Truncated regular vines in high dimensions with application to financial data

This is a preprint of an article published in the Canadian Journal of Statistics Vol. 40, No. 1, 2012, Pages 68–85, http://www.interscience.wiley.com/

E. C. Brechmann, C. Czado and K. Aas

Corresponding author: E. C. Brechmann

Address: Center for Mathematical Sciences
Technische Universität München
Boltzmannstr. 3
D-85747 Garching, Germany

Telephone: +49 89 28917439

E-mail: brechmann@ma.tum.de

Abstract

Using only bivariate copulas as building blocks, regular vine copulas 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 copula. As a special case, we consider the simplification of a canonical vine copula using a multivariate copula as previously treated by Heinen and Valdesogo (2009) and Valdesogo (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 canonical vines, simplified vines

1 Introduction

A copula is a multivariate distribution with standard uniform marginal distributions. While the literature on copulas is substantial, most of the research is still limited to the bivariate case. However, recently hierarchical copula-based structures have been proposed as an alternative to the standard copula methodology. 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 (2010, 2011), Czado et al. (2011), and Smith et al. (2010).

Pair-copula constructions are also called regular vine (R-vine) copulas. They are hierarchical in nature, the various levels (also called trees) corresponding to the incorporation of more variables in the conditioning sets, using bivariate copulas as simple building blocks, the so-called pair-copulas. Until now, the concentration has been on two special cases of R-vine copulas; drawable vine (D-vine) and canonical vine (C-vine) copulas. However, very recently, there has been considerable progress in constructing R-vine copulas even in general, using graph theoretic algorithms (Dißmann et al. 2011).

The growing interest for pair-copula constructions/R-vine copulas is probably due to their high flexibility, which makes them able to 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 R-vine copulas 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 an R-vine copula can be either *truncated* or *simplified*. By a pairwisely truncated R-vine copula at level K, we mean an R-vine copula where all pair-copulas with conditioning set equal to or larger than K are replaced by independence copulas. The subject of optimal truncation of vine copulas has previously been treated by Kurowicka (2011), who constructs R-vine copulas bottom-up beginning with the highest level and iteratively moving to the first level. Her approach is however based on Pearson product-moment correlations and therefore does not reflect non-elliptical dependence adequately. The approach suggested here is very different. It sequentially proceeds top-down and does not rely on the assumption of elliptical dependence. Additionally, our approach allows to identify independence of variables based on a statistical test in contrast to the ad-hoc procedure of Kurowicka (2011).

An R-vine copula is defined to be pairwisely simplified at level K if all pair-copulas with conditioning set equal to or larger than K instead are replaced by Gaussian copulas. We advocate using Gaussian copulas for the following reasons. They mean a simplification since they are easy to specify and faster to estimate than, e.g., t copulas. Moreover, they are easy to interpret in terms of the correlation parameter. Most common Archimedean copulas such as the Clayton or the Gumbel, on the other hand, have asymmetric tail dependence and therefore not suitable for simplification, since such asymmetry for a large number of pair-copulas is a very strict assumption. Thus the choice of the Gaussian copula as "neutral" copula is reasonable. It will be shown in our 19-dimensional application how specification and simulation times can be significantly improved using simplification with Gaussian copulas.

To identify the most appropriate truncation/simplification level, we use a heuristic procedure based on 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 copula specification corresponding to a 19-dimensional data set consisting of Norwegian and international market variables.

For the special case of a C-vine copula, 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.*

It should be noted that joint simplification of C-vine copulas previously has been treated by Heinen and Valdesogo (2009) and Valdesogo (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 Valdesogo (2009)). Pairwisely simplified R-vine copulas and the identification of truncation/simplification levels are not considered in these papers.

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

2 Multivariate copulas and regular vines

Consider a vector $\mathbf{X} = (X_1, ..., X_d)$ of random variables with a joint density function f. Sklar's theorem (Sklar 1959) states that every multivariate distribution F with marginals $F_1, ..., F_d$ can be written as

$$F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d)), \tag{2.1}$$

for some appropriate d-dimensional copula C. Using the chain rule, we further have for an absolutely continuous F with strictly increasing continuous marginals $F_1, ..., F_d$ that

$$f(x_1, ..., x_d) = \left[\prod_{k=1}^{d} f_k(x_k)\right] \times c(F_1(x_1), ..., F_d(x_d)).$$

Here $c(\cdot)$ denotes the copula density. More details on copula theory can be found in the books by Joe (1997) and Nelsen (2006).

In higher-dimensional practical applications the choice of adequate copulas is limited. Multivariate copulas such as elliptical or exchangeable Archimedean are rather restricted and often not appropriate for dependence modeling. Hence, there is a growing need for more flexible copulas.

The notion of a regular vine distribution was introduced by Bedford and Cooke (2001, 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 bivariate copula, a so-called *pair-copula*. These pair-copulas then constitute the building blocks of the joint regular vine distribution. According to Definition 4.4 of Kurowicka and Cooke (2006) a *regular vine* (R-vine) $\mathcal{V}$ on d variables consists of trees $T_1, ..., T_{d-1}$ with nodes N_i and edges E_i for $i = 1, ..., d-1$, which satisfy the following:

1. T_1 has nodes $N_1 = \{1, ..., d\}$ and edges E_1.
2. For $i = 2, ..., d-1$ 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.

