Stationary vine copula models for multivariate time series

Thomas Nagler*, Daniel Krüger, and Aleksey Min

August 13, 2020

Abstract

Multivariate time series exhibit two types of dependence: across variables and across time points. Vine copulas are graphical models for the dependence and can conveniently capture both types of dependence in the same model. We derive the maximal class of graph structures that guarantees stationarity under a condition called translation invariance. Translation invariance is not only a necessary condition for stationarity, but also the only condition we can reasonably check in practice. In this sense, the new model class characterizes all practically relevant vine structures for modeling stationary time series. We propose computationally efficient methods for estimation, simulation, prediction, and uncertainty quantification and show their validity by asymptotic results and simulations. The theoretical results allow for misspecified models and, even when specialized to the *iid* case, go beyond what is available in the literature. The new model class is illustrated by an application to forecasting returns of a portolio of 20 stocks, where they show excellent forecast performance. The paper is accompanied by an open source software implementation.

Keywords: pair-copula, dependence, bootstrap, forecasting, Markov chain, sequential maximum likelihood

1. Introduction

In multivariate time series there are two types of dependence: *cross-sectional* and *serial*. The first refers to the dependence between variables at a fixed point in time. The second refers to dependence of two random vectors at different points in time. Copulas are general models for stochastic dependence and have been used for both types. One line of research considers copula models for serial dependence in a univariate Markov processes

*Corresponding author, Mathematical Institute, Leiden University, Niels Bohrweg 3, 2333 CA Leiden, The Netherlands (email: mail@tnagler.com)

(including, Beare and Seo, 2014, Chen and Fan, 2006, Chen et al., 2009, Darsow et al., 1992, Ibragimov, 2009, Nasri et al., 2019). An orthogonal, but just as popular approach is to use classical univariate time series models for the marginal dynamics and model the cross-sectional dependence in their residuals by a copula (Nasri and Rémillard, 2019, Oh and Patton, 2017, Patton, 2006). See also Aas (2016), Patton (2009, 2012) for surveys in the context of financial and economic time series.

Copulas can also be used to capture both types of dependence in a single model (e.g., Rémillard et al., 2012, Simard and Rémillard, 2015). In this context, vine copulas (Aas et al., 2009, Bedford and Cooke, 2002) have been proven particularly useful. Vine copulas are graphical models that build a d-dimensional dependence structure from two-dimensional building blocks, called *pair-copulas*. The underlying graph structure consists of a nested sequence of trees, called *vine*. Each edge is associated with a pair-copula and each pair-copulas encodes a (possibly conditional) dependence between a pair of variables. Brechmann and Czado (2015), Smith (2015), and Beare and Seo (2015) proposed different vine structures suitable for time series models. The proposed models are quite similar. The vine graphs start with copies of a 'cross-sectional' tree that connects variables observed at the same point in time. Further, these trees are constrained to be either stars (Brechmann and Czado, 2015) or paths (Beare, 2010, Smith, 2015). Cross-sectional trees are then linked by a specific building plan. This leaves us with little flexibility but much potential for generalization.

Inspired by the three latter works, we propose more flexible vine models for stationary time series. But we approach the problem from the opposite direction. A motivating idea behind the previous models is that they (allegedly) guarantee stationarity of the model under a condition called *translation invariance*: pair-copulas stay the same when corresponding random variables are shifted in time. Translation invariance is a necessary condition for stationarity of the model and it is also the only practicable condition to check. We derive a characterization of the class of vines for which translation invariance is also sufficient for stationarity (Theorem 1 and Theorem 2). This class allows for general vine structures for cross-sectional dependence and leaves a lot of flexibility for linking them across time. The class includes the D-vine and M-vine models of Smith (2015) and Beare and Seo (2015) as special cases, but not the COPAR model of Brechmann and Czado (2015). This implies that, despite claims to the contrary, the COPAR model is not stationary in general (see Example 1).

For practical purposes, it is convenient to restrict to Markovian models. The Markov property can be shown to be equivalent to placing independence copulas at certain edges in the vine (Theorem 3). We discuss computationally efficient methods for parameter estimation, simulation, prediction, and uncertainty quantification in such models and establish their asymptotic validity (Theorems 4–5 and 6–7). Besides being valid for weakly dependent time series, they also apply to the *iid* case (where the Markov order is 0). Even then, the results go beyond what is currently available in the literature. All methodology is implemented in the open source R package `svines` (available at `https://github.com/tnagler/svines`), which is built on top the C++ library `rvinecopulib` (Nagler and Vatter, 2020a).

The article is structured as follows. Section 2 introduces some necessary concepts and notation for vine copula models and reviews previously proposed vine models for

stationary time series. The class of stationary vine models and some properties are derived in Section 3. Section 4 discusses estimation and model selection, Section 5 simulation and simulation-based predicition. All proofs are collected in the appendices. The models are illustrated with simulations in Section 6 and an application to financial time series in Section 7. Section 8 offers concluding remarks.

2. Background on vine copula models

2.1. Copulas

Copulas are models for the dependence in a random vector. By Sklar's theorem (Sklar, 1959), any multivariate distribution F with marginal distributions $F_1, \dots, F_d$ can be expressed as

$$F(x_1, \dots, x_d) = C\big\{F_1(x_1), \dots, F_d(x_d)\big\}, \qquad \text{for all } \boldsymbol{x} \in \mathbb{R}^d,$$

for some function $C\colon [0,1]^d \to [0,1]$ called *copula*. It characterizes the dependence in F because it determines how margins interact. If $\boldsymbol{X} \sim F$ is continuous, then C is the unique joint distribution function of the random vector $\boldsymbol{U} = (F_1(X_1), \dots, F_d(X_d))$. A similar formula can be stated for the density:

$$f(x_1, \dots, x_d) = c\big\{F_1(x_1), \dots, F_d(x_d)\big\} \times \prod_{k=1}^{d} f_k(x_k), \qquad \text{for all } \boldsymbol{x} \in \mathbb{R}^d,$$

where c is the density of C and called the *copula density*, and $f_1, \dots, f_d$ are the marginal densities.

2.2. Regular vines

Vine copula models are based on an idea of Joe (1996) to decompose the copula into a cascade of bivariate copulas. This decomposition is not unique, but all possible decomposition can be organized as a graphical model, called *regular vine (R-vine)* Bedford and Cooke (2001, 2002). We shall briefly outline the basics of R-vines; for more details on R-vines, we refer to Czado (2019), Dissmann et al. (2013), Joe (2014).

A regular vine is a sequence of nested trees. A tree (V, E) is a connected acyclic graph consisting of vertices V and edges E.

Definition 1. *A collection of trees $\mathcal{V} = (V_k, E_k)_{k=1}^{d-1}$ on a set V_1 with d elements is called R-vine if*

(i) *T_1 is a tree with vertices V_1 and edges E_1,*

(ii) *for $k = 2, \dots, d-1$, T_k is a tree with vertices $V_k = E_{k-1}$,*

(iii) *(*proximity condition*) for $k = 2, \dots, d-1$: if vertices $a, b \in V_k$ are connected by an edge $e \in E_k$, then the corresponding edges $a = \{a_1, a_2\}$, $b = \{b_1, b_2\} \in E_{k-1}$, must share a common vertex: $|a \cap b| = 1$.*

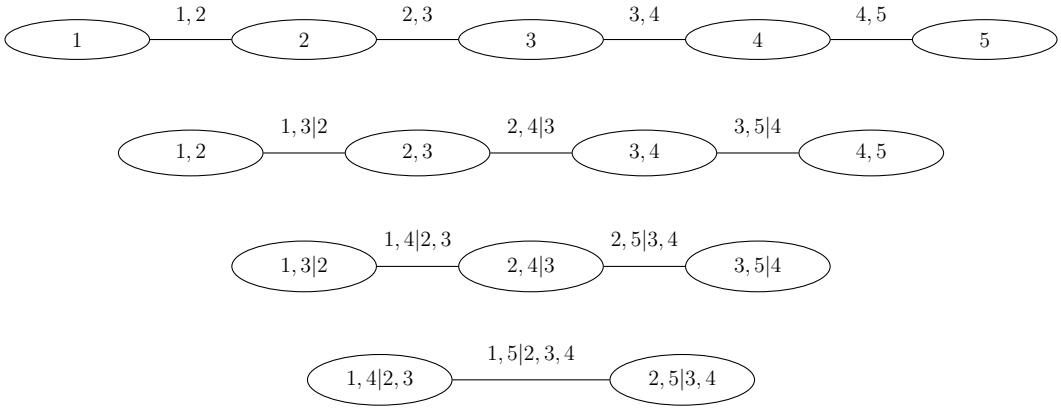

Figure 1: A five dimensional D-vine.

Special sub-classes of a regular vines are the so called D-vines (where each tree is a path) and C-vines (where each tree is a star). They represent two extreme cases. When a tree is a path, all trees at higher levels are fixed uniquely by the proximity condition (and paths themselves). When a tree is a star, the proximity condition poses no restrictions on the next tree. Some more nuanced properties of and valid operations on regular vines were derived by Morales Napoles et al. (2010), Joe et al. (2010) and Cooke et al. (2015).

The connection of regular vines to a decomposition of the dependence becomes apparent through a specific labeling of the edges. Each edge corresponds to a pair of random variables conditioned on some others. This is encoded in the *conditioned* and *conditioning sets* of an edge.

Definition 2. *The complete union of an edge $e \in E_k$ is given by*

$$\mathcal{U}_e = \{i \in V_1 |\ i \in e_1 \in e_2 \in \cdots \in e \text{ for some } (e_1, \ldots, e_{k-1}) \in E_1 \times \cdots \times E_{k-1}\}$$

and for a singleton $i \in V_1$ it is given by the singleton, i.e. $\mathcal{U}_i = \{i\}$.

Definition 3.

(i) The conditioning set *of an edge $e = (v_1, v_2)$ is $D_e = \mathcal{U}_{v_1} \cap \mathcal{U}_{v_2}$.*

(ii) The conditioned set *of an edge $e = (v_1, v_2)$ is defined as (a_e, b_e), where $a_e = \mathcal{U}_{v_1} \setminus D_e$ and $b_k = \mathcal{U}_{v_2} \setminus D_e$.*

We will then label an edge by $e = (a_e, b_e | D_e)$.

Any R-vine can be represented compactly in an upper-left triangular matrix $M = (m_{i,j})_{i,j=1,\ldots,d}$. The jth edge in the kth tree $e = (a_e, b_e|D_e)$ is encoded by the entries $(m_{d+1-j,j}, m_{k,j}|m_{k-1,j}, \ldots, m_{1,j})$. Less formally, the diagonal and kth element of the jth column form the conditioned set, the remaining entries above the kth element form the conditioning set. For details, we refer to Dissmann et al. (2013). For example, the R-vine

matrix of a five-dimensional D-vine is

$$M = \begin{bmatrix} 4 & 3 & 2 & 1 & 1 \\ 3 & 2 & 1 & 2 & 0 \\ 2 & 1 & 3 & 0 & 0 \\ 1 & 4 & 0 & 0 & 0 \\ 5 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and its graphical representation is shown in Figure 1. From M, we are able to reconstruct the trees of the regular vine and vice versa.

2.3. Vine copulas

A vine copula model identifies each edge of an R-vine with a bivariate copula. We shall write the model as $(\mathcal{V}, \mathcal{C}(\mathcal{V}))$, where $\mathcal{V} = (V_k, E_k)_{k=1}^{d-1}$ is the vine structure, d the number of variables, and $\mathcal{C}(\mathcal{V}) = \{c_e \colon e \in E_k, k = 1, \dots, d-1\}$ the set of associated bivariate copulas. As an example, consider the regular vine shown in Figure 1. The nodes in the first tree represent the random variables $U_1, \dots, U_5$. All edges connecting them are identified with a bivariate copula (or *pair-copula*). The edge (a_e, b_e) then encodes the dependence between U_{a_e} and U_{b_e}. In the second tree, the edges have labels $(a_e, b_e | D_e)$ and encode the dependence between U_{a_e} and U_{b_e} conditional on U_{D_e}. In the following trees, the number of conditioning variables increases.

Bedford and Cooke (2001) showed that the density of a such a model has a product form:

$$c(\boldsymbol{u}) = \prod_{m=1}^{d-1} \prod_{e \in E_k} c_{a_e, b_e | D_e} \left(u_{a_e|D_e}, u_{b_e|D_e} | \boldsymbol{u}_{D_e} \right),$$

where $u_{a_e|D_e} := C_{a_e|D_e}(u_{a_e}|\boldsymbol{u}_{D_e})$, $\boldsymbol{u}_{D_e} := (u_l)_{l \in D_e}$ is a subvector of $\boldsymbol{u} = (u_1, \dots, u_d) \in [0,1]^d$ and $C_{a_e|D_e}$ is the conditional distribution of U_{a_e} given U_{D_e}. For $e \in E_k$, such conditional distributions can be expressed recursively as

$$u_{a_e|D_e} = \frac{\partial C_{a_{e'}, b_{e'}|D_{e'}}(u_{a_{e'}|D_{e'}}, u_{b_{e'}|D_{e'}}|\boldsymbol{u}_{D_{e'}})}{\partial u_{b_{e'}|D_{e'}}},$$

where $e' \in E_{k-1}$, $a_e = a_{e'}$, $b_{e'} \in D_e$ and $D_{e'} = D_e \setminus b_{e'}$. At the end of the recursion, the right hand side involves an edge $e' \in E_1$, for which $u_{a_{e'}|D_{e'}} = u_{a_{e'}}$ and $u_{b_{e'}|D_{e'}} = u_{b_{e'}}$.

To make the model tractable, one commonly ignores the influence of $\boldsymbol{u}_{D_e}$ on the pair-copula density $c_{a_e, be|D_e}$. Under this assumption, the density simplifies to

$$c(\boldsymbol{u}) = \prod_{m=1}^{d-1} \prod_{e \in E_k} c_{a_e, be|D_e} \left(u_{a_e|D_e}, u_{b_e|D_e} \right).$$

Since each pair-copula can be modelled separately, simplified vine copulas remain quite flexible. For a more extensive treatment, we refer to Aas et al. (2009) and Czado (2019).

We further note that a similar factorization holds when some variables are discrete (see, Stöber, 2013, Section 2.1). Although both continuity and the simplifying assumption are irrelevant for our theoretical results, we will stick to the simplified, continuous case for notational convenience.

2.4. Vine copula models for multivariate time series

Now suppose $(\boldsymbol{X_t})_{t=1,\dots,n} = (X_{t,1}, \dots, X_{t,d})_{t=1,\dots,n}$ is a stationary time series, which we model by a vine copula. All existing regular vine models for multivariate time series follow the same idea (Beare and Seo, 2015, Brechmann and Czado, 2015, Smith, 2015). There is a vine capturing cross-sectional dependence of $\boldsymbol{X}_t \in \mathbb{R}^d$ for all time points $t = 1, \dots, n$. The first trees of the cross-sectional structures at time t and $t+1$ are then linked by one edge connecting a vertex from the structure at t to one vertex from the one at $t+1$. Because the time series is stationary, it is reasonable to assume that the cross-sectional structure and the linking vertices are time invariant.

The existing models (illustrated in Figures 2 to 4) make specific choices for the cross-sectional structure and connecting edge:

- *D-vine* of Smith (2015): (i) the cross-sectional structure is a D-vine, (ii) two D-vines at time points t and $t+1$ are connected at the two distinct variables that lie at opposite borders of the D-vine. Without loss of generality, we can assume that this is the first variable, i.e., an edge is added for $(X_{d,t}, X_{1,t+1})$. With these choices, there is only one global vine model satisfying proximity condition, which is a long D-vine spanning all variables at all time points.

- *M-vine* of Beare and Seo (2015): (i) the cross-sectional structure is a D-vine, (ii) two D-vines at time points t and $t+1$ are connected at one variable that lies at the border of the D-vine, i.e., an edge is added for $(X_{t,1}, X_{t+1,1})$. With the additional restriction that nodes of adjacent time points are connected first, this also fixes all further trees of the vine. (See also, Begin et al., 2020, for their connection to vector autoregressive models).

- *COPAR* of Brechmann and Czado (2015) : (i) the cross-sectional structure is a C-vine, (ii) two C-vines at time points t and $t+1$ are connected at the root node of the C-vine, i.e., an edge is added for $(X_{t,1}, X_{t+1,1})$. This leaves a lot of flexibility for higher trees and the authors settled on a specific set of rules. In particular, the model contains all edges of a D-vine on the variables $X_{1,1}, X_{2,1}, \dots, X_{n,1}$.

