Pair-copula constructions of multivariate copulas

Claudia Czado

1 Introduction

The famous Sklar's theorem (see [54]) allows to build multivariate distributions using a copula and marginal distributions. For the basic theory on copulas see the first chapter ([14]) or the books on copulas by Joe ([32]) and Nelson ([51]). Much emphasis has been put on the bivariate case and in [32] and [51] many examples of bivariate copula families are given. However the class of multivariate copulas utilized so far has been limited. Especially financial applications need flexible multivariate dependence structures in the center of the distribution as well as in tails. For value at risk (for a definition see for example [44]) calculations we need flexibility in the tails. One such measure are the upper and lower tail dependence parameter (for a definition see [15]), which coincide for (reflection) symmetric distributions. For example the Gaussian copula allows for an arbitrary correlation matrix with zero tail dependence, while the the multivariate t-copula has only a single degree of freedom parameter which drives the tail dependence parameter. Both the Gaussian and the t-copula are examples of an elliptical copula (see for example [18] and [20]).

In addition to elliptical copulas attention has focused on multivariate extensions of the Archimedian copulas. In this class we have fully and partially nested Archimedian copulas as discussed in [32], [56] and [52]. Hierarchical Archimedian copulas are considered in [52], while multiplicative Archimedian copulas are proposed in [50] and [43]. However these extensions require additional parameter restrictions and thus result in reduced flexibility for modeling dependence structures.

The first topic of this chapter is to present a general construction method for multivariate copulas using only bivariate copulas, which is called a pair-copula construction (PCC). This includes a simple derivation of PCC models such as D-vines and canonical vines. More general PCC's such as regular vines (see [4], [5] and

Claudia Czado
Department of Mathematics, Technische Universität München, Garching, Germanye-mail: cczado@ma.tum.de

[38]) are introduced and some of their properties are discussed. The second topic is to provide statistical inference methods, when only parametric bivariate copulas are used as building blocks in a PCC model. Here we present three methods one based on stepwise estimation, one on maximum likelihood and one on a Bayesian approach. Applications of these methods in the literature will be given. The next topic involves model selection within a specified PCC model. Application areas will be discussed next. We close with further extensions and open problems.

2 Pair copula constructions of D-vine , canonical and regular vine distributions

We assume that all joint, marginal and conditional distributions are absolutely continuous with corresponding densities. In 1996 Joe ([31]) gave the first pair-copula construction of a multivariate copula. He gave the construction in terms of distribution functions, while Bedford and Cooke (see [4] and [5]) expressed these constructions in terms of densities. They organized these constructions in a graphical way involving a sequence of nested trees, which they called regular vines. They also identified two popular subclasses of PCC models, which they call D-vines and canonical vines. Their results are developed in more detail in the book by Kurowicka and Cooke (see [38]). First we present easy derivations of D-vines and canonical vines before we introduce general regular vines.

2.1 Pair-copula constructions of D-vine and canonical vine distributions

The starting point for constructing multivariate distribution is the well known recursive decomposition of a multivariate density into products of conditional densities. For this let $(X_1,...,X_d)$ be a set of variables with joint distribution F and density f, respectively. Consider the decomposition

$$\begin{aligned} f(x_1,...,x_d) &= f(x_d|x_1,\cdots,x_{d-1})f(x_1,\cdots,x_{d-1}) \\ &= \cdots = \prod_{t=2}^{d} f(x_t|x_1,\cdots,x_{t-1}) \times f(x_1). \end{aligned} \tag{1}$$

Here $F(\cdot|\cdot)$ and later $f(\cdot|\cdot)$ denote conditional cdf's and densities, respectively.

As second ingredient we need Sklar's theorem for dimension $d = 2$ given by

$$f(x_1,x_2) = c_{12}(F_1(x_1),F_2(x_2)) \cdot f_1(x_1) \cdot f_2(x_2), \tag{2}$$

where $c_{12}(\cdot,\cdot)$ is an arbitrary bivariate copula density. Using (2) we can express the conditional density of X_1 given X_2 as

$$f(x_1|x_2) = c_{12}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1). \tag{3}$$

For distinct indices $i, j, i_1, \cdots, i_k$ with $i < j$ and $i_1 < \cdots < i_k$ we use the abbreviation

$$c_{i,j|i_1,\cdots,i_k} := c_{i,j|i_1,\cdots,i_k}(F(x_i|x_{i_1}, \cdots, x_{i_k}), F(x_j|x_{i_1}, \cdots, x_{i_k})). \tag{4}$$

Using (3) for the conditional distribution of (X_1, X_t) given $X_2, \cdots X_{t-1}$ we can express $f(x_t|x_1, \cdots, x_{t-1})$ recursively as

$$\begin{aligned} f(x_t|x_1, \cdots, x_{t-1}) &= c_{1,t|2,\cdots,t-1} \times f(x_t|x_2, \cdots, x_{t-1}) \\ &= [\prod_{s=1}^{t-2} c_{s,t|s+1,\cdots,t-1}] \times c_{(t-1),t} \times f_t(x_t) \end{aligned} \tag{5}$$

Using (5) in (1) and $s = i, t = i + j$ it follows that

$$\begin{aligned} f(x_1, \ldots, x_d) &= [\prod_{t=2}^{d} \prod_{s=1}^{t-2} c_{s,t|s+1,\cdots,t-1}] \cdot [\prod_{t=2}^{d} c_{(t-1),t}][\prod_{k=1}^{d} f_k(x_k)] \\ &= [\prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,(i+j)|(i+1)\cdots,(i+j-1)}] \cdot [\prod_{k=1}^{d} f_k(x_k)] \end{aligned} \tag{6}$$

Note that the decomposition (6) of the joint density consists of pair-copula densities $c_{i,j|i_1,\cdots,i_k}(\cdot,\cdot)$ evaluated at conditonal distribution functions $F(x_i|x_{i_1}, \cdots, x_{i_k})$ and $F(x_j|x_{i_1}, \cdots, x_{i_k})$ for specified indices $i, j, i_1, \cdots, i_k$ and marginal densities f_k. This is the reason why we call such a decomposition *pair-copula decomposition*. This class of decompositions was named by Bedford and Cooke a *D-vine distribution*.