To build up a statistical model on R-vine trees with the node set $\mathcal{N} := \{N_1, ..., N_{d-1}\}$ and the edge set $\mathcal{E} := \{E_1, ..., E_{d-1}\}$, 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 called the *conditioned nodes*, while $D(e)$ is the *conditioning set*. An *R-vine distribution* is defined as the distribution of the random vector $\boldsymbol{X}$ with conditional copula density of $(X_{j(e)}, X_{k(e)})$ given the variables $\boldsymbol{X}_{D(e)}$ specified as $c_{j(e),k(e)|D(e)}$[1] for the R-vine trees with node set $\mathcal{N}$ and edge set $\mathcal{E}$. $\boldsymbol{X}_{D(e)}$ denotes the subvector of $\boldsymbol{X}$ determined by the indices in $D(e)$. In Theorem 4.2 of Kurowicka and Cooke (2006) it is proven that the joint density of $\boldsymbol{X}$ is uniquely determined and given by

$$f(x_1, ..., x_d) = \left[\prod_{k=1}^{d} f_k(x_k)\right] \times \left[\prod_{i=1}^{d-1} \prod_{e \in E_i} c_{j(e),k(e)|D(e)}(F(x_{j(e)}|\boldsymbol{x}_{D(e)}), F(x_{k(e)}|\boldsymbol{x}_{D(e)}))\right], \quad (2.2)$$

where $\boldsymbol{x}_{D(e)}$ denotes the subvector of $\boldsymbol{x} = (x_1, ..., x_d)'$ determined by the indices in $D(e)$. The rightmost part of Equation (2.2), which involves $d(d-1)/2$ bivariate copula densities, is called an *R-vine copula*.

An example of a seven-dimensional R-vine tree specification together with its edge indices is given in the left panel of Figure 1. This tree specification was found for a subset of the financial data in Section 7[2] according to the selection criteria discussed at the end of this section.

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, and
- 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 $i, j, i_1, ..., i_k$ with $i < j$ and $i_1 < ... < i_k$ we use the abbreviation

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

Using this notation the D-vine density is given by

$$f(x_1, ..., x_d) = \left[\prod_{k=1}^{d} f_k(x_k)\right] \times \left[\prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,i+j|i+1,...,i+j-1}\right], \quad (2.3)$$

and the C-vine density by

$$f(x_1, ..., x_d) = \left[\prod_{k=1}^{d} f_k(x_k)\right] \times \left[\prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{j,j+i|1,...,j-1}\right]. \quad (2.4)$$

See Aas et al. (2009) for more on simulation, inference and applications of *C-* and *D-vine copulas*. A five-dimensional C-vine tree specification is shown in the right panel of Figure 1.

[1] As in Aas et al. (2009) it is assumed here that this conditional distribution is independent of the conditioning variable $\boldsymbol{X}_{D(e)}$. Hobæk Haff et al. (2010) call this the *simplified* PCC which must not be confused with simplification discussed in the following.

[2] The considered variables and their corresponding numbers in the left panel of Figure 1 are: V17 (1), V20 (2), V18 (3), V1 (4), V10 (5), V14 (6) and V15 (7); cp. Table 1.

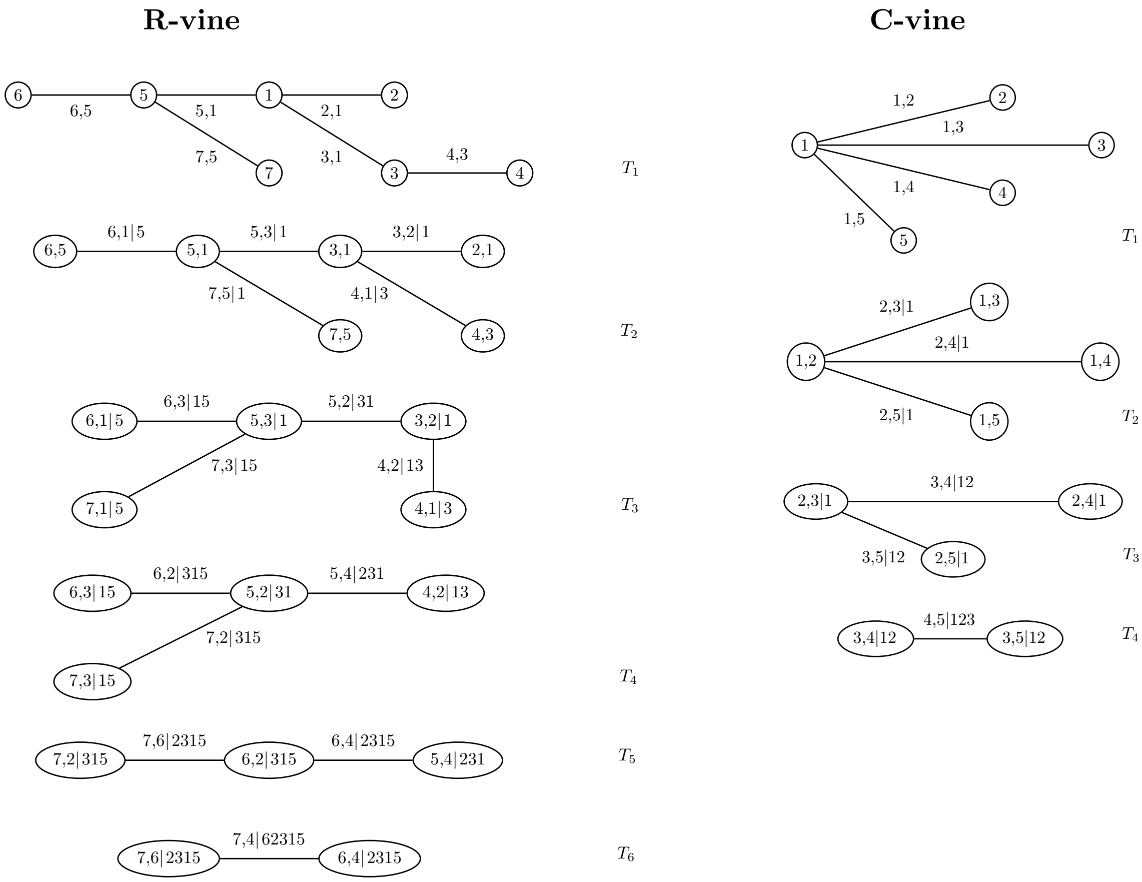

Figure 1: An R-vine tree specification on seven variables (left panel) and a C-vine tree specification on five variables (right panel) with edge indices.

