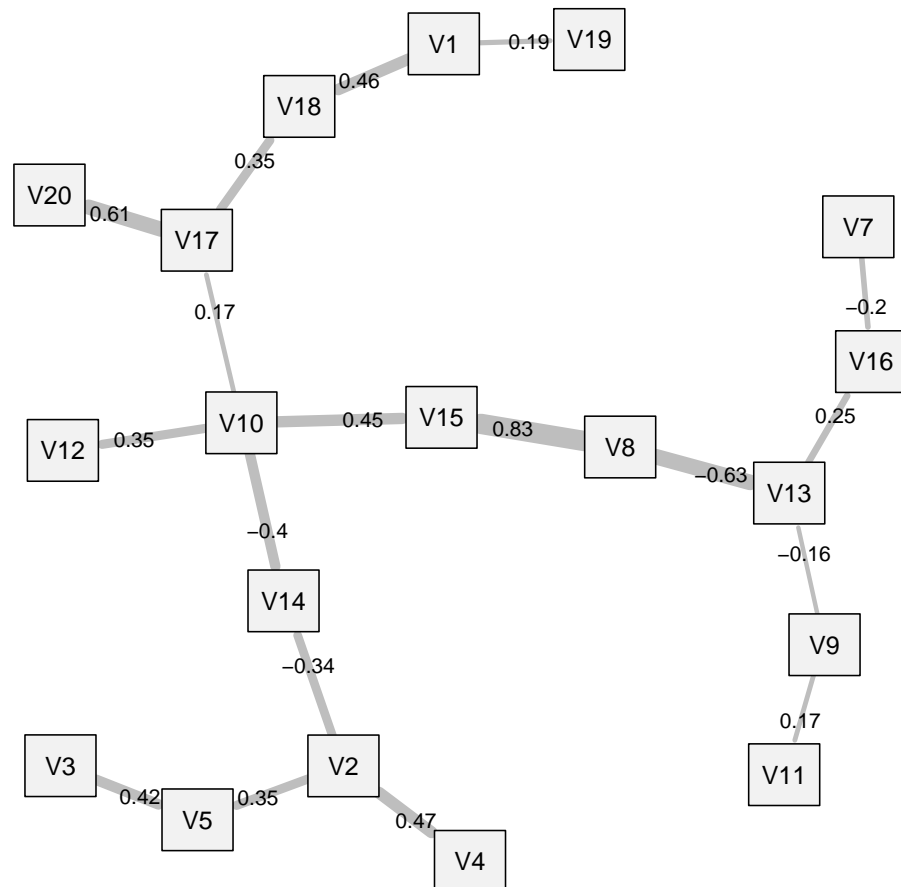


# Truncated regular vines in high dimensions

with application to financial data



**SAMBA/60/10**

**Eike C. Brechmann**

**Claudia Czado**

**Kjersti Aas**

**December 2010**

## Norwegian Computing Center

Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in the areas of information and communication technology and applied statistical modelling. The clients are a broad range of industrial, commercial and public service organizations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. The results of our projects may take the form of reports, software, prototypes, and short courses. A proof of the confidence and appreciation our clients have for us is given by the fact that most of our new contracts are signed with previous customers.

<b>Title</b>	<b>Truncated regular vines in high dimensions</b>
<b>Authors</b>	<b>Eike C. Brechmann , Claudia Czado , Kjersti Aas</b>
Date	December 2010
Publication number	SAMBA/60/10

### **Abstract**

Using only bivariate copulas as building blocks, regular vines constitute a flexible class of high-dimensional dependency models. However, the flexibility comes along with an exponentially increasing complexity in larger dimensions. In order to counteract this problem, we propose using statistical model selection techniques to either truncate or simplify a regular vine. As a special case, we consider the simplification of a canonical vine using a multivariate copula as previously treated by Heinen and Valdesogo (2009) and Robles (2009). We validate the proposed approaches by extensive simulation studies and use them to investigate a 19-dimensional financial data set of Norwegian and international market variables.

Keywords	Multivariate copula, regular vines, truncated vines, simplified vines
Target group	
Availability	
Project	SFI02-Risk
Project number	220302
Research field	Finance, insurance and commodity markets
Number of pages	43
© Copyright	Norwegian Computing Center

# Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
<b>2</b>	<b>Regular vines</b>	<b>7</b>
<b>3</b>	<b>Pairwise simplification and truncation</b>	<b>11</b>
<b>4</b>	<b>Joint simplification and truncation</b>	<b>13</b>
<b>5</b>	<b>Selection of truncation level in the general case</b>	<b>14</b>
<b>6</b>	<b>Selection of simplification level in the general case</b>	<b>18</b>
<b>7</b>	<b>Selection of simplification and truncation levels for C-vines</b>	<b>20</b>
<b>8</b>	<b>Simulation studies</b>	<b>22</b>
8.1	Pairwise simplification and truncation of R-vines	23
8.2	R-vine specification.	24
8.3	C-vine specification (joint simplification and truncation)	27
8.4	Summary	31
<b>9</b>	<b>Applications</b>	<b>32</b>
<b>10</b>	<b>Conclusions</b>	<b>39</b>
	<b>References</b>	<b>41</b>

# 1 Introduction

Recent developments in the area of multivariate modeling tend towards hierarchical, copula-based structures. One of the most promising of these structures is the pair-copula construction (PCC). The PCC was originally proposed by Joe (1996) and has further been explored by (Bedford and Cooke, 2001, 2002) and Kurowicka and Cooke (2006). After being set in an inferential context by Aas et al. (2009), the PCC has been used in various applications, see, e.g., Schirmacher and Schirmacher (2008), Chollete et al. (2009), Heinen and Valdesogo (2009), Berg and Aas (2009), Min and Czado (2009), Min and Czado (2010), and Smith et al. (2010).

Pair-copula constructions are also called regular vines (R-vines). They are hierarchical in nature, the various levels (also called trees) corresponding to the incorporation of more variables in the conditioning sets, using pair-copulas as simple building blocks. Until now, the concentration has been on two special cases of regular vines; drawable vines (D-vines) and canonical vines (C-vines). However, very recently, there has been considerable progress in constructing regular vines even in general, using graph theoretic algorithms (Dißmann, 2010).

The growing interest for pair-copula constructions/regular vines is probably due to their high flexibility. While built entirely from pair-copulas, they can model a wide range of complex dependencies. Nevertheless, these structures have some shortcomings. The most important being that the computational effort required to estimate all parameters grows exponentially with the dimension. For the regular vines to be really useful in practice, one need to be able to fit such structures to data with more than 20 dimensions. Hence, in this paper we treat the problem of determining whether a regular vine can be either *truncated* or *simplified*. By a pairwisely truncated regular vine at level  $K$ , we mean a regular vine where all pair-copulas with conditioning set equal to or larger than  $K$  are replaced by independence copulas. The subject of optimal truncation of vines has previously been treated by Kurowicka (2010), but the approach suggested there is very different from ours. A regular vine is pairwisely simplified at level  $K$  if the same pair-copulas instead are replaced by Gaussian copulas. Gaussian copulas mean a simplification since they are easier to specify than other copulas. Moreover, they are easy to interpret in terms of the correlation parameter. Gaussian copulas are also faster to estimate than, e.g., t-copulas.

To identify the most appropriate truncation/simplification level, we use statistical model selection methods; more specifically, AIC, BIC and the likelihood-ratio based test proposed by Vuong (1989). We first evaluate the performance of the different methods in a simulation study, and then we investigate whether it is

possible to simplify or truncate the R-vine specification corresponding to a 19-dimensional data set consisting of Norwegian and international market variables.

For the special case of a C-vine, the product of all pair-copulas with conditioning set equal to or larger than  $K$  (i.e., the pair-copulas involved in trees higher than  $K$ ) gives a  $(d - K)$ -variate copula, where  $d$  is the total number of variables. Hence, in this case one may in addition to the above-mentioned model selection methods, use copula goodness-of-fit tests to determine the truncation/simplification level. The first kind of methods are hereafter referred to as *pairwise truncation or simplification* and the latter as *joint truncation or simplification*. In our simulation study, we will compare the performance of the two types of approaches. It should be noted that simplification of C-vines previously have been treated by Heinen and Valdesogo (2009) and Robles (2009). There, it is referred to as “truncation”, while using our notation it would be called “simplification” (truncation in our meaning of the word is not explicitly discussed in Robles (2009)).

The rest of this paper is organized as follows. In Section 2 we provide necessary background on R-vines and their likelihood. In Section 3 we introduce the pairwise simplification and truncation of R-vines in general, while Section 4 treats the joint simplification and truncation of the special case of C-vines. The selection of an appropriate truncation and simplification level in the general case are discussed in Sections 5 and 6, respectively, while the same selection for the special case of C-vines is treated in Section 7. The performance of the different selection methods is studied in Section 8, while in Section 9 we apply the methodology in the context of a financial data set. Finally, Section 10 contains some concluding remarks.

## 2 Regular vines

The notion of a regular vine distribution was introduced by Bedford and Cooke (2001) and Bedford and Cooke (2002) and described in more detail in Kurowicka and Cooke (2006). It involves the specification of a sequence of trees where each edge corresponds to a pair-copula. These pair-copulas then constitute the building blocks of the joint vine distribution. According to Definition 4.4 of Kurowicka and Cooke (2006), a *regular vine* (R-vine) on  $d$  variables consists of trees  $T_1, \dots, T_{d-1}$  with nodes  $N_i$  and edges  $E_i$  for  $i = 1, \dots, d - 1$ , which satisfy the following:

1.  $T_1$  has nodes  $N_1 = \{1, \dots, d\}$  and edges  $E_1$ .
2. For  $i = 2, \dots, d - 1$  the tree  $T_i$  has nodes  $N_i = E_{i-1}$ .
3. (proximity condition) If two edges in tree  $T_i$  are to be joined by an edge in tree  $T_{i+1}$  they must share a common node.

An example of a seven-dimensional R-vine tree specification together with its edge indices is given in Figure 1 (the figure is taken from Dißmann (2010)).

