Stationary vine copula models for multivariate time series

Nagler, Thomas; Krüger, Daniel; Min, Aleksey

DOI
10.1016/j.jeconom.2021.11.015

Publication date
2022

Document Version
Accepted author manuscript

Published in
Journal of Econometrics

Citation (APA)
Nagler, T., Krüger, D., & Min, A. (2022). Stationary vine copula models for multivariate time series. *Journal of Econometrics*, *227*(2), 305-324. https://doi.org/10.1016/j.jeconom.2021.11.015

ARTICLE IN PRESS
Journal of Econometrics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Stationary vine copula models for multivariate time series

Thomas Nagler [a,*], Daniel Krüger [b], Aleksey Min [b]

[a] Delft Institute of Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628 CD Delft, Netherlands
[b] Department of Mathematics, Technical University of Munich, Boltzmannstr. 3, 85748, Garching, Germany

ARTICLE INFO

Article history:
Received 12 September 2020
Received in revised form 4 November 2021
Accepted 13 November 2021
Available online xxxx

Keywords:
Pair-copula
Dependence
Bootstrap
Forecasting
Markov chain
Sequential maximum likelihood

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 guarantee stationarity under a natural and verifiable condition called translation invariance. 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 portfolio of 20 stocks, where they show excellent forecast performance. The paper is accompanied by an open source software implementation.

1. Introduction

In multivariate time series there are two types of dependence: *cross-sectional* and *serial*. The first is dependence between variables at a fixed point in time. The second is dependence of two random vectors at different points in time. Copulas are general dependence models and have been used for both types. One line of research considers copula models for serial dependence in univariate Markov processes (including Darsow et al., 1992; Chen and Fan, 2006b; Chen et al., 2009; Ibragimov, 2009; Beare, 2010; Nasri et al., 2019). An orthogonal, but equally popular approach is to filter serial dependence by univariate time series models, like the ARMA-GARCH family, and model the cross-sectional dependence by a copula for the residuals (Patton, 2006; Hu, 2006; Chen and Fan, 2006a; Oh and Patton, 2017; Nasri and Rémillard, 2019; Chen et al., 2021). See also Patton (2009, 2012), Aas (2016) for surveys in the context of financial and economic time series.

Copulas can 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 (Bedford and Cooke, 2002; Aas et al., 2009) 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-copula encodes the (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 three models are quite similar. The vine graphs start with copies of a cross-sectional tree that connect variables observed at the same point in time. These trees are constrained to be either stars (Brechmann and Czado, 2015) or paths (Smith, 2015; Beare and Seo, 2015). Cross-sectional trees are then linked by a specific building plan.

* Corresponding author.
E-mail address: mail@tnagler.com (T. Nagler).

https://doi.org/10.1016/j.jeconom.2021.11.015

Inspired by the three latter works, we propose more flexible vine models for stationary time series (Section 3). But we approach the problem from the opposite direction. Previous models aim to guarantee stationarity of the model through a condition called *translation invariance* (Beare and Seo, 2015): pair-copulas stay the same when corresponding random variables are shifted in time. Translation invariance is a necessary condition for stationarity and the only practicable condition to check. We derive a characterization of the class of vines for which translation invariance is also sufficient for stationarity (Theorems 1 and 2). The class allows for general vine structures for cross-sectional dependence and leaves 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). Hence, the COPAR model is not stationary in general (see Example 1).

For practical purposes, it is convenient to restrict to Markovian models, which are easily obtained by placing independence copulas on most edges in the vine (Theorem 3). Parameters of such models can be estimated by adapting the popular sequential maximum-likelihood method to take time invariances into account. In Section 4, we show consistency and normality of parametric and semiparametric versions of this method (Theorems 4–7). By simulating from an estimated model, conditionally on the past, we can easily compute predictions. In Section 5, we translate this into a theoretical framework and prove consistency and asymptotic normality. The results also cover Monte Carlo integrals from estimated *iid* models, which are widely used but have not been formalized yet. We further propose a computationally efficient bootstrap method for both parameter estimates and predictions in Section 6. In Section 7, we apply the methodology to forecast portfolio returns, showing that our generalized models improve both performance and interpretability. Section 8 offers concluding remarks. Abstract mathematical results on general method-of-moment estimators, which empower all our main theorems, are stated in Appendix A. Additional illustrations, simulation results, and all proofs are provided in the supplementary materials. 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, 2020).

2. Multivariate time series based on vine copulas

2.1. Copulas

Copulas are models for the dependence in a random vector. By Sklar's theorem (Sklar, 1959), any multivariate distribution F of a d-dimensional random vector $\boldsymbol{X} = (X_1, \ldots, X_d)'$ with marginal distributions $F_1, \ldots, F_d$ can be expressed as

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

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

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

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

2.2. Regular vines

Vine copulas are a particularly flexible class of copula models. They are based on an idea of Joe (1996, 1997) 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)* (see, Bedford and Cooke, 2001, 2002). We shall briefly outline the basics of R-vines; for more details on R-vines, we refer to Dissmann et al. (2013), Joe (2014), Czado (2019). Additional illustrations of the following definitions can be found in Section S1 of the supplementary materials.

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 $k = 2, \ldots, d-1$, T_k is a tree with vertices $V_k = E_{k-1}$,
(iii) (*proximity condition*) for $k = 2, \ldots, d-1$: 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: $|a \cap b| = 1$.

Fig. 1 shows a graphical example of special regular vine, called D-vine. We call a vine a D-vine if each tree is a path. A tree is a path if each vertex is connected to at most two other vertices. Such a structure is most natural when there is a natural ordering (e.g., in time or space) of the variables. When one tree of the vine is a path, all trees at higher levels are fixed uniquely by the proximity condition. Another prominent sub-class are C-vines, where each tree is a star (see also

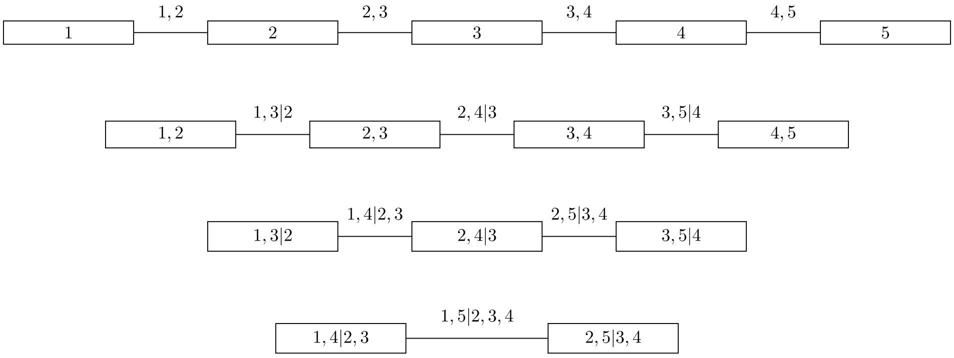

Fig. 1. A five-dimensional D-vine.