For R-vines in general, there are no expressions like (2.3) and (2.4). Hence, an efficient way of storing the indices of the pair-copulas required in the joint density expression (2.2) is needed. One such approach was recently proposed by Morales-Napoles (2010) and explored in more detail in Dißmann et al. (2011). It involves the specification of a lower triangular matrix $M = (m_{i,j}|i,j = 1,...,d) \in \{0,...,d\}^{d\times d}$ with $m_{i,i} = d - i + 1$. That is, the diagonal entries of M are the numbers $1,...,d$ 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

$$M = \begin{pmatrix} 7 & & & & & & \\ 4 & 6 & & & & & \\ 6 & 4 & 5 & & & & \\ 2 & 2 & 4 & 4 & & & \\ 3 & 3 & 2 & 2 & 3 & & \\ 1 & 1 & 3 & 1 & 2 & 2 & \\ 5 & 5 & 1 & 3 & 1 & 1 & 1 \end{pmatrix}, \tag{2.5}$$

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 (7,5), (6,5), (5,1) 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 (7,1|5), (6,1|5), etc. Proceeding like this, 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, i.e., (7,4|62315).

This matrix specification of R-vines at the same time directly allows for the derivation of the pair-copula decomposition (cp. Aas et al. (2009)) of the corresponding R-vine distribution. Let $M = (m_{i,j}|i, j = 1, ..., d)$ be an R-vine matrix corresponding to the R-vine $\mathcal{V}$. Then, according to Dißmann et al. (2011), the R-vine density is:

$$f(x_1, ..., x_d) = \left[\prod_{k=1}^{d} f_k(x_k)\right] \times \left[\prod_{j=d-1}^{1} \prod_{i=d}^{j+1} c_{m_{j,j},m_{i,j}|m_{i+1,j},...,m_{n,j}}\right], \tag{2.6}$$

where the pair-copulas have arguments $F(x_{m_{j,j}}|x_{m_{i+1,j}}, ..., x_{m_{n,j}})$ and $F(x_{m_{i,j}}|x_{m_{i+1,j}}, ..., x_{m_{n,j}})$.

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 τ or tail dependence. For this graph, we then find a maximum spanning tree (using the well-known algorithm of Prim (1957)), which is a tree on all nodes that maximizes the pairwise dependencies. Given this tree, we can now select pair-copulas, estimate parameters and compute transformed observations $F(x_{j(e)}|\boldsymbol{x}_{D(e)})$ for the next level, which in general are given by

$$F(x|\boldsymbol{v}) = \frac{\partial C_{xv_j|\boldsymbol{v}_{-j}}(F(x|\boldsymbol{v}_{-j}), F(v_j|\boldsymbol{v}_{-j}))}{\partial F(v_j|\boldsymbol{v}_{-j})}. \tag{2.7}$$

Here $C_{xv_j|\boldsymbol{v}_{-j}}$ is a bivariate copula, v_j is an arbitrary component of $\boldsymbol{v}$ and $\boldsymbol{v}_{-j}$ denotes the vector $\boldsymbol{v}$ excluding v_j.

At the second level, we repeat the first level procedure, and iterate until all trees are constructed and their pair-copulas sequentially estimated. See Dißmann et al. (2011) for more details, and Brechmann (2010, Section 3.2) for construction methods for the special cases of C- and D-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.

Unfortunately vine copulas estimated in this way are not robust against misspecification of the pair-copulas, as also noted by Hobæk Haff (2010). To the best of our knowledge reliable alternatives are however not yet available. Goodness-of-fit tests for each pair-copula term may help reduce this uncertainty. These tests are however computationally not feasible in higher dimensions, since the number of pair-copulas grows quadratically with the dimension. Moreover, a large scale simulation study in Brechmann (2010, Section 5.4) showed that copula selection using the AIC is more reliable than using goodness-of-fit tests.

3 Pairwise simplification and truncation

For R-vine copulas 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 copula increases with the dimension. Hence, we take a pragmatic approach. We do not attempt to find the best fitting R-vine copula, 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 copula, 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 copula 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 copula 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 copula becomes a Markov tree distribution, 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 a pairwisely truncated R-vine copula at level K by $\mathrm{tRV}(K)$ and pairwisely simplified K level ones by $\mathrm{sRV}(K)$. Further, let $\boldsymbol{\theta}_{\mathrm{tRV}(K)}$ be the pair-copula parameters of the truncated R-vine copula, i.e., $\boldsymbol{\theta}_{\mathrm{tRV}(K)} = \{\boldsymbol{\theta}_{j(e),k(e)|D(e)} :\ e \in E_i,\ i = 1, ..., K\}$, where $\boldsymbol{\theta}_{j(e),k(e)|D(e)}$ denotes the parameter(s) of the copula density $c_{j(e),k(e)|D(e)}$. Then, the density of a truncated R-vine copula at level K is given by

$$c_{\mathrm{tRV}(K)}(\boldsymbol{u}|\boldsymbol{\theta}_{\mathrm{tRV}(K)}) = \prod_{i=1}^{K}\prod_{e\in E_i} c_{j(e),k(e)|D(e)}(F(u_{j(e)}|\boldsymbol{u}_{D(e)}), F(u_{k(e)}|\boldsymbol{u}_{D(e)})), \tag{3.1}$$

where $\boldsymbol{u} = (u_1, ..., u_d)' \in [0, 1]^d$.