A second class of decompositions is possible, when one applies (3) to the conditional distribution of (X_{t-1}, X_t) given $X_1, \cdots, X_{t-2}$ to express $f(x_t|x_1, \cdots, x_{t-1})$ recursively. This yields the following expression

$$f(x_t|x_1, \cdots, x_{t-1}) = c_{t-1,t|1,\cdots,t-2} \times f(x_t|x_1, \cdots, x_{t-2}). \tag{7}$$

Using (7) instead of (5) in (1) and setting $j = t - k, j + i = t$ results in the following decomposition

$$\begin{aligned} f(x_1, ..., x_d) &= f(x_1) \times \left[\prod_{t=2}^{d} \prod_{k=1}^{t-1} c_{t-k,t|1,\cdots,t-k-1} \times f(x_t)\right] \\ &= \left[\prod_{t=2}^{d} \prod_{k=1}^{t-1} c_{t-k,t|1,\cdots,t-k-1}\right] \times [\prod_{k=1}^{d} f_k(x_k)] \\ &= \left[\prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{j,j+i|1,\cdots,j-1}\right] \times [\prod_{k=1}^{d} f_k(x_k)]. \end{aligned} \tag{8}$$

According to Bedford and Cooke this PCC is called a *canonical vine distribution*.

2.2 Regular vines distributions and copulas

Bedford and Cooke in [5] and [4] noticed that they can represent these pair-copula decompositions (6) and (8) graphical with a sequence of nested trees with undirected edges, which they call a vine tree. Edges in the trees denote the indices used for the conditional copula densities. Following [38] we recall for the convenience of the reader the definition of a regular vine. According to Definition 4.4 of [38] a *regular vine tree* on d variables consists of connected trees $T_1, \cdots T_{d-1}$ with nodes N_i and edges E_i for $i = 1, \cdots, d-1$, which satisfy the following

1. T_1 has nodes $N_1 = \{1, \cdots, d\}$ and edges E_1.
2. For $i = 2, \cdots, d-1$ the tree T_i has nodes $N_i = E_{i-1}$.
3. Two edges in tree T_i are joined in tree T_{i+1} if they share a common node in tree T_i

The edges in tree T_i will be denoted by $jk|D$ where $j < k$ and D is the conditioning set. Note that in contrast to [38] we order the conditioned set $\{j,k\}$ to make the order of the arguments in the bivariate copulas unique. If D is the empty set, we denote the edge by jk. The notation of an edge e in T_i will depend on the two edges in T_{i-1}, which have a common node in T_{i-1}. Denote these edges by $a = j(a), k(a)|D(a)$ and $b = j(b), k(b)|D(b)$ with $V(a) := \{j(a), k(a), D(a)\}$ and $V(b) := \{j(b), k(b), D(b)\}$, respectively. The nodes a and b in tree T_i are therefore joined by edge $e = j(e), k(e)|D(e)$, where

$$\begin{aligned}
j(e) &:= \min\{i : i \in (V(a) \cup V(b)) \setminus D(e)\} \\
k(e) &:= \max\{i : i \in (V(a) \cup V(b)) \setminus D(e)\} \\
D(e) &:= V(a) \cap V(b).
\end{aligned}$$

Two special vine tree specifications were identified by Bedford and Cooke, one they called *drawable vine trees* or short *D-vine trees*, while the other one is called *canonical vine trees* or short *C-vine trees*. They are defined as follows

A regular vine tree is called

- *D-vine tree* if each node in $T-1$ has at most 2 edges.
- *C-vine tree* if each tree T_i has a unique node with $d-i$ edges. The node with $d-1$ edges in tree T_1 is called the *root*.

In Figure 1 a graphical representation of a D-vine tree in five dimensions is given, while in Figure 2 we see a representation of a C-vine tree. For example the edge $e = 14|23$ in tree T_3 of Figure 1 is derived from edges $a = 13|2$ with $V(a) = \{1,2,3\}$ and $b = 24|3$ with $V(b) = \{2,3,4\}$. Note that $D(e) = \{2,3\}$, $j(e) = 1$ and $k = 4$.

To build up a statistical model on a regular vine tree with node set $\mathcal{N} := \{N_1, \cdots, N_{d-1}\}$ and edge set $\mathcal{E} := \{E_1, \cdots, E_{d-1}\}$ one associates each edge $e = j(e), k(e)|D(e)$ in E_i with a bivariate copula density $c_{j(e),k(e)|D(e)}$. Let $\mathbf{X}_{D(e)}$ be the sub random vector of $\mathbf{X}$ indicated by the indices contained in $D(e)$. A *vine distribution* is defined as the distribution of the random vector $\mathbf{X} := (X_1, \cdots, X_d)$ with marginal densities $f_k, k = 1, \cdots, d$ and the conditional density of $(X_{j(e)}, X_{k(e)})$ given the variables $\mathbf{X}_{D(e)}$ specified as $c_{j(e),k(e)|D(e)}$ for the regular vine tree with node set $\mathcal{N}$ and edge set $\mathcal{E}$. In Theorem 4.2 of [38] it is proven that the joint density of $\mathbf{X}$ is uniquely determined and given by

$$f(x_1, ..., x_d) = \prod_{r=1}^{d} f(x_r) \times \prod_{i=1}^{d-1} \prod_{e \in E_i} c_{j(e),k(e)|D(e)}(F(x_{j(e)}|\mathbf{x}_{D(e)}), F(x_{k(e)}|\mathbf{x}_{D(e)})), \quad (9)$$

where $\mathbf{x}_{D(e)}$ denotes the subvector of $\mathbf{x}$ indicated by the indices contained in $D(e)$. This is an analogue of the Hammersley-Clifford theorem for Markov random fields (see [8]) to vine distributions.