Figure S2 in the supplementary material). A tree is a star if there is one vertex that is connected to all remaining vertices. This structure is most natural when there is a single variable driving the others (e.g., a market factor driving individual stocks). In that case, the proximity condition poses no restrictions on the next tree.

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. We first need the definition of a *complete union*.

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. $\mathcal{U}_i = \{i\}$.

In other words, the complete union of an edge $e \in E_k$ is just the set of all vertices from the first tree T_1, which are involved in the iterative construction of the edge e.

Definition 3.

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

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

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

Definition 3 complements Definition 1 by relating each edge in the R-vine to a pair-wise conditional distribution, as we shall see in the following section. Note that the conditioning set D_e is empty for edges of the first tree level.

2.3. Vine copulas

By fixing the marginal distributions $F_1, \ldots, F_d$, it suffices to consider a random vector $\boldsymbol{U} = (F_1(X_1), \ldots, F_d(X_d))'$ with standard uniform margins to describe the dependence structure of $\boldsymbol{X}$. A vine copula model for $\boldsymbol{U}$ 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 $\mathcal{C}(\mathcal{V}) = \{c_e : e \in E_k, k = 1, \ldots, d-1\}$ the set of associated bivariate copulas. As an example, consider the regular vine shown in Fig. 1. The vertices in the first tree represent the random variables $U_1, \ldots, U_5$. 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 copula model for the vector $\boldsymbol{U}$ has a product form:

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

where $u_{a_e|D_e} := C_{a_e|D_e}(u_{a_e} \mid \boldsymbol{u}_{D_e})$, $\boldsymbol{u}_{D_e} := (u_l)_{l \in D_e}$ is a subvector of $\boldsymbol{u} = (u_1, \ldots, u_d) \in [0,1]^d$ and $C_{a_e|D_e}$ is the conditional distribution of U_{a_e} given $\boldsymbol{U}_{D_e}$. The functions $c_{a_e, b_e|D_e}$ are copula densities describing the dependence of U_{a_e} and U_{b_e} conditional on $\boldsymbol{U}_{D_e} = \boldsymbol{u}_e$. For every $e \in E_k$, the conditional distributions $C_{a_e|D_e}$ can be expressed recursively as

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

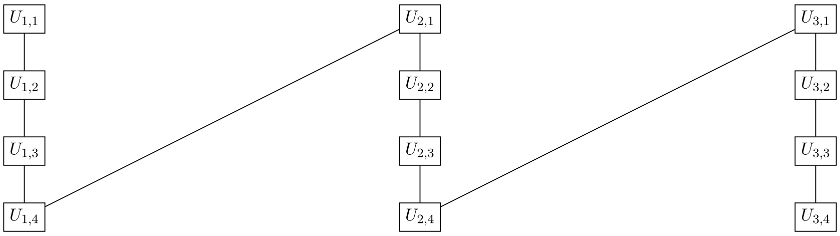

Fig. 2. Example for the first tree level of a four-dimensional D-vine on three time points.

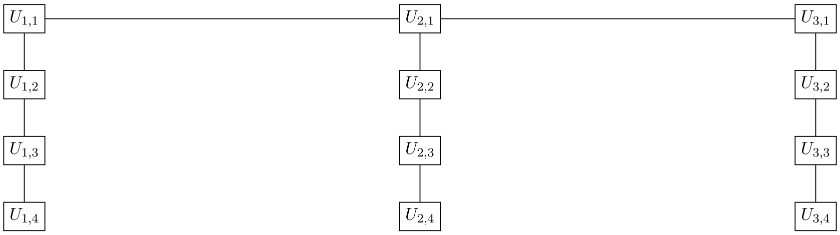

Fig. 3. Example for the first tree level of a four-dimensional M-vine on three time points.

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 $c_{a_e,b_e|D_e}$. Under this assumption, the density simplifies to

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

Since each pair-copula can be modeled separately, simplified vine copulas remain quite flexible. Conditional independence properties can be imposed by setting appropriate pair-copulas to the independence copula. This shall prove convenient when we construct Markovian time series models in Section 3.5. 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 immaterial for our theoretical results, we will stick to the simplified, continuous case for convenience. For a more extensive introduction to vine copulas, we refer to Aas et al. (2009) and Czado (2019).

2.4. Vine copula models for multivariate time series

Now suppose $(\boldsymbol{X}_t)_{t=1,\ldots,n} = (X_{t,1}, \ldots, X_{t,d})'_{t=1,\ldots,T}$ is a stationary time series, whose cross-sectional and temporal dependence structure we model by a vine copula. Throughout the paper, stationarity refers to strict stationarity. By fixing the stationary marginal distributions $F_1, \ldots, F_d$, it suffices to consider time series $(\boldsymbol{U}_t)_{t=1,\ldots,T} = (U_{t,1}, \ldots, U_{t,d})'_{t=1,\ldots,T}$ of marginally standard uniform variables, where $U_{t,1} = F_1(X_{t,1}), \ldots, U_{t,d} = F_d(X_{t,d})$. Note that the random variables $U_{t,j}$ in our model have two sub-indices. The first sub-index t indicates the time point and the second sub-index j determines the marginal variable. 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 $U_{t,i}$. In particular, the components of edge labels $e = (a_e, b_e|D_e)$ (i.e., a_e, b_e as well as elements of D_e) are tuples.

All existing regular vine models for multivariate time series follow the same idea (Beare and Seo, 2015; Smith, 2015; Brechmann and Czado, 2015). There is a vine capturing cross-sectional dependence of $\boldsymbol{U}_t \in \mathbb{R}^d$ for all time points $t = 1, \ldots, T$. 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 make specific choices for the cross-sectional structure and connecting edge. Their first tree for a four-dimensional model on three time points is illustrated in Figs. 2–4. Graphs of the first five trees $T_1, \ldots, T_5$ and additional details can be found in Section S2 of the supplementary materials. In short, there are three different models:

- *D-vine* of Smith (2015): (i) the cross-sectional structure is a D-vine, (ii) two cross-sectional D-vines at time points t and $t+1$ are connected at the two distinct vertices that lie at opposite borders of the D-vine trees. For the first tree, illustrated in Fig. 2, we can assume without loss of generality that these vertices are (t, d) and $(t+1, 1)$. With these choices, there is only one global vine model satisfying the proximity condition, which is a long D-vine spanning all variables at all time points.