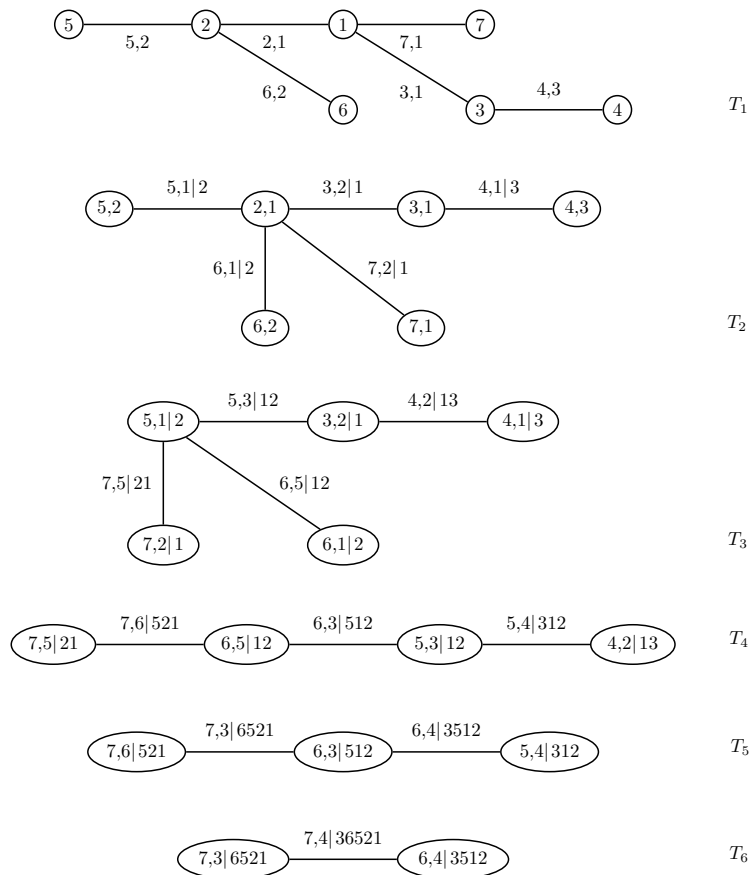


Figure 1. An R-vine tree on seven variables with edge indices

To build up a statistical model on R-vine trees with the node set  $\mathcal{N} := \{N_1, \dots, N_{d-1}\}$  and the edge set  $\mathcal{E} := \{E_1, \dots, E_{d-1}\}$ , one associates each edge  $e = j(e), k(e) | D(e)$  in  $E_i$  with a bivariate copula density  $c_{j(e), k(e) | D(e)}$ . The nodes  $j(e)$  and  $k(e)$  are denoted the *conditioned nodes*, while  $D(e)$  is the *conditioning set*. Let  $\mathbf{X}_{D(e)}$  be the subvector of  $\mathbf{X}$  determined by the indices contained in  $D(e)$ . A *regular vine distribution* is defined as the distribution of the random vector  $\mathbf{X} := (X_1, \dots, X_d)$  with marginal densities  $f_k, k = 1, \dots, d$ , 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 R-vine trees with node set  $\mathcal{N}$  and edge set  $\mathcal{E}$ . In Theorem 4.2 of Kurowicka and Cooke (2006) it is proven that the joint density of  $\mathbf{X}$  is uniquely determined and given by

$$f(x_1, \dots, x_d) = \left[ \prod_{k=1}^d f_k(x_k) \right] \times \left[ \prod_{i=1}^{d-1} \prod_{e \in E_i} c_{j(e), k(e) | D(e)}(F(x_{j(e)} | \mathbf{x}_{D(e)}), F(x_{k(e)} | \mathbf{x}_{D(e)})) \right], \quad (1)$$

where  $\mathbf{x}_{D(e)}$  denotes the subvector of  $\mathbf{x}$  determined by the indices contained in  $D(e)$ . For more details concerning the notation, see Czado (2010).

Until now, the concentration has been on two special cases of regular vines; drawable vines (D-vines) and canonical vines (C-vines). In particular an R-vine is called

- A *D-vine* if each node in  $T_1$  has a degree of at most 2, where the degree of a node denotes the number of connections or edges the node has to other nodes.
- A *C-vine* if each tree  $T_i$  has a unique node with degree  $d - i$ , the *root node*.

The corresponding R-vine distribution is called a D-vine or a C-vine distribution, respectively. For distinct indices  $i, j, i_1, \dots, i_k$  with  $i < j$  and  $i_1 < \dots < i_k$  we use the abbreviation

$$c_{i, j | i_1, \dots, i_k} := c_{i, j | i_1, \dots, i_k}(F(x_i | x_{i_1}, \dots, x_{i_k}), F(x_j | x_{i_1}, \dots, x_{i_k})). \quad (2)$$

This allows to write the joint densities of D- and C-vines as

- D-vine density

$$f(x_1, \dots, x_d) = \left[ \prod_{k=1}^d f_k(x_k) \right] \times \left[ \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i, i+j | i+1, \dots, i+j-1} \right] \quad (3)$$

- C-vine density

$$f(x_1, \dots, x_d) = \left[ \prod_{k=1}^d f_k(x_k) \right] \times \left[ \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{j, j+i | 1, \dots, j-1} \right]. \quad (4)$$



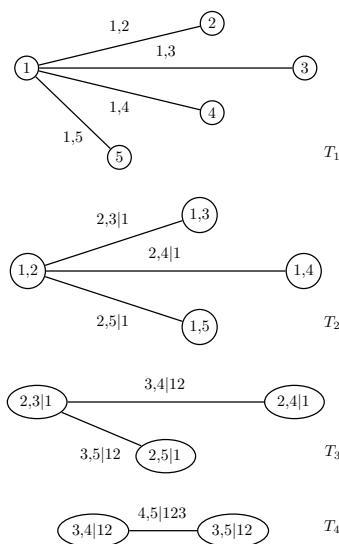


Figure 2. A C-vine tree on five variables with edge indices

See Aas et al. (2009) for more on simulation, inference and practical applications of C- and D-vines. A five-dimensional C-vine is shown in Figure 2.

For R-vines in general, there are no expressions like (3) and (4). Hence, an efficient way of storing the indices of the pair-copulas required in the joint density expression (1) is needed. One such approach was recently proposed by Kurowicka (2009) and explored in more detail in Dißmann (2010). It involves the specification of a lower triangular matrix  $M = (m_{i,j} | i, j = 1, \dots, d) \in \{0, \dots, d\}^{d \times d}$  with  $m_{i,i} = d - i + 1$ . That is, the diagonal entries of  $M$  are the numbers  $1, \dots, d$  in decreasing order. In this matrix, according to a rather technical condition, each row from the bottom up represents a tree, where the conditioned set is identified by a diagonal entry and by the corresponding column entry of the row under consideration, while the conditioning set is given by the column entries below this row. Corresponding copula types and parameters can conveniently be stored in matrices related to  $M$ .

The R-vine matrix corresponding to the R-vine in Figure 1 is

$$M = \begin{pmatrix} 7 & & & & & & & \\ 4 & 6 & & & & & & \\ 3 & 4 & 5 & & & & & \\ 6 & 3 & 4 & 4 & & & & \\ 5 & 5 & 3 & 2 & 3 & & & \\ 2 & 1 & 1 & 1 & 2 & 2 & & \\ 1 & 2 & 2 & 3 & 1 & 1 & 1 & \end{pmatrix},$$

where all other entries are zero. The bottom row of  $M$  corresponds to  $T_1$ , the second row from the bottom to  $T_2$ , and so on. To determine the edges in  $T_1$ , we

combine the numbers in the bottom row with the diagonal elements in the corresponding columns, i.e., the edges are  $(7,1)$ ,  $(6,2)$ ,  $(5,2)$  and so on. To determine the edges in  $T_2$ , we combine the numbers in the second row from the bottom with the diagonal elements in the corresponding columns, and condition on the elements in the bottom row, giving the edges  $(7,2|1)$ ,  $(6,1|2)$ , and so on. We proceed like this, and finally, the only edge in  $T_6$  is found by combining the two upper elements in the leftmost column of the matrix and condition on the remaining 5 entries in the same column. That is, the edge is  $(7,4|6521)$ .

The number of different possible R-vines in  $d$  dimensions is very large (Morales-Napoles et al., 2010). Hence, we need a way of selecting reasonable R-vine trees. Here, we will heuristically proceed as follows. We want to model the most important dependencies in the first trees. We therefore construct a graph on  $d$  nodes corresponding to the  $d$  variables, where all nodes are connected by a common edge, i.e., have  $d - 1$  neighbors. These edges have a weight according to a measure of pairwise dependence between the respective two variables, e.g., empirical Kendall's  $\tau$  or tail dependence. For this graph, we then find a maximum spanning tree (for this we use the well-known algorithm of Prim (1957)), which is a tree on all nodes which maximizes the pairwise dependencies. Given this tree, we can now select pair-copulas, estimate parameters and compute transformed observations  $F(x_{j(e)}|\mathbf{x}_{D(e)})$  for the next level. At the second level, we repeat the procedure from the first level, and the whole procedure is iterated until all trees are constructed and their pair-copulas sequentially estimated. See Dißmann (2010) for more details, and Brechmann (2010), Section 3.2 for construction methods for the special cases of D- and C-vines. In the latter case, the root node in each tree is found by choosing the variable with maximum sum of column entries in the matrix of pairwise dependencies.

### 3 Pairwise simplification and truncation

For the R-vine models to be useful for risk analysis of market portfolios, one needs to be able to fit the parameters of such models in the high-dimensional case, e.g., for 20-100 stocks. The computational effort needed to estimate all required parameters of an R-vine increases with the dimension. Hence, in this paper we take a pragmatic approach. We do not attempt to find the best fitting R-vine, but try to find the best fitting one under limited time and computational resources. To fix ideas, we want to allow for best possible specification of the first  $K$  trees in the R-vine, while higher order trees should only involve simple pair-copula terms, according to the idea that the most important dependencies are captured in the first trees.

Specifically, we denote an R-vine a *pairwisely simplified*  $K$  level one, if we replace all pair-copula terms which involve a conditioning set of size larger or equal to  $K$  by bivariate Gaussian copulas. Furthermore, we speak of a *pairwisely truncated* R-vine at level  $K$ , if all pair-copulas with conditioning set equal to, or larger than  $K$ , are set to bivariate independence copulas. If  $K = 1$ , the truncated R-vine becomes a Markov tree, where all conditional relationships are modeled as independent. Truncation may also be regarded as a special case of simplification, using Gaussian pair-copulas with correlation parameter equal to zero. Hence, it constitutes the greatest possible simplification.

In order to discuss the selection of simplification and truncation levels and propose appropriate procedures, we introduce some notation. First, we denote pairwisely truncated R-vines at level  $K$  by  $\text{tRV}(K)$  and pairwisely simplified  $K$  level ones by  $\text{sRV}(K)$ . Further, let  $\boldsymbol{\theta}_{\text{tRV}(K)}$  be the pair-copula parameters of the truncated R-vine, i.e.,  $\boldsymbol{\theta}_{\text{tRV}(K)} = \{\boldsymbol{\theta}_{j(e),k(e)|D(e)} : e \in E_i, i = 1, \dots, K\}$ , where  $\boldsymbol{\theta}_{j(e),k(e)|D(e)}$  denotes the parameter(s) of the copula  $c_{j(e),k(e)|D(e)}$ . Then, the density of a truncated R-vine at level  $K$  is given by

$$c_{\text{tRV}(K)}(\mathbf{u}|\boldsymbol{\theta}_T(K)) = \prod_{i=1}^K \prod_{e \in E_i} c_{j(e),k(e)|D(e)}, \quad (5)$$

where  $\mathbf{u} = (u_1, \dots, u_d)' \in [0, 1]^d$ , and the arguments of the copulas have been omitted for simplicity.

Note that the density of  $\text{tRV}(K)$  in (5) may be interpreted as a composite likelihood (see Lindsay (1988) and Varin et al. (2010)), where we compute an approximation to the joint density by setting  $c_{j(e),k(e)|D(e)} \equiv 1$  in higher order trees  $T_{K+1}, \dots, T_{d-1}$ , i.e., we are ignoring conditional pairwise dependencies. However, in our case, we know what we disregard when truncating, while this in general is

not the case in composite likelihood methods. From the theory of composite likelihood methods we directly may obtain consistency of the composite maximum likelihood estimates  $\hat{\boldsymbol{\theta}}_{\text{tRV}(K)}$  based on a sample from (5).