For the D-vine tree in Figure 1 the corresponding vine distribution has the joint density given by

$$\begin{aligned} f(x_1, \cdots, x_5) = {} & [\prod_{k=1}^{5} f_k(x_k)] \cdot c_{12} \cdot c_{23} \cdot c_{34} \\ & \times c_{45} \cdot c_{13|2} \cdot c_{24|3} \cdot c_{35|4} \cdot c_{14|23} \cdot c_{25|34} \cdot c_{15|234}, \end{aligned} \quad (10)$$

while the corresponding joint density for the C-vine distribution with tree representation (2) is given by

$$\begin{aligned} f(x_1, \cdots, x_5) = {} & [\prod_{k=1}^{5} f_k(x_k)] \cdot c_{12} \cdot c_{13} \cdot c_{14} \\ & \times \cdot c_{15} \cdot c_{23|1} \cdot c_{24|1} \cdot c_{25|1} \cdot c_{34|12} \cdot c_{35|12} \cdot c_{45|123}. \end{aligned}$$

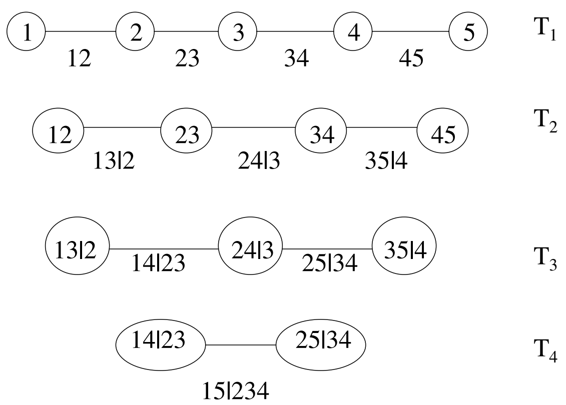

Fig. 1 A D-vine tree representation for $d = 5$.

Here we used the abbreviation defined in (4). Comparing (10) to (6), we see that (10) equals (6) for $d = 5$. Therefore we can identify (6) as the joint density of a D-vine distribution. The same is true for (8), i.e. (8) is the joint density of a C-vine distribution in five dimensions. We can of course use vine distributions to construct copulas, by just requiring that the marginal densities in (9) are univariate uniform densities.

This construction of multivariate distributions and copulas is very general and flexible, since we can use any bivariate copula as building block in the PCC model. In contrast to the extended multivariate Archimedian copulas no restriction to the Archimedian pair-copulas or further parameter restrictions are necessary. In finance the most commonly used pair-copulas are the Gaussian copula, the t-copula, the Clayton copula and Gumbel copula (see for example [1] for definitions and properties). One problem with the multivariate t-copula in financial applications is that we only have a single degree of freedom parameter which drives the tail dependence of all pairs of variables. In [2] it was first noticed that a PCC would overcome this problem and an application to financial stock data was given to demonstrate the superority of a D-vine copula with bivariate t-copulas as building blocks for the PCC over a multivariate t-copula approach. PCC models have been compared to alternative copula based models in [19] and [7] and again the PCC models performed very well among a large class of competitors.

To illustrate the model flexibility we consider a simple D-vine tree in 3 dimensions. Note that in three dimensions D-vines and C-vines coincide. The corresponding D-vine density with standard normal margins is therefore given by

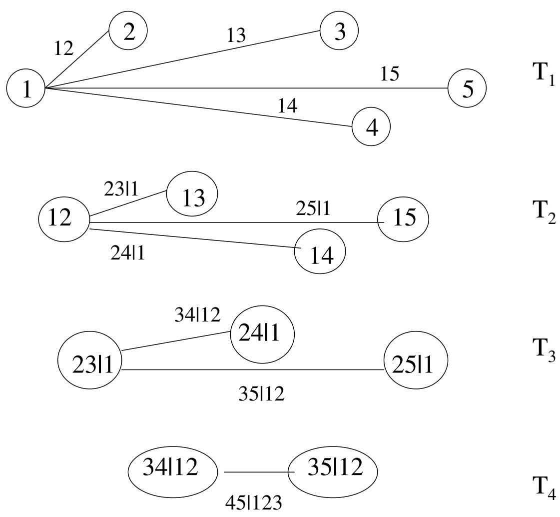

Fig. 2 A C-vine tree representation for $d = 5$.

$$c(x_1,x_2,x_3) = c_{12}(\Phi(x_1),\Phi(x_2)) \times c_{23}(\Phi(x_2),\Phi(x_3)) \times c_{13|2}(F(x_1,|x_2),F(x_3|x_2)) \times \phi(x_1) \times \phi(x_2) \times \phi(x_3), \quad (11)$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ denote the standard normal cdf and pdf, respectively. Here the pair-copulas c_{12}, c_{23} and $c_{13|2}$ will be chosen as either a bivariate Clayton ($C(\theta)$), bivariate Gumbel($G(\alpha)$) or bivariate Frank ($F(\eta)$) copula. The corresponding pair copula parameters are θ, α and η, respectively. We will use for example the abbreviation $DV(C(0.8),G(0.8),C(1))$ to denote the D-vine copula density (11), where c_{12} is $C(0.8)$, c_{23} is $G(0.8)$ and $c_{13|2}$ is $C(1)$. The bivariate marginal density for (X_1,X_2) and (X_2,X_3) are directly specified, while the bivariate marginal density for (X_1,X_3) needs to be computed by integrating (11) over the variable x_2. In Figure 3 density contours of (X_1,X_3) are plotted for four different choices. We see that a large variety of contour shapes are possible.

The tail behavior of vine copulas was investigated in [34].