The density in (3.1) may be interpreted as a composite likelihood (see Lindsay (1988) and Varin et al. (2010)). This can be shown as follows. The joint density in (2.6) may also be decomposed as

$$f(x_1, ..., x_d) = f_{m_{d,d}}(x_{m_{d,d}}) \times \prod_{i=2}^{d} f(x_{m_{d-i+1,d-i+1}}|x_{m_{d-i,d-i+1}}, ..., x_{m_{d,d-i+1}}), \tag{3.2}$$

where the order of the conditioning variables is fixed, and given by the column entries of the R-vine matrix M. If the R-vine copula is truncated at level K, the density in (3.2) reduces to

$$f(x_1, ..., x_d) = f_{m_{d,d}}(x_{m_{d,d}}) \times \prod_{i=2}^{d} f(x_{m_{d-i+1,d-i+1}}|x_{m_{d-K+1,d-i+1}}, ..., x_{m_{d,d-i+1}}), \tag{3.3}$$

since the pair-copulas in trees $T_{K+1}, ..., T_{d-1}$ are set to independence copulas. Such an approximate likelihood, where the conditioning set is a subset of the variables $X_{m_{d-K+1,d-i+1}}, ..., X_{m_{d,d-i+1}}$, has been proposed by Vecchia (1988) in the context of spatial models. For example, for $K = 1$ we obtain a Markov structure of order 1. If we set all margins and all pair-copulas to be Gaussian, (3.3) corresponds to the model of Vecchia (1988). In contrast to general composite likelihoods, the expression in (3.3) however does not require the choice of any weights and is in fact a valid probability density. The increasing conditional dependence in higher order trees directly leads to compatibility with composite likelihood methods. From their theory, we hence directly obtain consistency of the composite maximum likelihood estimates $\hat{\boldsymbol{\theta}}_{\mathrm{tRV}(K)}$ based on the density in (3.1).

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

$$c_{\mathrm{sRV}(K)}(\boldsymbol{u}|\boldsymbol{\theta}_{\mathrm{sRV}(K)}) = \left[\prod_{i=1}^{K}\prod_{e\in E_i} c_{j(e),k(e)|D(e)}\right] \times \left[\prod_{i=K+1}^{d-1}\prod_{e\in E_i} c^{\rho}_{j(e),k(e)|D(e)}\right], \tag{3.4}$$

where, $c^{\rho}_{j(e),k(e)|D(e)}$ denote Gaussian pair-copula densities with correlation parameter $\rho_{j(e),k(e)|D(e)}$, and the arguments of the copula densities have been omitted for simplicity. Further, $\boldsymbol{\theta}_{\mathrm{sRV}(K)}$ is the parameter set of $c_{\mathrm{sRV}(K)}$, i.e.,

$$\begin{aligned}\boldsymbol{\theta}_{\mathrm{sRV}(K)} &= \{\boldsymbol{\theta}_{j(e),k(e)|D(e)} :\ e \in E_i,\ i = 1, ..., K\}\\ &\quad \cup \{\rho_{j(e),k(e)|D(e)} :\ e \in E_i,\ i = K+1, ..., d-1\},\end{aligned} \tag{3.5}$$

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 5.1 and 5.2 we will develop heuristic 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 copula, 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 copula, all pair-copulas with a conditioning set larger than or equal to K dimensions can be modeled jointly by a $(d-K)$-dimensional copula as shown in Valdesogo (2009). 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 copula, while in the Gaussian case, the resulting C-vine copula is denoted as *jointly simplified*. In the following, jointly simplified K level C-vine copulas will be denoted by jsCV(K). Simplification of C-vine copulas has previously been treated by Heinen and Valdesogo (2009) and Valdesogo (2009) who refer to it as "truncation". Truncation in our meaning of the word is however not discussed in these papers.

For C-vine copulas the second component of the product in (3.4) reduces to a $(d-K)$-dimensional Gaussian copula. Hence, we obtain the density of a jointly simplified K level C-vine copula by rewriting (3.4) to

$$c_{\mathrm{jsCV}(K)}(\boldsymbol{u}|\boldsymbol{\theta}_{\mathrm{jsCV}(K)}) = \left[\prod_{j=1}^{K}\prod_{i=1}^{d-j} c_{j,j+i|1,...,j-1}\right] \times c^{\rho}_{K+1,...,d|1,...,K}(\cdot),$$

where $c^{\rho}_{(K+1),...,d|1,...,K}(\cdot)$ is a $(d-K)$-dimensional Gaussian copula density with arguments $F(x_{K+1}|x_1, ..., x_K)$, ..., $F(x_d|x_1, ..., x_d)$. The parameter set $\boldsymbol{\theta}_{\mathrm{jsCV}(K)}$ is defined similarly to (3.5) as

$$\boldsymbol{\theta}_{\mathrm{jsCV}(K)} = \{\boldsymbol{\theta}_{j,j+i|1,...,j-1} :\ j = 1, ..., K,\ i = 1, ..., d-j\} \cup \{\rho_{ij|1,...,K} :\ i, j = K+1, ..., d,\ i \neq j\},$$