There is obvious potential for generalization. First, we would like to allow for arbitrary R-vines in the cross-sectional structure. Second, we would like to connect two cross-sectional trees at arbitrary variables. Specific verisons of such models were constructed in preliminary work by Krüger (2018) (called *temporal vine (T-vine)*) and in unpublished work by Harry Joe. But where should we stop? In principle, we could take any $(d \times n)$-dimensional vine as a model for the vector $(\boldsymbol{X}_1, \dots, \boldsymbol{X}_n)$.

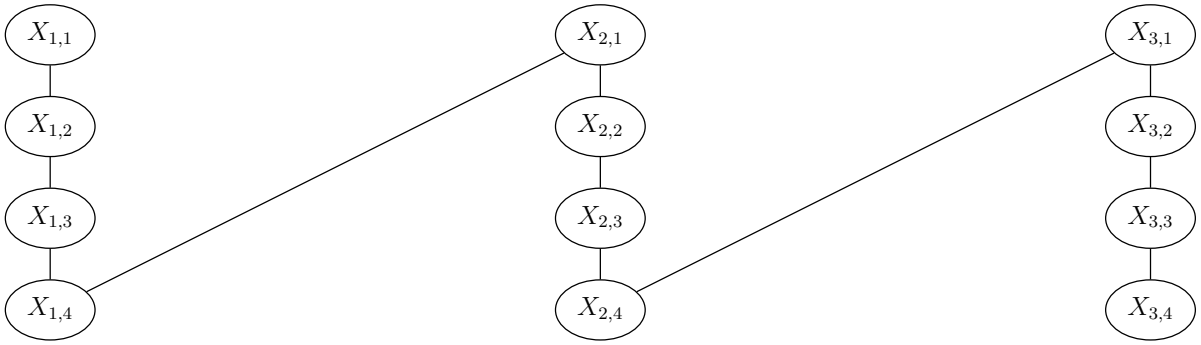

Figure 2: Example for the first tree level of a 4-dimensional D-vine on three time points.

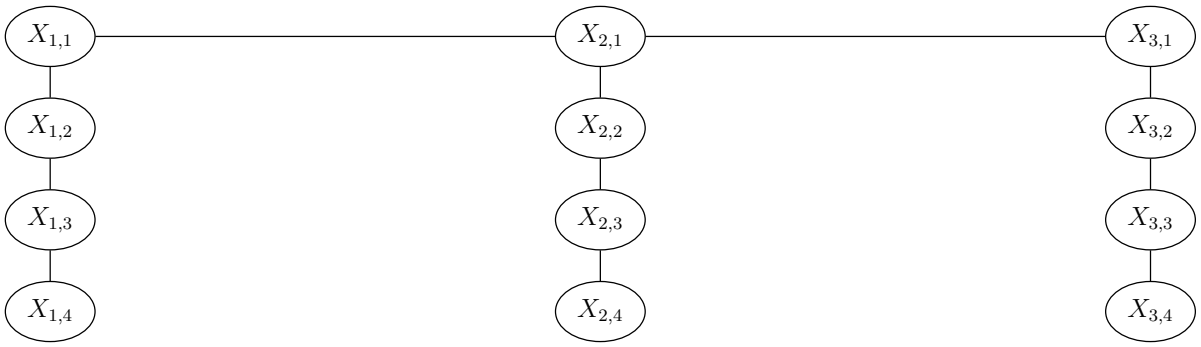

Figure 3: Example for the first tree level of a 4-dimensional M-vine on three time points.

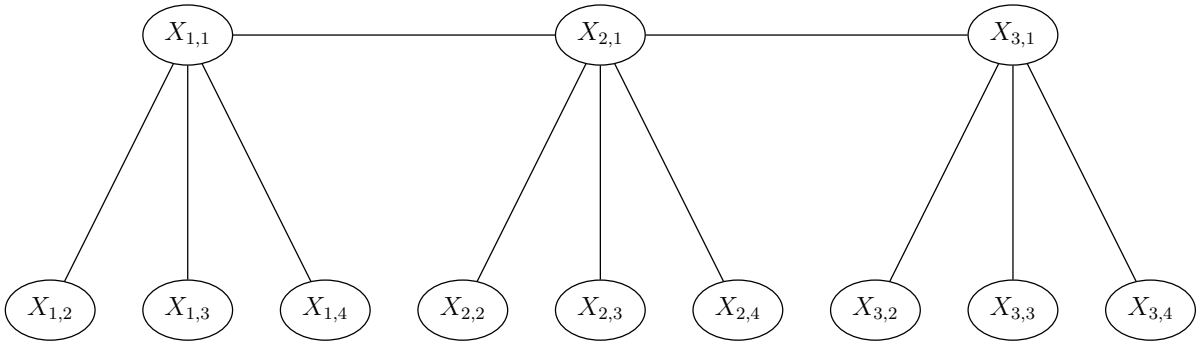

Figure 4: Example for the first tree level of a 4-dimensional COPAR on three time points.

3. Stationary vine copula models

The time series context is special. To facilitate inference, it is common to assume that the series is stationary, i.e., the distribution is invariant in time. When a time series is stationary, also its copula must satisfy certain invariances. This is a blessing and a curse: invariances reduce the complexity of the model, but not all structures guarantee stationarity under practicable conditions on the pair-copulas. We shall derive a generalization of the existing models that is maximally convenient in this sense. All proofs are collected in Appendix A.

3.1. Stationary time series

Let $(\mathcal{V}, \mathcal{C}(\mathcal{V}))$ be a vine copula model for the random vector $(\boldsymbol{U}_1^\top, \dots, \boldsymbol{U}_n^\top)^\top \in \mathbb{R}^{n \times d}$ and denote by c the corresponding copula density. Then $\boldsymbol{U}_1, \dots, \boldsymbol{U}_n \in \mathbb{R}^d$ is a strictly stationary time series if and only if $\boldsymbol{U}_{t_1}, \dots, \boldsymbol{U}_{t_m}$ and $\boldsymbol{U}_{t_1+\tau}, \dots, \boldsymbol{U}_{t_m+\tau}$ have the same joint distribution for all $1 \le t_1 < t_2 < \dots < t_m \le N$, $1 \le \tau \le N - \max_{j=1}^m t_j$, and $1 \le m \le N$.

For vine copulas, this condition can involve intractable functional equations. The reason is that only some pairwise (conditional) dependencies are explicit in the model. Explicit pairs are those that correspond to edges in the vine $\mathcal{V}$. All other dependencies are only implicit, i.e., they are characterized by the interplay of multiple pair-copulas. By only focusing on the explicit pairs, we see that *translation invariance* (Beare and Seo, 2015) is a necessary condition for stationarity.

Definition 4 (Translation invariance)**.** *A vine copula model $(\mathcal{V}, \mathcal{C}(\mathcal{V}))$ on the set $V_1 = \{1, \dots, n\} \times \{1, \dots, d\}$ is called* translation invariant *if it holds,*

$$c_{a_e, b_e | D_e} = c_{a_{e'}, b_{e'} | D_{e'}},$$

for all edges $e, e' \in \bigcup_{k=1}^{nd-1} E_k$ for which there is $\tau \in \mathbb{Z}$ such that

$$a_e = a_{e'} + (\tau, 0), \quad b_e = b_{e'} + (\tau, 0), \quad D_e = D_{e'} + (\tau, 0), \tag{1}$$

where the last equality is short for $D_e = \{v + (\tau, 0) \colon v \in D_{e'}\}$

Remark 1. *In the time series context, each vertex of a vine's first tree is identified with a tuple (t, i), where t is the time index and i is the variable index. The vertex (t, i) corresponds to the random variable $U_{t,i}$. The notation $e = e' + (\tau, 0)$ will be used short for (1) and indicates a shift in time by τ steps.*

For all non-explicit pairs, stationarity requires more complex integral equations to hold. Provided with sufficient computing power, they could be checked numerically for any given model. But even if it holds for a specific model, a slight change in the parameter of a single pair-copula may break it. This is problematic in practice. Hence, the practically relevant vine structures are those for which translation invariance is also a sufficient condition for stationarity.

3.2. Preliminaries

First we need some graph theoretic definitions. The first is a version of Definition 6 of Beare and Seo (2015).

Definition 5 (Restriction of vines)**.** *Let $\mathcal{V} = (V_k, E_k)_{k=1}^{nd-1}$ be a vine on $\{1\dots,n\} \times \{1,\dots,d\}$ and $V_1' = \{t,\dots,t+m\} \times \{1,\dots,d\}$ for some t, m with $1 \le t \le N$, $0 \le m \le N - t$. For all $k \ge 1$, define $E_k' = E_k \cap \binom{V_k'}{2}$ and $V_{k+1}' = E_k'$. Then the sequence of graphs $\mathcal{V}_{t,t+m} = (V_k', E_k')_{k=1}^{(m+1)d-1}$ is called* restriction of $\mathcal{V}$ *on the time points $t,\dots,t+m$.*

The graphical interpretation of this definition is straightforward. In the first tree of the regular vine $\mathcal{V}$, one deletes all vertices not in V_1' along with corresponding edges. In the remaining trees, all edges and vertices of the tree (V_k, E_k) affected by the deletion in the previous tree (V_{k-1}, E_{k-1}) are discarded. Note that the restriction $\mathcal{V}_{t,t+m} = (V_k', E_k')_{k=1}^{(m+1)d-1}$ is not necessarily a vine; the graphs (V_k', E_k') can be disconnected (hence, no trees). For example, if the first tree of the vine $\mathcal{V}$ contains a path $(1,i) - (3,i) - (2,i)$, the vertices $(1,i)$ and $(2,i)$ will be disconnected in $\mathcal{V}_{1,2}$.

Definition 6 (Translation of vines)**.** *Let $m \ge 0$, $\mathcal{V}_{t,t+m} = (V_{t,k}, E_{t,k})_{k=1}^{(m+1)d-1}$ be a vine on $\{t,\dots,t+\tau\} \times \{1,\dots,d\}$ and $\mathcal{V}_{s,s+m}$ be a vine on $\{s,\dots,s+\tau\} \times \{1,\dots,d\}$. We say that $\mathcal{V}_t$ is a* translation *of $\mathcal{V}_s$ (denoted by $\mathcal{V}_t \sim \mathcal{V}_s$) if there is a $\tau \in \mathbb{Z}$ such that for all $k = 1,\dots,d-1$ and edges $e \in E_{t,k}$, there is an edge $e' \in E_{s,k}$ such that $e = e' + (t-s, 0)$ (and vice versa).*

Remark 2. *We shall call two edges e, e' satisfying $e' + (\tau, 0)$ translations of another and write $e \sim e'$. This defines an equivalence relationship between edges.*

3.3. Stationary vines

Theorem 1. *Let $\mathcal{V}$ be a vine on the set $V_1 = \{1,\dots,n\} \times \{1,\dots,d\}$. Then the following statements are equivalent:*

(i) The vine copula model $(\mathcal{V}, \mathcal{C}(\mathcal{V}))$ is stationary for all translation invariant choices of $\mathcal{C}(\mathcal{V})$.

(ii) There are vines $\mathcal{V}^{(m)}, m = 0,\dots,n-1$, defined on $\{1,\dots,m+1\} \times \{1,\dots,d\}$, such that for all $m = 0,\dots,n-1$, $1 \le t \le n-m$,

$$\mathcal{V}_{t,t+m} \sim \mathcal{V}^{(m)}. \tag{2}$$

An important word in conditoin (i) is *all*. There are vines violating (ii) that are stationary for a specific choice of $\mathcal{C}(\mathcal{V})$. For example, $c_e \equiv 1$ for all edges always leads to a stationary model. But these structures are inpractical, because they limit the choices of $\mathcal{C}(\mathcal{V})$ to a restrictive and unknown set.

Condition (ii) can be seen as a graph theoretic notion of stationarity.

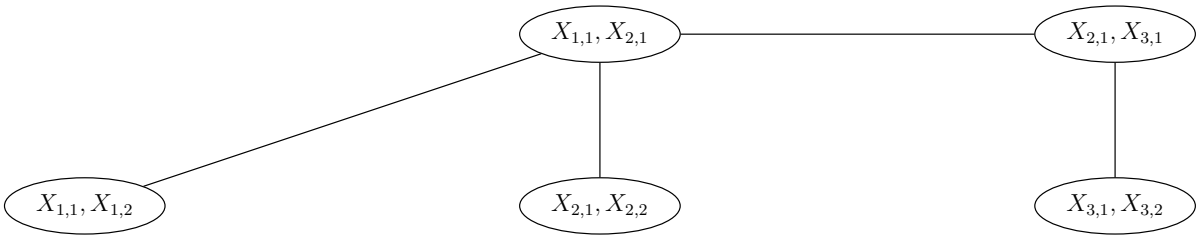

Figure 5: Example for the second tree level of a COPAR model with $d = 2, n = 3$.

Definition 7 (Stationary vines)**.** *A vine $\mathcal{V}$ on the set $V_1 = \{1, \dots, n\} \times \{1, \dots, d\}$ is called* stationary *if it satisfies condition (ii) of Theorem 1.*

It is easy to check that M-vines and D-vines of Beare and Seo (2015) and Smith (2015) are stationary. The structure of a COPAR model of Brechmann et al. (2012) is not stationary, however: the graph $\mathcal{V}_{t,t+1}$ is not a vine for $t \geq 2$ (the second level of the restricted graph is not a tree). This poses an additional constraint on the choice of pair copulas that went seemingly unnoticed.

Example 1. *Let us illustrate the tricky part of the proof Theorem 1 with the COPAR model for $d = 2$, $n = 3$. The second tree of the model is given in Figure 5 (see, Brechmann and Czado, 2015). For simplicity, we assume that all pair-copulas in trees $k = 1$ and $k \geq 3$ are independence. The restriction $\mathcal{V}_{2,3}$ of the model is obtained by deleting all nodes and edges where a time index 1 occurs. Clearly, $\mathcal{V}_{2,3}$ it is not a vine, because the node $(X_{2,1}, X_{2,2})$ is disconnected from the others. Now let us see why this is problematic. The joint copula density of nodes $(X_{2,1}, X_{2,2})$ and $(X_{2,1}, X_{3,1})$ equals the product of copulas associated with the edges along the path joining them, integrating over all intermediate nodes. That is,*

$$c_{(2,2),(3,1)|(2,1)}(u, v) = \int_0^1 c_{(1,1),(2,2)|(2,1)}(w, u) c_{(1,1),(3,1)|(2,1)}(w, v) dw.$$

By stationarity invariance, it must further hold

$$c_{(1,2),(2,1)|(1,1)}(u, v) = \int_0^1 c_{(1,1),(2,2)|(2,1)}(w, u) c_{(1,1),(3,1)|(2,1)}(w, v) dw.$$

The copula on the left hand side is an explicit dependence in the model, because it is associated with an edge in the graph (the leftmost one). Thus the equation contains three pair-copulas of the model that are not constrained by translation invariance. For most combinations of pair-copulas, the equality will not hold and the model is not stationary.

3.4. An explicit characterization of stationary vines

Stationary vines have a very specific structure. There is a d-dimensional vine $\mathcal{V}^{(0)}$ that contains only pairs for cross-sectional dependence. We will therefore call $\mathcal{V}^{(0)}$ the *cross-sectional structure* of $\mathcal{V}$. Next, there is a $2d$ dimensional vine $\mathcal{V}^{(1)}$ that nests two duplicates

of $\mathcal{V}^{(0)}$. Besides the cross-sectional parts, the vine contains d^2 pairs for dependence across two subsequent time points that are not yet constrained by translation invariance. A similar principal applies for vines $\mathcal{V}^{(m)}, m \geq 2$, with d^2 unconstrained edges entering in every step.

Stationary vines can also be characterized by a more explicit condition. Somewhat surprisingly, it suffices to pick a cross-sectional structure $\mathcal{V}^{(0)}$ and two permutations of $(1, \dots, d)$. The permutations determine how the first d trees of the cross-sectional structures are connected across two adjecent time points. However, the permutations are constrained by the choice of cross-sectional structure.

Definition 8 (Compatible permutations)**.** *We call a permutation $(i_1, \dots, i_d)$ of $(1, \dots, d)$* compatible *with a vine $\mathcal{V}$ on $\{1, \dots, d\}$ if for all $k = 2, \dots, d$, there is an edge $e \in E_{k-1}$ with conditioned set $\{i_k, i_r\}$ and conditioning set $\{i_1, \dots, i_{k-1}\} \setminus i_r$ for some $r \in \{1, \dots, k-1\}$.*

The first index of the permutation (i_1) is not constrained by compatibility. This further implies that any d-dimensional vine has at least d compatible permutations (see Lemma 1).

