.018, -0.002$ | $-0.64, -3.00$ | $22.17$ |
| | $\psi^0(T)\phi^5(B), \psi^5(T)\phi^4(B)$ | $-0.004, 0.001$ | $2.35, 5.23$ | |
| | $\phi^3(T)\phi^1(B), \psi^4(T)\phi^4(B)$ | $0.150, 0.109$ | $0.226, -7.30$ | |
| $0.5 < M < 0.75$ | $\psi^4(T)\phi^3(B), \phi^2(T)\phi^1(B)$ | $0.102, 0.056$ | $9.97, -3.31$ | $12.08$ |
| | $\phi^4(T)\phi^1(B), \phi^5(T)\phi^3(B)$ | $0.103, 0.106$ | $-0.25, 10.46$ | |
| | $\phi^5(T)\phi^1(B), \phi^1(T)\phi^3(B)$ | $0.118, -0.108, 0$ | $-269.93, -439.13$ | |
| $0.75 < M < 1$ | $\phi^4(T)\phi^1(B), \psi^1(T)\phi^4(B)$ | $-0.265, 0.156, 8$ | $247.49, -110.67$ | |
| | $\phi^5(T)\phi^5(B)$ | $0.093, 0.059$ | $373.59, -14.53$ | |
| | $\phi^3(T)\phi^5(B), \psi^4(T)\phi^5(B)$ | $-0.102, 0.096$ | $104.95, -29.99$ | $11.30$ |
| | $\phi^1(T)\phi^1(B), \phi^5(T)\phi^5(B)$ | $0.102, 0.069$ | $-108.01, -222.75$ | $12.86$ |
| | $\phi^3(T)\phi^5(B), \phi^3(T)\phi^5(B)$ | $0.021, 0.063$ | $-175.39, -15.69$ | |
bases, and log-likelihoods associated with each copula based on the OP and LMWs basis are given in Tables 10 and 11, respectively.

The log-likelihood of the overall pair-copula model using the OP (and LMWs) bases, derived by adding the log-likelihoods of the copulas constructed above, is 434.135 (and is 552.25 for LMWs bases). These values are considerably greater than the log-likelihoods

| Interval       | Bases                              | Log-likelihood |
|----------------|------------------------------------|----------------|
| $0 < M < 0.25, 0 < B < 0.25$ | $\phi_1 \phi_2, \phi_3 \phi_1, \phi_4 \phi_2, \phi_5 \phi_1, \phi_7 \phi_2, \phi_8 \phi_1$ | 10.64          |
| $0 < M < 0.25, 0.25 < B < 0.5$ | $\phi_1 \phi_4, \phi_3 \phi_3, \phi_4 \phi_5, \phi_5 \phi_3, \phi_6 \phi_5, \phi_8 \phi_3$ | 9.42           |
| $0 < M < 0.25, 0.5 < B < 0.75$ | $\phi_1 \phi_5, \phi_2 \phi_4, \phi_3 \phi_6, \phi_4 \phi_7, \phi_5 \phi_6, \phi_7 \phi_5$ | 13.67          |
| $0 < M < 0.25, 0.75 < B < 1$   | $\phi_1 \phi_6, \phi_2 \phi_5, \phi_3 \phi_7, \phi_4 \phi_8, \phi_5 \phi_7, \phi_6 \phi_5$ | 10.42          |
| $0.25 < M < 0.5, 0 < B < 0.25$ | $\phi_1 \phi_1, \phi_2 \phi_2, \phi_3 \phi_3, \phi_4 \phi_4, \phi_5 \phi_5, \phi_6 \phi_3$ | 12.99          |
| $0.25 < M < 0.5, 0.25 < B < 0.5$ | $\phi_3 \phi_1, \phi_4 \phi_2, \phi_5 \phi_3, \phi_6 \phi_4, \phi_7 \phi_5, \phi_8 \phi_6$ | 15.67          |
| $0.25 < M < 0.5, 0.5 < B < 0.75$ | $\phi_1 \phi_1, \phi_2 \phi_2, \phi_3 \phi_3, \phi_4 \phi_4, \phi_5 \phi_5, \phi_6 \phi_3$ | 10.56          |
| $0.25 < M < 0.5, 0.75 < B < 1$   | $\phi_1 \phi_1, \phi_2 \phi_2, \phi_3 \phi_3, \phi_4 \phi_4, \phi_5 \phi_5, \phi_6 \phi_3$ | 10.77          |
| $0.5 < M < 0.75, 0 < B < 0.25$  | $\phi_1 \phi_1, \phi_2 \phi_2, \phi_3 \phi_3, \phi_4 \phi_4, \phi_5 \phi_5, \phi_6 \phi_3$ | 9.89           |
| $0.5 < M < 0.75, 0.25 < B < 0.5$ | $\phi_1 \phi_1, \phi_2 \phi_2, \phi_3 \phi_3, \phi_4 \phi_4, \phi_5 \phi_5, \phi_6 \phi_3$ | 10.26          |
| $0.5 < M < 0.75, 0.5 < B < 0.75$ | $\phi_1 \phi_1, \phi_2 \phi_2, \phi_3 \phi_3, \phi_4 \phi_4, \phi_5 \phi_5, \phi_6 \phi_3$ | 14.01          |
| $0.5 < M < 0.75, 0.75 < B < 1$   | $\phi_1 \phi_1, \phi_2 \phi_2, \phi_3 \phi_3, \phi_4 \phi_4, \phi_5 \phi_5, \phi_6 \phi_3$ | 17.97          |
| $0.75 < M < 1, 0 < B < 0.25$   | $\phi_1 \phi_1, \phi_2 \phi_2, \phi_3 \phi_3, \phi_4 \phi_4, \phi_5 \phi_5, \phi_6 \phi_3$ | 11.17          |
| $0.75 < M < 1, 0.25 < B < 0.5$  | $\phi_1 \phi_1, \phi_2 \phi_2, \phi_3 \phi_3, \phi_4 \phi_4, \phi_5 \phi_5, \phi_6 \phi_3$ | 14.31          |
| $0.75 < M < 1, 0.5 < B < 0.75$  | $\phi_1 \phi_1, \phi_2 \phi_2, \phi_3 \phi_3, \phi_4 \phi_4, \phi_5 \phi_5, \phi_6 \phi_3$ | 10.61          |
| $0.75 < M < 1, 0.75 < B < 1$    | $\phi_1 \phi_1, \phi_2 \phi_2, \phi_3 \phi_3, \phi_4 \phi_4, \phi_5 \phi_5, \phi_6 \phi_3$ | 24.39          |
Table 12
Comparison between different models