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

Finally, note that for D-vine copulas, joint simplification as described above is not possible. The reason is that while C-vine copulas have a common conditioning set in each tree as shown in (2.4), this is not the case for the D-vine copula (see (2.3)). For instance, in a five-dimensional D-vine copula, the arguments to the pair-copula densities 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 and simplification levels

5.1 Selection of truncation level in the general case

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

$$l_{\mathrm{tRV}(K)}\left(\boldsymbol{\theta}_{\mathrm{tRV}(K)}|\boldsymbol{u}\right) = \sum_{i=1}^{n}\sum_{\ell=1}^{K}\sum_{e\in E_\ell} \ln\left[c_{j(e),k(e)|D(e)}\left(F(u_{i,j(e)}|\boldsymbol{u}_{i,D(e)}), F(u_{i,k(e)}|\boldsymbol{u}_{i,D(e)})|\boldsymbol{\theta}_{j(e),k(e)|D(e)}\right)\right], \tag{5.1}$$

where n is the number of data points $\boldsymbol{u}_i = (u_{i,1},...,u_{i,d})' \in [0,1]^d,\ i = 1,...,n$. From now on we assume that data has been transformed to the unit hypercube using the respective marginal distribution functions of the variables. In practice this is either done parametrically by selecting (and estimating) appropriate marginal distributions, or non-parametrically by using the empirical distribution functions. Here, we assume that the latter is the case, because it eliminates the risk of misspecification (and possible influences on the selection of the truncation or simplification level). For more details on these issues see Genest et al. (1995), Joe (1997) and Kim et al. (2007). However estimating the true margins by empirical ones alters the asymptotic distribution of the maximum likelihood (ML) estimates of the dependence parameters and hence influences the estimation of standard errors and critical values for hypothesis tests.

If there are sufficient computing resources, we can maximize the full log likelihood (5.1). Alternatively, one may use the stepwise/sequential ML-estimator originally proposed by Aas et al. (2009) and further explored by Hobæk Haff (2010) and Hobæk Haff (2011). Based on an extensive simulation study, the latter paper shows that the performance of the stepwise estimator is satisfactory compared to the full log likelihood method. To determine the sequential parameter estimates of $\boldsymbol{\theta}_{\mathrm{tRV}(K)}$, it is enough to only use transformed variables (2.7) up to tree K. These sequential estimates can then be used as starting values for maximizing $l_{\mathrm{tRV}(K)}(\boldsymbol{\theta}_{\mathrm{tRV}(K)}|\boldsymbol{u})$. For K small, the number of parameters to be maximized over is considerably reduced compared to a full R-vine copula specification.

We will start with $K = 1$ and fit a truncated R-vine copula (for $K = 0$ a pre-test of joint independence 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 tRV(K) to 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 tRV(K) is nested within 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

$$\begin{aligned}
AIC(\mathrm{tRV}(K)) &:= -2 \times l_{\mathrm{tRV}(K)}(\hat{\boldsymbol{\theta}}_{\mathrm{tRV}(K)}|\boldsymbol{u}) + 2 \times n_{\mathrm{tRV}(K)} \\
BIC(\mathrm{tRV}(K)) &:= -2 \times l_{\mathrm{tRV}(K)}(\hat{\boldsymbol{\theta}}_{\mathrm{tRV}(K)}|\boldsymbol{u}) + \ln(n) \times n_{\mathrm{tRV}(K)},
\end{aligned}$$

where $n_{\mathrm{tRV}(K)}$ denotes the dimension of $\boldsymbol{\theta}_{\mathrm{tRV}(K)}$. We choose the one of $\mathrm{tRV}(K)$ and $\mathrm{tRV}(K+1)$ with the smaller AIC or BIC value. If for some K_0 the smaller model is chosen, we stop, and declare $\mathrm{tRV}(K_0)$ as the best fitting model among the model sequence $\mathrm{tRV}(j),\ j = 0, ..., d-1$.

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

$$\nu := \frac{\frac{1}{n}\sum_{i=1}^{n} m_i}{\sqrt{\sum_{i=1}^{n}(m_i - \bar{m})^2}} > \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \tag{5.2}$$

where Φ^{-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.

When dealing with truncated R-vines, models are nested and a classical likelihood-ratio test could be used to compare $\mathrm{tRV}(K)$ and $\mathrm{tRV}(K+1)$. As we want to allow for misspecification of the models, the asymptotic distribution of the test is however hardly tractable (see Vuong (1989)). Moreover, a likelihood-ratio test cannot be corrected for the number of model parameters as conveniently as the Vuong test.

We therefore heuristically apply the Vuong test to compare $\mathrm{tRV}(K)$ (model f_1) and $\mathrm{tRV}(K+1)$ (model f_2). If $\nu \geq -\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$, we stop with $\mathrm{tRV}(K)$, since $\mathrm{tRV}(K)$ is preferred to, or indistinguishable from $\mathrm{tRV}(K+1)$, at level α. It thus determines the truncation level as the level K_0 for which $\mathrm{tRV}(K_0+1)$ does not provide a significant gain in the model fit. In the light of the regularity conditions of Vuong (1989), it is important to note that the log likelihood (5.1) is in fact a valid log likelihood under certain conditional independence conditions and that the sequential estimates are consistent and asymptotically normal as shown by Hobæk Haff (2010).