Theorem 2. *A vine $\mathcal{V}$ on $\{1, \dots, n\} \times \{1, \dots, d\}$ is stationary if and only if*

(i) there is a vine $\mathcal{V}^{(0)}$ on $\{0\} \times \{1, \dots, d\}$ such that $\mathcal{V}_t \sim \mathcal{V}^{(0)}$, for all $1 \leq t \leq N$,

(ii) there are two permutations $(i_1, \dots, i_d)$ and $(j_1, \dots, j_d)$ compatible with $\mathcal{V}^{(0)}$, such that

$$E_k = \bigcup_{t=1}^{n} E_k^{(0)} + (t, 0) \;\cup \bigcup_{t=1}^{n-1} \bigcup_{r=1}^{k} \left\{ e \colon a_e = (t, i_{k+1-r}), b_e = (t+1, j_r), D_e = \bigcup_{s=1}^{k-r} \{(t, i_s)\} \cup \bigcup_{s=1}^{r-1} \{(t+1, j_s)\} \right\}$$

for $k = 1, \dots, d$.

Remark 3. *The two permutations in (ii) can be equal.*

Remark 4. *In tree k, the index i_k connects the cross-sectional structure at time t with the variable j_1 at time $t+1$. Conversely, the index j_k connects the structure at time $t+1$ with the index i_1 at time t. We therefore call $(i_1, \dots, i_d)$* out-vertices *and $(j_1, \dots, j_d)$* in-vertices *of $\mathcal{V}^{(0)}$.*

Remark 5. *(i) If $\mathcal{V}^{(0)}$ is a D-vine with leaf i_1 and $(i_1, \dots, i_d) = (j_1, \dots, j_d)$, we obtain the M-vine model of Beare and Seo (2015).*

(ii) If $\mathcal{V}^{(0)}$ is a D-vine with leaf i_1 and $(i_1, \dots, i_d) = (j_d, \dots, j_1)$, we obtain the D-vine model of Smith (2015).

(iii) If we choose i_s, j_s, iteratively for $s \geq 2$ as the smallest compatible indices, we obtain the T-vine model of Krüger (2018).

$$
\begin{pmatrix}
* & * & * & \iddots & \iddots & (1,i_1) & * & * & * & (1,j_1) \\
* & * & \iddots & \iddots & \iddots & \vdots & * & * & \iddots & \\
* & \iddots & \iddots & \iddots & \iddots & (1,i_{d-1}) & * & (1,j_{d-1}) & & \\
(n-1,i_1) & \iddots & \iddots & \iddots & \iddots & (1,i_d) & (1,j_d) & & & \\
\vdots & \iddots & \iddots & \iddots & \iddots & (2,j_1) & & & & \\
\vdots & \iddots & \iddots & \iddots & \iddots & & & & & \\
(1,i_1) & \iddots & \iddots & \iddots & & & & & & \\
\vdots & \iddots & \iddots & & & & & & & \\
(1,i_d) & (n,j_{d-1}) & & & & & & & & \\
(n,j_d) & & & & & & & & &
\end{pmatrix}
$$

Figure 6: Matrix representation of stationary vines.

The proof relies on an induction argument, which we briefly illustrate. In the first tree level, cross-sectional trees are connected by a single edge connecting (i_1, t) to $(j_1, t+1)$. In the second tree, cross-sectional trees are connected by two edges. By the proximity condition, on of the nodes (i_1, t) and $(j_1, t+1)$ must be in the conditioning set. The conditioning set contains only one element, so the new edges must be $(i_2, t), (j_1, t+1)|(i_1, t)$ and $(i_1, t), (j_2, t+1)|(j_1, t+1)$, where $i_1 \neq i_2$ and $j_1 \neq j_2$. The same argument applies to higher tree levels.

Remark 6. *Theorem 2 implies that all stationary vines can be represented by a matrix of the form shown in Figure 6, where the stars correspond to the cross-sectional structure.*

3.5. Existence and uniqueness

The explicit characterization of Theorem 2 makes it easy to establish conditions for existence and uniqueness of a stationary vine. The first step is to show that a compatible permutation always exists.

Lemma 1. *For d-dimensional vine $\mathcal{V}$ and any $i_1 \in \{1, \dots, d\}$, there exists at least one permutation $(i_1, \dots, i_d)$ compatible with $\mathcal{V}$.*

Now the following result is an immediate conequence of Theorem 2 and Lemma 1.

Corollary 1.

(i) (Existence) For any vine $\mathcal{V}^*$, there exists a stationary vine with cross-sectional structure $\mathcal{V}^{(0)} = \mathcal{V}^*$.

(ii) (Uniqueness) Given a cross-sectional structure $V^{(0)}$ and two sequences of compatible in- and out-vertices, the stationary vine is unique.

	$N = 100, d = 5$	$N = 100,\ d = 20$	$N = 1\,000,\ d = 20$
general model	124 750	1 999 000	199 990 000
stationary model	2 485	39 790	399 790
stationary Markov(2) model	60	990	990
stationary Markov(1) model	35	590	590

Table 1: Number of distinct pair-copulas to specify for different vine models.

3.6. Markovian models

Stationarity is a convenient property because it limits model complexity. An arbitrary vine copula model for $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n \in \mathbb{R}^d$ requires to specify (or estimate) $nd(nd-1)/2 = O(n^2d^2)$ pair-copulas. In a stationary vine copula model, cross-sectional dependencies are associated with the same pair copulas for each time point. Similarly, serial dependencies are modeled with identical copulas for each lag. This significantly reduces the number of free pair-copulas in the model. We only need to specify $(n-1)d^2 + d(d-1)/2 = O(nd^2)$ of them, all other pair-copulas are constrained by translation invariance. When the time series contains more than a few dozen time points, this will still be too much.

Most popular time series models also satisfy the Markov property.

Definition 9. *A time series $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n \in \mathbb{R}^d$ is called Markov (process) of order p if for all $\boldsymbol{x} \in \mathbb{R}^d$,*

$$\mathbb{P}\big(\boldsymbol{X}_t \le \boldsymbol{x} \mid \boldsymbol{X}_{t-1}, \dots, \boldsymbol{X}_1\big) = \mathbb{P}\big(\boldsymbol{X}_t \le \boldsymbol{x} \mid \boldsymbol{X}_{t-1}, \dots, \boldsymbol{X}_{t-p}\big).$$

The Markov property limits complexity further. For the M-vine model, Beare and Seo (2015, Theorem 4) showed that it is equivalent to what they call *p-independence*. The same arguments apply for the more general class of stationary vines.

Theorem 3. *A vine copula model $(\mathcal{V}, \mathcal{C}(\mathcal{V}))$ on a stationary vine $\mathcal{V}$ is Markov of order p if and only if $\mathcal{C}(\mathcal{V})$ is* p-independent*, i.e., $c_e \equiv 1$ for all $e \notin \mathcal{V}_{t,t+p}$, $t = 1, \dots, n-p$.*

In a stationary Markov model of order p, the independence copula is assigned to all edges reflecting serial dependence of lags larger than p. This reduces the number of distinct pair copulas further to $pd^2 + d(d-1)/2 = O(pd^2)$. Table 1 shows the number of distinct copulas in an unrestricted model for the full time series, a stationary vine model, and a stationary vine model with Markov order $p = 1, 2$. We can see a significant reduction when imposing stationarity and the Markov property.

4. Estimation and model selection

Estimation of copula-based Markov chains was discussed earlier by Chen and Fan (2006) and Chen et al. (2009) for $d = 1$, $p = 1$. Rémillard et al. (2012) extended their results for $d \ge 1$ for the joint maximum-likelihood estimator (MLE) in a semiparametric model. Joint maximum-likelihood is unpopular for vine copula models, because they have many

parameters even in moderate dimension. Beare and Seo (2015) discussed a version of the popular stepwise maximum likelihood estimator (Aas et al., 2009) for M-vine copula models, but without theoretical gurantees. We shall introduce such a method for the more general class of stationary vines, prove its validity, and discuss selection of the vine structure and pair-copula families.

4.1. Parameter estimation

Suppose for the moment that the vine structure and copula families are known and we only need to estimate their parameters. In addition, the stationary marginal distributions $F_1, \dots, F_d$ must be estimated. We shall only deal with fully parametric models here.

We follow the common practice to estimate marginal models first before dealing with the copulas model. Suppose we are given parametric models $f_j(\cdot; \boldsymbol{\eta}_j)$, $j = 1, \dots, d$, for the marginal densities. Then the parameters can be estimated by the maximum-likelihood-type estimator

$$\widehat{\boldsymbol{\eta}}_j = \arg\max_{\boldsymbol{\eta}_j} \sum_{t=1}^{n} \ln f_j(X_{t,j}; \boldsymbol{\eta}_j), \quad j = 1, \dots, d. \tag{3}$$

This is different from the classical MLE because serial dependence is ignored. Given estimates of the marginal parameters, we then generate *pseudo-observations* from the copulas model via

$$\widehat{U}_{t,j} = F_j(X_{t,j}; \widehat{\boldsymbol{\eta}}_j), \quad t = 1, \dots, n, \quad j = 1, \dots, d.$$

Further, for all edges e in the vine, let $c_{[e]}(\cdot; \boldsymbol{\theta}_{[e]})$ be a parametric model with parameter $\boldsymbol{\theta}_{[e]}$. Because of translation invariance, many of the edges must have the same families and parameters. This is reflected by the notation $[e]$ which assigns a family $c_{[e]}(\cdot; \boldsymbol{\theta}_{[e]})$ and parameter $\boldsymbol{\theta}_{[e]}$ for the entire equivalence class $[e] = \{e' \colon e' \sim e\}$.

Recall from Section 2.3 that the joint density of the model involves conditional distributions of the form $C_{a_e|D_e}$ which can be expressed recursively. For an edge $e \in E_k$, denote by $S(a_e)$ the set of edges $e' \in \{E_1, \dots, E_{k-1}\}$ involved in this recursion and $\boldsymbol{\theta}_{S(a_e)} = (\boldsymbol{\theta}_{[e']})_{e' \in S(a_e)}$. Finally, write $[E_k] = \{[e] \colon e \in E_k\}$, $\boldsymbol{\theta}_{[E_k]} = (\boldsymbol{\theta}_{[e]})_{[e] \in [E_k]}$ and $\boldsymbol{\theta} = (\boldsymbol{\theta}_{[E_k]})_{k=1}^{(p+1)d-1}$ as the stacked parameter vector.

The joint (pseudo-)log-likelihood of a stationary vine copula model for $(\boldsymbol{X}_1, \dots, \boldsymbol{X}_n)$ is

$$\ell(\boldsymbol{\theta}) = \sum_{k=1}^{d(p+1)-1} \sum_{e \in E_k} \ln c_{[e]}\big\{C_{a_e|D_e}(\widehat{U}_{a_e} \mid \widehat{\boldsymbol{U}}_{D_e}; \boldsymbol{\theta}_{S(a_e)}), C_{b_e|D_e}(\widehat{U}_{b_e} \mid \widehat{\boldsymbol{U}}_{D_e}; \boldsymbol{\theta}_{S(b_e)}); \boldsymbol{\theta}_{[e]}\big\}.$$

The joint MLE $\arg\max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta})$ is often too demanding. The stepwise MLE of Aas et al. (2009) estimates the parameters of each pair-copula separately, starting from the first tree. We can adapt it to the setting of a Markov process of order p: for $k = 1, \dots, d(p+1) - 1$,

and for every $e' \in E_k$

$$\widehat{\boldsymbol{\theta}}_{[e']} = \arg\max_{\boldsymbol{\theta}_{[e']}} \sum_{e \sim e'} \ln c_{[e]}\big\{C_{a_e|D_e}(\widehat{U}_{a_e} \mid \widehat{\boldsymbol{U}}_{D_e}; \widehat{\boldsymbol{\theta}}_{S(a_e)}), C_{b_e|D_e}(\widehat{U}_{b_e} \mid \widehat{\boldsymbol{U}}_{D_e}; \widehat{\boldsymbol{\theta}}_{S(b_e)}); \boldsymbol{\theta}_{[e']}\big\}. \quad (4)$$

The vectors $\widehat{\boldsymbol{\theta}}_{S(a_e)}$, $\widehat{\boldsymbol{\theta}}_{S(b_e)}$ on the right hand side above only contain parameter estimates from previous trees, i.e., ones that we already found in an earlier iteration. This procedure is a multi-step extension of the inference for margins method (Joe, 2005, Joe and Xu, 1996) to the time series context.

The following asymptotic results allow for a possibly misspecified model. The limiting values $(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*)$ are the *pseudo-true* values defined as the stepwise maximizers of (3) and (10), but taking expectations of the sums. If the model is correctly specified, they agree with the true parameters. The following results and their proofs require some technical notation and assumptions, which are deferred to Appendix B to avoid distraction.

Theorem 4. *Under (A1)–(A7), the stepwise MLE $(\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\theta}})$ exists and $(\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\theta}}) \to_p (\boldsymbol{\eta}^*, \boldsymbol{\theta}^*)$.*

Theorem 5. *Suppose (A1)–(A8) hold and Then $\|(\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\theta}}) - (\boldsymbol{\eta}^*, \boldsymbol{\theta}^*)\| = O_p(n^{-1/2})$ and*

$$\sqrt{n}\begin{pmatrix} \widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^* \\ \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \end{pmatrix} \xrightarrow{d} \mathcal{N}\big(\mathbf{0}, \boldsymbol{J}^{-1}(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*)\boldsymbol{I}(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*)\boldsymbol{J}^{-1}(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*)^\top\big),$$

where $\boldsymbol{I}$ and $\boldsymbol{J}$ are defined in Appendix B.1.

The asymptotic covariance can be estimated from the data by numerical differentiation of the marginal and pair-copula log-likelihoods.

Remark 7. *Most assumptions of the theorems are standard regularity conditions, but two of them stick out. Assumption (A1) requires ergodicity of $(\boldsymbol{X}_t)_{t=1}^{\infty}$. For $p = 1, d = 1$, it is sufficient that the copula density is strictly positive on a set of measure 1 (Longla and Peligrad, 2012, Proposition 2). The proof can be easily extended to $p, d \geq 1$, which leads to the mild sufficient condition that all pair-copula densities are strictly positive on $(0, 1)^2$. Assumption (A8) (only required for Theorem 5) strengthens this to a condition on the decay of β-mixing coefficients. This is more problematic. For $p = 1, d = 1$, a sizeable literature (including Beare, 2010, 2012, Chen and Fan, 2006, Chen et al., 2009, Longla and Peligrad, 2012) suggests that all popular parametric models exhibit exponentially decaying mixing coefficients, which is stronger than necessary. However, extending these results to the multivariate case is nontrivial and poses an important open problem.*

The above results extend the existing literature in various ways. A similar result was obtained by Rémillard et al. (2012) for Markov models with nonparametric margins, $p = 1$, and a joint MLE for the copula parameters. It does not apply to the stepwise MLE commonly used in vine copula models, however. The only known results for the stepwise MLE were provided by Hobæk Haff (2013) in the *iid* ($p = 0$) case. These results are formulated for D-vines and assume nonparametric margins. Despite the wide use of the fully parametric stepwise MLE, the above theorems thus appear to be the first published

results for such models — even in the *iid* case. An extension to semi-parametric or fully nonparametric methods is possible, but beyond the scope of this paper; see Hobæk Haff (2013) and Nagler and Czado (2016) for similar results in the *iid* case.

Furthermore, the results of Hobæk Haff (2013) assume a correctly specified parametric model. This assumption is especially questionable in view of the common simplifying assumption (see Section 2.3). Not only do we need to find the correct family for all pair-copulas, but the simplifying assumption must hold as well. Even more, we must find a vine structure under which the simplifying assumption is true. This is extremely difficult because the number of possible structures grows exponentially in the dimension (Morales Napoles et al., 2010). Hence, practitioners must rely on heuristics. We shall see in the following section that such heuristics often fail to find such a structure, even when the model generating the data is simplified.

4.2. Model selection

Parametric models are easy to select in a fully stepwise procedure. The most common selection criteria are AIC or BIC. For the margins, these criteria are computed from the maximal pair-copula log-likelihood on the right hand side of (3). Then we continue stepwise with the pair-copulas, where AIC and BIC are computed from the maximal log-likelihood on the right hand side of (4).

For the vine structure, we propose a heuristic similar extending the popular algorithm of Dissmann et al. (2013). Its idea is to capture the strongest dependencies as early as possible in the tree structure. In addition to the cross-sectional structure $\mathcal{V}^{(0)}$, we also need to select the in- and out-vertices $(j_1, \dots, j_d)$ and $(i_1, \dots, i_d)$.