In general the conditional pair-copula densities in (9) might depend on the conditioning values $x_{i_1},\cdots,x_{i_k}$, however in this paper we assume the restriction that $c_{i,j|i_1,\cdots,i_k}(\cdot,\cdot)$ do not depend on $x_{i_1},\cdots,x_{i_k}$. This means that the decomposition (9) captures the dependency on the conditioning values solely through the arguments

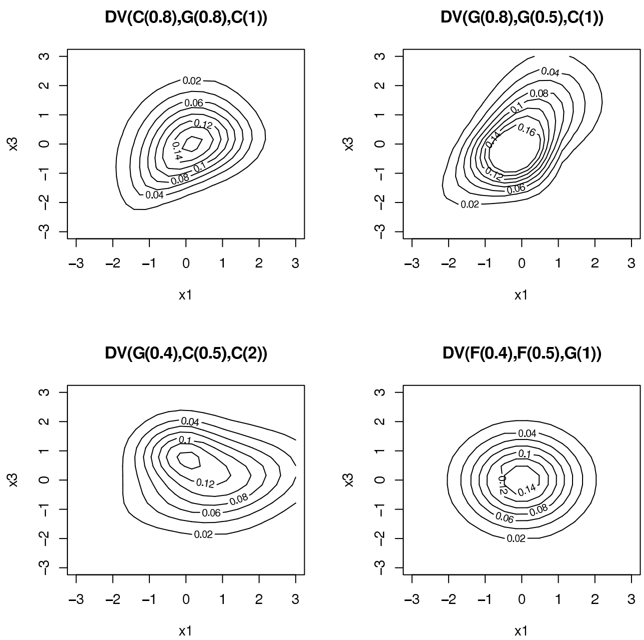

Fig. 3 Density contours of (X_1,X_3) for different three dimensional D-vine distributions with standard normal margins

$F(x_i|x_{i_1},\cdots,x_{i_k})$ and $F(x_j|x_{i_1},\cdots,x_{i_k})$. In a recent paper [24], the authors investigate under which conditions the decomposition of the form (6) for three dimensions satisfies the above restriction as well as what are the effects if this restriction is not satisfied on the value of risk. They claim that this restriction is not so severe. Therefore we consider in the following only decompositions in which conditional pair-copula densities do not depend on the conditioning variables.

For example [13] contains a regular vine density of the form (9) involving foreign exchange rates. It also considers a C-vine model. Finally we want to mention that two well known multivariate copulas can be recovered using vine copulas. The first one is the multivariate Gauss copula and the second one is the multivariate t-copula, which was shown in detail in Section 2 of [13].

3 Estimation methods for regular vine copulas

For estimation of regular vine parameters Kurowicka and Cooke in [38] followed a nonstandard way involving the determinant of the correlation matrix for the random vector distributed according to a regular vine. Using bivariate normal copulas with conditional correlations in the specification of the PCC model results in a multivariate normal distribution. Here one has to use the facts that partial and conditional correlations are equal for elliptical distributions (see [3]) and that conditional distributions of normals are normal with a covariance independent of the conditioning value. Bedford and Cooke in [4] provided a one-to-one relationship between unconditional and partial correlations for Gaussian distributions. Further the determinant of the correlation matrix can be expressed in terms of partial correlations. In the case of Gaussian random vectors the distribution for the determinant of an empirical version of the correlation matrix is known (see Theorem 5.1 of [38]), however bootstrapping would be necessary for other regular vine specification to determine the distribution. Further it is unclear how useful the determinant of the induced correlation matrix is for statistical inference.

Aas et.al. in [2] were the first to consider more standard estimation methods such as stepwise and maximum likelihood estimation (MLE), which we will discuss now. Emphasis here is on the estimation of vine copula parameters, i.e we want to estimate the parameters of the joint density (9), when all marginals are uniform based on an i.i.d. sample from such a density.

Since we have an explicit expression for the joint density the likelihood is easily derived (see [2] for explicit expressions for C- and D-vine copulas). These expressions however involve conditional cdf's, for which we need expressions as well. Joe in [31] showed that for $v \in D$ and $D_{-v} := D \setminus v$

$$F(x_j|\mathbf{x}_D) = \frac{\partial C_{x_j,x_v|D_{-v}}(F(x_j|\mathbf{x}_{D_{-v}}),F(x_v|\mathbf{x}_{D_{-v}}))}{\partial F(x_v|\mathbf{x}_{D_{-v}})}. \tag{12}$$

For the special case where $D = \{v\}$ it follows that

$$F(x_j|x_v) = \frac{\partial C_{x_j,x_v}(F(x_j),F(x_v))}{\partial F(x_v)}.$$

In the case of uniform margins this simplifies further for a parameterized copula cdf $C_{jv}(x_j,x_v) = C_{jv}(x_j,x_v|\theta_{jv})$ to

$$h(x_j|x_v,\theta_{jv}) := \frac{\partial C_{j,v}(x_j,x_v|\theta_{jv})}{\partial x_v}. \tag{13}$$

We can use (12) to express conditional cdf's where D contains more than one element. Following [13] it follows for $v \in D$