Algorithm 1 describes the truncation procedure based on the Vuong test. Truncation using information criteria proceeds in the same way, where we need to compute only the contribution from tree T_{j+1} to the AIC of $\mathrm{tRV}(K+1)$. This is due to the fact that the AICs of $\mathrm{tRV}(K)$ and $\mathrm{tRV}(K+1)$ are equal with the exception of the contribution from tree T_{j+1}. Since all copulas in tree T_{j+1} of $\mathrm{tRV}(K)$ are independence copulas, the contribution from tree T_{j+1} to the AIC of $\mathrm{tRV}(K)$ is zero. Hence, if the contribution from tree T_{j+1} to the AIC of $\mathrm{tRV}(K+1)$ 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.

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

Input: Observations of d variables, significance level α.

for $j = 0, ..., d-2$ **do**
  Specify tRV($j+1$) by additionally constructing tree T_{j+1} with appropriate pair-copulas.
  Perform a Vuong test for tRV(j) (model f_1) and tRV($j+1$) (model f_2), i.e., determine test statistic ν as in (5.2), possibly with Akaike or Schwarz correction.
  if $\nu \geq -\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$ **then**
    Truncate the R-vine copula at level $K = j$, i.e., exit the loop with tRV(j).
  end if
end for

Output: Pairwisely truncated K level R-vine copula, or fully specified R-vine copula, if no truncation is possible.

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

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

Figure 2: Pair-copula density terms of five-dimensional D-vine copulas truncated after the second tree T_2 (smaller model) and after the third tree T_3 (larger model), respectively, where $\pi_{ij|D}$ denote densities of independence copulas.

Example 1 (Pairwise truncation of R-vine copulas.)**.** *We consider a five-dimensional D-vine copula. 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 copula 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 2. Note that this is not an exact model comparison between a truncated D-vine copula 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.*

5.2 Selection of simplification level in the general case

Selection of simplification levels, i.e., model selection between sRV(K) and sRV($K+1$), proceeds in essentially the same way as for truncation. However, on the contrary to tRV(K) and tRV($K+1$), the models sRV(K) and sRV($K+1$) are not nested, in general $\boldsymbol{\theta}_{\mathrm{sRV}(K)} \not\subset \boldsymbol{\theta}_{\mathrm{sRV}(K+1)}$. Non-nested models may be compared using the Vuong test under the assumption that models are not equal[3]. 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

[3]A pre-test for overlapping (partially nested) models as outlined in Vuong (1989) is not performed, since it is numerically not feasible and the assumption of unequal models is sensible here.

Algorithm 2 Simplification of R-vine copulas based on the Vuong test.

Input: Observations of d variables, significance level α.

1: **for** $j = 0, ..., d-2$ **do**

2: Specify sRV(j) by constructing higher order trees $T_{j+1}, ..., T_{d-1}$ with bivariate Gaussian copulas.

3: Specify sRV($j+1$) by additionally constructing tree T_{j+1} with appropriate pair-copulas, and by constructing higher order trees $T_{j+2}, ..., T_{d-1}$ with bivariate Gaussian copulas.

4: Perform a Vuong test for sRV(j) (model f_1) and sRV($j+1$) (model f_2), i.e., determine test statistic ν as in (5.2), possibly with Akaike or Schwarz correction.

5: **if** $\nu \geq -\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$ **then**

6: Simplify the R-vine copula at level $K = j$, i.e., exit the loop with sRV(j).

7: **end if**

8: **end for**

Output: Pairwisely simplified K level R-vine copula, or fully specified R-vine copula, if no simplification is possible.

T_{d-1} are equal in models sRV(K) and 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 procedure based on the Vuong test is outlined in Algorithm 2. Simplification using information criteria proceeds similarly but under the assumption that trees $T_{j+2}, ..., T_{d-1}$ 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 $C^{\rho}_{15|234}$ and $\tilde{C}^{\rho}_{15|234}$ in T_4 of the smaller and the larger model, respectively, are equal if we work with AIC and BIC.

5.3 Selection of simplification and truncation levels for C-vine copulas

The procedures described in Sections 5.1 and 5.2 may of course be used also for the special case of a C-vine copula. 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. If the p-value of an independence test is larger than a preliminarily chosen level, we truncate the structure at level K. Also for the purpose of simplification, if the p-value of an appropriate copula goodness-of-fit test for the multivariate Gaussian copula is larger than a preliminarily chosen level, we simplify the structure at level K. As previously described, this way of truncation/simplification is denoted *joint* truncation/simplification. Algorithm 3 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.

6 Simulation studies

We have evaluated the performance of the heuristic simplification and truncation procedures presented in Section 5 in extensive simulation studies. A thorough description of the results can be found in the supplementary material. To summarize the main findings, the procedures based on the Vuong test with or without correction for the number of parameters should be used in most

Algorithm 3 Joint simplification of C-vine copulas.

Input: Observations of d variables, significance level α.

1: **for** $j = 0, ..., d-2$ **do**

2: Perform a copula goodness-of-fit test for 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-value $> \alpha$ **then**

4: Simplify the C-vine copula at level $K = j$, i.e., exit the loop with jsCV(j).

5: **end if**

6: Specify tree T_{j+1} with appropriate pair-copulas.

7: **end for**

Output: Jointly simplified K level C-vine copula, or fully specified C-vine copula, if no simplification is possible.

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 identify truncation and especially simplification too late, but they are very fast compared to the Vuong test (in a 52-dimensional example, AIC identified the truncation/simplification level 44%/80% faster than the Vuong test). Parsimonious models can be obtained by using the Vuong test with Schwarz correction.

7 Application

In this section, we analyze a 19-dimensional data set consisting of Norwegian and international financial variables. See Table 1 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 copula increases with the dimension. Hence, the aim of the work presented here was to investigate whether simplification or truncation of the R-vine copula specification corresponding to this 19-dimensional data set is possible.