We first focus on the cross-sectional structure. We compute the (absolute) empirical Kendall's τ between all pairs of variables and find the maximum spanning tree. Then we find the optimal in-/out-vertices by computing all pair-wise empirical Kendall's τ between the original time series and a lagged version and choose the edge with maximal $|\tau|$. Now the first tree is completely specififed. Then we estimate all parameters in this tree and generate pseudo-observations for the next. We again build a maximum spanning tree for the cross-sectional part with the proximity condition as side constraints. Then we find the compatible in-/out-vertices by maximizing absolute Kendall's τ for the corresponding edges in Theorem 2 (ii). We continue this way until the first d trees are selected, which completely determines the remaining structure.

The procedure can be simplified for M- and D-vines by imposing appropriate constraints on the cross-sectional structure (which becomes a shortest path problem) and the in-/out-vertices (where only i_1 is a free parameter). It must be emphasized that all these methods are heuristic and give no guarantees to find the optimal model in their class.

5. Simulation and prediction

Vine copula models are quite complex and rarely allow closed-form expressions of conditional means, quantiles, or the predictive distribution. It is fairly easy to simulate from vine copula models and approximate such quantities by Monte Carlo methods.

Algorithm 1 Simulating K time steps from a p-Markovian vine copula model conditional on the past.

Input: A p-independent, stationary vine copula model $(\mathcal{V}, \mathcal{C}(\mathcal{V}))$, stationary marginal distributions $F_1, \dots, F_d$, and past observations $\boldsymbol{X}_{t-p+1}, \dots, \boldsymbol{X}_t$.
Output: Samples $\boldsymbol{X}_{t+1}, \dots, \boldsymbol{X}_{t+K}$ from the conditional distribution given $\boldsymbol{X}_{t-p+1}, \dots, \boldsymbol{X}_t$.

(i) Set $U_{t-p+k,j} = F_j(X_{t-p+k,j})$ for $k = 1, \dots, p$, $j = 1, \dots, d$.

(ii) **for** $k = 0, \dots, K-1$:

a) Set $(\boldsymbol{W}_1, \dots, \boldsymbol{W}_p) = R_{\mathcal{C}(\mathcal{V}_{t-p+1,t})}(\boldsymbol{U}_{t-p+1+k}, \dots, \boldsymbol{U}_{t+k})$.

b) Simulate $\boldsymbol{W}_{p+1} \sim \text{Uniform}([0,1]^d)$.

c) Set $(\boldsymbol{U}_{t-p+1+k}, \dots, \boldsymbol{U}_{t+k+1}) = R^{-1}_{\mathcal{C}(\mathcal{V}_{t-p+1,t+1})}(\boldsymbol{W}_1, \dots, \boldsymbol{W}_{p+1})$.

end for

(iii) Set $X_{t+k,j} = F_j^{-1}(U_{t+k,j})$ for $k = 1, \dots, K$, $j = 1, \dots, d$.

But the standard simulation algorithm (e.g., Czado, 2019, Chapter 6) poses unnecessary computational demands in the time series context.

5.1. Simulation

Markovian models can be simulated much more efficiently than general vine copula models. An efficient algorithm generalizing ideas of Kraus and Czado (2017) is given in Algorithm 1.

It is based on the *Rosenblatt transform*, whose inverse also appears in the standard algorithm. The Rosenblatt transform applies certain conditional distribution functions to a random vector to turn it into independent uniforms. Conversely, the inverse Rosenblatt transform turns independent uniforms into a vector with arbitrary joint distribution. Let $M = (m_{i,j})_{i,j=1}^d$ be the R-vine matrix corresponding to a d-dimensional vine copula model $(\mathcal{V}, \mathcal{C}(\mathcal{V}))$. Our version of the Rosenblatt transform is defined as $R_{\mathcal{C}(\mathcal{V})} \colon [0,1]^d \mapsto [0,1]^d$ with

$$\big(R_{\mathcal{C}(\mathcal{V})}(\boldsymbol{u})\big)_{m_{d+1-j,j}} = C_{m_{d+1-j,j}|m_{d-j,j},\dots,m_{1,j}}(u_{m_{d+1-j,j}} \mid \boldsymbol{u}_{m_{d-j,j},\dots,m_{1,j}}), \quad j = 1, \dots, d.$$

The conditional distributions in this formula can be computed recursively from pair-copulas in $\mathcal{C}(\mathcal{V})$ as explained in Section 2.3. The inverse transformation is

$$\big(R^{-1}_{\mathcal{C}(\mathcal{V})}(\boldsymbol{u})\big)_{m_{d+1-j,j}} = C^{-1}_{m_{d+1-j,j}|m_{d-j,j},\dots,m_{1,j}}(u_{m_{d+1-j,j}} \mid \boldsymbol{u}_{m_{d-j,j},\dots,m_{1,j}}), \quad j = 1, \dots, d,$$

which can also be computed recursively from pair-copulas in $\mathcal{C}(\mathcal{V})$, see Czado (2019).

Algorithm 1 can also be used to simulate unconditionally by giving it an unconditional

$(\boldsymbol{U}_{t-p+1}, \dots, \boldsymbol{U}_t)$ from the vine copula model $(\mathcal{V}_{t-p+1,t}, \mathcal{C}(\mathcal{V}_{t-p+1,t}))$ as an input. Note that in step (ii) c), only U_{t+k+1} changes; all other entries on the left remain unchanged by the definition of the (inverse) Rosenblatt transform.

5.2. Prediction

With the ability to simulate conditionally on the past, it easy to compute predicitions for all sorts of quantities, like conditional means or quantiles. Suppose there is an *identifying function* ψ_μ, such that the quantity of interest μ^* can be found as solution of the estimating equation

$$\mathrm{E}_{(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*)}\{\psi_{\mu^*}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+k}) \mid \boldsymbol{X}_{t-1} = \boldsymbol{x}_{t-1}, \dots, \boldsymbol{X}_{t-p} = \boldsymbol{x}_{t-p}\} = 0,$$

where $\mathrm{E}_{(\boldsymbol{\eta}*, \boldsymbol{\theta}*)}$ denotes expectation with respect to the pseudo-true model defined in Section 4.1. This framework covers conditional means, probabilities, quantiles, expectiles, and more; see, e.g., Nagler and Vatter (2020b).

The conditional expectation above can be approximated to an arbitrary accuracy by Monte Carlo integration. Using Algorithm 1, simulate N *iid* replicates $(\boldsymbol{X}_t^{(i)}, \dots, \boldsymbol{X}_{t+k}^{(i)})_{i=1}^N$ from the (estimated) conditional distribution of $(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+k})$ given $\boldsymbol{X}_{t-1} = \boldsymbol{x}_{t-1}, \dots, \boldsymbol{X}_{t-p} = \boldsymbol{x}_{t-p}$ with parameters $(\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\theta}})$ (as defined in Section 4.1). Then define the estimator $\widehat{\mu}$ as the solution of the approximate equation

$$\frac{1}{N} \sum_{i=1}^{N} \psi_{\widehat{\mu}}(\boldsymbol{X}_t^{(i)}, \dots, \boldsymbol{X}_{t+k}^{(i)}) = 0. \tag{5}$$

The following results account for the fact that we simulate from an estimated model.

Theorem 6. *Suppose $N \to \infty$ and that conditions (A1)–(A7) and (A9)–(A10) hold. Then $\widehat{\mu} \to_p \mu^*$.*

Theorem 7. *Suppose $n = o(N)$ and that conditions (A1)–(A10) hold. Then $\widehat{\mu} - \mu^* = O_p(n^{-1/2})$ and*

$$\sqrt{n}(\widehat{\mu} - \mu^*) \to_d \mathcal{N}\big(\mathbf{0}, \boldsymbol{\kappa}(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*, \mu^*)^\top \boldsymbol{J}^{-1}(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*) \boldsymbol{I}(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*) \boldsymbol{J}^{-1}(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*)^\top \boldsymbol{\kappa}(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*, \mu^*)\big),$$

where $\boldsymbol{I}, \boldsymbol{J}$ are defined in Appendix B.1 and $\boldsymbol{\kappa}$ in Appendix B.6.

Simulation-based prediction from vine copula models is used widely in the last decade, despite a lack of theoretical justification. A consistency result for extreme quantile estimation in semiparametric *iid* models was previously established by (Gong et al., 2015, Theorem 1). In contrast, the results above focus on fully parametric models and a allow for a generic predicition target. In addition, Theorem 7 characterizes a distributional limit for such predicitions. This is of practical importance because it allows to properly assess estimation/prediction uncertainty.

In principle, also the vectors $\boldsymbol{\kappa}(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*, \mu^*)$ can be computed using numerical differentiation. But because the prediction method involves simulation, this technique is computa-

tionally demanding. Parametric or block bootstrap methods (Genest and Rémillard, 2008, Künsch, 1989) are general alternatives for approximating the asymptotic distribution, but require to refit the model many times and, hence, are similarly demanding. We propose a more convenient bootstrap method based on the asymptotic distribution of the parameter vector:

(i) Simulate *iid* random vectors $(\widehat{\boldsymbol{\eta}}_r, \widehat{\boldsymbol{\theta}}_r), r = 1, \dots, R$ from the asymptotic distribution given in Theorem 5.

(ii) For each $r = 1, \dots, R$, compute $\widehat{\mu}_r$ using (5), where $\boldsymbol{X}_t^{(i)}, \dots, \boldsymbol{X}_{t+k}^{(i)}, i = 1, \dots, N$ are simulated conditionally from the model with parameters $(\widehat{\boldsymbol{\eta}}_r, \widehat{\boldsymbol{\theta}}_r)$.

The simulations in the rth step should be done independently of those in other steps. Using the same arguments as in the proof of Theorem 7, we can show that $\widehat{\mu}_r$ has the same limiting distribution as $\widehat{\mu}$. In particular, the proposed bootstrap provides valid inferences even under a misspecified model (in contrast to the classical parametric bootstrap).

Theorem 8. *If $n = o(N)$ and conditions (A1)–(A10) hold,*

$$\sqrt{n}(\widehat{\mu}_1 - \widehat{\mu}, \dots, \widehat{\mu}_r - \widehat{\mu}) \to_d (Z_1, \dots, Z_R),$$

where $Z_1, \dots, Z_R$ are independent copies of a random variable with the same distribution as in Theorem 7.

6. Numerical experiments

To validate the methodology proposed in the previous sections, we work with the following default setup: We repeatedly simulate time series of length $n = 100, 400, 1\,000, 2\,500$ from models of varying size (cross-sectional dimension $d = 5, 10$, Markov order $p = 1, 2$), containing either only Gaussian or only Gumbel pair-copulas and standard normal margins. The pair-copula parameters are set such that the Kendall's τ in tree k equals τ_0/k with either $\tau_0 = 0.2$ (weak dependence) or $\tau_0 = 0.7$ (strong dependence). This decay in dependence strength serves two purposes. First, it provides controls the strength of dependence for implicit pairs. Our specific choice produces models where the implicit pairs have a Kendall's τ of approximately the same magnitude than the explicit ones. Second, it reflects the common practice to work with vine structure that capture strongest dependencies early (see Section 4.2).

6.1. Parameter estimation

The stepwise estimation procedure of Section 4.1 is validated in Figure 7. The x-axis shows the sample size and the y-axis the average estimation error for a rescaled version of Kendall's τ: $\sum_k \sum_{e \in E_k} |k\widehat{\tau}_e - \tau_0| / \sum_k |E_k|$. The rescaling of $\widehat{\tau}_e$ by k ensures that errors have comparable scale across trees. Both axes are on a logarithmic scale, so that $\sqrt{n}$-convergence would show as a straight line with slope $-1/2$. As a guide, a dotted line with slope $-1/2$ is added to each plot. We see that $\sqrt{n}$-convergence is achieved in all scenarios.

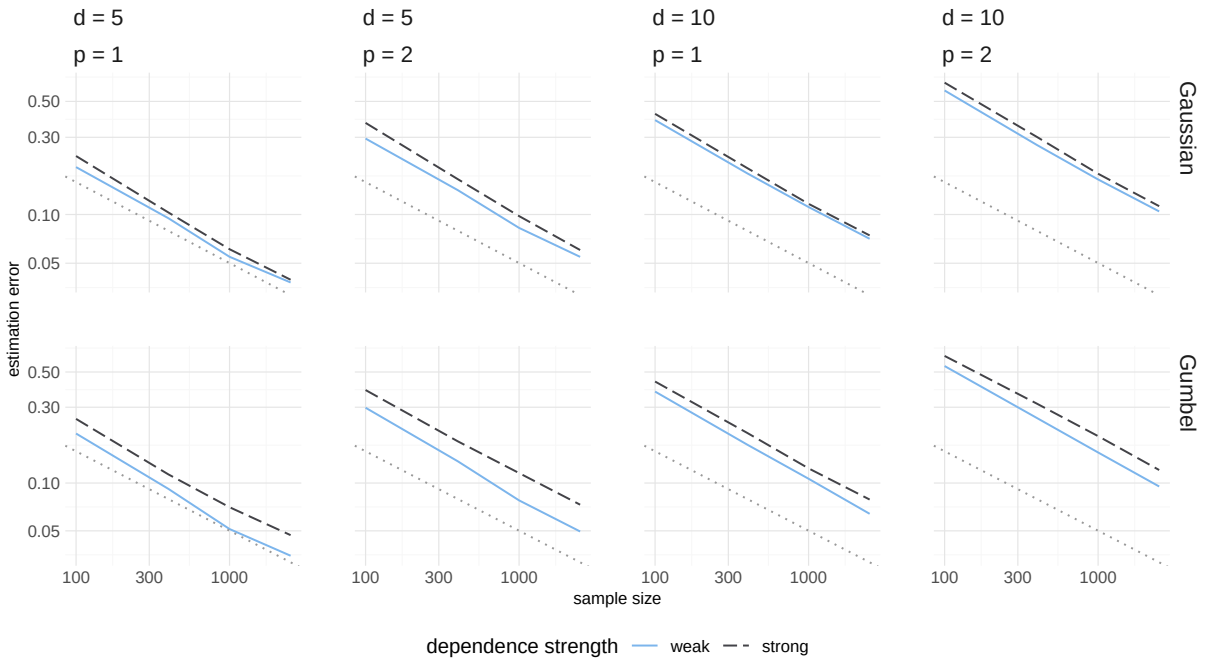

Figure 7: Estimation error $\sum_k \sum_{e \in E_k} |k\widehat{\tau}_e - \tau_0| / \sum_k |E_k|$ of the stepwise estimator in an S-vine model with Gaussian margins. Both axes have logarithm scale. The errors decay with rate $n^{-1/2}$ if they are parallel to the dotted line.

6.2. Uncertainty quantification

To validate the bootstrap method proposed in Section 5.2, we simulate from the same models as before, but now the goal is to estimate the conditional 90%-quantile of $\mu_t = (X_{t,1}, \cdots + X_{t,d})/d$ given the past. We fit the models, estimate μ_t with (5), and construct 90%-confidence intervals using the bootstrap method. The figure shows the coverage probability on the y-axis with the target level of 0.9 indicated by the dotted line. In most scenarios the coverage is approximately right, even a bit conservative for Gaussian models. An exception are the larger models consisting of Gumbel copulas with strong dependence (bottom row, dashed lines). Here the coverage is too low but appears to catch up as the sample size increases.