The density of a simplified  $K$  level R-vine is given by

$$c_{\text{sRV}(K)}(\mathbf{u}|\boldsymbol{\theta}_{\text{sRV}(K)}) = \left[ \prod_{i=1}^K \prod_{e \in E_i} c_{j(e),k(e)|D(e)} \right] \times \left[ \prod_{i=K+1}^{d-1} \prod_{e \in E_i} c_{j(e),k(e)|D(e)}^\rho \right], \quad (6)$$

where,  $c_{j(e),k(e)|D(e)}^\rho$  denote Gaussian pair copulas with correlation parameter  $\rho_{j(e),k(e)|D(e)}$ . Further,  $\boldsymbol{\theta}_{\text{sRV}(K)}$  is the parameter set of  $c_{\text{sRV}(K)}$ , i.e.,

$$\begin{aligned} \boldsymbol{\theta}_{\text{sRV}(K)} = & \{ \boldsymbol{\theta}_{j(e),k(e)|D(e)} : e \in E_i, i = 1, \dots, K \} \\ & \cup \{ \rho_{j(e),k(e)|D(e)} : e \in E_i, i = K + 1, \dots, d - 1 \}, \end{aligned} \quad (7)$$

with  $\boldsymbol{\theta}_{j(e),k(e)|D(e)}$  denoting the parameter(s) of the copula  $c_{j(e),k(e)|D(e)}$ .

In Sections 5 and 6 we will develop procedures for the selection of truncation and simplification levels, respectively, but first, in Section 4 we will describe the special case of a C-vine, for which joint simplification of the remaining  $d - K$  trees is possible.

## 4 Joint simplification and truncation

If we consider the special case of a C-vine, all pair-copulas with a conditioning set larger than or equal to  $K$  dimensions can be modeled jointly by a  $(d - K)$ -dimensional copula. Typically we will choose a simple shape for this  $(d - K)$ -dimensional copula, such as the independence copula or multivariate Gaussian copula. In the case of an independence copula we speak of a *jointly truncated* C-vine, while in the other case the resulting C-vine is denoted *jointly simplified*. In the following, jointly simplified  $K$  level C-vines will be denoted by  $\text{jsCV}(K)$ . It should be noted that simplification of C-vines previously has been treated by Heinen and Valdesogo (2009) and Robles (2009). There, it is referred to as “truncation”, while it using our notation would be denoted “simplification” (truncation in our meaning of the word is not explicitly discussed in Robles (2009)).

For C-vines the second component of the product in (6) reduces to a  $(d - K)$ -dimensional Gaussian copula. Hence, we obtain the density of a jointly simplified  $K$  level C-vine by rewriting (6) to

$$c_{\text{jsCV}(K)}(\mathbf{u}|\boldsymbol{\theta}_J(K)) = \left[ \prod_{j=1}^K \prod_{i=1}^{d-j} c_{j,j+i|1,\dots,j-1} \right] \times c_{K+1,\dots,d|1,\dots,K}^{\rho},$$

where  $c_{(K+1),\dots,d|1,\dots,K}^{\rho}$  denotes a  $(d - K)$ -dimensional Gaussian copula. The parameter set  $\boldsymbol{\theta}_{\text{jsCV}(K)}$  is defined similarly to (7) as

$$\begin{aligned} \boldsymbol{\theta}_{\text{jsCV}(K)} &= \{ \boldsymbol{\theta}_{j,j+i|1,\dots,j-1} : j = 1, \dots, K, i = 1, \dots, d - j \} \cup \\ &= \{ \rho_{ij|1,\dots,K} : i, j = K + 1, \dots, d, i \neq j \}, \end{aligned}$$

where  $\boldsymbol{\theta}_{j,j+i|1,\dots,j-1}$  are the parameters of the pair-copulas  $c_{j,j+i|1,\dots,j-1}$ , while  $\rho_{ij|1,\dots,K}$  denote the entries of the correlation matrix of the multivariate Gaussian copula  $c_{K+1,\dots,d|1,\dots,K}^{\rho}$ .

Finally, note that for D-vines, joint simplification as described above is not possible. The reason is that while C-vines have a common conditioning set in each tree as shown in (4), this is not the case for the D-vine (see (3)). For instance, in a five-dimensional D-vine, the arguments to the pair-copulas in tree  $T_2$  are  $F(x_1|x_2)$ ,  $F(x_3|x_2)$ ,  $F(x_2|x_3)$ ,  $F(x_4|x_3)$ ,  $F(x_3|x_4)$  and  $F(x_5|x_4)$ . Crosswise relationships such as  $F(x_2|x_3)$  and  $F(x_3|x_2)$  complicate the situation.

## 5 Selection of truncation level in the general case

We will now consider the selection of the truncation level in the general case. Note that  $\text{tRV}(K)$  is nested in  $\text{tRV}(K + 1)$ , since  $\boldsymbol{\theta}_{\text{tRV}(K)} \subset \boldsymbol{\theta}_{\text{tRV}(K+1)}$ . The log likelihood for  $\text{tRV}(K)$  is given by

$$l_{\text{tRV}(K)}(\boldsymbol{\theta}_{\text{tRV}(K)}|\mathbf{u}) = \sum_{i=1}^n \sum_{\ell=1}^K \sum_{e \in E_\ell} \ln [c_{j(e),k(e)|D(e)}(F(u_{i,j(e)}|\mathbf{x}_{i,D(e)}), F(x_{i,k(e)}|\mathbf{x}_{i,D(e)})|\boldsymbol{\theta}_{j(e),k(e)|D(e)})], \quad (8)$$

where  $n$  is the number of data points. If there are sufficient computing resources, we can maximize the full log likelihood (8). Alternatively, one may use the step-wise/sequential ML-estimator originally proposed by Aas et al. (2009) and further explored by Hobæk Haff (2010). To determine the sequential parameter estimates of  $\boldsymbol{\theta}_{\text{tRV}(K)}$ , it is enough to only use transformed variables up to tree  $K$ . These sequential estimates can then be used as starting values for maximizing  $l_{\text{tRV}(K)}(\boldsymbol{\theta}_{\text{tRV}(K)}|\mathbf{u})$ . We denote by  $\hat{\boldsymbol{\theta}}_{\text{tRV}(K)}$  the full maximum likelihood estimator and by  $\tilde{\boldsymbol{\theta}}_{\text{tRV}(K)}$  the sequential estimator of  $\boldsymbol{\theta}_{\text{tRV}(K)}$ , respectively. Note that for  $K$  small, the number of parameters to be maximized over is considerably reduced compared to a full R-vine specification.

We will start with  $K = 1$  and fit a truncated R-vine (for  $K = 0$  a pre-test of joint independence or normality can be performed). We thereafter increase  $K$  by one and assess how much gain we get by fitting the extra tree. If the gain is negligible we stop and use the resulting specification. If the gain is large enough, we increase  $K$  by one again, and proceed in this way until we have reached a truncation level  $K_0$ , which either gives a sufficient fit, or we have reached the computational time frame we allowed for the estimation process.

To assess whether there is gain to move from model  $\text{tRV}(K)$  to  $\text{tRV}(K + 1)$ , we now consider two kinds of statistical model selection techniques; AIC/BIC and the likelihood-ratio based test proposed by Vuong (1989). First, since  $\text{tRV}(K)$  is nested within  $\text{tRV}(K + 1)$ , we can compare the AIC or BIC values of the two models to quantify the marginal gain of an additionally fitted tree. In particular, these quantities are given by

$$\begin{aligned} AIC(\text{tRV}(K)) &:= -2 \times l_T^K(\hat{\boldsymbol{\theta}}_{\text{tRV}(K)}|\mathbf{u}) + 2 \times n_{\text{tRV}(K)} \\ BIC(\text{tRV}(K)) &:= -2 \times l_T^K(\hat{\boldsymbol{\theta}}_{\text{tRV}(K)}|\mathbf{u}) + \ln(n) \times n_{\text{tRV}(K)}, \end{aligned} \quad (9)$$

where  $n_{\text{tRV}(K)}$  denotes the dimension of  $\boldsymbol{\theta}_{\text{tRV}(K)}$ . We choose the one of  $\text{tRV}(K)$  and  $\text{tRV}(K + 1)$  with the smaller AIC or BIC value. If for some  $K_0$  the smaller

model is chosen, we stop, and declare  $\text{tRV}(K_0)$  as the best fitting model among the model sequence  $\text{tRV}(j)$ ,  $j = 0, \dots, d - 1$ .

Alternatively, we can use the likelihood-ratio based test proposed by Vuong (1989). In order to compare two competing models  $f_1$  and  $f_2$  with estimated parameters  $\hat{\theta}_1$  and  $\hat{\theta}_2$ , respectively, we compute the standardized sum,  $\nu$ , of the log differences of their pointwise likelihoods  $m_i := \log \left[ \frac{f_1(x_i|\hat{\theta}_1)}{f_2(x_i|\hat{\theta}_2)} \right]$  for observations  $x_i$ ,  $i = 1, \dots, n$ .  $\nu$  is shown to be asymptotically standard normal, leading to the following test. We prefer model 1 to model 2 at level  $\alpha$  if

$$\nu := \frac{\frac{1}{n} \sum_{i=1}^n m_i}{\sqrt{\sum_{i=1}^n (m_i - \bar{m})^2}} > \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right), \quad (10)$$

where  $\Phi^{-1}$  denotes the inverse of the standard normal distribution function. If  $\nu < -\Phi^{-1} \left( 1 - \frac{\alpha}{2} \right)$  we choose model 2. If, however,  $|\nu| \leq \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right)$ , no decision among the models is possible. Like AIC and BIC, the Vuong test statistic may be corrected for the number of parameters used in the models. There are two possible corrections; the Akaike and the Schwarz corrections, which correspond to the penalty terms in the AIC and the BIC, respectively.

We apply the Vuong test to compare  $\text{tRV}(K)$  (model  $f_1$ ) and  $\text{tRV}(K + 1)$  (model  $f_2$ ). If  $\nu \geq -\Phi^{-1} \left( 1 - \frac{\alpha}{2} \right)$ , we stop with  $\text{tRV}(K)$ , since  $\text{tRV}(K)$  is preferred to, or indistinguishable from  $\text{tRV}(K + 1)$ , at level  $\alpha$ . Hence, since the Vuong test attaches a significance level to its decision between the two models, it determines the truncation level as the level  $K_0$  for which  $\text{tRV}(K_0 + 1)$  does not provide a *statistically significant* gain in the model fit.

Note that even if AIC and BIC as well as the Vuong test require maximum likelihood estimates for their asymptotics, we will use sequential estimates as an approximation, since full maximum likelihood estimates are computationally quite demanding to obtain even for truncated R-vine models with rather small  $K$ . Sequential estimates have shown to provide a reasonably good approximation to the full maximum likelihood estimates, see Brechmann (2010), Section 11.3.7 and Appendix B and Hobæk Haff (2010). Algorithms 1 and 2 describe the truncation procedures. Algorithm 2 is stated in terms of the AIC, but the BIC can of course be used as well. In line 4 of Algorithm 2 we compute only the contribution from tree  $T_{j+1}$  to the AIC of  $\text{tRV}(K + 1)$ . This is due to the fact that the AICs of  $\text{tRV}(K)$  and  $\text{tRV}(K + 1)$  are equal with the exception of the contribution from tree  $T_{j+1}$ . Since all copulas in tree  $T_{j+1}$  of  $\text{tRV}(K)$  are independence copulas, the contribution from tree  $T_{j+1}$  to the AIC of  $\text{tRV}(K)$  is zero. Hence, if the contribution from tree  $T_{j+1}$  to the AIC of  $\text{tRV}(K + 1)$  is greater than zero, we truncate at level  $j$ .