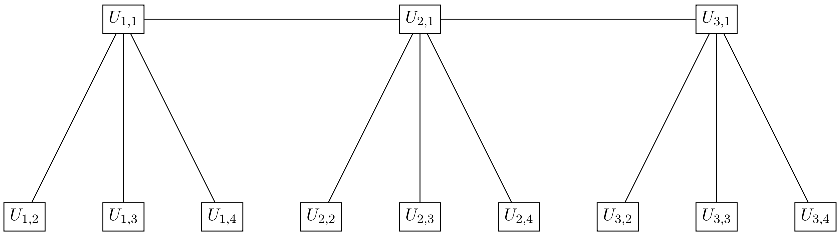

Fig. 4. Example for the first tree level of a four-dimensional COPAR on three time points.

- *M-vine* of Beare and Seo (2015): (i) the cross-sectional structure is a D-vine, (ii) two cross-sectional D-vines at time points t and $t+1$ are connected at one vertex that lies at the same border of the D-vine trees. For the first tree, illustrated in Fig. 3, we may assume without loss of generality that the vertices $(t, 1)$ and $(t+1, 1)$ are connected. With the additional restriction that vertices of adjacent time points are connected first, this also uniquely fixes all further trees of the vine.
- *COPAR* of Brechmann and Czado (2015): (i) the cross-sectional structure is a C-vine, (ii) the first trees of two C-vines at time points t and $t+1$ are connected at the root vertex of the C-vine. For the first tree, illustrated in Fig. 4, we may assume without loss of generality that vertices $(t, 1)$ and $(t+1, 1)$ are connected. 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 $U_{1,1}, U_{2,1}, \ldots, U_{T,1}$.

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 versions of such models were constructed in preliminary work by Krüger (2018) (called *temporal vine*) and in unpublished work by Harry Joe. But where should we stop? In principle, we could take any $(T \times d)$-dimensional vine as a model for the vector $(\boldsymbol{U}_1, \ldots, \boldsymbol{U}_T)$. We address this question comprehensively in the following section.

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 vine 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 Section S5 of the supplementary material.

3.1. Stationary time series

As explained in Section 2.4, any stationary time series can be transformed into one with uniform marginal distributions by the probability integral transform. To ease our exposition, we shall therefore assume uniform marginal distributions in what follows. Let $(\mathcal{V}, \mathcal{C}(\mathcal{V}))$ be a vine copula model for the random vector $(\boldsymbol{U}_1', \ldots, \boldsymbol{U}_T')' \in \mathbb{R}^{Td}$. The time series $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_T \in \mathbb{R}^d$ is strictly stationary if and only if $\boldsymbol{U}_{t_1}, \ldots, \boldsymbol{U}_{t_m}$ and $\boldsymbol{U}_{t_1+\tau}, \ldots, \boldsymbol{U}_{t_m+\tau}$ have the same joint distribution for all $1 \le t_1 < t_2 < \cdots < t_m \le T$, $1 \le \tau \le T - \max_{j=1}^m t_j$, and $1 \le m \le T$.

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*, defined below, is a necessary condition for stationarity. Recall that, in the time series context, the elements of the set V_1 of a vine are tuples (t, j) with time index $t = 1, \ldots, T$ and variable index $j = 1, \ldots, d$.

Definition 4 (*Translation Invariance*). A vine copula model $(\mathcal{V}, \mathcal{C}(\mathcal{V}))$ on the set $V_1 = \{1, \ldots, T\} \times \{1, \ldots, d\}$ is called *translation invariant* if $c_{a_e, b_e | D_e} = c_{a_{e'}, b_{e'} | D_{e'}}$ holds for all edges $e, e' \in \bigcup_{k=1}^{Td-1} E_k$ 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 $D_e = \{v + (\tau, 0) : v \in D_{e'}\}$.

Remark 1. The notation $e = e' + (\tau, 0)$ will be used short for (1) and indicates a shift in time by τ steps. For example, if $a_{e'} = (t, j)$ then $a_e = (t + \tau, j)$ and similarly for b_e and D_e.

Translation invariance was formally defined by Beare and Seo (2015), and also used as an implicit motivation for the models of Brechmann and Czado (2015) and Smith (2015). For example, in the M-vine of Beare and Seo (2015) shown in Fig. 3, the copulas associated with the edges $(1, 1) - (1, 2)$ and $(2, 1) - (2, 2)$ must be the same. As mentioned above, this is only a necessary condition for stationarity. 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. These structures are characterized in what follows.

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 $\mathcal{V} = (V_k, E_k)_{k=1}^{Td-1}$ be a vine on $\{1 \dots, T\} \times \{1, \dots, d\}$ and $V_1' = \{t, \dots, t+m\} \times \{1, \dots, d\}$ for some t, m with $1 \leq t \leq T$, $0 \leq m \leq T - t$. For all $k \geq 1$, 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 $t, \dots, t+m$.

Simply put: to restrict a vine on time points t to $t+m$, we delete all edges and vertices where time indices outside the range $[t, t+m]$ appear in the labels. 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 $\mathcal{V}_{1,2}$.

The *translation* of a vine $\mathcal{V}$ is obtained by shifting all vertices and edges in time by the same amount.

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

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

Additional illustrations of these concepts are provided in Section S1 of the supplementary material.

3.3. Stationary vines

The last definitions are key to ensure stationarity in vine copula models. If for all time points s, t and gap m, the restriction of a vine on $t, \dots, t+m$ is a translation of the restriction on $s, \dots, s+m$, translation invariance will guarantee stationarity.

Theorem 1. *Let $\mathcal{V}$ be a vine on the set $V_1 = \{1, \dots, T\} \times \{1, \dots, d\}$. 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 $\mathcal{V}^{(m)}$, $m = 0, \dots, T-1$, defined on $\{1, \dots, m+1\} \times \{1, \dots, d\}$, such that for all $m = 0, \dots, T-1$, $1 \leq t \leq T - m$,

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

An important word in condition (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 impractical, 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 for vine structures.

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

S-vines have a distinctive 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 these 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 (i.e., going from $\mathcal{V}^{(m-1)}$ to $\mathcal{V}^{(m)}$). An illustrative example for a five-dimensional S-vine on three time points is given in Section S2.5 of the supplementary materials.