| Type of copula                                           | Log-likelihood |
|---------------------------------------------------------|----------------|
| Gaussian copula (Aas et al. 2009)                       | 263.505,2      |
| t copula (Aas et al. 2009)                              | 291.801,4      |
| Minimum information copula based on polynomial basis    | 388.859        |
| (Bedford et al., 2013a)                                 |                |
| Minimum information copula based on orthonormal         | 434.135        |
| polynomial                                              |                |
| Minimum information copula based on LMWs                | 552.25         |

of the fitted pair-copula models to the data using the Gaussian copula, t-copula, and the approximated pair-copula model using the ordinary polynomial bases (with log-likelihood equals to 388.859).

Now, we compare our method with the other methods used to approximate the multivariate distribution fitted to the Norwegian financial returns data. In order to make a comparison we use the log-likelihood of the approximated density function by the method presented in this article and other approaches reported in Aas et. (2009) and Bedford et al. (2013a). The log-likelihood of the overall pair-copula model using the OP and LMWs bases are 434.135 and 552.25, respectively. These values are much greater than the ones corresponding to the t-copula examined by Aas et al. (2009) of 291.801 and the minimum information copula based on the ordinary polynomial bases of Bedford et al. (2013a) of 388.859. Note that, if we only use five bases to approximate the multivariate density of interest, the log-likelihoods associated with OP and LMWs bases will be 429.3982 and 446.235, respectively, which are still clearly larger than the model proposed by Bedford et al. (2013a) based on the six ordinary polynomial bases. We have computed the log-likelihood of the data sample for five different copula models used on the same vine structure: The Gaussian copula, the t-copula used by Aas et al. (2009), the minimum information copula using the ordinary polynomial bases and our approximated copulas. We illustrate the corresponding results in Table 12.

5.1. Sampling from a Minimal Information Vine and Validation by Simulation

Simulation from vines is briefly discussed in Bedford and Cooke (2001, 2002), and Kurowicka and Cooke (2006). Aas et al. (2009) present a Simulation method from a pair-copulas decomposed model fitted to the Norwegian financial data. In this section, we adapt the sampling algorithm proposed in Kurowicka and Cooke (2006) to draw a sample from the constructed minimally informative vine proposed in this article, and make comparisons between correlations in the simulated data and in the original data.