Before we move on to selection of the simplification level in the general case, we turn to an illustrative example.

smaller model				larger model			
$T_1 : c_{12}$	$c_{23}$	$c_{34}$	$c_{45}$	$T_1 : c_{12}$	$c_{23}$	$c_{34}$	$c_{45}$
$T_2 :$	$c_{13 2}$	$c_{24 3}$	$c_{35 4}$	$T_2 :$	$c_{13 2}$	$c_{24 3}$	$c_{35 4}$
$T_3 :$	$\pi_{14 23}$	$\pi_{25 34}$		$T_3 :$	$c_{14 23}$	$c_{25 34}$	
$T_4 :$	$\pi_{15 234}$			$T_4 :$	$\pi_{15 234}$		

Figure 3. Pair-copula terms of five-dimensional D-vines truncated after the second tree  $T_2$  (smaller model) and after the third tree  $T_3$  (larger model), respectively, where  $\pi_{ij|D}$  denote independence copulas.

**Example 1 (Pairwise truncation of R-vines.)** *We consider a five-dimensional D-vine. Assume that we have already appropriately specified the pair-copulas of the first two trees  $T_1$  and  $T_2$ . We now want to determine whether the D-vine can be truncated or simplified at level 2. This is done by measuring the marginal gain of a third tree  $T_3$ . If the marginal gain is too small either in terms of AIC/BIC or as determined by a Vuong test, we truncate at level  $K = 2$ . Hence, we simply have to compare the smaller model ( $T_1 + T_2$ ) to the larger model ( $T_1 + T_2 + T_3$ ) as illustrated in Figure 3. Note that this is not an exact model comparison between a truncated D-vine and a fully specified one ( $T_1 + T_2 + T_3 + T_4$ ), but only an approximation to the truth, since possible dependencies in the fourth tree  $T_4$  are ignored in the comparison. However, under the assumption that most dependencies are captured in the first trees, this should be a reasonable approximation.*

**Require:** Observations of  $d$  variables, significance level  $\alpha$ .

- 1: **for**  $j = 0, \dots, d - 2$  **do**
- 2:   Specify  $\text{tRV}(j + 1)$  by additionally constructing tree  $T_{j+1}$  with appropriate pair-copulas..
- 3:   Perform a Vuong test for  $\text{tRV}(j)$  (model 1) and  $\text{tRV}(j + 1)$  (model 2), i.e., determine test statistic  $\nu$  as in (10), possibly with Akaike or Schwarz correction.
- 4:   **if**  $\nu \geq -\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$  **then**
- 5:     Truncate the R-vine at level  $K = j$ , i.e., exit the loop with  $\text{tRV}(j)$ .
- 6:   **end if**
- 7: **end for**

**Ensure:** Pairwisely truncated  $K$  level R-vine, or fully specified R-vine, if no truncation is possible.

Algorithm 1. Truncation of R-vines based on the Vuong test.



**Require:** Observations of  $d$  variables.

- 1: **for**  $j = 0, \dots, d - 2$  **do**
- 2:   Specify  $\text{tRV}(j + 1)$  by additionally constructing tree  $T_{j+1}$  with appropriate pair-copulas.
- 3:   Compute the AIC for  $\text{tRV}(j + 1)$  based on tree  $T_{j+1}$ .
- 4:   **if** this  $AIC > 0$  **then**
- 5:     Truncate the R-vine at level  $K = j$ , i.e., exit the loop with  $\text{tRV}(j)$ .
- 6:   **end if**
- 7: **end for**

**Ensure:** Pairwisely truncated  $K$  level R-vine, or fully specified R-vine, if no truncation is possible.

Algorithm 2. Truncation of R-vines using information criteria.

## 6 Selection of simplification level in the general case

Selection of simplification levels, i.e., model selection between  $\text{sRV}(K)$  and  $\text{sRV}(K + 1)$ , proceeds in essentially the same way as for truncation. However, on the contrary to  $\text{tRV}(K)$  and  $\text{tRV}(K + 1)$ , the models  $\text{sRV}(K)$  and  $\text{sRV}(K + 1)$  are not nested, in general  $\theta_{\text{sRV}(K)} \not\subset \theta_{\text{sRV}(K+1)}$ . Non-nested models may be compared using the Vuong test. If we use AIC or BIC, however, we have to deal with an increased variability (Ripley, 2008, pp. 34-35). Since we build models according to the paradigm that the most important dependencies are captured in the first trees, we assume that the specifications of trees  $T_{K+2}$  to  $T_{d-1}$  are equal in models  $\text{sRV}(K)$  and  $\text{sRV}(K + 1)$  if we work with AIC and BIC. Moreover, we know that trees  $T_1$  to  $T_K$  are the same. Hence, we have achieved “as much nestedness as possible”, since only tree  $T_{K+1}$  is different in both models.

The simplification procedures are outlined in Algorithms 3 and 4. Moreover,  $\widehat{\text{sRV}}(j + 1)$  denotes a simplified R-vine model under the assumption that trees  $T_{j+2}, \dots, T_{d-1}$  are specified with bivariate Gaussian copulas according to the discussion of nestedness above. In Example 1 this means that we assume that the Gaussian pair copulas  $c_{15|234}^\rho$  and  $\tilde{c}_{15|234}^\rho$  in  $T_4$  of the smaller and the larger model, respectively, are equal if we work with AIC and BIC.

**Require:** Observations of  $d$  variables, significance level  $\alpha$ .

- 1: **for**  $j = 0, \dots, d - 2$  **do**
- 2:   Specify  $\text{sRV}(j)$  by constructing higher order trees  $T_{j+1}, \dots, T_{d-1}$  with bivariate Gaussian copulas.
- 3:   Specify  $\text{sRV}(j + 1)$  by additionally constructing tree  $T_{j+1}$  with appropriate pair-copulas, and by constructing higher order trees  $T_{j+2}, \dots, T_{d-1}$  with bivariate Gaussian copulas.
- 4:   Perform a Vuong test for  $\text{sRV}(j)$  (model 1) and  $\text{sRV}(j + 1)$  (model 2), i.e., determine test statistic  $\nu$  as in (10), possibly with Akaike or Schwarz correction.
- 5:   **if**  $\nu \geq -\Phi^{-1}(1 - \frac{\alpha}{2})$  **then**
- 6:     Simplify the R-vine at level  $K = j$ , i.e., exit the loop with  $\text{sRV}(j)$ .
- 7:   **end if**
- 8: **end for**

**Ensure:** Pairwisely simplified  $K$  level R-vine, or fully specified R-vine, if no simplification is possible.

Algorithm 3. Simplification of R-vines based on the Vuong test.

**Require:** Observations of  $d$  variables.

- 1: **for**  $j = 0, \dots, d - 2$  **do**
- 2: Specify  $\text{sRV}(j)$  by constructing tree  $T_{j+1}$  with bivariate Gaussian copulas (the remaining trees can be ignored primarily).
- 3: Specify  $\widehat{\text{sRV}}(j + 1)$  by constructing tree  $T_{j+1}$  with appropriate pair-copulas.
- 4: Compute the AIC for  $\text{sRV}(j)$  (model 1) and  $\widehat{\text{sRV}}(j + 1)$  (model 2) based on tree  $T_{j+1}$ .
- 5: **if**  $AIC_1 < AIC_2$  **then**
- 6: Simplify the R-vine at level  $K = j$  by specifying the remaining Gaussian pair-copulas for trees  $T_{j+2}, \dots, T_{d-1}$ , i.e., exit the loop with  $\text{sRV}(j)$ .
- 7: **end if**
- 8: **end for**

**Ensure:** Pairwisely simplified  $K$  level R-vine, or fully specified R-vine, if no simplification is possible.

Algorithm 4. Simplification of R-vines using information criteria.

# 7 Selection of simplification and truncation levels for C-vines

The procedures described in Sections 5 and 6 may of course be used also for the special case of a C-vine. However, in this case, we may alternatively use multivariate independence tests or copula goodness-of-fit tests to determine whether we can truncate or simplify the model at level  $K$ , respectively. As previously described, this way of truncation or simplification is denoted *joint* truncation or simplification. In the simulation study in Section 8 we will examine joint truncation using the independence test based on the empirical copula process studied in Genest and Rémillard (2004), as well as using the simpler test based on a multivariate extension of Spearman’s  $\rho$  as proposed by Quessy (2009). If the p-value of one of these tests is larger than a preliminarily chosen level, we truncate the structure at level  $K$ . Also for the purpose of simplification, we will use two different tests; the blanket test based on the empirical copula process studied by Genest and Rémillard (2008), and the test based on Rosenblatt’s transformation, suggested by Breymann et al. (2003). The first-mentioned test was shown to perform very well in the studies Berg (2009) and Genest et al. (2009), while the second test presumably works well for testing the Gaussian copula against copulas with heavy tails. If the p-value of one of these tests is larger than a preliminarily chosen level, we simplify the structure at level  $K$ . Algorithm 5 outlines the joint simplification procedure, while joint truncation proceeds in exactly the same way but with the use of an independence test in the second line.

An example is given in the following.

**Example 2 (Joint truncation and simplification of C-vines.)** *We now consider a five-dimensional C-vine similar to the D-vine discussed in Example 1. Again assume that we have already appropriately specified the first two trees  $T_1$  and  $T_2$  and now want to investigate whether we can truncate or simplify at level  $K = 2$ , i.e., if the model to the left in Figure 4 is more appropriate than the one to the right.*

truncated model					simplified model				
$T_1 :$	$c_{12}$	$c_{13}$	$c_{14}$	$c_{15}$	$T_1 :$	$c_{12}$	$c_{13}$	$c_{14}$	$c_{15}$
$T_2 :$	$c_{23 1}$	$c_{24 1}$	$c_{25 1}$		$T_2 :$	$c_{23 1}$	$c_{24 1}$	$c_{25 1}$	
$T_{3/4} :$	$\pi_{345 12}$				$T_{3/4} :$	$c_{345 12}^\rho$			

Figure 4. Pair and multivariate copula terms of five-dimensional C-vines jointly truncated and jointly simplified after the second tree  $T_2$ , respectively, where  $\pi_{345|12}$  denotes a three-dimensional independence copula and  $c_{345|12}^\rho$  a three-dimensional Gaussian copula.

To solve this problem, we test whether the transformed observations from tree  $T_2$  are

jointly independent (in the case of truncation), or whether they follow a multivariate (here: trivariate) Gaussian copula  $c_{345|12}^{\rho}$  (in the case of simplification). If the  $p$ -value of the respective test is larger than the pre-specified significance level  $\alpha$ , then the null hypothesis that the transformed observations are jointly independent/normal cannot be rejected. Hence we truncate/simplify the C-vine at level  $K = 2$ .

**Require:** Observations of  $d$  variables, significance level  $\alpha$ .

- 1: **for**  $j = 0, \dots, d - 2$  **do**
- 2: Perform a copula goodness-of-fit test for  $\text{jsCV}(j)$  to test if the transformed observations from tree  $T_j$  can be appropriately modeled with a  $(d - j)$ -dimensional Gaussian copula.
- 3: **if**  $p\text{-value} > \alpha$  **then**
- 4: Simplify the C-vine at level  $K = j$ , i.e., exit the loop with  $\text{jsCV}(j)$ .
- 5: **end if**
- 6: Specify tree  $T_{j+1}$  with appropriate pair-copulas.
- 7: **end for**

**Ensure:** Jointly simplified  $K$  level C-vine, or fully specified C-vine, if no simplification is possible.

Algorithm 5. Joint simplification of C-vines.