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

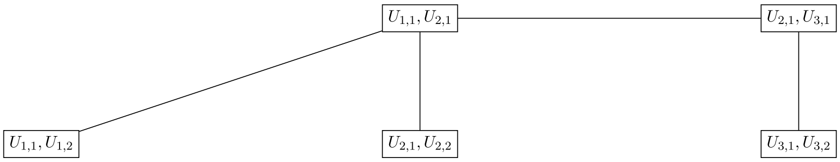

Fig. 5. Example for the second tree level of a COPAR model with $d = 2, T = 3$.

Example 1. Let us illustrate the tricky part of the proof Theorem 1 with the COPAR model for $d = 2$, $T = 3$. The second tree of the model is given in Fig. 5, see also Figure S11 in the supplement for the remaining trees. For simplicity, we assume that all pair-copulas in trees $k = 1$ and $k \geq 3$ are independence copulas. The restriction $\mathcal{V}_{2,3}$ of the model is obtained by deleting all vertices and edges where a time index 1 occurs. Clearly, $\mathcal{V}_{2,3}$ is not a vine, because the vertex $(U_{2,1}, U_{2,2})$ is disconnected from the others. Now let us see why this is problematic. The joint copula density of vertices $(U_{2,1}, U_{2,2})$ and $(U_{2,1}, U_{3,1})$ equals the product of copulas associated with the edges along the path joining them, integrating over all intermediate vertices. 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 translation 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 does not hold and the model is not stationary.

3.4. An explicit characterization of stationary vines

Stationary vines can also be characterized more explicitly. 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 adjacent time points. The first permutation, called *in-vertices*, determines how trees of one cross-sectional vine are connected to trees of the preceding time point; the second permutation, called *out-vertices*, regulates the connections to the succeeding time point. The permutations are constrained by the choice of cross-sectional structure. For simplicity, we omit the time index t in the following definition.

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 $\{i_1, \dots, i_{k-1}\}\backslash i_r$ for some $r \in \{1, \dots, k-1\}$.

The first index of the permutation (i_1) is not constrained by compatibility, but the remaining ones are. A permutation is only compatible if the vine contains the edge $\{i_2, i_1\}$ in the first tree. Further, the vine must contain an edge with either (a) conditioned set $\{i_3, i_1\}$ and conditioning set $\{i_2\}$, or (b) an edge with conditioned set $\{i_3, i_2\}$ and conditioning set $\{i_1\}$, etc. Because i_1 is unconstrained, this also implies that any d-dimensional vine has at least d compatible permutations (see Lemma 1 below). The following theorem shows that S-vines are characterized by the cross-sectional structure and two compatible permutations.

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

(i) there is a vine $\mathcal{V}^{(0)}$ on $\{0\} \times \{1, \dots, d\}$ such that $\mathcal{V}_{t,t} \sim \mathcal{V}^{(0)}$, for all $1 \leq t \leq T$,
(ii) there are two permutations $(i_1, \dots, i_d)$ and $(j_1, \dots, j_d)$ compatible with $\mathcal{V}^{(0)}$, such that

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

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

As mentioned earlier, S-vines generalize previous models:

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

	$T = 100, d = 5$	$T = 100, d = 20$	$T = 1\,000, d = 20$
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

(i) If $\mathcal{V}^{(0)}$ is a D-vine and $(i_1, \ldots, i_d) = (j_1, \ldots, j_d)$, we obtain the M-vine of Beare and Seo (2015).
(ii) If $\mathcal{V}^{(0)}$ is a D-vine and $(i_1, \ldots, i_d) = (j_d, \ldots, j_1)$, we obtain the long D-vine 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).

Compared to the models of Beare and Seo (2015) and Smith (2015), S-vines do not require the cross-sectional structure to be D-vines. Further, we have some degree of freedom in how we connect variables across different time points. This relaxation can improve both interpretability and performance of associated copula models, as illustrated in Section 7 and Section S5 of the supplementary materials. The explicit characterization of Theorem 2 also 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 any d-dimensional vine $\mathcal{V}$ and any $i_1 \in \{1, \ldots, d\}$, there exists at least one permutation $(i_1, \ldots, i_d)$ compatible with $\mathcal{V}$.*

Now the following result is an immediate consequence 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 $\mathcal{V}^{(0)}$ and two sequences of compatible in- and out-vertices, the stationary vine is unique.

3.5. Markovian models

Stationarity is a convenient property because it limits model complexity. An arbitrary vine copula model for $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_T \in [0, 1]^d$ requires to specify (or estimate) $Td(Td-1)/2 = O(T^2d^2)$ 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 $(T-1)d^2 + d(d-1)/2 = O(Td^2)$ of them, all other pair-copulas are constrained by translation invariance. When the time series contains more than a few dozen time points, this is still too much. Most popular time series models also satisfy the Markov property:

Definition 9. A time series $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_T \in [0, 1]^d$ is called Markov (process) of order p if for all $\boldsymbol{u} \in [0, 1]^d$,

$$\mathrm{P}\big(\boldsymbol{U}_t \leq \boldsymbol{u} \mid \boldsymbol{U}_{t-1}, \ldots, \boldsymbol{U}_1\big) = \mathrm{P}\big(\boldsymbol{U}_t \leq \boldsymbol{u} \mid \boldsymbol{U}_{t-1}, \ldots, \boldsymbol{U}_{t-p}\big).$$

The Markov property limits complexity further. For the M-vine model, Beare and Seo (2015, Theorem 4) showed that it is equivalent to an independence constraint on the pair-copulas, used similarly by the Markovian models of Brechmann and Czado (2015) and Smith (2015). The same arguments apply for the 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 $c_e \equiv 1$ for all $e \notin \mathcal{V}_{t,t+p}$, $t = 1, \ldots, T-p$.*

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. Parameter estimation

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 step-wise maximum likelihood estimator (Aas et al., 2009) for M-vine copula models, but without theoretical guarantees. We shall introduce such a method for the more general class of stationary vines and prove its validity, allowing for either parametric or nonparametric marginal models. The step-wise method is fast also for large models, but incurs a small loss in efficiency according to Hobæk Haff (2013).

4.1. Estimation of marginal models

We follow the common practice to estimate marginal models first. Given estimates $\widehat{F}_1, \ldots, \widehat{F}_d$ of the marginal distributions, the copula parameters can then be estimated based on 'pseudo-observations' $\widehat{U}_{t,j} = \widehat{F}_j(X_{t,j})$, $t = 1, \ldots, T$, $j = 1, \ldots d$.