Before analyzing the dependence in the data set, we selected appropriate ARMA-GARCH time series models for the univariate margins (see the supplementary material for more details). 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. In the light of results due to Chen and Fan (2006), the method of maximum pseudo likelihood is consistent even when time series are fitted to the margins.

For the sake of reference, we first fit a full R-vine copula to this data set, using the approach described in Section 2. We use Kendall's τ's as edge weights, and pair-copulas are selected from a range of 11 bivariate families using AIC: independence copula, Gaussian, t, Clayton, rotated Clayton (90°), Gumbel, rotated Gumbel (90°), Frank, Joe, Clayton-Gumbel (BB1), Joe-Clayton (BB7). For more information on copula types, see, e.g., Nelsen (2006) or Joe (1997). The independence copula is chosen according to the bivariate independence test based on Kendall's τ as described in Genest and Favre (2007). If the p-value is larger than 5%, the independence copula is chosen to obtain more parsimonious models and therefore results in an additional inherent truncation.

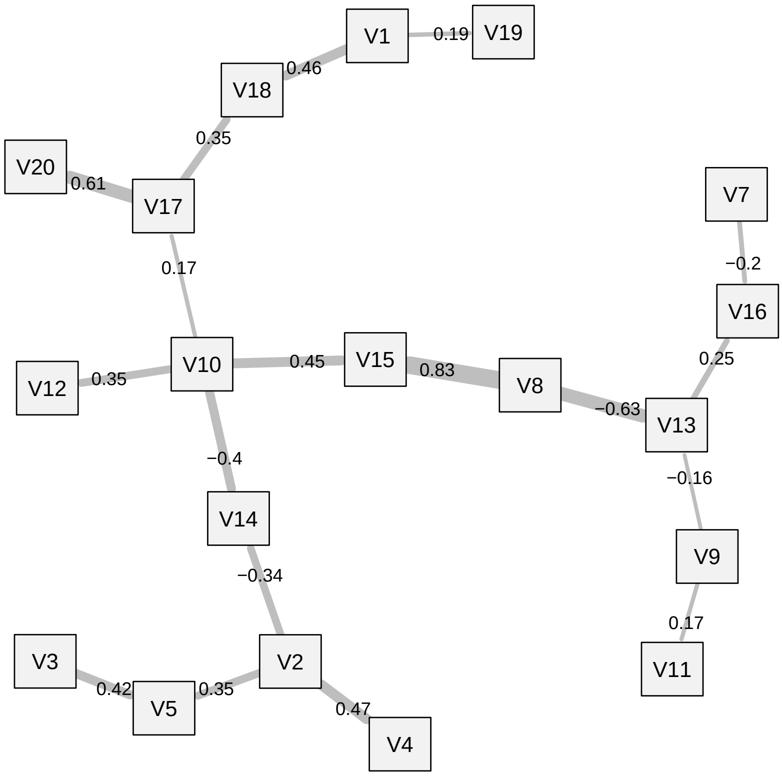

Figure 3: First tree of the full R-vine copula model for the Norwegian financial data set. The edge labels indicate empirical Kendall's τ's between the respective variables.

Figure 3 shows the first tree of the fitted R-vine copula. See Table 1 for the correspondence between the IDs and the variable descriptions. The edge labels represent the empirical Kendall's τ's between the respective variables. The corresponding R-vine matrix specifications with copula types and parameters can be found in the supplementary material. In particular, the first tree pair-copula terms identify strong to medium tail dependence and some asymmetries. Dependencies modeled in higher order trees are much weaker.

In economical terms, the tree in Figure 3 has an evident interpretation. It identifies 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). 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 (the 5-year German Government Rate and the MSCI World index), while the interest rate and exchange rate clusters are connected by V14 and V2 (the WGBI bond index and the USD-NOK exchange rate).

In addition to the R-vine copula with different copulas for different pairs, we also fitted an R-vine copula with t copulas for all pairs. The BIC-values in the two upper rows of Table 2 show however that the R-vine copula with mixed copulas is superior to the one with only t copulas.

Having fitted the full R-vine copula, we apply the different statistical model selection criteria from Sections 5.1 and 5.2 to investigate whether truncation and/or simplification of this R-vine is possible. Table 2 shows the results. We report log likelihood values, the number of

ID	description	ID	description
V1	Norwegian Financial Index	V12	5-year US Government Rate
V2	USD-NOK exchange rate	V13	Norwegian bond index (BRIX)
V3	EURO-NOK exchange rate	V14	Citigroup World Government Bond Index (WGBI)
V4	YEN-NOK exchange rate		
V5	GBP-NOK exchange rate	V15	Norwegian 6-year Swap Rate
V7	3-month Norwegian Inter Bank Offered Rate	V16	ST2X - Government Bond Index (fix modified duration of 0.5 years)
V8	Norwegian 5-year Swap Rate	V17	Morgan Stanley World Index (MSCI)
V9	3-month Euro Interbank Offered Rate	V18	OSEBX - Oslo Stock Exchange main index
V10	5-year German Government Rate	V19	Oslo Stock Exchange Real Estate Index
V11	3-month US Libor Rate	V20	S&P Hedge Fund Index

Table 1: Variables of the Norwegian financial data set.

parameters[4] 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 τ'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 copula may be truncated at level 6, or even at level 4, depending on the desired level of parsimonity (and of course at the expense of accuracy).