## 8 Simulation studies

Figure 5 shows a flow chart which highlights the main heuristical steps for each level when fitting an R-vine:

1. Tree construction using the maximum spanning tree algorithm for R-vines (or by identifying root nodes for C-vines).
2. Pair copula type selection (e.g., using copula goodness-of-fit tests or AIC comparisons).
3. Pair copula parameter estimation (usually via maximum likelihood estimation or alternatively by inversion of Kendall's  $\tau$ ).
4. Investigating whether truncation/simplification is possible.

Typically steps 2 and 3 are jointly performed, since, e.g., for AIC comparisons the parametric copula has to be estimated.

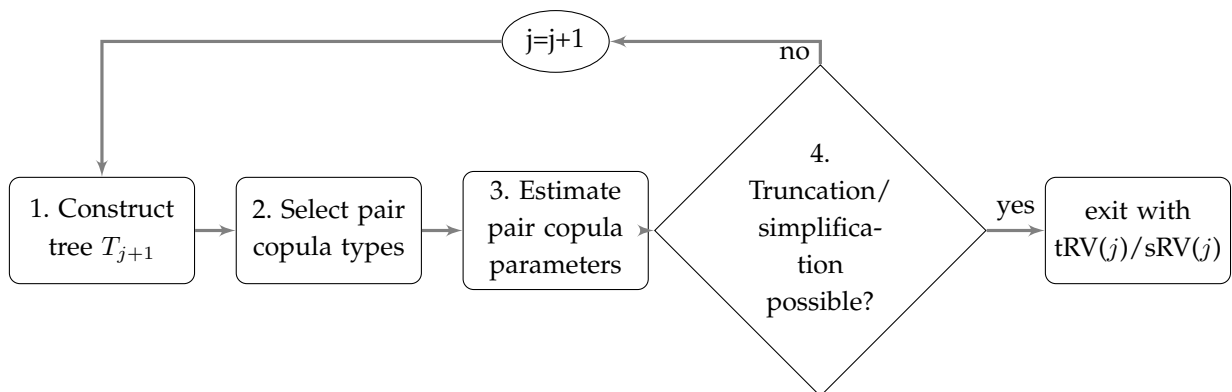


Figure 5. Flow chart of the truncation and simplification procedures described in Algorithms 1-4.

We have evaluated the performance of the simplification and truncation procedures presented in Sections 5-7 in three extensive simulation studies. The first study concerns pairwise simplification and truncation of R-vines and is described in Section 8.1. This study focuses solely on the tests proposed for identifying simplification and truncation levels, e.g. the step 4 above. In contrast, the second study, described in Section 8.2, examines the full heuristical specification procedure for R-vines shown in Figure 5. Finally, in the last study, treated in Section 8.3, we have investigated whether the tailor-made procedure for C-vines described in Section 7 works well.

## 8.1 Pairwise simplification and truncation of R-vines

To examine the performance of the pairwise simplification and truncation methods proposed in Sections 5 and 6, we focus on step 4 of the general R-vine specification procedure shown in Figure 5. That is, we fix the structure of the R-vine (step 1) and perform steps 2-4 only. In step 2, we select pair-copulas from a range of 11 bivariate families: independence copula, Gaussian, t, Clayton, rotated Clayton (90 degrees), Gumbel, rotated Gumbel (90 degrees), Frank, Joe, Clayton-Gumbel (BB1), Joe-Clayton (BB7). For more information on copula types, see, e.g., Nelsen (2006) or Joe (1997).

The study is based on the seven-dimensional R-vine shown in Figure 1. We choose four different scenarios for the types of copulas in the structure:

- **Scenario 1:** Clayton copulas in  $T_1$  and  $T_2$  with Kendall's  $\tau$  values equal to 0.6 and 0.3 in  $T_1$  and  $T_2$ , respectively.
- **Scenario 2:** t copulas in  $T_1$  and  $T_2$  with Kendall's  $\tau$  values equal to 0.6 and 0.3 in  $T_1$  and  $T_2$ , respectively and corresponding degrees of freedom values equal to 3 and 7.
- **Scenario 3:** Mixed copulas (t, Clayton, Gumbel) in  $T_1$  and  $T_2$  with with Kendall's  $\tau$  values equal to 0.6 and 0.3 in  $T_1$  and  $T_2$ , respectively and corresponding degrees of freedom values equal to 3 and 7<sup>1</sup>.
- **Scenario 4:** t copulas in  $T_1$  and  $T_2$  with mixed Kendall's  $\tau$  values between 0.6 and 0.8 in  $T_1$  and between 0.2 and 0.4 in  $T_2$ . The degrees of freedom parameter is chosen between 3 and 5 in  $T_1$  and between 7 and 9 in  $T_2$ .

In all four scenarios, the trees  $T_3, \dots, T_6$  are specified with independence copulas in the truncation setting and with Gaussian copulas in the simplification setting. That is, the correct truncation/simplification level  $\bar{K}$  is 2 in all our experiments. In the simplification setting, the correlation parameter of the Gaussian copula was set to 0.25 in  $T_3$ , 0.20 in  $T_4$ , 0.15 in  $T_5$  and 0.10 in  $T_6$ , respectively.

We simulated  $n \in \{500, 1000\}$  observations for each scenario and used  $R = 100$  repetitions. For each scenario/repetition/ $n$  we determined the truncation or simplification level  $K_0$  based on five different criteria:

- the Vuong test without correction,
- the Vuong test with Akaike correction (*V.Akaike*),
- the Vuong test with Schwarz correction (*V.Schwarz*),

---

1. Pairs with t copulas: (2,1), (3,1), (3,2|1). Pairs with Clayton copulas: (4,3), (6,2), (5,1|2), (7,2|1). Pairs with Gumbel copulas: (5,2), (7,1), (4,1|3), (6,1|2).

- AIC,
- BIC.

For the Vuong tests we set  $\alpha = 0.05$ .

The performance of the different criteria is assessed using a scoring method. For each repetition, the score is increased by 1 if  $\bar{K}$  is identified correctly, and by 0.5 if  $|\bar{K} - K_0| = 1$ . If the fitted model is appropriate, this score should be high (the maximum value is 100).

The resulting scores are shown in Table 1. Evidently the Vuong test without correction identifies truncation and simplification levels most accurately. Truncation levels are also very well determined by the Vuong tests with corrections. However, for scenario 2, and especially for scenario 4, the Schwarz correction term leads to earlier simplifications, and hence reduced scores. Generally, truncation is easier than simplification, and for this task AIC and BIC also perform quite well. However, when it comes to the issue of simplification, using AIC and BIC, the simplification level is generally determined too late.

To summarize, the Vuong test seems to be the best criterion, while more parsimonious models can be obtained by using Akaike or Schwarz corrections. For truncation, AIC and BIC are also valid alternatives, while they are less reliable for simplification.

## 8.2 R-vine specification

In practice the “true” structure of an R-vine is hardly ever known. Hence it is important to evaluate the complete heuristical R-vine specification procedure outlined in Figure 5. In particular, the use of the maximum spanning tree algorithm in step 1 almost certainly leads to the selection of other than the “true” variable pairs. Take for instance the R-vine in scenario 4. For this R-vine there actually are three pairs of variables at the first level that correspond to higher Kendall’s  $\tau$  values than some of the true pairs. This can be verified using the formula for the partial Kendall’s  $\tau$ , see e.g. Simon (1977):

$$\tau_{xy|z} = (\tau_{xy} - \tau_{xz}\tau_{yz}) / \left( \sqrt{1 - \tau_{xz}^2} \sqrt{1 - \tau_{yz}^2} \right).$$

Hence, the maximum spanning tree algorithm very likely selects some of these pairs instead of the true ones. Having selected other pairs than the “true ones” at the first level, may lead to a completely different tree in the end. Hence, instead of aiming at *reconstructing* the “true” underlying model, we try to find the best *fitting* model. Ideally speaking, this is of course the reconstructed true model.

This study is also based on the seven-dimensional R-vine shown in Figure 1 with the four different pair-copula scenarios described in Section 8.1. Again we sim-



scenario	$n$	Vuong	V.Akaike	V.Schwarz	AIC	BIC
truncation						
1	500	100.0	100.0	100.0	87.0	95.0
	1000	99.5	100.0	100.0	94.5	99.5
2	500	99.5	100.0	100.0	86.5	94.5
	1000	100.0	100.0	100.0	91.0	97.5
3	500	100.0	100.0	100.0	93.0	97.5
	1000	100.0	100.0	100.0	89.5	96.5
4	500	100.0	100.0	100.0	84.5	97.0
	1000	100.0	100.0	100.0	87.0	97.0
simplification						
1	500	100.0	100.0	100.0	25.5	42.5
	1000	99.5	99.5	99.5	50.5	68.5
2	500	98.0	95.0	66.5	23.0	40.5
	1000	100.0	100.0	99.5	56.0	71.0
3	500	99.5	100.0	100.0	25.5	42.0
	1000	100.0	100.0	100.0	48.5	65.0
4	500	94.5	83.0	56.5	20.0	36.5
	1000	100.0	99.5	86.5	50.0	65.0

Table 1. Truncation and simplification scores for the four scenarios described in Section 8.1. Each column corresponds to a different truncation or simplification procedure, respectively.

ulated  $n \in \{500, 1000\}$  observations for each scenario and used  $R = 100$  repetitions. The R-vine for each scenario/repetition/value of  $n$  was found using the procedure in Figure 5 with the same five criteria for determining the truncation/simplification levels as in Section 8.1 and selecting among the same 11 pair-copula families. For each scenario/repetition/value of  $n$  we also fitted a fully specified R-vine. The full R-vine represents the best obtainable model by our model construction principles described in Section 2, and hence serves as a benchmark model.

To assess the performance of the complete heuristical R-vine specification procedure outlined in Figure 5, we used 3 different evaluation measures:

1. *Simplification/truncation level score*: For each repetition, the score is increased by 1 if  $\bar{K}$  is identified correctly, and by 0.5 if  $|\bar{K} - K_0| = 1$ . If the fitted model is appropriate, this score should be high. The maximum value is 100. This is the same measure as was used in Section 8.1.
2. *Mean KLIC*: For each repetition, we compute the empirical Kullback-Leibler distance (Kullback and Leibler, 1951) between the fitted model and the true model. The measure is then the mean over all repetitions<sup>2</sup>.
3. *Vuong test without correction*: For each repetition we compare the true model and the fitted (simplified or truncated) model using the Vuong test without correction and count the number of times  $\nu \leq \Phi^{-1}(1 - \frac{\alpha}{2})$  with  $\alpha = 0.05$ , i.e. the number of times we do *not* prefer the true model (see Section 5). This measure serves a similar purpose as the KLIC, but it attaches a significance level to its decision.

More details on the evaluation criteria and the simulation studies in general can be found in Brechmann (2010), Chapter 10. The results are shown in Tables 2 (truncation) and 3 (simplification), where the structure is the same for both tables. The first two columns contain the scenario number and the number of observations, respectively. The third column gives the names of the evaluation measures, while the fourth column provides the values of these evaluation measures for the case when we have fitted a full R-vine. Finally, the last five columns contain the values of the same evaluation measures, but for truncated/simplified R-vines based on the five different criteria.