$$\begin{aligned}
F(x_j|\mathbf{x}_D) &= \int_{-\infty}^{x_j} c_{jv|D_{-v}}(F(u_j|\mathbf{x}_{D_{-v}}),F(x_v|\mathbf{x}_{D_{-v}}))\, f(u_j|\mathbf{x}_{D_{-v}})\,\mathrm{d}u_j \\
&= \int_{-\infty}^{x_j} \frac{\partial^2 C_{jv|D_{-v}}(F(u_j|\mathbf{x}_{D_{-v}}),F(x_v|\mathbf{x}_{D_{-v}}))}{\partial F(u_j|\mathbf{x}_{D_{-v}})\,\partial F(x_v|\mathbf{x}_{D_{-v}})}\, \frac{\partial F(u_j|\mathbf{x}_{D_{-v}})}{\partial u_j}\,\mathrm{d}u_j \\
&= \frac{1}{\partial F(x_v|\mathbf{x}_{D_{-v}})} \int_{-\infty}^{x_j} \underbrace{\frac{\partial^2 C_{jv|D_{-v}}(F(u_j|\mathbf{x}_{D_{-v}}),F(x_v|\mathbf{x}_{D_{-v}}))}{\partial F(u_j|\mathbf{x}_{D_{-v}})}\, \frac{\partial F(u_j|\mathbf{x}_{D_{-v}})}{\partial u_j}}_{\frac{\partial}{\partial u_j} C_{jv|D_{-v}}(F(u_j|\mathbf{x}_{D_{-v}}),F(x_v|\mathbf{x}_{D_{-v}}))}\,\mathrm{d}u_j \\
&= \frac{\partial}{\partial F(x_v|\mathbf{x}_{D_{-v}})} C_{jv|D_{-v}}(F(x_j|\mathbf{x}_{D_{-v}}),F(x_v|\mathbf{x}_{D_{-v}})) \\
&= \frac{\partial}{\partial \eta} C_{jv|D_{-v}}(F(x_j|\mathbf{x}_{D_{-v}}),\eta)|_{\eta=F(x_v|\mathbf{x}_{D_{-v}})} \\
&= h(F(x_j|\mathbf{x}_{D_{-v}})|F(x_v|\mathbf{x}_{D_{-v}})|\theta_{jv|D_{-v}}).
\end{aligned}$$

This shows that the conditional cdf's with conditioning set D can be build up recursively using the h-function from conditional cdf's with lower dimensional conditioning set. Overall this allows an recursive determination of the likelihood. The inverses of these h-functions can be used to facilitate sampling from D and C vines using a conditional approach (see [2] and [40]).

However the number of parameters grows quadratically in the dimension d, since there $\frac{d\times(d-1)}{2}$ different pair-copulas to be parametrized. Therefore it is useful to consider a stepwise estimation approach, where we estimate the parameters from the first tree to the last one sequentially. In an initial step estimate the parameters corresponding to the pair-copulas in the first tree using any method you prefer. For example the correlation parameter ρ of a bivariate t-copula pair is estimated using Kendall's τ and in second part the degree of freedom parameter ν is maximized using the estimated ρ. For the copula parameters identified in the second tree, one first has to transform the data with the h function required for the appropriate conditional cdf using estimated parameters to determine realizations needed in the second tree.

For example we want to estimate the parameters of copula $c_{13|2}$. First transform the observations $\{u_{1,t}, u_{2,t}, u_{3,t}, t = 1, \cdots, n\}$ to $u_{1|2,t} := h(u_{1,t}|u_{2,t}, \hat{\theta}_{12})$ and $u_{3|2,t} := h(u_{3,t}|u_{2,t}, \hat{\theta}_{23})$, where $\hat{\theta}_{12}$ and $\hat{\theta}_{23}$ are the estimated parameters in the first tree. Now estimate $\theta_{13|2}$ based on $\{u_{1|2,t}, u_{3|2,t}; t = 1, \cdots, n\}$. Continue sequentially with this procedure until all copula parameters of all trees are estimated. Note for trees T_i with $i \geq 2$ recursive applications of the h functions are needed to transform to the appropriate conditional cdf.

This stepwise estimation gives parameter estimates, but so far the asymptotic distribution of the stepwise estimates has not been determined, therefore the use of these parameters as starting values is more appropriate. In contrast MLE's of the pair-copulas are efficient under regularity conditions with asymptotic variance-covariance given by the inverse of the Fisher information matrix. However it is difficult to determine the Fisher information matrix, so one uses in general the observed Hessian matrix instead. Again the Hessian matrix corresponding to a sample from (9) is difficult to express analytically but simple to approximate numerically. It may however happen that this numerical approximation might not yield a positive definite variance-covariance matrix. In this case further numerical manipulations are necessary. This is the reason why in the first paper on ML estimation ([2]) no estimated standard errors where given. In subsequent papers (see [13] and [19]) these have been added. In most papers D- and C-vine copula parameters are estimated, at the moment only [13] considers a regular vine copula.

These difficulties have been noted by Min and Czado (see [47]), which instead propose to follow a Bayesian approach. Here parameters are estimated using Markov Chain Monte Carlo (MCMC) methods (see for example [9]) and they employ the Metropolis Hastings algorithm (see [28] and [46]) for D-vines with pair-copulas to be chosen as a bivariate t-copula. Interval estimates are provided by credible intervals. This approach can also be easily extended to credible intervals for functions of the copula parameters. Examples for such functions are tail dependence coefficients, λ- function of [21] and value at risk. In particular the λ- function can be used to assess model fit. Credible intervals for the tail dependence coefficient for of pairs with bivariate t-copula and the corresponding λ-function are provided in [47].

4 Model selection among vine specifications

The number of different D- and C-vines is very large. In [2] Aas et. al show that for a C-vine decomposition on d nodes there are $d!/2$ distinct C-vine trees and this is also the number of distinct D-vine trees. For regular vine trees the number is even larger (see [49]). This means that we need additional structure to select reasonable vine trees. First it might be reasonable to restrict to C- and D-vine trees. A C-vine tree might be reasonable if a there is a variable which drives all other variables. This might be the case if one considers foreign exchange rates. In all other cases a D-vine tree might be enough to consider.

For the order in the trees corresponding to a D-vine copula Aas et. al in [2] put the strongest bivariate dependencies in the first tree of the D-vine tree specification. Strongest bivariate dependencies within the copula distribution might be measured by Kendall's τ or the tail dependence coefficient λ, which is a function of the chosen bivariate copula.

Another approach is to choose a vine tree distribution with the smallest partial correlation in the last tree. However this requires basically a Gaussian tree distribution, since conditional correlations are only easily estimated for a Gaussian distribution, where partial correlations and conditional correlations are equal and the partial correlations have a one-to-one correspondence to unconditional correlations. This approach for Gaussian D-vine copulas has been described for example in Chapter 5 of [41]. For regular vines this problem has been considered in [36].