6.3. Model selection

To assess the model selection heuristic proposed in Section 4.2, we use a randomized version of the default setup. In each iteration, we draw a random S-vine structure. For the pair-copulas, we draw with equal probability from the Gaussian and Gumbel copulas and rotate them randomly at 0, 90, 180 and 270 with equal probability. The parameters of pair copulas are set according to a randomly drawn Kendall's τs from a Beta distribution with parameters $\alpha = 5\tau_0/(k - \tau_0)$ and $\beta = 5$ with $\tau_0 = 0.2$ (weak dependence) or $\tau_0 = 0.7$ (strong dependence). The expected value of Kendall's τ in tree k is then τ_0/k as before. We generate a time series from the model and run the selection heuristics outlined above, allowing for all parametric families implemented in `rvinecopulib` (Nagler and Vatter,

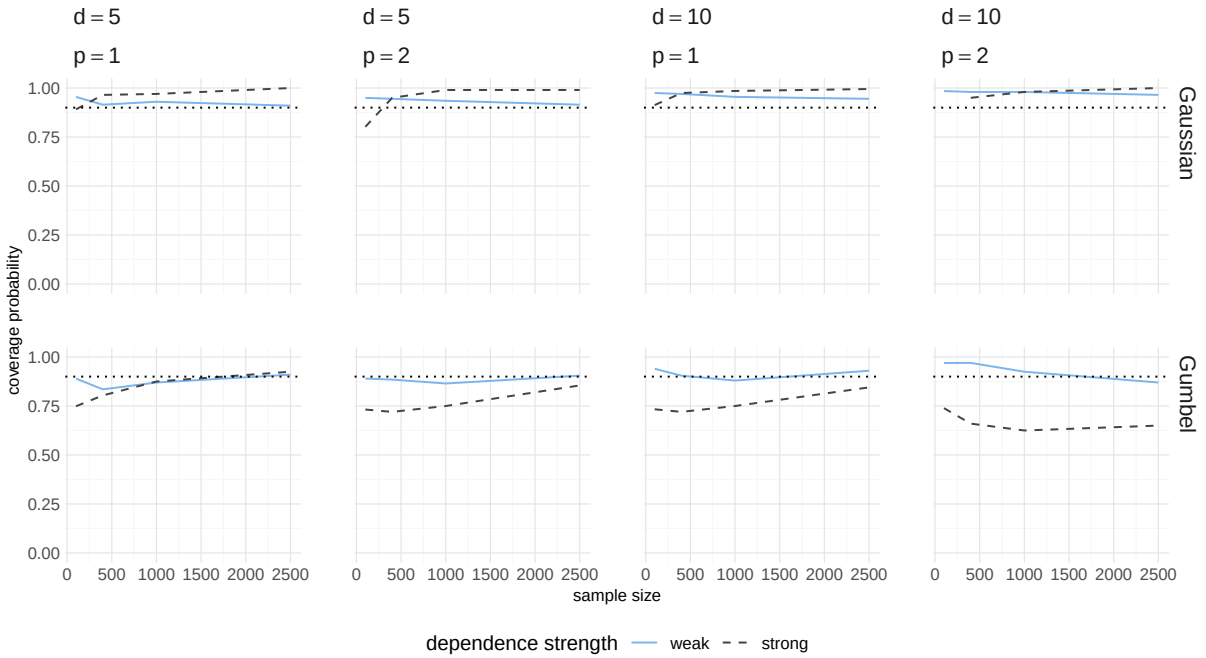

Figure 8: Coverage probabilities of bootstrapped 90%-confidence intervals for estimating the 90%-quantile of $(X_{t,1} + \cdots + X_{t,d})/d$.

2020a).

The x-axis in Figure 9 shows the sample size and the y-axis the KL-divergence between the true and estimate models. Again both axes are on a logarithmic scale and we would expect to see straight lines in a correctly specified model. However, none of the lines are straight, indicating that the heuristics rarely identify the structure that generated the data. This underlines how important it is to base inference on results and tools that allow for misspecification. The M- and D-vine models are hardly distinguishible, but the S-vine model performs best in all scenarios. The gap increases with increasing dependence strength and sample size. This suggests that the added flexibility of the general stationary model can make a difference in applications.

7. Application

Vine copula models are widely used in finance, in particular for modeling cross-sectional dependence in time series of financial returns (Aas, 2016). The most common approach is to model marginal series with ARMA/GARCH-models and the cross-sectional dependence of their residuals with a vine copula. Stationary vine copula models are different; they incorporate both serial and cross-sectional dependence in a single vine copula model.

We consider daily stock returns of 20 companies retrieved from Yahoo Finance[1]. These companies belong to several industry branches and can be found in Table 2. The data covers the time slot from 1st January 2015 until 31st December 2019, containing in total

[1]https://de.finance.yahoo.com/

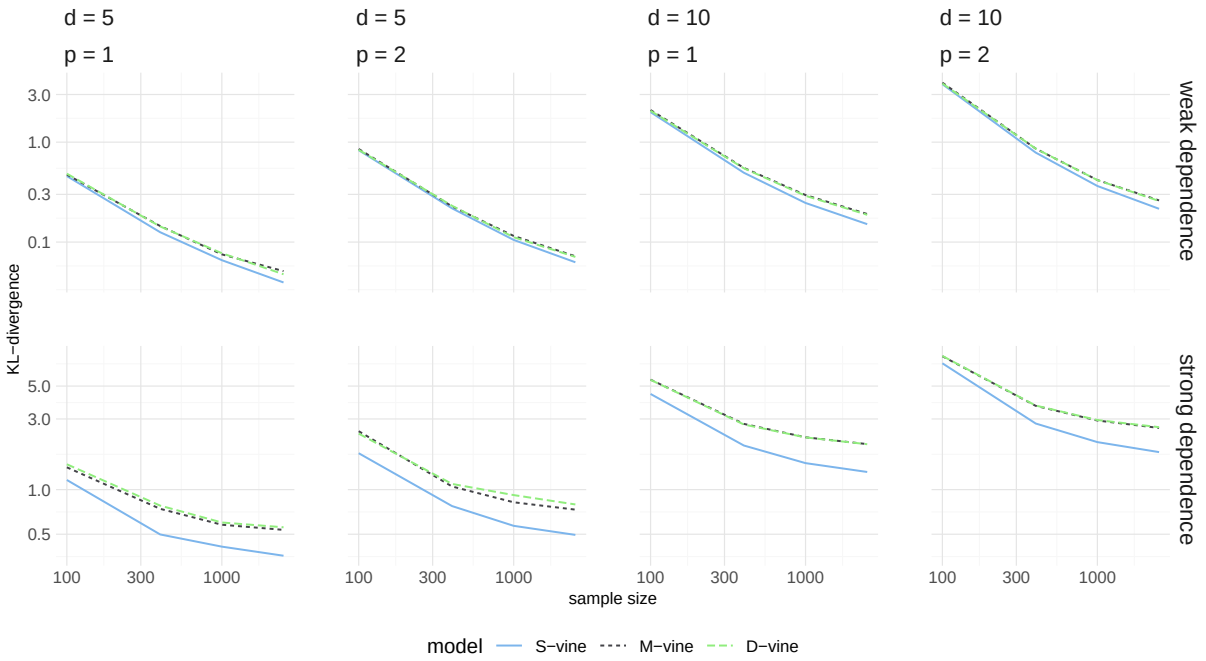

Figure 9: Averaged empirical KL-divergence for S-, M- and D-vine models

Coding	Company	Industry Branch	Coding	Company	Industry Branch
1	Allianz	Insurance	11	Microsoft	IT
2	AXA	Insurance	12	Apple	IT
3	Generali	Insurance	13	Amazon	IT/Consumer goods
4	MetLife	Insurance	14	Alphabet	IT
5	Prudential	Insurance	15	Alibaba	IT/Consumer goods
6	Ping An	Insurance	16	Exxon	Oil and gas
7	BMW	Automotive	17	Shell	Oil and gas
8	General Motors	Automotive	18	PetroChina	Oil and gas
9	Toyota	Automotive	19	Airbus	Aerospace
10	Hyundai	Automotive	20	Boeing	Aerospace

Table 2: Companies, their coding, and industry branches.

1296 trading days.

7.1. In-sample analysis

We start with an in-sample illustration of models fit to the whole data set. We first fit skew-t distributions to the individual time series of each company. We then apply the probability integral transform to obtain *pseudo-observations* of the copula model. We consider the M-vine, D-vine, and a general stationary (S-)vine models from the previous section, each with Markov orders $p = 1$. (Higher order models were fit in preliminary experiments, but did not improve fit/performance.)

In Figure 10 we illustrate the first trees of the M- and D-vine obtained via the previously described approach. We observe that the cross-sectional D-vine for both approaches is

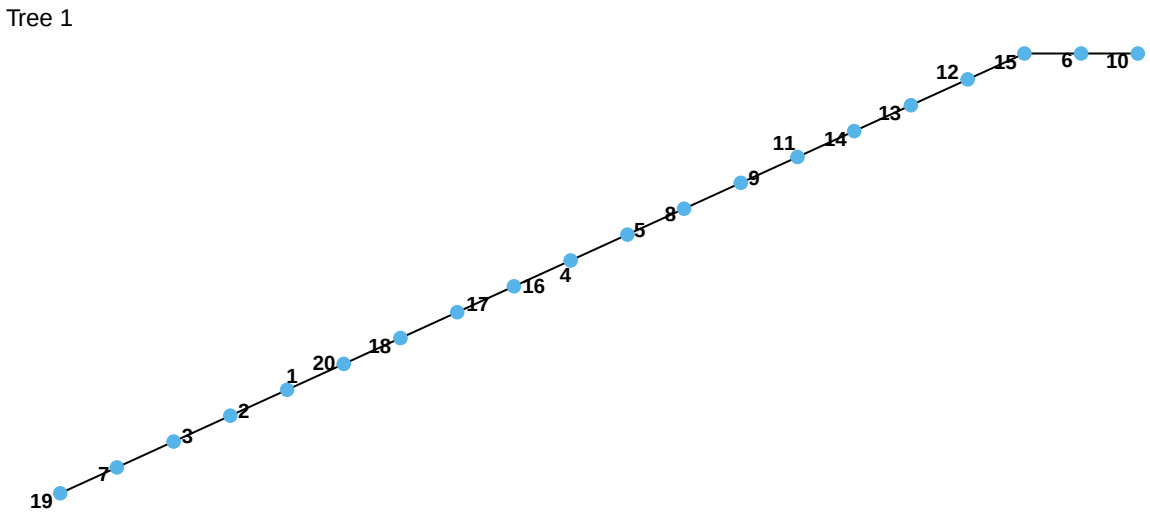

Figure 10: First tree of M- and D-vines fitted on the whole data set. Trees across time-steps are connected at $(i_1, j_1) = (8, 8)$ (with $\widehat{\tau} \approx 0.02$) for the M-vine , and $(i_1, j_1) = (19, 10)$ (with $\widehat{\tau} \approx 0.05$) for the D-vine.

described by a path $19 - \cdots - 10$. The M-vine makes the serial connection by an edge linking the same stock from time t to time $t + 1$. In this case, the connection is $(i_1, j_1) = (19, 19)$ (General Motors→General Motors) which has an empirical Kendall's τ of 0.02. The only other viable choice would have been $(10, 10)$ (Hyundai→Hyundai), but it had a lower Kendall's τ of around 0.01. The D-vine connects two opposites ends of the path, here from Hyundai (19) to General Motors (10) ($\widehat{\tau} = 0.05$).

The corresponding tree of the S-vine can be seen in Figure 11. The cross-sectional connection is described by a regular vine. We can identify some clusters of industry branches: IT (variables 11–15), insurance (1–5), and oil and gas (16–18). Interestingly, regional factors seem to be more important than the branch for aerospace and automotive stocks, however. The European manufacturers BMW (7) and Airbus (19) are attached to the European insurance cluster (1–3). American counterparts General Motors (8) and Boeing (20) are linked to the American insurances MetLife and Prudential (4, 5). Some of these links can also be identified from the M-/D-vine structure in Figure 10, but not as prominently. This plus in interpretability is one of the big advantages of using general R-vines as the cross-sectional structure.

The inter-serial connection of the S-vine is made at $(i_1, j_2) = (15, 6)$ (Alibaba→Ping An) with an empirical Kendall's τ of 0.16. The dependence here is much stronger than for the serial connections of the M- and D-vine models. This reflects the greater flexibilty of the S-vine model. Recall that compatibility does not restrict the connection in the first tree. We are thus free to choose from all possible in-/out- pairs. The linking edge is interesting in itself. First, it links to different companies across subsequent time points. Hence, this dependence must be stronger than any inter-serial dependence of a single stock. Second, it links Alibaba, a Chinese IT/Consumer goods company, to Ping An, a Chinese insurance company, which makes sense economically. Further, this link did not appear in the cross-sectional parts of either of the vine models. So while the cross-sectional

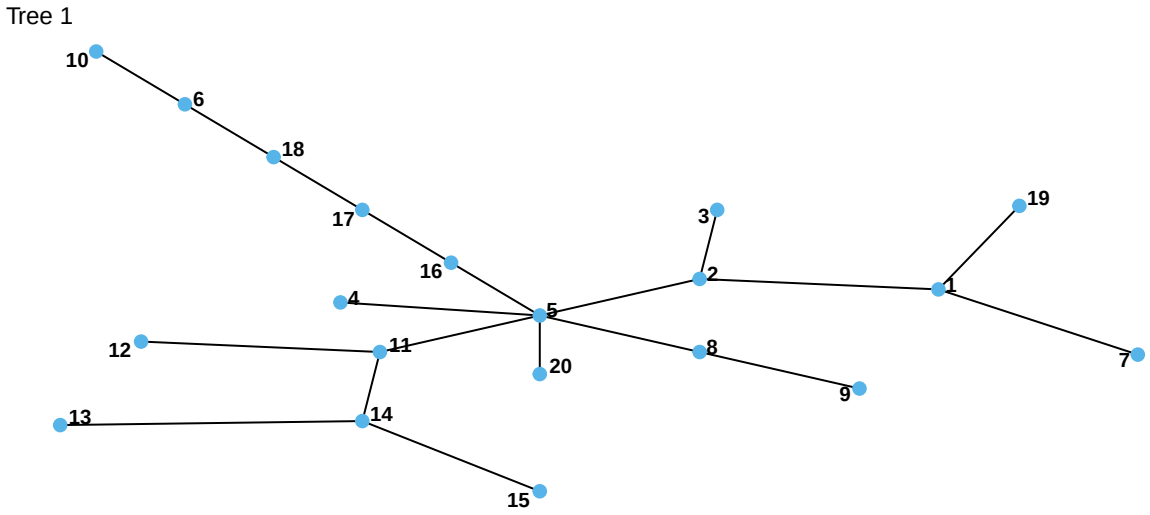

Figure 11: First tree of an S-vine fitted on the whole data set. Trees across time-steps are connected at $(i_1, j_1) = (15, 6)$ (with $\widehat{\tau} \approx 0.16$).

S-vine	M-vine	D-vine	VAR	GARCH-vine	DCC-GARCH
-163 371	-163 257	-163 258	-156 360	-162 279	-159 857

Table 3: Aikaike's information criterion for the three vine copula time series models.

dependence between the companies is comparably weak, their inter-temporal dependence is still quite strong.

The fit of the models is compared by AIC in Table 3. We also include three popular competitor models: a *vector autoregressive* (VAR) model of order 1 estimated by the `vars` R package (Pfaff, 2008), a combination of ARMA-GARCH marginal models (with skew-t residuals) with a vine copula for the residuals (using `rugarch` and `rvinecopulib`, Ghalanos, 2020, Nagler and Vatter, 2020a), and the DCC-GARCH model of Engle and Sheppard (2001) based on a multivariate t distribution (using `rmgarch`, Ghalanos, 2019). The ARMA-GARCH orders are selected for each marginal series individually by AIC. The VAR model clearly performs worst, since it cannot account for heteroscadicity. We further see that the vine models outperform the GARCH-vine and DCC-GACH models. The S-vine provides the best fit.

7.2. Out-of-sample predictions

We now compare the forecasting abilities by a backtest. We fit all models on three years' data (one year has 252 trading days). On each of the following days, we make predictions for the cumulative portfolio return over the next day or week and compare them to the observed data. Every half year the models are fitted again on three years' data.

Our predictions take the form of a Monte-Carlo sample drawn from the predictive distribution. They are evaluated with three types of measures:

- `CRPS`: The *continuous ranked probability score* of Gneiting and Raftery (2007).