We observe the following from the results in Table 2:

---

2.  $\overline{KLIC}(h_0) := \frac{1}{R} \sum_{r=1}^R \left[ \frac{1}{n} \sum_{i=1}^n h_0(\mathbf{u}_i^{0(r)}) \log \left[ h_0(\mathbf{u}_i^{0(r)}) \right] - \frac{1}{n} \sum_{i=1}^n h_0(\mathbf{u}_i^{0(r)}) \log \left[ f_r(\mathbf{u}_i^{0(r)} | \hat{\boldsymbol{\theta}}_r) \right] \right],$

where  $\mathbf{u}_i^{0(r)}$  is the  $i$ -th simulated observation from the true vine model with density  $h_0$ , while the density of the approximating vine model in the  $r$ -th repetition is given by  $f_r$  with estimated parameters  $\hat{\boldsymbol{\theta}}_r$ . Note that this empirical measure can also be negative in contrast to the theoretical KLIC.

1. For all scenarios, except scenario 4,  $n=500$ , the truncation level score is quite low. Hence, for most of the scenarios/repetitions, the identified truncation level is significantly higher than 2.
2. The high truncation levels are also verified by comparing the mean KLIC values for the full and truncated models. The mean KLIC values for the truncated models are in most cases quite close to those of the full models, indicating a low degree of truncation.
3. Another interesting fact about the mean KLIC values for the full models, is that they are negative for scenario 2,  $n=500$  and scenario 4,  $n=500$ . This means, that for these two scenarios, the R-vine specification procedure outlined in Figure 5 in average actually lead to R-vines that give a better fit to the simulations from the true model, than the true model itself.
4. Hence, even though the truncation level score for scenario 2,  $n=500$  is as low as 35.0, the low mean KLIC value and the high value for the Vuong test actually indicates that for the majority of the repetitions, the estimated models are better than the true model.
5. The mean KLIC values for scenario 3 are larger than those for the other scenarios. We believe the reason for this to be the mixed copula types at the first level. It is known from Joe et al. (2010) that if all the copulas at the first level are lower (upper) tail dependent, then the copulas of the remaining pairs also are lower (upper) tail dependent. Hence, if other pairs than the true ones are chosen for scenario 1, 2 or 4, these pairs at least have the same sort of dependence as the true pairs, that is, they are lower tail dependent in scenario 1 and both upper and lower tail dependent in scenarios 2 and 4. However, if other pairs than the true ones are chosen for scenario 3, these may have a dependency structure that, if we are unlucky, does not fit very well with any of the 11 pair-copula families used in this study. Hence, the resulting full R-vine may differentiate substantially from the true one.

The results in Table 3 are similar to those in Table 2. However, for all scenarios and criteria, the simplification level scores are higher than the truncation level scores. That is, simplification seems generally to take place earlier than truncation. Moreover, the mean KLIC values all over are higher in the simplification case than in the truncation case.

### 8.3 C-vine specification (joint simplification and truncation)

In this study we investigate whether the procedure for joint truncation and simplification of C-vines described in Section 7 works well. Instead of the seven-dimensional R-vine used in Section 8.2, we here use a ten-dimensional C-vine.

scenario	$n$	eval. crit.	truncation procedure					
			full	Vuong	V.Akaike	V.Schwarz	AIC	BIC
1	500	truncation level	-	22.5	23.5	26.0	19.0	21.0
		mean KLIC	0.022	0.030	0.033	0.035	0.026	0.026
		Vuong test	70	64	63	61	67	67
	1000	truncation level	-	15.5	15.5	17.5	13.0	15.0
		mean KLIC	0.029	0.037	0.037	0.041	0.036	0.037
		Vuong test	32	31	31	31	32	31
2	500	truncation level	-	32.5	33.5	35.0	30.0	30.5
		mean KLIC	-0.003	0.005	0.006	0.010	0.001	0.001
		Vuong test	98	92	91	88	95	95
	1000	truncation level	-	31.0	31.5	33.0	26.0	30.5
		mean KLIC	0.010	0.012	0.012	0.016	0.016	0.012
		Vuong test	80	79	79	76	79	79
3	500	truncation level	-	42.0	42.5	43.5	34.5	38.5
		mean KLIC	0.127	0.130	0.130	0.131	0.129	0.130
		Vuong test	26	25	25	25	26	25
	1000	truncation level	-	48.5	48.5	49.0	39.5	43.5
		mean KLIC	0.111	0.111	0.111	0.112	0.111	0.111
		Vuong test	16	15	15	15	16	15
4	500	truncation level	-	61.5	74.0	95.5	49.5	53.5
		mean KLIC	-0.001	0.005	0.011	0.027	0.002	0.003
		Vuong test	99	96	91	74	99	97
	1000	truncation level	-	49.5	50.5	68.5	42.5	47.5
		mean KLIC	0.007	0.009	0.009	0.017	0.008	0.009
		Vuong test	85	79	78	63	82	80

Table 2. Simulation results for the four stepwise truncation scenarios described in Section 8.2. Each column corresponds to a different truncation procedure.

scenario	n	eval. crit.	simplification procedure					AIC	BIC
			full	Vuong	V.Akaike	V.Schwarz			
1	500	simplification level	-	41.5	42.0	45.5	0.5	1.0	
		mean KLIC	0.016	0.036	0.037	0.044	0.016	0.017	
		Vuong test	91	70	68	62	90	89	
	1000	simplification level	-	23.5	23.0	28.5	2.0	4.5	
		mean KLIC	0.033	0.043	0.043	0.046	0.033	0.034	
		Vuong test	28	20	20	16	28	27	
2	500	simplification level	-	64.0	68.0	72.5	3.5	13.5	
		mean KLIC	-0.005	0.022	0.029	0.058	-0.005	-0.001	
		Vuong test	97	79	72	51	97	93	
	1000	simplification level	-	41.5	52.0	68.5	8.5	16.0	
		mean KLIC	0.004	0.011	0.017	0.026	0.005	0.007	
		Vuong test	87	78	67	55	86	84	
3	500	simplification level	-	64.5	66.0	70.5	5.5	11.5	
		mean KLIC	0.174	0.200	0.200	0.207	0.174	0.176	
		Vuong test	7	7	7	7	7	7	
	1000	simplification level	-	52.5	56.0	62.0	15.5	17.0	
		mean KLIC	0.117	0.124	0.125	0.129	0.117	0.118	
		Vuong test	16	16	16	16	16	16	
4	500	simplification level	-	88.5	87.5	74.5	27.0	32.0	
		mean KLIC	0.011	0.024	0.029	0.046	0.012	0.014	
		Vuong test	93	78	74	53	90	88	
	1000	simplification level	-	69.0	77.0	88.5	14.5	26.5	
		mean KLIC	0.024	0.031	0.033	0.039	0.025	0.026	
		Vuong test	51	29	29	21	47	41	

Table 3. Simulation results for the four stepwise simplification scenarios described in Section 8.2. Each column corresponds to a different simplification procedure.

However, the truncation/simplification level is still chosen to be  $\bar{K} = 2$ , and we also for this structure specify four different scenarios for the types of copulas:

- **Scenario 1:** Clayton copulas in  $T_1$  and  $T_2$  with Kendall's  $\tau$  values equal to 0.6 and 0.3 in  $T_1$  and  $T_2$ , respectively.
- **Scenario 2:** t copulas in  $T_1$  and  $T_2$  with Kendall's  $\tau$  values equal to 0.6 and 0.3 in  $T_1$  and  $T_2$ , respectively and corresponding degrees of freedom values equal to 3 and 7.
- **Scenario 3:** Mixed copulas (t, Clayton, Gumbel)<sup>3</sup> in  $T_1$  and  $T_2$  with with Kendall's  $\tau$  values equal to 0.6 and 0.3 in  $T_1$  and  $T_2$ , respectively and corresponding degrees of freedom values equal to 3 and 7.
- **Scenario 4:** t copulas in  $T_1$  and  $T_2$  with mixed Kendall's  $\tau$  values between 0.6 and 0.8 in  $T_1$  and between 0.2 and 0.4 in  $T_2$ . The degrees of freedom parameter is chosen between 3 and 5 in  $T_1$  and between 7 and 9 in  $T_2$ .

In all four scenarios, the trees  $T_3, \dots, T_9$  are specified with independence copulas in the truncation setting and with Gaussian copulas in the simplification setting. In the simplification setting, the correlation parameter of the Gaussian copula was 0.25 in  $T_3$ , 0.20 in  $T_4$  and  $T_5$ , 0.15 in  $T_6$  and  $T_7$  and 0.10 in  $T_8$  and  $T_9$ , respectively.

Like for the study described in Section 8.2 we simulated  $n \in \{500, 1000\}$  observations for each scenario and used  $R = 100$  repetitions. For each scenario/repetition/ $n$  we fitted an C-vine using the procedure described in Section 2 (using Kendall's  $\tau$ 's as weights and selecting copula type using AIC). In this study, the truncation/simplification level  $K_0$  was determined in two different ways. We either used

- a "stepwise" procedure with the same five criteria as in Section 8.2, or
- multivariate independence and copula goodness-of-fit tests.

In the latter case, we used four different tests; two for testing for independence:

- Test based on empirical copula process (Genest and Rémillard, 2004)
- Test based on a multivariate extension of Spearman's  $\rho$  (Quessy, 2009)

and two for testing for normality:

- The blanket test based on the empirical copula process (Genest and Rémillard, 2008)
- The test based on Rosenblatt's transformation (Breymann et al., 2003).

---

3. Pairs with t copulas: (1,4), (1,7), (1,10), (2,5|1), (2,8|1). Pairs with Clayton copulas: (1,3), (1,6), (1,9), (2,4|1), (2,7|1), (2,10|1). Pairs with Gumbel copulas: (1,2), (1,5), (1,8), (2,3|1), (2,6|1), (2,9|1).

All results from this study can be found in Brechmann (2010), Section 10.2. Here, we only give the main findings which are:

1. The performance of the stepwise methods based on the Vuong test is stronger than that of the multivariate goodness-of-fit tests, despite the fact that joint simplification and truncation use all available information in order to determine simplification/truncation levels.
2. Joint simplification using the goodness-of-fit test based on the Rosenblatt transformation is always superior to the one based on the empirical copula process. Especially for scenarios 2 and 4, the test based on the empirical copula process fails completely (this can particularly be seen in terms of a large mean KLIC).
3. Both multivariate independence tests under consideration for joint truncation perform well, but the truncation test based on the Spearman's  $\rho$  seems to be slightly more accurate than that based on the empirical copula process.

## 8.4 Summary

To summarize, the procedures based on the Vuong test with or without correction for the number of parameters should be used in most cases, even for joint simplification/truncation where tailor-made procedures are available. The procedures based on AIC/BIC can be regarded as “quick and dirty” alternatives. These criteria tend to indentify simplification/truncation too late, but they are very fast compared to the the Vuong test (in a 52-dimensional example, AIC identified the truncation/simplification level 44%/80% faster than the Vuong test). Moreover, the resulting model usually is at least close to the true model.

## 9 Applications

In this section, we analyze a 19-dimensional data set consisting of Norwegian and international financial variables. See Table 4 for a description. The variables constitute the market portfolio of a large Norwegian financial institution and hence, it is very important to correctly model the dependencies between them. The observed time period is from 3/25/2003 to 3/26/2008, resulting in 1107 daily observations. As previously stated, the computational effort needed to estimate all required parameters of an R-vine increases with the dimension. Hence, the aim of the work presented here was to investigate whether simplification or truncation of the R-vine specification corresponding to this 19-dimensional data set is possible.