Once a vine tree specification is chosen one needs to select the pair-copula terms of the vine distribution. For this Aas et. al. in [2] suggest to follow a stepwise approach. First they consider the pairs of variables involved in the first tree and apply a goodness-of-fit (GOF) test (see [6] and [22] for a review of such tests) for each such pair when the copula family varies and pick the family which gives the best fit for this pair. Now transform the data in the way described for the stepwise estimation and continue with the conditional pairs of the next tree within the vine specification. For K pair-copula families this involves fitting and testing $K \times \frac{d\times(d-1)}{2}$ bivariate models. Mendes et.al. (see [45]) are using this stepwise approach for the analysis of Brazilian financial stocks, while Shirmacher and Shirmacher in [53] use bivariate chi^2 display to select the bivariate pair copula family. While this is a feasible first approach, there are obvious problems with the choice of the pair-copulas in the higher trees, since the transformed data only gives an approximation to the conditional cdf's. This uncertainty is ignored and it increases as one moves up the different trees. In addition the critical values of the GOF tests are difficult to obtain, if the test is applied not directly on an i.i.d sample of the copula, but to rank transformed standardized residuals after applying an appropriate marginal model. In this case bootstrapping might be needed.

If one wants to avoid this stepwise approach, one could attempt to apply GOF tests directly on the full d dimensional sample. However if one allows for for K alternative families of pair copulas, this would involve fitting $K^{\frac{d\times(d-1)}{2}}$ models, which is excessive even when K and d are small. Alternatively we consider now Bayesian approaches, which can traverse large model spaces without having to visit all models.

We start with the following subproblem: Once a vine tree specification is selected one is interested in the possibility of reducing the vine distribution further by identifying (conditional) independencies present in the data. For a vine distribution with a single pair-copula family this means that we want to identify pair-copula terms, which can be replaced by a bivariate independence copula. This task is easiest being accomplished by using reversible jump MCMC (RJMCMC) first discussed in [23]. This approach was followed by Min and Czado in [48] who investigate D-vines with bivariate t-copulas. The RJMCMC algorithm is developed and implemented for arbitrary dimensions. In a second approach suggested by Smith et. al. (see [55]) selection indicators for each pair-copula are introduced to select between the chosen pair-copula family and the independence copula. Again a Bayesian approach is followed and an appropriate MCMC algorithm is developed. In [55] the performance of the method is tested when selection is between independence copulas and either Gauss, Clayton or Gumbel pair copulas. It is evident that both the RJMCMC or the selection indicator approach can be extended to choose between different pair-copula families and this is topic of current research.

5 Applications of vine distributions

One application area of vines is to positive definite matrices and correlation matrices. We have a one-to-one relationship between partial and unconditional correlations. Therefore the values of the partial correlations are unstricted in $[-1,1]$ and can be chosen independently, while still inducing positive definite matrices. This was used in applications to linear algebra (see [37] and [37]). Random generation of correlation matrices are considered in [33] using D-vines and more general using regular vines in [42]. Random distributions of correlation matrices are useful as prior choices in a Bayesian setup. This choice was used in [41].

Another area of applications are in the area of distributions on a directed acyclic graph (DAG) or some times also called Bayesian belief network (BBN). These distributions are specified through conditional independence statements described through the graph. For Gaussian and discrete DAG's see for example [11]. Models for variables in $[0,1]$ which are only characterized through possibly conditional rank (Spearman's) correlations on the arcs of the DAG are called nonparametric BBN's (see [39]). In [26] connections between nonparametric BBN's and a series of D-vine distributions are established. Choose for each rank correlation a copula which realizes all rank correlations in [-1,1] and where a zero rank correlation induces independence. With this copula choice it is shown in [26] that the joint distribution on the BBN is uniquely induced by the rank correlations. In the case of normal BBN's the structure of the graph can be learned by removing arcs as long as the determinant of the rank correlation matrix determined by the partial correlations is close to the empirical rank correlation matrix (see [27]). Note that partial and conditional rank correlations are equal. Mixed continuous and discrete BBN's are discussed in [25].

Pair copula constructions have found their applications in the analysis of financial data. Here one starts with appropriate time series models such as ARMA-GARCH or skewed-t GARCH models. The corresponding standardized residuals of each margin are now considered as an i.i.d sample over time. The dependency across margins is modeled with a a vine distribution. Here one can follow a parametric or nonparametric approach to estimate marginal and copula parameters. The simplest estimation approach is a two step approach. For the first step estimate marginal parameters separately for each margin and form standardized residuals r_{it}. In the parametric approach the distribution of the standardized residuals for each margin is assumed to be known. Let $F_i(\cdot|\hat{\theta}_m)$ denote this distribution, where $\hat{\theta}_i^m$ are the estimated marginal parameters for margin i. Now define the probability integral transforms $u_{it} := F^{-1}(r_i t|\hat{\theta}_i^m)$. For the nonparametric approach one uses the empirical distribution function of $\{r_{it}, t = 1, \cdots, T\}$ instead of $F_i(\cdot|\hat{\theta}_m)$. In both approaches the data $\mathbf{u}_t = (u_{1t}, \cdots, u_{dt}), t = 1, \cdots, T$ is assumed to be an i.i.d sample from a regular vine distribution. In a second step the copula parameters are estimated based on the data $\mathbf{u}_t, t = 1, \cdots, T$ using one of the estimation methods discussed in (3). This two-step estimation procedure using the nonparametric approach has been followed by [2] and [13], while the parametric approach was used by [45] and [30] together with the stepwise procedure for selecting and estimating the copula parameters. In [10] also a parametric two step approach was followed in a regime switching setup using the EM algorithm. A truncated C vine copula was introduced in [29] and a stepwise estimation procedure to estimate the C-vine parameters is followed. The usefulness of this formulation was demonstrated in a portfolio of over 90 stocks, however standard errors are not provided.