Suppose we are given parametric models $f_j(\cdot; \boldsymbol{\eta}_j)$, $j = 1, \ldots, d$, for the marginal densities. Then the parameters can be estimated by the maximum-likelihood-type estimator

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

Given estimates of the marginal parameters, we then generate pseudo-observations from the copulas model via $\widehat{U}_{t,j}^{(P)} = F_j(X_{t,j}; \widehat{\boldsymbol{\eta}}_j)$, $t = 1, \ldots, T$, $j = 1, \ldots, d$.

We can also consider semiparametric copula models by estimating the marginal distributions by empirical distribution functions $\widehat{F}_j(x) = \sum_{t=1}^{T} \mathbb{1}(X_{t,j} \leq x)/(T+1)$. This leads to the pseudo-observations $\widehat{U}_{t,j}^{(SP)} = \widehat{F}_j(X_{t,j})$, $t = 1, \ldots, T$, $j = 1, \ldots, d$. In what follows, pseudo-observations $\widehat{U}_{t,j}$ are used generically in place of $\widehat{U}_{t,j}^{(P)}$ or $\widehat{U}_{t,j}^{(SP)}$.

4.2. Estimation of copula parameters

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 $[e] = \{e' : e' \sim e\}$. 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. We again write $C_{a_{[e]}|D_{[e]}}$ to highlight the invariance of the function with respect to shifts in time. For an edge $e \in E_k$, denote by $S_a(e)$ the set of edges $e' \in \{E_1, \ldots, E_{k-1}\}$ involved in this recursion and $\boldsymbol{\theta}_{S_a([e])} = (\boldsymbol{\theta}_{[e']})_{e' \in S_a(e)}$. Finally, write $[E_k] = \{[e] : e \in E_k\}$, $\boldsymbol{\theta}_{[E_k]} = (\boldsymbol{\theta}_{[e]})_{[e] \in [E_k]}$ 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 $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T)$ is

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

where $(\widehat{\boldsymbol{U}}_1, \ldots, \widehat{\boldsymbol{U}}_T)$ can be either the parametric or nonparametric pseudo-observations. The joint MLE, $\arg\max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta})$, is often too demanding. The step-wise 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 $k = 1, \ldots, d(p+1) - 1$ and every $e' \in E_k$

$$\widehat{\boldsymbol{\theta}}_{[e']} = \arg\max_{\boldsymbol{\theta}_{[e']}} \sum_{e \sim e'} \ln c_{[e]}\Big\{C_{a_{[e]}|D_{[e]}}(\widehat{U}_{a_e} \mid \widehat{\boldsymbol{U}}_{D_e}; \widehat{\boldsymbol{\theta}}_{S_a([e])}), C_{b_{[e]}|D_{[e]}}(\widehat{U}_{b_e} \mid \widehat{\boldsymbol{U}}_{D_e}; \widehat{\boldsymbol{\theta}}_{S_b([e])}); \boldsymbol{\theta}_{[e']}\Big\}. \tag{4}$$

The vectors $\widehat{\boldsymbol{\theta}}_{S(a_{[e]})}, \widehat{\boldsymbol{\theta}}_{S(b_{[e]})}$ in (4) only contain parameter estimates from previous trees, i.e., ones that were already found in earlier iterations.

4.3. Asymptotic results

In what follows, we establish consistency and asymptotic normality of the parametric and semiparametric parameter estimates. Their proofs are given in Section S6 of the supplementary material. All results are derived as consequences of the more general Theorems A.1 and A.2 given in Appendix A. A discussion of the results is given at the end of this section.

We shall assume in the following that the series $(\boldsymbol{X}_t)_{t \in \mathbb{Z}}$ is strictly stationary and absolutely regular. 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 $(0, 1)^2$. In what follows, $\| \cdot \|$ denotes the Euclidean norm.

4.3.1. Parametric estimator

To state the asymptotic results, it is convenient to introduce some more notation. For S-vines of Markov order p, one can check that any edge $e \in E_k$, $1 \leq k \leq d(p+1) - 1$, the set $\{a_e, b_e, D_e\}$ only contains variables at most p time points apart. More precisely, if $a_e = (t_1, j_1)$, $b_e = (t_2, j_2)$ and $t = \min\{t_1, t_2\}$, then $(x_{a_e}, x_{b_e}, \boldsymbol{x}_{D_e})$ is a sub-vector of $(\boldsymbol{x}_t, \ldots, \boldsymbol{x}_{t+p})$. Hence, we denote

$$F_{[e],1,\boldsymbol{\eta},\boldsymbol{\theta}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{1+p}) = F_{a_{[e]}|D_{[e]}}(x_{a_{[e]}} \mid \boldsymbol{x}_{D_{[e]}}; \boldsymbol{\eta}, \boldsymbol{\theta}_{S_a([e])}),$$
$$F_{[e],2,\boldsymbol{\eta},\boldsymbol{\theta}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{1+p}) = F_{b_{[e]}|D_{[e]}}(x_{b_{[e]}} \mid \boldsymbol{x}_{D_{[e]}}; \boldsymbol{\eta}, \boldsymbol{\theta}_{S_b([e])}),$$

and

$$\begin{aligned}
\boldsymbol{s}_{j,\eta_j}(\boldsymbol{x}_t, \dots, \boldsymbol{x}_{t+p}) &= \nabla_{\eta_j} \ln f_j(x_{t,j}),\\
\boldsymbol{s}_{[e],\eta,\theta}(\boldsymbol{x}_t, \dots, \boldsymbol{x}_{t+p}) &= \nabla_{\theta_{[e]}} \ln c_{[e]}\{F_{[e],1,\eta,\theta}(\boldsymbol{x}_t, \dots, \boldsymbol{x}_{t+p}), F_{[e],2,\eta,\theta}(\boldsymbol{x}_t, \dots, \boldsymbol{x}_{t+p}); \boldsymbol{\theta}_{[e]}\},
\end{aligned}$$

and define

$$\boldsymbol{\phi}^{(P)}_{\eta,\theta} = \begin{pmatrix} \big(\boldsymbol{s}_{j,\eta_j}\big)_{j=1,\dots,d} \\ \big(\boldsymbol{s}_{[e],\eta,\theta}\big)_{[e]\in[E_k],k=1,\dots,(p+1)d} \end{pmatrix}.$$

Up to a finite number of terms, the parametric step-wise MLE $(\widehat{\boldsymbol{\eta}}^{(P)}, \widehat{\boldsymbol{\theta}}^{(P)})$ is then defined as the solution of the estimating equation

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

To allow for misspecified parametric models, we further define pseudo-true values $(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*)$ via

$$\mathbb{E}\{\boldsymbol{\phi}^{(P)}_{\eta^*,\theta^*}(\boldsymbol{X}_1, \dots, \boldsymbol{X}_{1+p})\} = 0,$$

and note that they agree with the true parameters if the model is correctly specified. Here and in the sequel, all expectations are taken with respect to the unknown true distribution of $\boldsymbol{X}_1, \dots, \boldsymbol{X}_{1+p}$.