Before analyzing the dependence in the data set, we selected appropriate time series models for the univariate margins. For the stock indices and exchange rates, we used a constant-mean model, while for the other variables either an AR(1)-model (long interest rates and bond indices) or an ARMA(1,1)-model (short interest rates and real estate index) were chosen. Further, for all variables, except for the 3-month interest rates, a GARCH(1,1)-model was used to model the volatility. The error distribution of the GARCH-model was chosen to be the standardized normal-inverse Gaussian (NIG) distribution (Barndorff-Nielsen, 1997) for all variables except for the real estate and hedge fond indices, for which the Student's t-distribution was used instead. After filtering the original returns with the chosen univariate models, the standardized residual vectors are converted to uniform pseudo-observations using their empirical distribution functions. More details about the marginal models can be found in Brechmann (2010).

For the sake of reference, we first fit a full R-vine to this data set, using the approach described in 2. Like for the simulated data, Kendall's  $\tau$ 's are used as edge weights, and pair-copulas are selected from a range of 11 bivariate families using AIC.

Figure 6 shows the first two trees of the fitted R-vine. See Table 4 for the correspondence between the IDs and the variable names. The edge labels represent the Kendall's  $\tau$ 's between the respective variables. The chosen copula types and estimated parameters for the first two trees are given in Table 5. For the t copula, the first and second parameter components are the correlation and degrees of freedom parameters respectively. For the BB1 and BB7 copulas, the two components are  $\theta$  and  $\delta$  as in the definition in Joe (1997). The corresponding R-vine matrices can be found in Brechmann (2010). In economical terms,  $T_1$  has an evident interpretation. It is constituted of three clusters of economically similar variables. The first cluster consists of the stock indices, the hedge fond index and the real



estate index (variables V1, V17, V18, V19 and V20). The second cluster consists of the interest rates and the bond indices (V7, V8, V9, V10, V11, V12, V13, V14, V15 and V16), and finally, the exchange rates (V2, V3, V4, and V5) constitute the third cluster. The stock and interest rate clusters are linked through the variables V10 and V17 (i.e., the 5-year German Government Rate and the MSCI World index), while the interest rate and exchange rate clusters are connected through V14 and V2 (i.e., the WGBI bond index and the USD-NOK exchange rate).

In addition to the R-vine with different copulas for different pairs, we also fitted an R-vine with t copulas for all pairs. However, the BIC-values in the two upper rows of Table 6 show that the R-vine with mixed copulas is superior to the the one with only t copulas.

Having fitted the full R-vines, we apply the different statistical model selection criteria from Sections 5 and 6 to investigate whether truncation and/or simplification of the R-vine specification corresponding to this data set are possible. Table 6 shows the results. We report log likelihood values, the number of parameters<sup>4</sup> and BIC for the truncated/simplified models obtained using the different criteria (truncation/simplification based on AIC and BIC turned out to give the same results for this data set). In addition, the table shows the test statistics of Vuong tests (with and without Schwarz correction) with respect to the null hypothesis that the fully specified model and simplified/truncated model are equivalent. Test statistics indicated by “\*” imply that the null hypothesis cannot be rejected at the 5% level or that the simplified/truncated model is even superior.

If we first turn to the truncation results, they show that truncation at level 6 seems to give a slightly better model than truncation at level 4. The hypothesis that the fully specified model and the truncated model are equivalent is however rejected for both tRV(4) and tRV(6), meaning that there still seems to be significant dependencies after tree  $T_6$ . In (Brechmann, 2010, Section 11.2.2) we have studied the model tRV(4) in more detail by considering, among others, joint tail behavior, copula Q-Q plots and Kendall’s  $\tau$ ’s of simulated observations. The results showed that although this model did not fully reproduce the observed data characteristics, it may be viewed as an adequate specification for the data. Hence, we conclude that the most important dependencies in this data set are actually captured in the first four to six trees, meaning that the corresponding R-vine may be truncated at level 6, or even at level 4, depending on the desired level of parsimony (and of course at the expense of accuracy).

---

4. Note that the number of parameters of a full R-vine with  $d$  variables is  $d(d-1)/2$  if all pair-copulas have one parameter each. The reason why the number of parameters shown in Table 6 is much smaller than this, is that many of the pair-copulas in the full R-vine (both the one with mixed copulas and the one with only t-copulas) are estimated to be independence copulas.

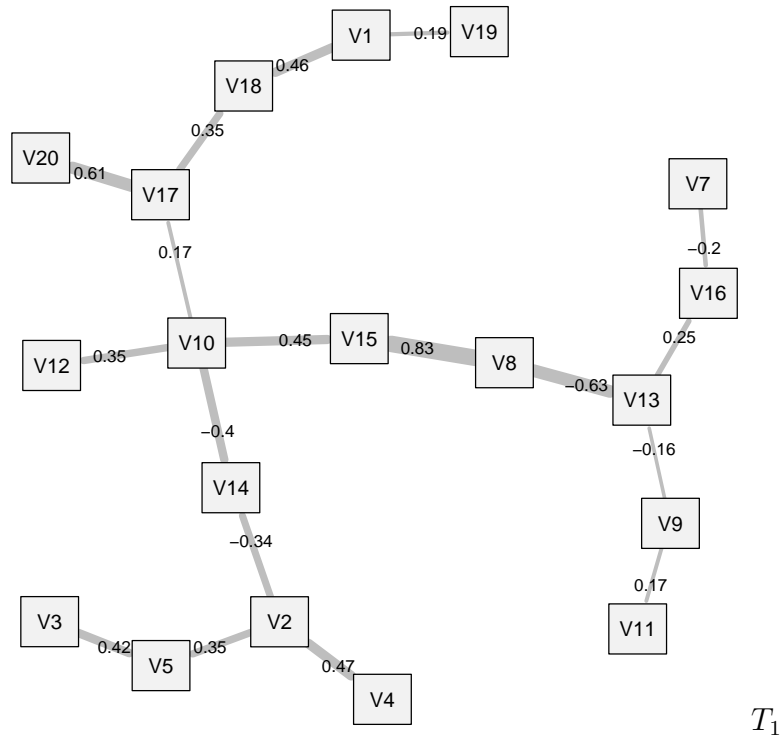
ID	variable name	description
V1	FINX	Norwegian Financial Index
V2	USDNOK	USD-NOK exchange rate
V3	EURNOK	EURO-NOK exchange rate
V4	Yen	YEN-NOK exchange rate
V5	GBP	GBP-NOK exchange rate
V7	NIBOR3M	3-month Norwegian Inter Bank Offered Rate
V8	NIBOR5Y	Norwegian 5-year Swap Rate
V9	EUR3M	3-month Euro Interbank Offered Rate
V10	EUR5Y	5-year German Government Rate
V11	USD3M	3-month US Libor Rate
V12	USD5Y	5-year US Government Rate
V13	NoOblig	Norwegian bond index (BRIX)
V14	IntOblig	Citigroup World Government Bond Index (WGBI)
V15	Anleggsoblig	Norwegian 6-year Swap Rate
V16	Pengemarked	ST2X - Government Bond Index (fix modified duration of 0.5 years)
V17	IntAksjer	Morgan Stanley World Index (MSCI)
V18	NoAksjer	OSEBX - Oslo Stock Exchange main index
V19	Eiendom	Oslo Stock Exchange Real Estate Index
V20	Hedgefond	S&P Hedge Fund Index

Table 4. Variables of the Norwegian financial data set.

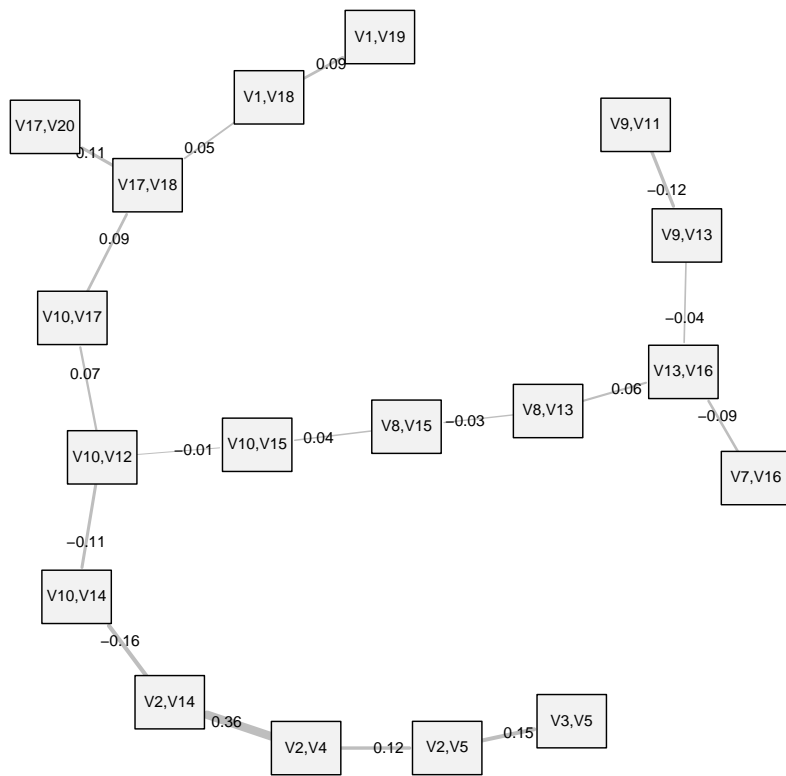
As far as simplification is concerned, sRV(6) seems to be slightly better than sRV(2) in terms of BIC. However, the hypothesis that the fully specified model and the simplified model are equivalent is not rejected for any of the models. Based on this, and also on a more thorough study of sRV(2) in (Brechmann, 2010, Section 11.2.2) we conclude that all important asymmetric dependencies as well as tail dependencies seems to be captured in the first two trees. Hence, simplification at level 2 seems appropriate.

The  $d$ -dimensional t-copula with one common degrees of freedom parameter is currently the state-of-the-art approach for modeling financial return data. A number of papers, such as Mashal and Zeevi (2002), have shown that the fit of this copula is generally superior to that of other  $d$ -dimensional copulae for such data. Hence, we wanted also to compare the truncated and simplified R-vines to this structure. The parameters of the t-copula were estimated in two steps. First, the correlation matrix of the t-copula was determined by inversion of bivariate Kendall's  $\tau$ -values and then the degrees of freedom parameter was found using maximum likelihood estimation. As indicated by the results in Table 6, the t-copula is statistically equivalent or even inferior to all truncated and simplified R-vine models, in particular when taking the number of parameters into account. The latter is due to the fact that the t-copula requires the specification of the full correlation matrix, while the R-vine specifications might be reduced with bivariate independence tests. Further, it can be noticed that the likelihood of the full R-vine model is larger than that of the t-copula, meaning that it is important to model asymmetric dependencies (see Table 5).

Finally, computing times relative to the full R-vine model are shown in Table 7. If we first turn to the sequential estimation, we see that the AIC/BIC based procedures confirm their naming as "quick and dirty". The procedures based on the Vuong test require more time, since the comparisons are more complex and require the estimation of additional trees in the case of simplification. Using the full maximum likelihood estimation, however, the truncated and simplified R-vines can be fitted much faster than a full R-vine. Moreover, simulation from the truncated/simplified models is of course more computationally efficient than from the full R-vine model.