Kurowicka and Cooke (2006) proposed a Sampling algorithm from the fitted vine using the cumulative approach. The sampling strategy is as follows:

1. sample \( u_i; i = 1, \ldots, n \) independent uniform on \([0,1]\);
2. \( x_1 = u_1; \)
3. \( x_2 = F_{2|1}^{-1}(u_2 | x_1); \)
Table 13
Correlation coefficients of the original data and the simulated data generated from the approximated multivariate distribution using OP bases

|     | Original M | Original B | Original S | Simulated using OP Bases M | Simulated using OP Bases B | Simulated using OP Bases S |
|-----|------------|------------|------------|----------------------------|----------------------------|----------------------------|
| T   | −0.2292    | 0.3177     | −0.22895   | −0.19941                   | 0.31685                    |
|     | 0.4803     | −0.0551    | 0.48101    | −0.05493                   |
|     | −0.1557    |            | −0.156101  |

4. \( x_3 = F_{3|1,2}^{-1}(u_3 | x_1, x_2); \)
5. \( \ldots = \ldots \)
6. \( x_n = F_{n|1,...,n-1}^{-1}(u_n | x_1, \ldots, x_{n-1}); \)

where \( x_i \) is realization values of \( X_i \).

We can validate our approximation of the fitted multivariate distribution associated with the Norwegian financial returns data by the method introduced in this article. We start our simulation by discretizing the distribution function of each variable into 200 points to numerically approximate the corresponding multivariate distribution. The discretized distribution functions are inverted numerically, and the rank correlations between any two pair of adjacent variables and the conditional correlations are calculated based on 5000 simulations. We then compare them with ones calculated from the original data. The results are summarized in Tables 13 and 14 and show strong consistency. Pseudo code for the simulation is given in Algorithm 1. Note that we use some generic functions in the several steps. The first function is \( Cond(x, f, G) \) which finds the distribution function or density conditional on the value \( x \) from a joint density \( f \) over a grid of points in the unit square defined by vector \( G = (g_1, \ldots, g_k) \). This is also given in Algorithm 3 of Bedford et al. (2013a). The second function that we need to address is \( Bin_m(x, f) \) which selects the correct conditional copula density from \( m \) candidates \( f \) based on which bin \( x \) falls into.

Given these functions, the simulation takes a similar form to that given in Kurowicka and Cooke (2006) and explained above. The joint minimum information copula densities, which are inputs to this algorithm, are evaluated on the discretized grid of points.

Table 14
Correlation coefficients of the original data and the simulated data generated from the approximated multivariate distribution using LMWs bases

|     | Original M | Original B | Original S | Simulated using LMWs Bases M | Simulated using LMWs Bases B | Simulated using LMWs Bases S |
|-----|------------|------------|------------|----------------------------|----------------------------|----------------------------|
| T   | −0.2292    | 0.3177     | −0.22915   | −0.19912                   | 0.31691                    |
|     | 0.4803     | −0.0551    | 0.48101    | −0.05508                   |
|     | −0.1557    |            | −0.15591   |
Algorithm 1 Algorithm to simulate from a 4 dimensional D-vine in $(X_1, X_2, X_3, X_4)$ given uniform marginals and minimum information copulas.

Sample $u_1, \ldots, u_4$ where $u_i \sim U(0, 1)$

$x_1 = u_1$

$F_{2|1} = \text{Cond}(x_1, f_{12}, G)$

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

$f_{1|2} = \text{Cond}(x_2, f_{12}, G)$

$f_{3|2} = \text{Cond}(x_2, f_{23}, G)$

$f'_{1|2} = \text{Bin}_4(x_2, f_{13|2})$

$f_{1|2} = f_{1|2} f_{3|2} f'_{1|2}$

$F_{3|12} = \text{Cond}(x_1, f_{13|2}, G)$

$x_3 = F_{3|12}^{-1}(u_3)$

$f_{4|3} = \text{Cond}(x_3, f_{34}, G)$

$f_{2|3} = \text{Cond}(x_3, f_{23}, G)$

$f'_{24|3} = \text{Bin}_4(x_3, f_{24|3})$

$f_{24|3} = f_{2|3} f_{4|3} f'_{24|3}$

$f_{4|23} = \text{Cond}(x_2, f_{24|3}, G)$

$f_{1|23} = \text{Cond}(x_3, f_{13|2}, G)$

$f'_{14|23} = \text{Bin}_{16}(x_2, x_3, f_{14|23})$

$f_{14|23} = f_{1|23} f_{4|23} f'_{14|23}$

$F_{4|123} = \text{Cond}(x_1, f_{14|23}, G)$

$x_4 = F_{4|123}^{-1}(u_4)$

6. Conclusion

In this article, we extend the novel method originally presented by Bedford et al. (2013a) to approximate a multivariate distribution by any vine structure to any degree of approximation. The main idea to implement this approximation method is to use the minimum information copulas that can be determined to any required degree of precision based on the available data. The approximation method used in this article is quite flexible and easy to implement. In order to approximate a multivariate distribution for the observed data, one only needs to specify a vine structure, a basis family, and the expected values for the certain functions associated with some constraints on each pairwise copula. Bedford et al. (2013a) 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 (unlike the situation with Bayesian networks, where not all structures can be used to model a given distribution), we show 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 copulas, 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 copulas. In addition, the choice of vine structure becomes more significant when we truncate class of copulas 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 approximation level $\epsilon$. This approximation can be made more accurate by adding more bases to achieve the desired level of approximation $\epsilon$ (further details can be found in Bedford et al., 2013a).