We impose the following regularity conditions:

(P1) The pseudo-true values $(\boldsymbol{\eta}^*, \boldsymbol{\theta}^*)$ lie in the interior of $\mathcal{H} \times \Theta$ and for every $\epsilon > 0$,

$$\inf_{\|\eta-\eta^*\|+\|\theta-\theta^*\|>\epsilon} \left\|\mathbb{E}\left\{\boldsymbol{\phi}^{(P)}_{\eta,\theta}(\boldsymbol{X}_1, \dots, \boldsymbol{X}_{1+p})\right\}\right\| > 0.$$

(P2) The function $\boldsymbol{\phi}^{(P)}_{\eta,\theta}$ is continuously differentiable with respect to $(\boldsymbol{\eta}, \boldsymbol{\theta})$ and satisfies

$$\mathbb{E}\left[\sup_{\eta\in\tilde{\mathcal{H}}}\sup_{\theta\in\tilde{\Theta}}\left\{\left\|\boldsymbol{\phi}^{(P)}_{\eta,\theta}(\boldsymbol{X}_1, \dots, \boldsymbol{X}_{1+p})\right\| + \left\|\nabla_{(\eta,\theta)}\boldsymbol{\phi}^{(P)}_{\eta,\theta}(\boldsymbol{X}_1, \dots, \boldsymbol{X}_{1+p})\right\|\right\}\right] < \infty$$

for any compact $\tilde{\mathcal{H}} \times \tilde{\Theta} \subseteq \mathcal{H} \times \Theta$.

(P3) The matrix $\boldsymbol{J}_{\eta^*,\theta^*} = \mathbb{E}\left\{\nabla'_{(\eta^*,\theta^*)}\boldsymbol{\phi}^{(P)}_{\eta^*,\theta^*}(\boldsymbol{X}_1, \dots, \boldsymbol{X}_{1+p})\right\}$ is invertible.

(P4) The β-mixing coefficients of $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ satisfy $\sum_{t=0}^{\infty}\int_0^{\beta(t)} Q^2(u)du < \infty$, where Q is the inverse survival function of $\|\boldsymbol{\phi}^{(P)}_{\eta^*,\theta^*}(\boldsymbol{X}_1, \dots, \boldsymbol{X}_{t+p})\|$.

Condition (P1) ensures identifiability of the model parameters, (P2) and (P3) are standard regularity condition for maximum-likelihood methods. Condition (P4) quantifies a trade-off between moments of the 'score' function $\boldsymbol{\phi}^{(P)}_{\eta^*,\theta^*}$ and the mixing rate, see the discussion in Section 4.3.3.

Theorem 4. *Under (P1)–(P2), it holds* $(\widehat{\boldsymbol{\eta}}^{(P)}, \widehat{\boldsymbol{\theta}}^{(P)}) \to_p (\boldsymbol{\eta}^*, \boldsymbol{\theta}^*)$.

Theorem 5. *Under (P1)–(P4), it holds* $\|(\widehat{\boldsymbol{\eta}}^{(P)}, \widehat{\boldsymbol{\theta}}^{(P)}) - (\boldsymbol{\eta}^*, \boldsymbol{\theta}^*)\| = O_p(T^{-1/2})$ *and*

$$\sqrt{T}\begin{pmatrix}\widehat{\boldsymbol{\eta}}^{(P)} - \boldsymbol{\eta}^* \\ \widehat{\boldsymbol{\theta}}^{(P)} - \boldsymbol{\theta}^*\end{pmatrix} \xrightarrow{d} \mathcal{N}\left\{\boldsymbol{0}, \boldsymbol{J}^{-1}_{\eta^*,\theta^*}\boldsymbol{I}_{\eta^*,\theta^*}(\boldsymbol{J}^{-1}_{\eta^*,\theta^*})'\right\},$$

where $\boldsymbol{I}_{\eta^*,\theta^*} = \sum_{t=1}^{\infty}\{1 + \mathbb{1}(t \geq 2)\}\mathbb{E}\left\{\boldsymbol{\phi}^{(P)}_{\eta^*,\theta^*}(\boldsymbol{X}_1, \dots, \boldsymbol{X}_{1+p})\boldsymbol{\phi}^{(P)}_{\eta^*,\theta^*}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+p})'\right\}$.

Note that *iid* models are included as a special case:

Corollary 2. *If* $(\boldsymbol{X}_t)_{t\in\mathbb{N}}$ *is* iid *and* $p = 0$, *then Theorems 4 and 5 hold with* $\boldsymbol{I}_{\eta^*,\theta^*} = \mathbb{E}\left\{\boldsymbol{\phi}^{(P)}_{\eta^*,\theta^*}(\boldsymbol{X}_1)\boldsymbol{\phi}^{(P)}_{\eta^*,\theta^*}(\boldsymbol{X}_1)'\right\}$. *If the latter expectation exists, condition (P4) can be dropped.*

4.3.2. Semiparametric estimator

Similarly to the parametric case, we define

$$\begin{aligned}
C_{[e],1,\theta}(\boldsymbol{u}_t, \dots, \boldsymbol{u}_{t+p}) &= C_{a_{[e]}|D_{[e]}}(u_{a_e} \mid \boldsymbol{u}_{D_e}; \boldsymbol{\theta}_{S_a([e])}),\\
C_{[e],2,\theta}(\boldsymbol{u}_t, \dots, \boldsymbol{u}_{t+p}) &= C_{b_{[e]}|D_{[e]}}(u_{a_e} \mid \boldsymbol{u}_{D_e}; \boldsymbol{\theta}_{S_b([e])}),
\end{aligned}$$

and

$$\boldsymbol{s}_{[e],\boldsymbol{\theta}}(\boldsymbol{u}_1,\dots,\boldsymbol{u}_{1+p}) = \nabla_{\boldsymbol{\theta}_{[e]}} \ln c_{[e]}\Big\{C_{[e],1,\boldsymbol{\theta}}(\boldsymbol{u}_t,\dots,\boldsymbol{u}_{t+p}), C_{[e],2,\boldsymbol{\theta}}(\boldsymbol{u}_t,\dots,\boldsymbol{u}_{t+p}); \boldsymbol{\theta}_{[e]}\Big\}.$$