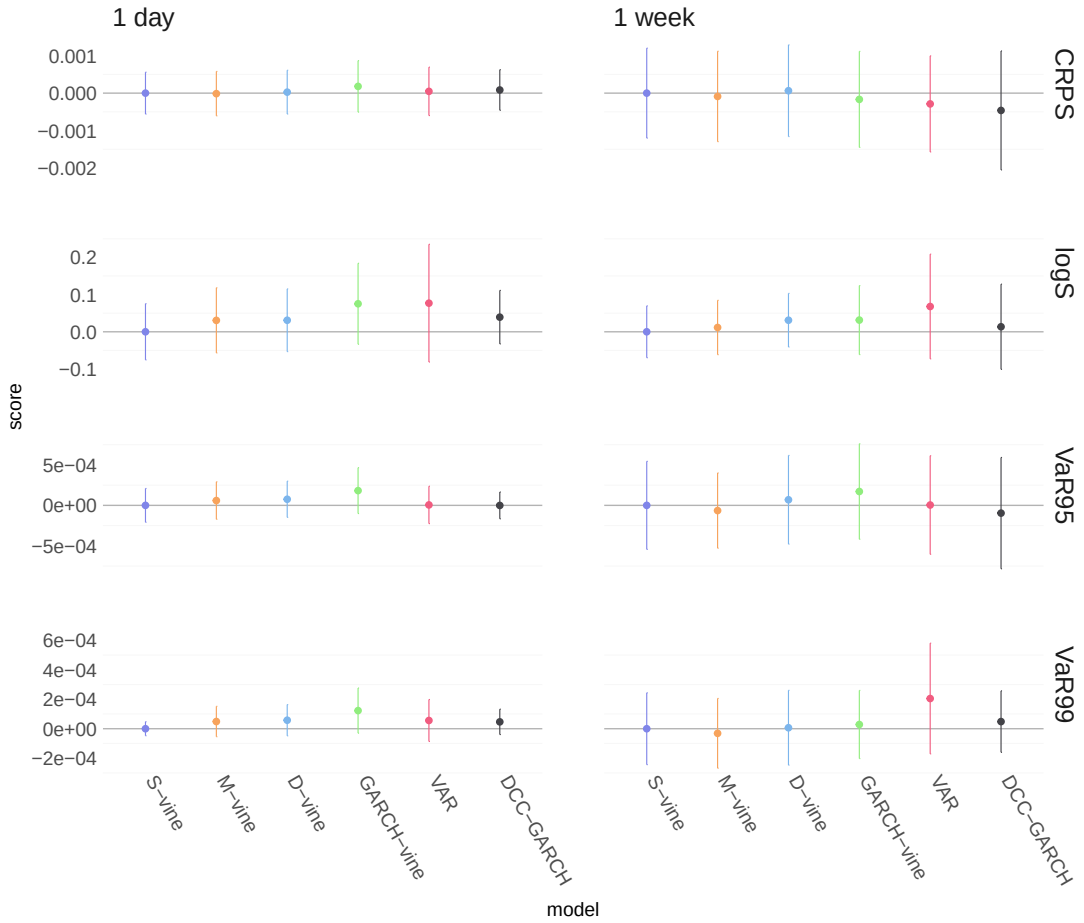

Figure 12: Forecast performance of various time series models. Dots are mean performance, error bars indicate 90%-confidence intervals (accounting for 30 lags of autocorrelation). The left panel orresponds to 1-day-ahead, the right to 1-week-ahead forecasts. Scores are centered such that S-vine has score 0.

- `logS`: The negative predictive log-likelihood.
- `VaR95`, `VaR99`: The *check-loss* known from quantile-regression (e.g., Koenker and Xiao, 2002) computed for predicted quantiles at levels 0.05 and 0.01. Such quantiles are popular risk measures in banking and insurance, where they are called *Value-at-Risk (VaR)*.

`CRPS` and `logS` are computed with the `scoringRules` R package (Jordan et al., 2019), VaRs as empirical quantiles of the Monte-Carlo sample. The measures are averaged across 1000 randomly sampled portfolios. The first 19 weights are drawn uniformly from a Uniform$(-0.15, 0.25)$ distribution and the 20th set such that weights sum up to one.

The forecast performance is shown in Figure 12. The dots are the average measure over the full period, the errorbars indicate 90%-confidence intervals (adjusted for serial dependence). The left panel corresponds to 1-day-ahead, the right to 1-week-ahead forecasts. Scores are centered such that S-vine has score 0. Some observations:

- A general observation is that uncertainty (as indicated by the confidence intervals) is rather larger compared to the differences between models. Everything that follows should therefore be taken with a grain of salt.

- The three stationary vine models perform similarly in all scenarios. The S-vine and M-vine tend to perform slightly better than the long D-vine. The S-vine is uniformly best for 1-day-ahead forecasts, but slightly outperformed by the M-vine for 1-week-ahead forecasts (except for `logS`). After all, the vine structures are found by heuristics and there is no guarantee that the best is found.

- For 1-day-ahead forecasts, the S-vine performs best for all measures, especially for extreme quantiles and the predicitive log-likelihood (`logS`). For 1-week-ahead forecasts it is outperformed by most models for `CRPS` and for `VaR95` by the DCC-GARCH. It compares favorably for the other measures.

We conclude that stationary vine models provide good forecasts for financical time series. This is somewhat remarkable since, in contrast to the vine models, the GARCH-vine and DCC-GARCH models were specifically designed for such data. Recently, some copula families have been specifically designed for modeling serial dependence in economic time series (e.g., Bladt and McNeil, 2020, Loaiza-Maya et al., 2018), but were not used in this article. We expect that incorporating such families will lead to a further increase in performance.

8. Discussion

This work deals with vine copula models for the joint distribution of a stationary time series. We derived the maximal class of vine structures that guarantee stationarity under practicable conditions. The underlying principle is intuitive: we start with a vine model for the dependence at a specific time point and connect copies of this model serially in a way that preserves time ordering. This class includes previously proposed models of Beare and Seo (2015) and Smith (2015) as special cases. The COPAR model of Brechmann and Czado (2015) was shown to be inadequate in this sense because it fails to guarantee stationarity under simple conditions. The simulations and application suggest that the added flexibility leads to improvements over the previous models. Another benefit is the greater interpretability of the model strucutre. But more importantly, our contribution gives a final answer in the search for vine copula models suitable for stationary time series.

Building on earlier ideas, we developed methods for parameter estimation, model selection, simulation, prediction, and uncertainty quantification in such models. All methods are designed with computational efficiency in mind, such that the full modeling pipeline runs in no more than a few minutes on a customary laptop. Except for (the hopeless case of) model selection, we provide theoretical justifications for the methodology in the form of asymptotic results. To the best of our knowledge, these are the first results applicable to vine copula models under serial dependence. Even when specialized to the *iid* case, they extend the existing literature in several ways. This provides post-hoc

justification for what is already practiced widely: stepwise estimation and simulation-based inference in fully parametric, but usually misspecified R-vine models. The proposed bootstrap procedure appears to be new and shall prove useful beyond the time series.

Despite confirmatory numerical experiments, a limitation of the results is an assumption on the mixing properties of the time series (required only for the asymptotic distribution). Judging from earlier work in a narrower context, we do not believe this poses a serious issue. However, we do not yet know any easily verifiable sufficient conditions. Investigating the mixing properties of stationary vine copulas — and multivariate copula models more generally — shall be an interesting path for future research.

A. Proofs of graph theoretic results

A.1. Proof of Theorem 1

It is easy to see that (ii) together with translation invariance implies that the model is stationary. Now consider the reverse implication. Condition (2) is obviously true for $m = n - 1$ and $t = n - m = 1$. Now take $m = n - 2$.

We first show that $\mathcal{V}_{1,n-1}$ must be a vine, i.e., a sequence of trees satisfying the proximity condition. The proximity condition cannot be violated because $\mathcal{V}$ is a vine. We thus need to show that the restriction $\mathcal{V}_{1,n-1} = (V_{1,n-1,k}, E_{1,n-1,k})_{k=1,\dots,(n-1)d-1}$ is a sequence of trees. Trees are connected, acyclic graphs. $\mathcal{V}_{1,n-1}$ cannot contain cycles, because it is constructed from deleting nodes and edges out of the sequence of trees $\mathcal{V}$. It thus remains to show that the graphs $(V_{1,n-1,k}, E_{1,n-1,k})_{k=1,\dots,(n-1)d-1}$ are connected. We will prove this by contradiction.

Let $k \geq 1$ be the smallest level where the graph $(V_{1,n-1,k}, E_{1,n-1,k})$ is not connected and denote by E_k the corresponding edge set of the complete vine $\mathcal{V}$. Recall that the kth graph (V_k, E_k) of the whole vine $\mathcal{V}$ is a tree and therefore connected. For the restricted graph to be disconnected, (V_k, E_k) must contain a path $P = (e_1, \dots, e_{\ell+1})$, $1 \leq \ell \leq d$, connecting vertices $(v, v_1, \dots, v_\ell, v')$ with $v, v' \in V_{1,n-1,k}$ and $v_1, \dots, v_\ell \notin V_{1,n-1,k}$. Suppose that all pair-copulas in trees below and above the kth level are independence copulas. Then the joint density of $(U_v, U_{v'})$ can be written as

$$\int_{[0,1]^\ell} c_{e_1}(u_v, u_{v_1}) \times \cdots \times c_{e_{\ell+1}}(u_{v_\ell}, u_{v'}) du_{v_1} \cdots du_{v_\ell}. \tag{6}$$

Now consider the density of $(U_w, U_{w'}) = (U_{v+(1,0)}, U_{v'+(1,0)})$. Since (V_k, E_k) is connected, there must be another path $P' = (e'_1, \dots, e'_{L+1})$,$L \geq 0$, connecting vertices $(w, w_1, \dots, w_L, w')$. Hence, the joint density of $(U_{v+(1,0)}, U_{v'+(1,0)})$ is

$$\int_{[0,1]^L} c_{e'_1}(u_w, u_{w_1}) \times \cdots \times c_{e'_{L+1}}(u_{w_L}, u_{w'}) du_{w_1} \cdots du_{w_L}. \tag{7}$$

For the model to be stationary, (6) and (7) must be equal for all $(u_v, u_{v'}) = (u_w, u_{w'}) \in (0,1)^2$. Because there is no time point $n+1$, the edges e'_1 and e'_{L+1} cannot be translations of e_1 and $e_{\ell+1}$. Therefore, translation invariance of $\mathcal{C}(\mathcal{V})$ does not imply equality of (6)

and (7), which contradicts our premise. Hence, $\mathcal{V}_{1,n-1}$ must be a vine. That $\mathcal{V}_{2,n-2}$ is also a vine follows from symmetric arguments.

It remains to show that $\mathcal{V}_{1,n-1} \sim \mathcal{V}_{2,n}$. If $\mathcal{V}_{1,n-1} \nsim \mathcal{V}_{2,n}$, then there is an edge $e_1 \in E_{k,1,n-1}$ for which there is no translation in $E_{k,2,n}$. Similarly, there must be an edge $e_2 \in E_{k,2,n}$ for which there is no translation in $E_{k,1,n-1}$. Suppose all edges in $\mathcal{V}$ except the translations of e_1 and e_2 are independence copulas. Then the joint densities of $(\boldsymbol{U}_1, \dots, \boldsymbol{U}_{n-1})$ and $(\boldsymbol{U}_2, \dots, \boldsymbol{U}_N)$ are

$$c_{1,n-1}(\boldsymbol{u}) = \prod_{e \sim e_1} c_e(u_{a_e}, u_{b_e}), \quad \text{and} \quad c_{2,n}(\boldsymbol{u}) = \prod_{e \sim e_2} c_e(u_{a_e}, u_{b_e}).$$

For all $\boldsymbol{u} \in [0,1]^{d(n-1)}$, it must hold $c_{1,n-1}(\boldsymbol{u}) = c_{2,n}(\boldsymbol{u})$, but because $e_1 \nsim e_2$, this is not ensured by translation invariance, which is a contradiction. Hence, $\mathcal{V}_{1,n-1} \sim \mathcal{V}_{2,n}$.

We have shown that (ii) holds for $m = n-2$. This and (i) also imply that (i) holds for vine copula models $(V_{1,n-1}, \mathcal{C}(V_{1,n-1}))$ and $(\mathcal{V}_{2,n}, \mathcal{C}(V_{2,n}))$. Since these are the only properties we used, we can inductively show that (ii) must hold for all $m = n-3, \dots, 1$. □

A.2. Proof of Theorem 2

It is easy to check that a vine satisfying (i)–(ii) is stationary. It is sufficient to specify the edges of the first d trees, since the dth tree is a path. This fact and the proximity condition fix all edges in later trees.

Now consider the reverse implication. Let $\mathcal{V}$ be a stationary vine. In the first tree of $\mathcal{V}$ there can only be one edge between cross-sectional trees at adjacent time points, since E_1 must not contain cycles. We denote this edge by $\{(t, i_1), (t+1, j_1)\}$. Now suppose that there is some $k \leq d-1$ such that (ii) holds for all k' with $1 \leq k' \leq k$. We will show that it must also hold for $k+1$.

First observe that the edges

$$\bigcup_{r=1}^{k} \Big\{ e \colon a_e = (t, i_{k+1-r}), b_e = (t+1, j_r), D_e = \bigcup_{s=1}^{k-r} \{(t, i_s)\} \cup \bigcup_{s=1}^{r-1} \{(t+1, j_s)\} \Big\} \subset E_k$$

form a path. This together with the proximity condition fixes edges $\{e_{r,t} \colon r = 2, \dots, k, t = 1, \dots, n-1\} \subset E_{k+1}$ with

$$a_{e_r} = (t, i_{k+2-r}), \quad b_{e_r} = (t+1, j_r), \quad D_{e_r} = \bigcup_{s=1}^{k+1-r} \{(t, i_s)\} \cup \bigcup_{s=1}^{r-1} \{(t+1, j_s)\}.$$

For E_{k+1} to form a tree, there must be two more edges: one connects the path to an edge in $E_{k+1}^{(0)} + (t, 0)$, the other connects it to an edge in $E_{k+1}^{(0)} + (t+1, 0)$. By the proximity condition, these edges must have the form

$$a_e = (t, i_{k+1}), \quad b_e = (t+1, j_1), \quad D_e = \bigcup_{s=1}^{k} \{(t, i_s)\}$$

and

$$a_{e'} = (t, i_1), \quad b_{e'} = (t+1, j_{k+1}), \quad D_{e'} = \bigcup_{s=1}^{k} \{(t+1, j_s)\},$$

where $i_{k+1} \notin \bigcup_{s=1}^{k} \{i_s\}$ and $j_{k+1} \notin \bigcup_{s=1}^{k} \{j_s\}$. Furthermore, these edges are only permitted by the proximity condition if $E_k^{(0)}$ contains edges e, e' with

$$\begin{aligned} a_e &= i_{k+1}, \quad & b_e &= i_r, \quad & D_e &= \{i_1, \dots, i_k\} \setminus i_r, \\ a_{e'} &= j_{k+1}, \quad & b_{e'} &= j_{r'}, \quad & D_{e'} &= \{j_1, \dots, j_k\} \setminus j_{r'}. \end{aligned}$$

for some $r, r' \in \{1, \dots, k\}$. Hence (ii) holds for all $1 \leq k \leq d$ and the two permutations must be compatible with $\mathcal{V}^{(0)}$. □

A.3. Proof of Lemma 1

Let $i_1 \in \{1, \dots, d\}$ be arbitrary. Because $\mathcal{V}$ is a vine, its first tree clearly contains an edge with conditioned set $\{i_1, i_2\}$ for some $i_2 \neq i_1$. Now suppose we have found $2 \leq k < d-1$ indices $i_1, \dots, i_k$ indices that do not violate the condition in Definition 8. In particular, there is $e \in E_{k-1}$ with conditioned set $\{i_k, i_r\}$ and conditioning set $\{i_1, \dots, i_{k-1}\} \setminus i_r$ for some $r \in \{1, \dots, k-1\}$. In the $(k+1)$th level, this edge becomes a vertex. Because $\mathcal{V}$ is a vine, there must be an edge leaving this vertex. This edge must have conditioned set $\{i, j\}$ and conditioning set $\{i_1, \dots, i_{k-1}\} \setminus i$ for some $j \notin \{i_1, \dots, i_k\}$ and $i \in \{i_k, i_r\}$. Setting $i_{k+1} = j$, we see that the condition is also satisfied for $k+1$. □

B. Proofs of asymptotic results

B.1. Notation

We first need some additional notation. Recall that $\widehat{U}_{t,j} = F_j(X_{t,j}; \widehat{\boldsymbol{\eta}}_j)$ and write

$$F_{a_e|D_e}(X_{a_e} \mid \boldsymbol{X}_{D_e}; \widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\theta}}_{S(a_e)}) = C_{a_e|D_e}(\widehat{U}_{a_e} \mid \widehat{\boldsymbol{U}}_{D_e}; \widehat{\boldsymbol{\theta}}_{S(a_e)}),$$

so we can express the sum in (4) as

$$\sum_{e \sim e'} \ln c_{[e]}\big\{F_{a_e|D_e}(X_{a_e} \mid \boldsymbol{X}_{D_e}; \widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\theta}}_{S(a_e)}), F_{b_e|D_e}(X_{b_e} \mid \boldsymbol{X}_{D_e}; \widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\theta}}_{S(b_e)}); \boldsymbol{\theta}_{[e']}\big\}. \tag{8}$$