Joint estimation approaches of both marginal and copula parameters are rare. First examples are D-vine copulas with pair t copula pairs and AR(1) margins are investigated in [12], while Gaussian D-vines copulas with regression are treated in [41]. In both papers a Bayesian approach was followed.

Finally a first application to a geostatistical continuous Markov mesh model was provided by [35] involving a 4 dimensional D-vine.

6 Summary and open problems

Pair copula constructions provide a powerful tool to construct flexible multivariate distributions which can be used to model complex dependencies. Especially the modeling power of financial statistical models is enormously increased. While there has been progress in developing inference methods methods, more needs to be done in the area of estimation, model selection and adaptation the special data structures.

In the area of estimation, the lack of standard errors for the stepwise estimators of the copula parameters is evident. Here the appropriate asymptotic theory has to be developed and fast computation implementation has to be provided. Another estimation problem is to be solved are fast joint estimators of marginal and copula parameters, since the now common two step estimation procedures are not efficient. Here marginals models as applied in financial statistics need to be considered. For joint estimation the Bayesian approach seems to be the most promising. For financial applications a joint Bayesian inference provides natural tools to assess the variability of value of risk estimates.

The problem of selecting the appropriate vine tree specification and the appropriate pair copula family is a challenging problem area. While in the past the flexibility was limited, the model flexibility is now so large, that one has to consider additional structures for the selection. Restrictions such as provided by truncated canonical vines in [29] are promising but need to be fully explored. Bayesian techniques such as RJMCMC or model indicators need to be studied further in the context of selecting pair copula families. Finally the problem of prividing effective non-nested model selection criteria needs to be considered.

Finally adaptation to special data structures will enhance the applicability of this model construction method. Here we name the necessity to allow for time varying copula parameters. First approaches of such models are given in [29] and [30]. However only stepwise estimates without standard errors are provided, while full ML or Bayesian estimates are not investigated. In insurance applications one can is often faced with multivariate counts such as claim counts or multivariate zero truncated claim severities. A sampling method involving vines for multivariate counts has been developed in [16] and [17]. To model dependencies among multivariate discrete or censored variables using vine distributions is another interesting research area. Finally the development of statistical models on graphs or geostatistical structures and their inference involving the pair-copula construction method is an interesting and challenging area.

Acknowledgements Claudia Czado is supported by the Deutsche Forschungsgemeinschaft. (CZ86 1-3)

References

1. Aas, K.: Modelling the dependence structure of financial assets: A survey of four copulas. Tech. Rep. SAMBA/22/04, Norsk Regnesentral (2004)
2. Aas, K., Czado, C., Frigessi, A., Bakken, H.: Pair-copula constructions of multiple dependence. Insurance, Mathematics and Economics **44**, 182–198 (2009)
3. Baba, K., Sibuya, M.: Equivalence of partial and conditional correlation coefficients. Journal of Japanese Statistical Society **35**, 1–19 (2005)
4. Bedford, T., Cooke, R.M.: Probability density decomposition for conditionally dependent random variables modeled by vines. Annals of Mathematics and Artificial Intelligence **32**, 245–268 (2001)
5. Bedford, T., Cooke, R.M.: Vines - a new graphical model for dependent random variables. Annals of Statistics **30**(4), 1031–1068 (2002)
6. Berg, D.: Copula goodness-of-fit testing: an overview and power comparison. The European Journal of Finance pp. 1466–4364 (2009)
7. Berg, D., Aas, K.: Models for construction of higher-dimensional dependence: A comparison study (2009). Forthcoming in The European Journal of Finance
8. Besag, J.: Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society, Series B **36**, 192–236 (1974)
9. Chib, S.: Markov chain Monte Carlo methods: Computation and inference. In: Handbook of Econometrics, pp. 3569–3649. North Holland (2001)
10. Chollete, L., Heinen, A., Valdesogo, A.: Modeling international financial returns with a multivariate regime switching copula (2008). Preprint
11. Cowell, R., Dawid, A., S., L., D., S.: Probabilistic networks and experts systems. Springer, New York (1999)
12. Czado, C., Gärtner, F., Min, A.: Analysis of australian electricity loads using joint bayesian inference of d-vines with autoregressive margins. Tech. rep., Zentrum Mathematik, Technische Universität München (2009). URL http://www-m4.ma.tum.de/Papers/index.html