For a generic vector of functions $\boldsymbol{G} = (G_1,\dots,G_d)$ write $\boldsymbol{G}(\boldsymbol{x}) = \big(G_1(x_1),\dots,G_d(x_d)\big)$. Let $\widehat{\boldsymbol{F}} = (\widehat{F}_1,\dots,\widehat{F}_d)$ be the vector of empirical distribution functions and $\boldsymbol{F} = (F_1,\dots,F_d)$ be the true distributions. Setting $\boldsymbol{\phi}_{\boldsymbol{\theta}}^{(SP)} = (\boldsymbol{s}_{[e],\boldsymbol{\theta}})_{[e]\in[E_k],k=1,\dots,(p+1)d}$, the semiparametric step-wise MLE $\widehat{\boldsymbol{\theta}}^{(SP)}$ is then defined as the solution of the estimating equation

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\phi}_{\boldsymbol{\theta}}^{(SP)}\{\widehat{\boldsymbol{F}}(\boldsymbol{X}_t),\dots,\widehat{\boldsymbol{F}}(\boldsymbol{X}_{t+p})\} = 0.$$

Further, we define the pseudo-true value $\boldsymbol{\theta}^*$ via

$$\mathbb{E}\Big[\boldsymbol{\phi}_{\boldsymbol{\theta}^*}^{(SP)}\{\boldsymbol{F}(\boldsymbol{X}_1),\dots,\boldsymbol{F}(\boldsymbol{X}_{1+p})\}\Big] = 0.$$

For $u \in (0,1)$ and some $\gamma \in [0,1)$, define the weight function $w(u) = u^\gamma(1-u)^\gamma$ and set $\mathcal{F}_\delta = \times_{j=1}^{d}\{G\colon \mathbb{R} \to [0,1], \sup_x |F_j(x) - G(x)|/w\{F_j(x)\} \le \delta\}$.

(SP1) The pseudo-true value $\boldsymbol{\theta}^*$ lies in the interior of Θ and for every $\epsilon > 0$,

$$\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|>\epsilon}\Big\|\mathbb{E}\Big[\boldsymbol{\phi}_{\boldsymbol{\theta}}^{(SP)}\{\boldsymbol{F}(\boldsymbol{X}_1),\dots,\boldsymbol{F}(\boldsymbol{X}_{1+p})\}\Big]\Big\| > 0.$$

(SP2) The functions $\boldsymbol{\phi}_{\boldsymbol{\theta}}^{(SP)}$ are continuously differentiable with respect to $\boldsymbol{\theta}$ and its arguments and there is $\delta > 0$ such that

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\tilde{\Theta}}\sup_{\boldsymbol{G}\in\mathcal{F}_\delta}\left\{\Big\|\boldsymbol{\phi}_{\boldsymbol{\theta}}^{(SP)}\{\boldsymbol{F}(\boldsymbol{X}_1),\dots,\boldsymbol{F}(\boldsymbol{X}_{1+p})\}\Big\| + \Big\|\nabla_{\boldsymbol{\theta}}\boldsymbol{\phi}_{\boldsymbol{\theta}}^{(SP)}\{\boldsymbol{G}(\boldsymbol{X}_1),\dots,\boldsymbol{G}(\boldsymbol{X}_{1+p})\}\Big\|\right\}\right] < \infty,$$

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\tilde{\Theta}}\sup_{\boldsymbol{G}\in\mathcal{F}_\delta}\left\|\frac{\partial}{\partial\{G_j(X_{t,j})\}}\boldsymbol{\phi}_{\boldsymbol{\theta}}^{(SP)}\{\boldsymbol{G}(\boldsymbol{X}_1),\dots,\boldsymbol{G}(\boldsymbol{X}_{1+p})\}w\{F_j(X_{t,j})\}\right\|\right] < \infty$$

for any compact $\tilde{\Theta} \subseteq \Theta$, $t = 1,\dots,1+p$, $j = 1,\dots,d$.

(SP3) The mixed derivatives $\nabla_{\boldsymbol{\theta}}\nabla'_{\boldsymbol{u}}\boldsymbol{\phi}_{\boldsymbol{\theta}}^{(SP)}(\boldsymbol{u})$ are continuous in $\boldsymbol{u} \in (0,1)^{(p+1)d}$ and $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}^*$ and there is $\delta > 0$ such that for all $t = 1,\dots,1+p$, $j = 1,\dots,d$,

$$\mathbb{E}\left[\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|\le\delta}\sup_{\boldsymbol{G}\in\mathcal{F}_\delta}\left\|\frac{\partial}{\partial\{G_j(X_{t,j})\}}\nabla_{\boldsymbol{\theta}}\boldsymbol{\phi}_{\boldsymbol{\theta}}^{(SP)}\{\boldsymbol{G}(\boldsymbol{X}_1),\dots,\boldsymbol{G}(\boldsymbol{X}_{1+p})\}w\{F_j(X_{t,j})\}\right\|\right] < \infty.$$

(SP4) The matrix $\boldsymbol{J}_{\boldsymbol{\theta}^*} = \mathbb{E}\Big[\nabla'_{\boldsymbol{\theta}^*}\boldsymbol{\phi}_{\boldsymbol{\theta}^*}^{(SP)}\{\boldsymbol{F}(\boldsymbol{X}_1),\dots,\boldsymbol{F}(\boldsymbol{X}_{1+p})\}\Big]$ is invertible.

(SP5) For $\gamma \in [0,1/2)$, the β-mixing coefficients of $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ satisfy $\beta(t) = O(t^{-a})$ with $a > 1/(1-2\gamma)$ and it holds $\sum_{t=0}^{\infty}\int_0^{\beta(t)} Q^2(u)du < \infty$, where Q is the inverse survival function of $\|\boldsymbol{\phi}_{\boldsymbol{\theta}^*}^{(SP)}\{\boldsymbol{F}(\boldsymbol{X}_1),\dots,\boldsymbol{F}(\boldsymbol{X}_{1+p})\}\|$.