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 even for sRV(2). 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) tail dependencies seem 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

[4]Note that the number of parameters of a full R-vine copula 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 2 is much smaller than this, is that many of the pair-copulas in the full R-vine copula (both the one with mixed copulas and the one with only t copulas) are estimated to be independence copulas. The R-vine copula with only t copulas would otherwise have 342 parameters rather than 104!

proc. type	procedure	level	log likeli-hood	no. of param.	BIC	V. stat. w.r.t. full model	V. stat. (Schwarz) w.r.t. full model	V. stat. w.r.t. multi. t cop.	V. stat. (Schwarz) w.r.t. multi. 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 2: R-vine copula specifications of the Norwegian financial data set obtained by different procedures (full maximum likelihood estimation). Test statistics with a "*" imply that the considered model is indistinguishable from or superior to the full R-vine copula 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-vine copulas as well as the R-vine copula with only pair t copulas.

model		estimation (part I: sequential)	estimation (part II: full MLE)	estimation (part I & II)	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 3: Computing times relative to the full R-vine copula model for the models identified in Table 2.

that of other d-dimensional copulas for such data. Hence, we wanted also to compare the truncated and simplified R-vine copulas 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 τ-values, and then the degrees of freedom parameter was found using maximum likelihood estimation. As indicated by the results in Table 2, the t copula is statistically equivalent or even inferior to all truncated and simplified R-vine copula 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 even the full R-vine copula might be reduced with bivariate independence tests, and hence leads to more parsimonious copula models.

Finally, computing times relative to the full R-vine model are shown in Table 3. 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.

8 Conclusions

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

We also considered truncating or simplifying the special case of a C-vine copula. 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, simulations showed that our procedures developed for the general R-vine copula 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 copula 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 copula 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) tail dependencies are captured in the first two trees.

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

Acknowledgement

We like to thank Arnoldo Frigessi and Ingrid Hobæk Haff for contributing valuable ideas and information. We acknowledge the helpful comments of the referees, which further improved the manuscript. The work conducted by Kjersti Aas is sponsored by Statistics for Innovation, (sfi)^2. Claudia Czado is supported by the DFG (German Research Foundation) grant CZ 86/1-3. The numerical computations were performed on a Linux cluster supported by DFG grant INST 95/919-1 FUGG.

References

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

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

gence 32, 245–268.

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

Berg, D. and K. Aas (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.

Chen, X. and Y. Fan (2006). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics 135*, 125–154.

Chollete, L., A. Heinen, and A. Valdesogo (2009). Modeling international financial returns with a multivariate regime switching copula. *Journal of Financial Econometrics 7*, 437–480.

Czado, C., U. Schepsmeier, and A. Min (2011). Maximum likelihood estimation of mixed C-vines with application to exchange rates. *To appear in Statistical Modelling*.

Dißmann, J., E. C. Brechmann, C. Czado, and D. Kurowicka (2011). Selecting and estimating regular vine copulae and application to financial returns. *Submitted for publication*.

Genest, C. and A.-C. Favre (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering 12*, 347–368.

Genest, C., K. Ghoudi, and L. Rivest (1995). A semi-parametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika 82*, 543–552.

Heinen, A. and A. Valdesogo (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 for publication*.

Hobæk Haff, I. (2011). Comparison of estimators for pair-copula constructions. *To appear in the Journal of Multivariate Analysis*.

Hobæk Haff, I., K. Aas, and A. Frigessi (2010). On the simplified pair-copula construction - simply useful or too simplistic? *Journal of Multivariate Analysis 101*(5), 1296–1310.

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 (Ed.), *Distributions with Fixed Marginals and Related Topics*.

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

Kim, G., M. J. Silvapulle, and P. Silvapulle (2007). Comparison of semiparametric and parametric methods for estimating copulas. *Computational Statistics & Data Analysis 51*(6), 2836–2850.

Kurowicka, D. (2011). Optimal truncation of vines. In D. Kurowicka and H. Joe (Eds.), *Dependence Modeling: Vine Copula Handbook*. World Scientific Publishing Co.

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

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics 80*, 220–239.

Mashal, R. and A. Zeevi (2002). Beyond correlation: Extreme co-movements between financial assets. Technical report, Columbia University.

Min, A. and C. Czado (2010). Bayesian inference for multivariate copulas using pair-copula constructions. *Journal of Financial Econometrics 8*(4), 511–546.

Min, A. and C. Czado (2011). Bayesian model selection for multivariate copulas using pair-copula constructions. *To appear in the Canadian Journal of Statistics*.

Morales-Napoles, O. (2010). Counting vines. In D. Kurowicka and H. Joe (Eds.), *Dependence Modeling: Vine Copula Handbook*. World Scientific Publishing Co.

Morales-Napoles, O., R. Cooke, and D. Kurowicka (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* (2nd ed.). New York: Springer.

Prim, R. C. (1957). Shortest connection networks and some generalizations. *Bell System Technical Journal 36*, 1389–1401.

Ripley, B. D. (2008). *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press.

Schirmacher, D. and E. Schirmacher (2008). Multivariate dependence modeling using pair-copulas. Technical report, Society of Acturaries: 2008 Enterprise Risk Management Symposium, April 14-16, Chicago.

Sklar, A. (1959). Fonctions dé repartition á n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris 8*, 229–231.

Smith, M., A. Min, C. Czado, and C. Almeida (2010). Modeling longitudinal data using a pair-copula decomposition of serial dependence. *To appear in the Journal of the American Statistical Association*.

Valdesogo, A. (2009). *Multivariate volatility models using copulas*. Ph. D. thesis, Center for Operations Research and Econometrics, Université Catholique de Louvain, Belgium.

Varin, C., N. Reid, and D. Firth (2010). An overview of composite likelihood methods. *To appear in Statistica Sinica*.

Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society. Series B 50*(2), 297–312.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica 57*, 307–333.