13. Czado, C., Min, A., Baumann, T., Dakovic, R.: Pair-copula constructions for modeling exchange rate dependence (2008). Preprint, available under http://www-m4.ma.tum.de/Papers/index.html
14. Durante, F., Sempi, C.: Copula theory: an introduction. In: F. Durante, W. Härdle, P. Jaworki, T. Rychlik (eds.) Workshop on Copula Theory and its Applications, pp. ???–??? Springer, Dortrecht (NL) (2010)
15. Embrechts, P., Lindskog, F., McNeil, A.: Modelling dependence with copulas and applications to risk management. In: S.T.Rachev (ed.) Handbook of Heavy Tailed Distributions in Finance. Elsevier, North-Holland (2003)
16. Erhardt, V., Czado, C.: A method for approximately sampling high-dimensional count variables with prespecified pearson correlation. (2008). Preprint, available under http://www-m4.ma.tum.de/Papers/index.html
17. Erhardt, V., Czado, C.: A method for approximately sampling high-dimensional count variables with prespecified pearson correlation - a comparison to norta (2008). Preprint, available under http://www-m4.ma.tum.de/Papers/index.html
18. Fang, H.B., Fang, K., Kotz, S.: The meta-elliptical distributions with given marginals. Journal of Multivariate Analysis **82**(1), 1–16 (2002)
19. Fischer, M., Köck, C., Schlüter, S., Weigert, F.: An empirical analysis of multivariate copula models. Quantitative Finance p. DOI 10.1080/14697680802595650 (2009)
20. Frahm, G., Junker, M., Szimayer, A.: Elliptical copulas:applicability and limitations. Statist. Probab. Lett. **63**(3), 275–286 (2003)
21. Genest, C., Rivest, L.P.: Statistical inference procedures for bivariate Archimedean copulas. J. Amer. Statist. Assoc. **88**(423), 1034–1043 (1993)
22. Genest, C., Rmillard, B., Beaudoin, D.: Omnibus goodness-of-fit tests for copulas: A review and a power study. Insurance: Mathematics and Economics **44**, 199–213 (2009)
23. Green, P.: Reversible jump markov chain monte carlo computation and Bayesian model determination. Biometrika **82**, 711–732 (1995)
24. Haff, I.H., Aas, K., Frigessi, A.: On the simplified pair-copula construction - simply useful or too simplistic? Tech. rep., Norwegian Computing Center, Oslo (2009)
25. Hanea, A., Kurowicka, D.: Mixed non-parametric continuous and discrete bayesian belief networks. In: Advances in mathematical modeling for reliability. IOS Press, Amsterdam (2008)
26. Hanea, A., Kurowicka, D., Cooke, R.: Hybrid method for quantifying and analyzing bayesian belief networks. Quality and Realiability Engineering **22**(6), 709–729 (2006)
27. Hanea, A., Kurowicka, D., Cooke, R., Ababei, D.: Mining and visualizing ordinal data with non-parametric continuous bbn's. Computational Statistics and Data Analysis p. doi:10.1016/j.csda.2008.09.032 (2008)
28. Hastings, W.: Monte Carlo sampling methods using Markov chains and their applications. Biometrika **57**, 97–109 (1970)
29. Heinen, A., Valdesogo, A.: Asymmetric capm dependence for large dimensions: The canonical vine autoregressive copula model (2008). Preprint
30. Heinen, A., Valdesogo, A.: Dynamic d vines (2009). Preprint
31. Joe, H.: Families of m-variate distributions with given margins and m(m-1)/2 bivariate dependence parameters. In: L. Rüschendorf and B. Schweizer and M. D. Taylor (ed.) Distributions with Fixed Marginals and Related Topics (1996)
32. Joe, H.: Multivariate Models and Dependence Concepts. Chapman & Hall, London (1997)
33. Joe, H.: Generating random correlation matrices based on partial correlations. Journal of Multivariate Analysis **97**, 2177–2189 (2006)
34. Joe, H., Li, H., Nikoloulopoulos, A.: Tail dependence and vine copulas. Tech. rep., Department of Mathematics, Washington State University, Pullman, WA, USA (2008)
35. Kolbjornsen, O., Stien, M.: The d-vine creation of non-gaussian random fields (2008). Preprint
36. Kurowicka, D.: Optimal truncation of vines (2009). Preprint
37. Kurowicka, D., Cooke, R.: A parametrization of positive definite matrices in terms of partial correlation vines. Linear Algebra and its Applications **372**, 225–251 (2003)
38. Kurowicka, D., Cooke, R.: Uncertainty analysis with high dimensional dependence modelling. Wiley, Chichester (2006)

39. Kurowicka, D., Cooke, R.M.: Distribution-free continuous bayesian belief. In: Modern statistical and mathematical methods in reliabiliy. World Scientific (2005)
40. Kurowicka, D., Cooke, R.M.: Sampling algorithms for generating joint uniform distributions using the vine-copula method. Computational Statistics & Data Analysis **51**(6), 2889–2906 (2007)
41. Lanzendörfer, J.N.: Joint estimation of parameters in multivariate normal regression with correlated errors using pair-copula constructions and an application to finance (2009). Diploma Thesis, Center for Mathematical Sciences, Technische Universität München
42. Lewandowski, D., Kurowicka, D., Joe, H.: Generating random correlation matrices based on vines and extended onion method. Journal of Multivariate Analysis **100**, 1989–2001 (2009)
43. Liebscher, E.: Modelling and estimation of multivariate copulas. Tech. rep., Working Paper, University of Applied Sciences Merseburg, Germany (2006)
44. McNeil, A.J., Frey, R., Embrechts, P.: Quantitative Risk Management: Concepts, Techniques and Tools. Princeton University Press (2006)
45. de Melo Mendes, B.V., Semeraro, M.M., Leal, R.C.: Pair-copulas modeling in finance. Tech. rep., IM/COPPEAD, Federal University at Rio de Janeiro, Brazil (2009)
46. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equations of state calculations by fast computing machines. Journal of Chemical Phisics **21**, 1087–1092 (1953)
47. Min, A., Czado, C.: Bayesian inference for multivariate copulas using pair-copula constructions (2008). Preprint, available under http://www-m4.ma.tum.de/Papers/index.html
48. Min, A., Czado, C.: Bayesian model selection for multivariate copulas using pair-copula constructions (2009). Preprint
49. Morales-Napoles, O., Cooke, R.M., Kurowicka, D.: About the number of vines and regular vines on n nodes (2009). Submitted to Discrete Applied Mathematics
50. Morillas, P.: A method to obtain new copulas from a given one. Metrika **61**, 169–184 (2005)
51. Nelson, D.: ARCH models as diffusion approximations. Journal of Econometrics **45**, 7–38. (1990)
52. Savu, C., Trede, M.: Hierarchical Archimedean copulas. In: International Conference on High Frequency Finance, Konstanz, Germany, May (2006)
53. Schirmacher, D., Schirmacher, E.: Multivariate dependence modeling using pair-copulas. Tech. rep., Society of Acturaries: 2008 Enterprise Risk Management Symposium, April 14-16, Chicago (2008). URL http://www.soa.org/library/monographs/other-monographs/2008/april/2008-erm-toc.aspx
54. Sklar, A.: Fonctions dé repartition á n dimensions et leurs marges. Publ. Inst. Stat. Univ. Paris **8**, 229–231 (1959)
55. Smith, M., Min, A., Czado, C., Almeida, C.: Modeling longitudinal data using a pair-copula decomposition of serial dependence (2009). Preprint
56. Whelan, N.: Sampling from Archimedian copulas. Quantitative Finance **4**, 339–352 (2004)