Let us rewrite this in a more convenient way. Recall that each vertex in the first tree of the vine is identified by a tuple $(t, j) \in \{1, \dots, n\} \times \{1, \dots, d\}$. For all $e \sim e'$, each vertex (t, j) appearing in the conditioned or conditioning set of e must have a corresponding vertex (t', j) playing an equivalent role in e' for some $t' \neq t$. Let $e_1 \sim e'$ be the edge where a time index 1 appears in the complete union and let K be the largest time index

appearing. Defining

$$a_{e,t} = a_{e_1} + (t-1,0), \quad b_{e,t} = b_{e_1} + (t-1,0), \quad D_{e,t} = D_{e_1} + (t-1,0),$$

we denote

$$\begin{aligned} &\chi_{t,[e]}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+K}; \boldsymbol{\theta}_{[e]}, \boldsymbol{\eta}, \boldsymbol{\theta}_{S(a_e,b_e)}) \\ &= \ln c_{[e]}\big\{F_{a_{e,t}|D_{e,t}}(X_{a_{e,t}} \mid \boldsymbol{X}_{D_{e,t}}; \boldsymbol{\eta}, \boldsymbol{\theta}_{S(a_{e,t})}), F_{b_{e,t}|D_{e,t}}(X_{b_{e,t}} \mid \boldsymbol{X}_{D_{e,t}}; \boldsymbol{\eta}, \boldsymbol{\theta}_{S(b_e)}); \boldsymbol{\theta}_{[e]}\big\}, \end{aligned} \tag{9}$$

where $S(a_e, b_e) = S(a_e) \cup S(b_e)$. Now the pseudo-true values $(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*)$ are defined iteratively as

$$\boldsymbol{\eta}^* = \arg\max_{\boldsymbol{\eta}} \sum_{j=1}^{d} \mathbb{E}\{\ln f_j(X_{t,j}; \boldsymbol{\eta}_j)\},$$

and for $k = 1, \dots, (p+1)d - 1$,

$$\boldsymbol{\theta}^*_{[E_k]} = \arg\max_{\boldsymbol{\theta}_{[e]}} \sum_{[e] \in [E_k]} \mathbb{E}\{\chi_{t,[e]}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+K}; \boldsymbol{\theta}_{[e]}, \boldsymbol{\eta}^*, \boldsymbol{\theta}^*_{S(a_e,b_e)})\}.$$

Define further

$$\boldsymbol{\phi}_t(\boldsymbol{\eta}, \boldsymbol{\theta}) = \boldsymbol{\phi}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+p}; \boldsymbol{\eta}, \boldsymbol{\theta}) = \begin{pmatrix} \big(\nabla_{\boldsymbol{\eta}_j} \ln f_j(X_{t,j}; \boldsymbol{\eta}_j)\big)_{j=1}^{d} \\ \big(\nabla_{\boldsymbol{\theta}_{[e]}} \chi_{t,[e]}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+K}; \boldsymbol{\theta}_{[e]}, \boldsymbol{\eta}, \boldsymbol{\theta}_{S(a_e,b_e)})\big)_{[e] \in [E_1]} \\ \vdots \\ \big(\nabla_{\boldsymbol{\theta}_{[e]}} \chi_{t,[e]}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+K}; \boldsymbol{\theta}_{[e]}, \boldsymbol{\eta}, \boldsymbol{\theta}_{S(a_e,b_e)})\big)_{[e] \in [E_{(p+1)d-1}]} \end{pmatrix}$$

$$\boldsymbol{J}(\boldsymbol{\eta}, \boldsymbol{\theta}) = \mathbb{E}\big\{\nabla^{\top}_{(\boldsymbol{\eta},\boldsymbol{\theta})} \boldsymbol{\phi}_t(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+p}; \boldsymbol{\eta}, \boldsymbol{\theta})\big\}$$

$$\boldsymbol{I}(\boldsymbol{\eta}, \boldsymbol{\theta}) = \sum_{t=-\infty}^{\infty} \mathbb{C}\mathrm{ov}\big\{\boldsymbol{\phi}_0(\boldsymbol{\eta}, \boldsymbol{\theta}), \boldsymbol{\phi}_t(\boldsymbol{\eta}, \boldsymbol{\theta})\big\}.$$

B.2. Assumptions

(A1) The series $(\boldsymbol{X}_t)_{t=1}^{\infty}$ is ergodic.

(A2) The pseudo-true values $(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*)$ lie in the interior of the compact parameter space $\mathcal{H} \times \Theta \subseteq \mathbb{R}^s$ for some $s \in \mathbb{N}$.

(A3) For some $\delta > 0$,

$$\sum_{j=1}^{d} \mathbb{E}\{\ln f_j(X_{t,j}; \boldsymbol{\eta}_j^*)\} > \sup_{\|\boldsymbol{\eta} - \boldsymbol{\eta}^*\| > \delta} \sum_{j=1}^{d} \mathbb{E}\{\ln f_j(X_{t,j}; \boldsymbol{\eta}_j)\},$$

and $k = 1, \dots, (p+1)d - 1$,

$$\sum_{[e]\in[E_k]} \mathbb{E}\{\chi_{t,[e]}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+K}; \boldsymbol{\theta}^*_{[e]}, \boldsymbol{\eta}^*, \boldsymbol{\theta}^*_{S(a_e,b_e)})\}$$
$$> \sup_{\|\boldsymbol{\theta}_{[E_k]} - \boldsymbol{\theta}^*_{[E_k]}\| > \delta} \sum_{[e]\in[E_k]} \mathbb{E}\{\chi_{t,[e]}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+K}; \boldsymbol{\theta}_{[e]}, \boldsymbol{\eta}^*, \boldsymbol{\theta}^*_{S(a_e,b_e)})\}.$$

(A4) The matrix $\boldsymbol{J}(\boldsymbol{\eta}, \boldsymbol{\theta})$ is continuous in a neighborhood of $(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*)$.

(A5) There are functions G_j such that for all $j = 1, \dots, d$, $\mathbb{E}\{G_j(X_{t,j})\} < \infty$ and

$$|\ln f_j(X_{t,j}; \boldsymbol{\eta}_j) - \ln f_j(X_{t,j}; \boldsymbol{\eta}'_j)| \le G_j(X_{t,j}) \|\boldsymbol{\eta}_j - \boldsymbol{\eta}'_j\|,$$

for all $\boldsymbol{\eta}_j, \boldsymbol{\eta}'_j \in \mathcal{H}_j$.

(A6) There are functions $G_{[e]}$ such that for all edges e in the vine, $\mathbb{E}\{G_{[e]}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+K})\} < \infty$ and

$$|\chi_{t,[e]}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+K}; \boldsymbol{\theta}_{[e]}, \boldsymbol{\eta}^*, \boldsymbol{\theta}^*_{S(a_e,b_e)}) - \chi_{t,[e]}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+K}; \boldsymbol{\theta}'_{[e]}, \boldsymbol{\eta}^*, \boldsymbol{\theta}^*_{S(a_e,b_e)})|$$
$$\le G_{[e]}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+K}) \|\boldsymbol{\theta}_{[e]} - \boldsymbol{\theta}'_{[e]}\|,$$

and

$$\sup_{(\boldsymbol{\eta}, \boldsymbol{\theta}_{[e]}, \boldsymbol{\theta}_{S(a_e,b_e)})} \left\|\nabla_{(\boldsymbol{\eta}, \boldsymbol{\theta}_{S(a_e,b_e)})} \chi_{t,[e]}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+K}; \boldsymbol{\theta}_{[e]}, \boldsymbol{\eta}, \boldsymbol{\theta}_{S(a_e,b_e)})\right\| \le G_{[e]}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+K}).$$

(A7) The map $\boldsymbol{\phi}_t$ is continuously differentiable in a neighborhood of $(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*)$. Further, there is a function G such that $\mathbb{E}\{G(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+K})\} < \infty$ and

$$\|\nabla^\top_{(\boldsymbol{\eta}, \boldsymbol{\theta})} \boldsymbol{\phi}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+p}; \boldsymbol{\eta}, \boldsymbol{\theta}) - \boldsymbol{\phi}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+p}; \boldsymbol{\eta}', \boldsymbol{\theta}')|$$
$$\le G_{[e]}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+K}) \|(\boldsymbol{\eta}, \boldsymbol{\theta}) - (\boldsymbol{\eta}', \boldsymbol{\theta}')\|$$

for all $(\boldsymbol{\eta}, \boldsymbol{\theta}), (\boldsymbol{\eta}', \boldsymbol{\theta}')$ in some δ-ball around $(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*)$.

(A8) For some $\gamma > 2$, the sequence $\boldsymbol{X}_t$ is β-mixing with rate at least $t^{-\gamma/(\gamma-2)}$ and $\mathbb{E}\{\|\phi_t(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*)\|^\gamma\} < \infty$.

(A9) There is a strictly positive function Ψ such that $\Psi(\boldsymbol{x}) \ge \sup_{\mu\in\mathcal{M}} |\psi_\mu(\boldsymbol{x})|$ for all $\boldsymbol{x} \in \mathbb{R}^{dk}$ and for some $\delta > 0$, $\gamma > 2$,

$$\sup_{\|(\boldsymbol{\eta}, \boldsymbol{\theta}) - (\boldsymbol{\eta}^*, \boldsymbol{\theta}^*)\| < \delta} \mathbb{E}_{(\boldsymbol{\eta}, \boldsymbol{\theta})}\{\Psi(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+k})^\gamma\} < \infty,$$

where $\mathbb{E}_{(\boldsymbol{\eta}, \boldsymbol{\theta})}$ denotes expectation with respect to the probability measure $\mathbb{P}_{(\boldsymbol{\eta}, \boldsymbol{\theta})}$ induced by the vine copula model with parameters $(\boldsymbol{\eta}, \boldsymbol{\theta})$. Furthermore, there is

$s < 2$ such that for every $\epsilon > 0$,

$$\sup_{\|(\boldsymbol{\eta},\boldsymbol{\theta})-(\boldsymbol{\eta}^*,\boldsymbol{\theta}^*)\|<\delta} \ln N_{[]}\big(\epsilon, \{\psi_\mu \colon \mu \in \mathcal{M}\}, L_2(\mathbb{P}_{(\boldsymbol{\eta},\boldsymbol{\theta})})\big) \leq \epsilon^{-s}.$$

For a definition of the bracketing number $N_{[]}$, see e.g., van der Vaart and Wellner (1996, Definition 2.1.6).

(A10) The map $(\boldsymbol{\eta}, \boldsymbol{\theta}, \mu) \mapsto \mathbb{E}_{(\boldsymbol{\eta},\boldsymbol{\theta})}\{\psi_\mu(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+k})\}$ is twice continuously differentiable in a neighborhood of $(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*, \mu^*)$ and $|\nabla_\mu \mathbb{E}_{(\boldsymbol{\eta}^*,\boldsymbol{\theta}^*)}\{\psi_{\mu^*}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+k})\}| > 0$.

B.3. Proof of Theorem 4

For $j = 1, \dots, d$, let

$$M_{n,j}(\boldsymbol{\eta}_j) = \frac{1}{n}\sum_{t=1}^{n} \ln f_j(X_{t,j}; \boldsymbol{\eta}_j), \quad M_j(\boldsymbol{\eta}_j) = \mathbb{E}\{\ln f_j(X_{t,j}; \boldsymbol{\eta}_j)\}.$$

Since $\mathcal{H}_j$ is compact and $\ln f_j$ continuous by (A2) and (A7), $M_{n,j}$ has a maximum for every n and, by definition, $M_{n,j}(\widehat{\boldsymbol{\eta}}_j) \geq \sup_{\boldsymbol{\eta}_j} M_n(\boldsymbol{\eta}_j)$. We will show that $\sup_{\boldsymbol{\eta}_j} |M_n(\boldsymbol{\eta}_j) - M(\boldsymbol{\eta}_j)| \to_p 0$.

Let $\epsilon > 0$ be arbitrary. Since $\mathcal{H}_j$ is a compact subset of a Euclidean space, there are finitely many $\boldsymbol{\eta}_j^{(1)}, \dots, \boldsymbol{\eta}_j^{(L)} \in \mathcal{H}_j$ such that for any $\boldsymbol{\eta}_j \in H_j$, there is ℓ such that $\|\boldsymbol{\eta}_j - \boldsymbol{\eta}_j^{(\ell)}\| < \epsilon$. Then (A5) and the triangle inequality imply

$$\begin{aligned}
&\sup_{\boldsymbol{\eta}_j} |M_{n,j}(\boldsymbol{\eta}_j) - M_j(\boldsymbol{\eta}_j)| \\
&\leq \max_{1\leq\ell\leq L} |M_{n,j}(\boldsymbol{\eta}_j^{(\ell)}) - M_j(\boldsymbol{\eta}_j^{(\ell)})| + \frac{1}{n}\sum_{t=1}^{n} G_j(X_{t,j})\epsilon + \mathbb{E}\{G_j(X_{t,j})\}\epsilon.
\end{aligned}$$

Because $(X_{t,j})_{t=1}^n$ is ergodic by (A1), it holds

$$M_{n,j}(\boldsymbol{\eta}_j^{(\ell)}) \to_p M_j(\boldsymbol{\eta}_j^{(\ell)}), \qquad \frac{1}{n}\sum_{t=1}^{n} G_j(X_{t,j}) \to_p \mathbb{E}\{G_j(X_{t,j})\}.$$

Thus $\sup_{\boldsymbol{\eta}_j} |M_{n,j}(\boldsymbol{\eta}_j) - M_j(\boldsymbol{\eta}_j)| \to_p \bar{M} \leq 2\mathbb{E}\{G(X_{t_j})\}\epsilon$ in probability, and because ϵ was arbitrary, it must hold $\sup_{\boldsymbol{\eta}_j} |M_{n,j}(\boldsymbol{\eta}_j) - M_j(\boldsymbol{\eta}_j)| \to_p 0$. Then by (A3) and Corollary 3.2.3 of van der Vaart and Wellner (1996), $\widehat{\boldsymbol{\eta}}_j \to_p \boldsymbol{\eta}_j^*$.

Similar arguments apply for the pair-copula parameters. The only complication is that (4) involves previously estimated parameters. Suppose we want to estimate a parameter $\boldsymbol{\theta}_{[e']}$ with $e' \in E_k, k \geq 1$. Let either $k = 1$ or $k \geq 2$ with $\widehat{\boldsymbol{\theta}}_{[e']} \to_p \boldsymbol{\theta}_{[e']}$ for all $e' \in E_1, \dots, E_{k-1}$. We want to estimate a parameter $\boldsymbol{\theta}_{[e]}$ with $e \in E_k$. Using the notation

in (9) and a Taylor expansion, we see that (8) equals

$$\begin{aligned}
&\sum_{t=1}^{n-K} \chi_{t,[e]}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+K}; \boldsymbol{\theta}_{[e]}, \widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\theta}}_{S(a_e,b_e)}) \\
&= \sum_{t=1}^{n-K} \chi_{t,[e]}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+K}; \boldsymbol{\theta}_{[e]}, \boldsymbol{\eta}^*, \boldsymbol{\theta}^*_{S(a_e,b_e)}) \\
&\quad + \sum_{t=1}^{n-K} \left\| \nabla_{(\boldsymbol{\eta}, \boldsymbol{\theta}_{S(a_e,b_e)})} \chi_{t,[e]}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+K}; \boldsymbol{\theta}_{[e]}, \boldsymbol{\eta}', \boldsymbol{\theta}'_{S(a_e,b_e)}) \right\| \times o_p(1),
\end{aligned}$$

for some $(\boldsymbol{\eta}', \boldsymbol{\theta}_{S(a_e,b_e)})$ on the segment between $(\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\theta}}_{S(a_e,b_e)})$ and $(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*_{S(a_e,b_e)})$. Because $\boldsymbol{X}_t$ is ergodic and using (A6), we see that the second term above is of order $o_p(n)$ uniformly in $\boldsymbol{\theta}_{[e]}$.