$T_1$



$T_2$

Figure 6. First two trees of the full R-vine model for the Norwegian financial data set. The edge labels indicate empirical Kendall's  $\tau$ 's between the respective variables.

variable pair	copula type	copula parameters 1 <sup>st</sup> comp.	2 <sup>nd</sup> comp.	Kendall's $\tau$	lower tail dep.	upper tail dep.
Tree 1						
V13-V16	BB7	1.28	0.43	0.27	0.20	0.28
V7-V16	t	-0.33	8.02	-0.21	0.00	0.00
V9-V13	t	-0.24	12.71	-0.15	0.00	0.00
V8-V13	t	-0.84	5.01	-0.63	0.00	0.00
V8-V15	t	0.96	3.04	0.82	0.79	0.79
V10-V15	t	0.64	7.13	0.44	0.22	0.22
V10-V12	t	0.53	6.25	0.36	0.18	0.18
V10-V17	t	0.25	12.30	0.16	0.01	0.01
V17-V18	BB1	0.47	1.22	0.34	0.30	0.23
V17-V20	BB1	0.98	1.61	0.58	0.64	0.46
V10-V14	t	-0.58	6.85	-0.40	0.00	0.00
V2-V14	t	-0.51	12.01	-0.34	0.00	0.00
V1-V18	Gauss	0.66	-	0.46	0.00	0.00
V9-V11	t	0.28	6.92	0.18	0.07	0.07
V2-V4	t	0.67	9.13	0.47	0.19	0.19
V2-V5	Gauss	0.53	-	0.36	0.00	0.00
V3-V5	BB1	0.23	1.53	0.42	0.14	0.43
V1-V19	Gauss	0.29	-	0.19	0.00	0.00
Tree 2						
V7-V13 V16	Frank	-0.82	-	-0.09	0.00	0.00
V9-V16 V13	t	-0.06	16.42	-0.04	0.00	0.00
V8-V16 V13	Frank	0.57	-	0.06	0.00	0.00
V13-V15 V8	Indep.	-	-	0.00	0.00	0.00
V8-V10 V15	Indep.	-	-	0.00	0.00	0.00
V12-V15 V10	Indep.	-	-	0.00	0.00	0.00
V12-V17 V10	t	0.11	10.46	0.07	0.01	0.01
V10-V18 V17	Gauss	0.15	-	0.10	0.00	0.00
V18-V20 V17	Clayton	0.22	-	0.10	0.04	0.00
V12-V14 V10	t	-0.18	9.48	-0.12	0.00	0.00
V2-V10 V14	t	-0.27	21.52	-0.17	0.00	0.00
V1-V17 V18	Gauss	0.08	-	0.05	0.00	0.00
V11-V13 V9	Frank	-1.05	-	-0.12	0.00	0.00
V4-V14 V3	t	0.53	8.64	0.36	0.12	0.12
V4-V5 V2	t	0.20	26.48	0.13	0.00	0.00
V2-V3 V5	Gumbel	1.16	-	0.14	0.00	0.18
V18-V19 V1	Gauss	0.14	-	0.09	0.00	0.00

Table 5. Copula types and parameters for the first two trees of the full R-vine model for the Norwegian financial data set. The last columns indicate Kendall's  $\tau$ 's and tail dependence parameters as implied by the parameters.

proc. type	procedure	level	log likel	no. of param	BIC	V. stat. w.r.t full	V. stat. (Schwarz) w.r.t full	V. stat. w.r.t t copula	V. stat. (Schwarz) w.r.t t cop.
full model		-	6390.75	92	-12130.22	-	-	-1.93*	-10.22*
pair t copulas		-	6378.33	104	-12020.42	0.82*	3.61	-1.71*	-9.44*
multivariate t copula		-	6324.98	172	-11432.34	1.93*	10.22	-	-
trunc.	Vuong	6	6274.47	77	-12003.83	7.25	3.94	1.34*	-7.59*
	V.Schwarz	4	6234.05	68	-11986.72	7.97	3.65	2.39	-7.28*
	AIC/BIC	6	6274.47	77	-12003.83	7.25	3.94	1.34*	-7.59*
simpl.	Vuong	2	6350.09	84	-12105.52	3.19	0.97*	-0.75*	-10.07*
	V.Schwarz	2	6350.09	84	-12105.52	3.19	0.97*	-0.75*	-10.07*
	AIC/BIC	6	6373.80	88	-12124.63	2.46	0.41*	-1.41*	-10.02*

Table 6. R-vine specifications of the Norwegian financial data set obtained from different procedures (full maximum likelihood estimation). Test statistics indicated by “\*” imply that the considered model is indistinguishable from or superior to the full R-vine model (columns 7 and 8) or the multivariate t copula (columns 9 and 10), respectively, at the 5% level. Models considered are the truncated/simplified R-vines as well as the R-vine with only pair t copulas.

model	Sequential estimation	Full ML estimation	Seq + ML estimation	simulation	
pair t copulas	0.74	1.27	1.25	1.04	
trunc.	Vuong	1.18	0.47	0.50	0.89
	V.Schwarz	1.03	0.34	0.37	0.81
	AIC/BIC	0.63	0.47	0.48	0.90
simpl.	Vuong	2.16	0.47	0.53	0.73
	V.Schwarz	2.12	0.46	0.52	0.73
	AIC/BIC	0.72	0.58	0.59	0.92

Table 7. Computing times relative to the full R-vine model for the models identified in Table 6.

## 10 Conclusions

In this paper we have considered the problem of determining whether R-vines can be pairwise truncated or alternatively, simplified with Gaussian pair-copulas, after a certain tree. In an extensive simulation study different procedures for truncation and simplification were proposed and evaluated. The results showed that procedures based on the Vuong test performed particularly well.

We have also considered truncating or simplifying the special case of a C-vine. In this case, the remaining dependencies may be captured by a multivariate copula; the independence copula for the truncation alternative and the Gaussian copula for the simplification one. Hence, simplification/truncation levels may be determined using a multivariate copula goodness-of-fit-test. However, the simulation study showed that our procedures developed for the general R-vine overall seemed to detect the simplification/truncation levels more accurately than the multivariate goodness-of-fit-tests.

Finally, we have investigated whether it is possible to simplify or truncate the R-vine specification corresponding to a 19-dimensional data set consisting of Norwegian and international market variables. This study showed that the most important dependencies in the Norwegian data set are captured in the first 4-6 trees, meaning that the corresponding R-vine may be truncated at level 6, or even at level 4. Moreover, simplification at level 2 seemed to be appropriate, indicating that all important asymmetric dependencies as well as tail dependencies are captured in the first two trees.

To summarize, the methods discussed in this paper allow to efficiently construct R-vine models even in higher dimensions and under time or resource restrictions. As such, R-vine models constitute a flexible and powerful class of high-dimensional dependency models, available for a wide range of applications.

# Acknowledgements

We like to thank Arnaldo Frigessi and Ingrid Hobæk Haff for contributing valuable ideas and information. The work conducted by Kjersti Aas is sponsored by Statistics for Innovation, (sfi)<sup>2</sup>. Claudia Czado is supported by the DFG (German Research Foundation) grant CZ 86/1-3.



# References

- Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance, Mathematics and Economics*, 44:182–198.
- Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics*, 24:1–13.
- Bedford, T. and Cooke, R. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32:245–268.
- Bedford, T. and Cooke, R. (2002). Vines - a new graphical model for dependent random variables. *Annals of Statistics*, 30(4):1031–1068.
- Berg, D. (2009). Copula goodness-of-fit testing: an overview and power comparison. *The European Journal of Finance*, 15:675–701.
- Berg, D. and Aas, K. (2009). Models for construction of higher-dimensional dependence: A comparison study. *European Journal of Finance*, 15:639–659.
- Brechmann, E. C. (2010). Truncated and simplified regular vines and their applications. Diploma Thesis, Faculty of Mathematics, Technische Universität München, Germany.
- Breymann, W., Dias, A., and Embrechts, P. (2003). Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance*, 1:1–14.
- Chollete, L., Heinen, A., and Valdesogo, A. (2009). Modeling international financial returns with a multivariate regime switching copula. *Journal of Financial Econometrics*, 7:437–480.
- Czado, C. (2010). Pair-copula constructions of multivariate copulas. In Jaworski, P., Durante, F., Härdle, W., and Rychlik, T., editors, *Copula Theory and Its Applications*. Springer, Berlin.
- Dißmann, J. (2010). Selecting regular vines and their applications. Diploma Thesis, Faculty of Mathematics, Technische Universität München, Germany.
- Genest, C. and Rémillard, B. (2004). Tests for independence and randomness based on the empirical copula process. *Test*, 13(2):335–369.
- Genest, C. and Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'Institut Henri Poincaré: Probabilités et Statistiques*, 44:1096–1127.

- Genest, C., Rémillard, B., and Beaudoin, D. (2009). Omnibus goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44:199–213.
- Heinen, A. and Valdesogo, A. (2009). Asymmetric CAPM dependence for large dimensions: the canonical vine autoregressive model. CORE discussion papers 2009069, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE).
- Hobæk Haff, I. (2010). Estimating the parameters of a pair-copula construction. Submitted.
- Joe, H. (1996). 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, editor, *Distributions with Fixed Marginals and Related Topics*.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.
- Joe, H., Li, H., and Nikoloulopoulos, A. (2010). Tail dependence functions and vine copulas. *Journal of Multivariate Analysis*, 101(1):252–270.
- Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.
- Kurowicka, D. (2009). Optimal truncation of vines. Preprint.
- Kurowicka, D. (2010). Optimal truncation of vines. In Kurowicka, D. and Joe, H., editors, *DEPENDENCE MODELING: Handbook on Vine Copulae*. World Scientific Publishing Co.
- Kurowicka, D. and Cooke, R. (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*. Wiley, Chichester.
- Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics*, 80:220–239.
- Mashal, R. and Zeevi, A. (2002). Beyond correlation: Extreme co-movements between financial assets. Technical report, Columbia University.
- Min, A. and Czado, C. (2009). Bayesian model selection for multivariate copulas using pair-copula constructions. In revision for the Canadian Journal of Statistics.
- Min, A. and Czado, C. (2010). Bayesian inference for multivariate copulas using pair-copula constructions. Accepted for publication in *Journal of Financial Econometrics*.

- Morales-Napoles, O., Cooke, R., and Kurowicka, D. (2010). About the number of vines and regular vines on  $n$  nodes. Submitted to *Linear Algebra and its Applications*.
- Nelsen, R. (2006). *An Introduction to Copulas*. Springer, New York, 2nd edition.
- Prim, R. C. (1957). Shortest connection networks and some generalizations. *Bell System Technical Journal*, 36:1389–1401.
- Quessy, J.-F. (2009). Theoretical efficiency comparisons of independence tests based on multivariate versions of Spearman's rho. *Metrika*, 70:315–338.
- Ripley, B. D. (2008). *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge.
- Robles, A. V. (2009). *Multivariate volatility models using copulas*. PhD thesis, Center for Operations Research and Econometrics, Université Catholique de Louvain, Belgium.
- Schirmacher, D. and Schirmacher, E. (2008). Multivariate dependence modeling using pair-copulas. Technical report, Society of Actuaries: 2008 Enterprise Risk Management Symposium, April 14-16, Chicago. Available from: <http://www.soa.org/library/monographs/other-monographs/2008/april/2008-erm-toc.aspx>.
- Simon, G. (1977). A nonparametric test of total independence based on Kendall's tau. *Biometrika*, 64:277–282.
- Smith, M., Min, A., Czado, C., and Almeida, C. (2010). Modeling longitudinal data using a pair-copula decomposition of serial dependence. In revision for the *Journal of the American Statistical Association*.
- Varin, C., Reid, N., and Firth, D. (2010). An overview of composite likelihood methods. To appear in *Statistica Sinica*.
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57:307–333.