Our focus in this article was to introduce more flexible and computationally faster bases family. We concentrate on the OP and LMWs bases. The OP and LMWs bases exhibit an appropriate property, which makes the distribution approximation faster in the sense that adding a new element to their expansion does not change coefficient of the already found shorter expansion which is not the case for the nonorthonormal basis (e.g., the ordinary polynomial bases considered by Bedford et al., 2013a) where any new item has in general nonzero projection on previous items. It means that the already found coefficients of the expansion would have to be changed. In other words, applying these bases is so important from three main aspects: firstly, less computation time is required to approximate the minimum information copula of interest; secondly, the fitted models to the data using the minimum information copulas based on these bases are more reliable in the sense that their log-likelihoods are considerably larger than the log-likelihood of the alternative models proposed in the previous studies as shown above; thirdly, the approximations made in this article are robust in the sense that they are not sensitive to the initial values chosen for the parameter values.

In addition to these properties, our method has another nice property that it can be used to build arbitrarily good approximations to the original distribution. One of the most clear sources of potential error in our approximation is the choice of base where it is convenient to take a low number of functions $h_j$. The terms chosen in both OP and LMWs would generate asymmetric copulas which seems to have great impact in modeling general datasets. The use of large numbers of functions does give more accuracy, at the cost of considerable extra computation at the construction stage but at no extra cost at the sampling stage. Indeed, we can approximate the requested model more precisely using less numbers of the bases proposed in this article and with smaller computation time than the alternative methods. In fact, the generalization made in this article gives natural ways to generate asymmetric copulas, and simple ways to specify nonconstant conditional correlations (or other moments). At moment, we are investigating some alternative methods to the stepwise method used in this article to find the most optimal basis functions in a sense that with smaller number of these bases, we would get the largest log-likelihood.

The method used in this article is very flexible and any functions can be used to construct the minimum information copulas. This method can be used for modeling more complex applications at which basis functions should be computed in computer codes. Due to numerous evaluation of these function to construct the minimum information distribution, the computation and then approximation will be infeasible. One suggestion to ease the computation and reduce the complexity of model is to use the Gaussian process emulators.

Acknowledgments

The authors wish to acknowledge the reviewer for his/her helpful and insightful comments to the manuscript. The authors also wish to acknowledge discussions with Professor Tim Bedford and Dr. Kevin Wilson.

References

Aas, K., Czado, K. C., Frigessi, A., Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance, Mathematics and Economics* 44: 182–198.
Alpert, B., Beylkin, G., Gines, D., Vozovoi, L. (2002). Adaptive solution of partial differential equations in multiwavelet bases. *Journal of Computational Physics* 182: 149–190.

Bedford, T. (2006). Interactive expert assignment of minimally-informative copulae. *Applied Probability Trust* 2: 1–5.

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., Cooke, R. M. (2002). Random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence* 32: 245–268.

Bedford, T., Daneshkhah, A., Wilson, K. (2013a). Approximate uncertainty modeling with vine copulas. To appear in *Risk Analysis*.

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., Wilson, K. J., Daneshkhah, A. (2013b). Uncertainty modelling on coupled models using minimum information methods. Submitted *Reliability Engineering and System Safety*.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31: 307–327.

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

Embrechts, P., Lindskog, F., McNeil, A. J. (2003). Modelling dependence with copulas and applications to risk management. In: *Handbook of Heavy Tailed Distributions in Finance*. Amsterdam: Elsevier/North-Holland, volume 31, pp. 307–327.

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

Gui, W. (2009). *Adaptive Series Estimators for Copula Densities*. PhD thesis, Florida State University College of Arte and Sciences.

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

Kurowicka, D., Cooke, R. (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*. England: John Wiley.

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

Lagarias, J. C., Reeds, J. A., Wright, M. H., Wright, P. E. (1998). Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal of Optimization*, 9(1):112–147.

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

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

Locke, B., Peter, A. (2012). Multiwavelet density estimation, submitted *Computational Statistics and Data Analysis*.

Shamsi, M., Razzaghi, M. (2005). Solution of Hallen’s integral equation using multiwavelets. *Computer Physics Communications* 168: 187–197.