Now defining

$$M_n(\boldsymbol{\theta}_{[e]}) = \frac{1}{n} \sum_{t=1}^{n-K} \chi_{t,[e]}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+K}; \boldsymbol{\theta}_{[e]}, \boldsymbol{\eta}^*, \boldsymbol{\theta}^*_{S(a_e,b_e)}), \quad M_{[e]}(\boldsymbol{\theta}_{[e]}) = \mathbb{E}\{M_{n,[e]}(\boldsymbol{\theta}_{[e]})\},$$

we can use (A6) and proceed as for the marginal estimates to show

$$\sup_{\boldsymbol{\theta}_{[e]} \in \Theta_{[e]}} |M_{n,[e]}(\boldsymbol{\theta}_{[e]}) - M_{[e]}(\boldsymbol{\theta}_{[e]})| \to_p 0$$

and $\widehat{\boldsymbol{\theta}}_{[e]} \to_p \boldsymbol{\theta}^*_{[e]}$. □

B.4. Proof of Theorem 5

Since the maxima are over smooth functions, (3) implies

$$\sum_{t=1}^{n} \nabla_{\boldsymbol{\eta}_j} \ln f_j(X_{t,j}; \widehat{\boldsymbol{\eta}}_j) = 0,$$

and (4) implies

$$\frac{1}{n} \sum_{t=1}^{n} \nabla_{\boldsymbol{\theta}_{[e]}} \chi_{t,[e]}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+K}; \widehat{\boldsymbol{\theta}}_{[e]}, \widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\theta}}_{S(a_e,b_e)}) = 0, \tag{10}$$

where $\chi_{t,[e]}$ was defined in (9). (Technically, the sum above should contain $n - K$ terms, but the difference is asymptotically negligible and ignored in what follows). Then the estimator $(\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\theta}})$ can be expressed as the solution of the system of equations

$$\frac{1}{n} \sum_{t=1}^{n} \boldsymbol{\phi}_t(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+p}; \widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\theta}}) = \mathbf{0}.$$

A Taylor expansion yields

$$\begin{aligned}\mathbf{0} = &\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{\phi}_t(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+p}; \boldsymbol{\eta}^*, \boldsymbol{\theta}^*) \\ &+ \frac{1}{n}\sum_{t=1}^{n} \nabla^\top_{(\boldsymbol{\eta},\boldsymbol{\theta})} \boldsymbol{\phi}_t(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+p}; \boldsymbol{\eta}', \boldsymbol{\theta}') \begin{pmatrix} \widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^* \\ \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \end{pmatrix},\end{aligned}$$

for some $(\boldsymbol{\eta}', \boldsymbol{\theta}')$ on the segment between $(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*)$ and $(\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\theta}})$.

With similar arguments as in the proof of Theorem 4 and (A7), we can show that

$$\sup_{\|(\boldsymbol{\eta},\boldsymbol{\theta})-(\boldsymbol{\eta}^*,\boldsymbol{\theta}^*)\|<\delta} \left\| \frac{1}{n}\sum_{t=1}^{n} \nabla^\top_{(\boldsymbol{\eta},\boldsymbol{\theta})} \boldsymbol{\phi}_t(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+p}; \boldsymbol{\eta}, \boldsymbol{\theta}) - \boldsymbol{J}(\boldsymbol{\eta}, \boldsymbol{\theta}) \right\| \to_p 0,$$

for some $\delta > 0$. Since furthermore $\boldsymbol{J}(\boldsymbol{\eta}, \boldsymbol{\theta})$ is continuous and $(\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\theta}}) \to_p (\boldsymbol{\eta}^*, \boldsymbol{\theta}^*)$, it holds

$$\frac{1}{n}\sum_{t=1}^{n} \nabla^\top_{(\boldsymbol{\eta},\boldsymbol{\theta})} \boldsymbol{\phi}_t(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+p}; \boldsymbol{\eta}', \boldsymbol{\theta}') \to_p \boldsymbol{J}(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*),$$

Hence,

$$\begin{pmatrix} \widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^* \\ \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \end{pmatrix} = -\boldsymbol{J}^{-1}(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*) \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{\phi}_t(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+p}; \boldsymbol{\eta}^*, \boldsymbol{\theta}^*)\{1 + o_p(1)\}.$$

Then the result follows from (A8), Slutsky's lemma, and the central limit theorem for mixing random variables (e.g., Bosq, 2012, Theorem 1.7). □

B.5. Proof of Theorem 6

Observe that the random vectors $\boldsymbol{X}_t^{(i)}$ appearing in (5) come from a model with random parameter values $(\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\theta}})$. Let us make that more explicit in by writing (5) as

$$\mathbb{P}^{(N)}_{(\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\theta}})} \psi_{\widehat{\mu}} = 0,$$

where $\mathbb{P}^{(N)}_{(\boldsymbol{\eta},\boldsymbol{\theta})}$ is the empirical measure corresponding to N *iid* samples from the model model $\mathbb{P}_{(\boldsymbol{\eta},\boldsymbol{\theta})}$. By (A9) and Theorem 2.8.4 of van der Vaart and Wellner (1996), the average above converges to its expectation, uniformly in μ and the model $\mathbb{P}_{(\boldsymbol{\eta},\boldsymbol{\theta})}$ generating $\boldsymbol{X}_t^{(i)}$. More precisely, for any $\epsilon > 0$ and some $\delta > 0$,

$$\lim_{N\to\infty} \sup_{\|(\boldsymbol{\eta},\boldsymbol{\theta})-(\boldsymbol{\eta}^*,\boldsymbol{\theta}^*)\|<\delta} \mathbb{P}_{(\boldsymbol{\eta},\boldsymbol{\theta})}\left(N^{1/2} \sup_{\mu\in\mathcal{M}} \left|\mathbb{P}^{(N)}_{(\boldsymbol{\eta},\boldsymbol{\theta})}\psi_\mu - \mathbb{P}_{(\boldsymbol{\eta},\boldsymbol{\theta})}\psi_\mu\right| > \epsilon \right) = 0,$$

where

$$\mathbb{P}_{(\boldsymbol{\eta},\boldsymbol{\theta})}\psi_\mu = \mathbb{E}_{(\boldsymbol{\eta},\boldsymbol{\theta})}\big\{\psi_\mu(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+k}\big\}.$$

So (5) is equivalent to solving

$$\mathbb{P}_{(\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\theta}})}\psi_{\widehat{\mu}} + O_p(N^{-1/2}) = 0. \tag{11}$$

Using $N \to \infty$ and Theorem 4 yields

$$\mathbb{P}_{(\boldsymbol{\eta}^*,\boldsymbol{\theta}^*)}\psi_{\mu^*} + \big(\nabla_\mu \mathbb{P}_{(\boldsymbol{\eta}^*,\boldsymbol{\theta}^*)}\psi_{\mu^*}\big)(\widehat{\mu} - \mu^*) + o_p(1) = 0$$

which implies $\widehat{\mu} \to_p \mu^*$ because $\mathbb{P}_{(\boldsymbol{\eta}^*,\boldsymbol{\theta}^*)}\psi_{\mu^*} = 0$ by definition and $\big|\nabla_\mu \mathbb{P}_{(\boldsymbol{\eta}^*,\boldsymbol{\theta}^*)}\psi_{\mu^*}\big| > 0$ by (A10). □

B.6. Proof of Theorem 7

We continue from (11), but use one more term in the expansion. With $n = o(N)$ and $\|(\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\theta}}) - (\boldsymbol{\eta}^*,\boldsymbol{\theta}^*)\|^2 = o_p(n^{-1/2})$ (Theorem 5), we get

$$\mathbb{P}_{(\boldsymbol{\eta}^*,\boldsymbol{\theta}^*)}\psi_{\mu^*} + \big(\nabla_{(\boldsymbol{\eta},\boldsymbol{\theta})}\mathbb{P}_{(\boldsymbol{\eta}^*,\boldsymbol{\theta}^*)}\psi_{\mu^*}\big)\begin{pmatrix}\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^* \\ \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\end{pmatrix} + \big(\nabla_\mu \mathbb{P}_{(\boldsymbol{\eta}^*,\boldsymbol{\theta}^*)}\psi_{\mu^*}\big)(\widehat{\mu} - \mu^*) + o_p(n^{-1/2}) = 0,$$

or equivalently,

$$\sqrt{n}(\widehat{\mu} - \mu^*) = \big(\nabla_\mu \mathbb{P}_{(\boldsymbol{\eta}^*,\boldsymbol{\theta}^*)}\psi_{\mu^*}\big)^{-1}\big(\nabla_{(\boldsymbol{\eta},\boldsymbol{\theta})}\mathbb{P}_{(\boldsymbol{\eta}^*,\boldsymbol{\theta}^*)}\psi_{\mu^*}\big)\begin{pmatrix}\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^* \\ \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\end{pmatrix} + o_p(1).$$

Defining

$$\boldsymbol{\kappa}(\boldsymbol{\eta},\boldsymbol{\theta},\mu) = \big(\nabla_\mu \mathbb{P}_{(\boldsymbol{\eta}^*,\boldsymbol{\theta}^*)}\psi_{\mu^*}\big)^{-1}\big(\nabla_{(\boldsymbol{\eta},\boldsymbol{\theta})}\mathbb{P}_{(\boldsymbol{\eta}^*,\boldsymbol{\theta}^*)}\psi_{\mu^*}\big),$$

the claim then follows from Theorem 5 and Slutsky's lemma. □

References

Aas, K. (2016). Pair-copula constructions for financial applications: A review. *Econometrics 4*(4), 43.

Aas, K., C. Czado, A. Frigessi, and H. Bakken (2009). Pair-copula constructions of multiple dependence. *Insurance Math. Econom. 44*(2), 182–198.

Beare, B. K. (2010). Copulas and temporal dependence. *Econometrica 78*(1), 395–410.

Beare, B. K. (2012). Archimedean copulas and temporal dependence. *Econometric Theory 28*(6), 1165–1185.

Beare, B. K. and J. Seo (2014). Time irreversible copula-based markov models. *Econometric Theory 30*(5), 923–960.

Beare, B. K. and J. Seo (2015). Vine copula specifications for stationary multivariate Markov chains. *J. Time Series Anal. 36*(2), 228–246.

Bedford, T. and R. M. Cooke (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Ann. Math. Artif. Intell. 32*(1-4), 245–268. Representations of uncertainty.

Bedford, T. and R. M. Cooke (2002). Vines—a new graphical model for dependent random variables. *Ann. Statist. 30*(4), 1031–1068.

Begin, É., P. Dutilleul, C. Beaulieu, and T. Bouezmarni (2020). M-vine decomposition and var (1) models. *Statistics & Probability Letters 158*, 108660.

Bladt, M. and A. J. McNeil (2020). Time series copula models using d-vines and v-transforms: an alternative to garch modelling. *arXiv preprint arXiv:2006.11088*.

Bosq, D. (2012). *Nonparametric statistics for stochastic processes: estimation and prediction*, Volume 110. Springer Science & Business Media.

Brechmann, E. C. and C. Czado (2015). Copar—multivariate time series modeling using the copula autoregressive model. *Applied Stochastic Models in Business and Industry 31*(4), 495–514.

Brechmann, E. C., C. Czado, and K. Aas (2012). Truncated regular vines in high dimensions with application to financial data. *Canad. J. Statist. 40*(1), 68–85.

Chen, X. and Y. Fan (2006). Estimation of copula-based semiparametric time series models. *Journal of Econometrics 130*(2), 307–335.

Chen, X., W. B. Wu, and Y. Yi (2009, 12). Efficient estimation of copula-based semiparametric markov models. *Ann. Statist. 37*(6B), 4214–4253.

Cooke, R. M., D. Kurowicka, and K. Wilson (2015). Sampling, conditionalizing, counting, merging, searching regular vines. *Journal of Multivariate Analysis 138*, 4–18.

Czado, C. (2019). Analyzing dependent data with vine copulas. *Lecture Notes in Statistics, Springer*.

Darsow, W. F., B. Nguyen, and E. T. Olsen (1992). Copulas and Markov processes. *Illinois J. Math. 36*(4), 600–642.

Dissmann, J., E. C. Brechmann, C. Czado, and D. Kurowicka (2013). Selecting and estimating regular vine copulae and application to financial returns. *Comput. Statist. Data Anal. 59*, 52–69.

Engle, R. F. and K. Sheppard (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate garch. Technical report, National Bureau of Economic Research.

Genest, C. and B. Rémillard (2008, 12). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Ann. Inst. H. Poincaré Probab. Statist. 44*(6), 1096–1127.

Ghalanos, A. (2019). *rmgarch: Multivariate GARCH models.* R package version 1.3-7.

Ghalanos, A. (2020). *rugarch: Univariate GARCH models.* R package version 1.4-4.

Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association 102*(477), 359–378.

Gong, J., Y. Li, L. Peng, and Q. Yao (2015). Estimation of extreme quantiles for functions of dependent random variables. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 77*(5), 1001–1024.

Hobæk Haff, I. (2013, 05). Parameter estimation for pair-copula constructions. *Bernoulli 19*(2), 462–491.

Ibragimov, R. (2009). Copula-based dependence characterizations for higher-order Markov processes. *Econometric Theory 25*, 819–846.

Joe, H. (1996). Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In *Distributions with fixed marginals and related topics (Seattle, WA, 1993)*, Volume 28 of *IMS Lecture Notes Monogr. Ser.*, pp. 120–141. Inst. Math. Statist., Hayward, CA.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of multivariate Analysis 94*(2), 401–419.

Joe, H. (2014). *Dependence modeling with copulas.* CRC press.

Joe, H., R. M. Cooke, and D. Kurowicka (2010). Regular vines: generation algorithm and number of equivalence classes. In *Dependence Modeling: Vine Copula Handbook*, pp. 219–231. World Scientific.

Joe, H. and J. J. Xu (1996). The estimation method of inference functions for margins for multivariate models.

Jordan, A., F. Krüger, and S. Lerch (2019). Evaluating probabilistic forecasts with scoringRules. *Journal of Statistical Software 90*(12), 1–37.

Koenker, R. and Z. Xiao (2002). Inference on the quantile regression process. *Econometrica 70*(4), 1583–1612.

Kraus, D. and C. Czado (2017). D-vine copula based quantile regression. *Computational Statistics & Data Analysis 110*, 1–18.

Krüger, D. (2018). General vine copula models for stationary multivariate time series. Master thesis, Technical University of Munich.

Künsch, H. R. (1989, 09). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics 17*(3), 1217–1241.

Loaiza-Maya, R., M. S. Smith, and W. Maneesoonthorn (2018). Time series copulas for heteroskedastic data. *Journal of Applied Econometrics 33*(3), 332–354.

Longla, M. and M. Peligrad (2012). Some aspects of modeling dependence in copula-based markov chains. *Journal of Multivariate Analysis 111*, 234–240.

Morales Napoles, O., R. M. Cooke, and D. Kurowicka (2010). About the number of vines and regular vines on n nodes. Technical report, url: http://resolver.tudelft.nl/uuid:912abf55-8112-48d2-9cca-323f7f6aecc7, TU Delft.

Nagler, T. and C. Czado (2016). Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas. *Journal of Multivariate Analysis 151*, 69–89.

Nagler, T. and T. Vatter (2020a). *rvinecopulib: High Performance Algorithms for Vine Copula Modeling*. R package version 0.5.2.1.0.

Nagler, T. and T. Vatter (2020b). Solving estimating equations with copulas. *arXiv preprint arXiv:1801.10576*.

Nasri, B. R. and B. N. Rémillard (2019). Copula-based dynamic models for multivariate time series. *Journal of Multivariate Analysis 172*, 107 – 121. Dependence Models.

Nasri, B. R., B. N. Rémillard, and T. Bouezmarni (2019). Semi-parametric copula-based models under non-stationarity. *Journal of Multivariate Analysis 173*, 347 – 365.

Oh, D. H. and A. J. Patton (2017). Modeling dependence in high dimensions with factor copulas. *Journal of Business & Economic Statistics 35*(1), 139–154.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International economic review 47*(2), 527–556.

Patton, A. J. (2009). *Copula–Based Models for Financial Time Series*, pp. 767–785. Berlin, Heidelberg: Springer Berlin Heidelberg.

Patton, A. J. (2012). A review of copula models for economic time series. *Journal of Multivariate Analysis 110*, 4 – 18. Special Issue on Copula Modeling and Dependence.

Pfaff, B. (2008). Var, svar and svec models: Implementation within R package vars. *Journal of Statistical Software 27*(4).

Rémillard, B., N. Papageorgiou, and F. Soustra (2012). Copula-based semiparametric models for multivariate time series. *Journal of Multivariate Analysis 110*, 30–42.

Simard, C. and B. Rémillard (2015). Forecasting time series with multivariate copulas. *Dependence Modeling 3*(1).

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris 8*, 229–231.

Smith, M. S. (2015). Copula modelling of dependence in multivariate time series. *International Journal of Forecasting 31*(3), 815–833.

Stöber, J. (2013). *Regular vine copulas with the simplifying assumption, time-variation, and mixed discrete and continuous margins.* Ph. D. thesis, Technical University of Munich.

van der Vaart, A. W. and J. A. Wellner (1996). Weak convergence. In *Weak convergence and empirical processes*, pp. 16–28. Springer.