Similar to the parametric case, (SP1) ensures identifiability of the model parameters and (SP4) is a standard regularity condition. Conditions (SP2)–(SP3) are more involved due to the suprema over function classes $\mathcal{F}_\delta$. This is typical for semiparametric copulas models (see, Genest et al., 1995; Tsukahara, 2005; Chen and Fan, 2006b; Hobæk Haff, 2013; Chen et al., 2020). Derivatives of copula functions tend to blow up in the corners of the unit hypercube. This can be offset by exploiting stronger convergence properties of the empirical margins in these corners. The function w is used to strengthen the metric accordingly. (SP5) quantifies the trade-off between moments of the 'score' function $\boldsymbol{\phi}_{\boldsymbol{\theta}^*}^{(SP)}$ and the mixing rate, see also our discussion below.

Theorem 6. *Under (SP1)–(SP2), it holds* $\widehat{\boldsymbol{\theta}}^{(SP)} \to_p \boldsymbol{\theta}^*$.

Theorem 7. *Under (SP1)–(SP5), it holds* $\|\widehat{\boldsymbol{\theta}}^{(SP)} - \boldsymbol{\theta}^*\| = O_p(T^{-1/2})$ *and*

$$T^{1/2}(\widehat{\boldsymbol{\theta}}^{(SP)} - \boldsymbol{\theta}^*) \xrightarrow{d} \mathcal{N}\big\{\boldsymbol{0}, \boldsymbol{J}_{\boldsymbol{\theta}^*}^{-1}\boldsymbol{I}_{\boldsymbol{\theta}^*}(\boldsymbol{J}_{\boldsymbol{\theta}^*}^{-1})'\big\},$$

where $\boldsymbol{I}_{\boldsymbol{\theta}^*} = \mathbb{E}(\boldsymbol{Z}_1\boldsymbol{Z}_1') + 2\sum_{t=2}^{\infty}\mathbb{E}(\boldsymbol{Z}_1\boldsymbol{Z}_t')$ *and*

$$\boldsymbol{Z}_t = \boldsymbol{\phi}_{\boldsymbol{\theta}^*}^{(SP)}\{\boldsymbol{F}(\boldsymbol{X}_t),\dots,\boldsymbol{F}(\boldsymbol{X}_{t+p})\} + \boldsymbol{D}(\boldsymbol{X}_t),$$

$$\boldsymbol{D}(x_1,\dots,x_d) = \sum_{j=1}^{d}\sum_{t=1}^{1+p}\mathbb{E}\left[\{\mathbb{1}(x_j \le X_{t,j}) - F_j(X_{t,j})\}\frac{\partial\boldsymbol{\phi}_{\boldsymbol{\theta}^*}^{(SP)}\{\boldsymbol{F}(\boldsymbol{X}_1),\dots,\boldsymbol{F}(\boldsymbol{X}_{1+p})\}}{\partial\{F_j(X_{t,j})\}}\right].$$

Corollary 3. *If* $(\boldsymbol{X}_t)_{t\in\mathbb{N}}$ *is* iid *and* $p = 0$, *then Theorems 6 and 7 hold with*

$$\boldsymbol{I}_{\boldsymbol{\theta}^*} = \mathbb{E}\left([\boldsymbol{\phi}^{(SP)}_{\boldsymbol{\theta}^*}\{\boldsymbol{F}(\boldsymbol{X}_1)\} + \boldsymbol{D}(\boldsymbol{X}_1)][\boldsymbol{\phi}^{(SP)}_{\boldsymbol{\theta}^*}\{\boldsymbol{F}(\boldsymbol{X}_1)\} + \boldsymbol{D}(\boldsymbol{X}_1)]'\right).$$

If $\mathbb{E}[\|\boldsymbol{\phi}_{\boldsymbol{\theta}^*}\{\boldsymbol{F}(\boldsymbol{X}_t)\}\|^2] < \infty$, *condition (SP5) can be dropped.*

4.3.3. Discussion

The $\sqrt{T}$-convergence predicted by our theorems is confirmed in numerical experiments in Section S4.1 of the supplementary materials. The results extend the existing literature in various ways. The results on the fully parametric sequential MLE (Theorems 4 and 5) appear to be new — even in the *iid* case. Joe (2005) provided a similar result when there are only two steps: one for the marginal parameters and one for the copula parameters. For semiparametric models (Theorems 6 and 7), a similar result was obtained by Rémillard et al. (2012) for $p = 1$, and a joint MLE for the copula parameters. It does not apply to the step-wise MLE commonly used in vine copula models, however. The only known results for the semiparametric step-wise MLE were provided by Hobæk Haff (2013) in the *iid* ($p = 0$) case. These results assume a D-vine structure and correctly specified copula model. The latter assumption is especially questionable in view of the common simplifying assumption (see Section 2.3).

Also the results in Chen and Fan (2006b) are obtained as a special case with $d = 1, p = 1$. The regularity conditions here are slightly weaker than theirs, and also than those of Tsukahara (2005) and Hobæk Haff (2013) in the *iid* case. Specifically, conditions (SP2)–(SP3) require a first moment uniformly in $\bar{\Theta} \times \mathcal{F}_\delta$, whereas previous results require a second moment. A higher order moment constraint is only imposed on $\boldsymbol{\phi}^{(SP)}_{\boldsymbol{\theta}^*}\{\boldsymbol{F}(\boldsymbol{X}_1), \dots, \boldsymbol{F}(\boldsymbol{X}_{1+p})\}$ and only at the single point $(\boldsymbol{\theta}^*, \boldsymbol{F})$ via (SP5). Conditions (P4) and (SP5) ensure existence of the asymptotic covariance and are, to the best of our knowledge, the weakest known for β-mixing time series. Writing generically $\boldsymbol{\phi}$ for either $\boldsymbol{\phi}^{(P)}_{\boldsymbol{\eta}^*,\boldsymbol{\theta}^*}$ in (P4) or $\boldsymbol{\phi}^{(SP)}_{\boldsymbol{\theta}^*}$ in (SP5), the conditions are satisfied in each of the following cases (see, Rio, 2017, Section 1.4):

(i) $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is *iid* and $\mathbb{E}\left\{\|\boldsymbol{\phi}(\boldsymbol{X}_1, \dots, \boldsymbol{X}_{1+p})\|^2\right\} < \infty$.

(ii) There is $b \in (0, 1)$ such that $\beta(t) = O(b^t)$ and

$$\mathbb{E}\left\{\|\boldsymbol{\phi}(\boldsymbol{X}_1, \dots, \boldsymbol{X}_{1+p})\|^2 \ln(1 + \|\boldsymbol{\phi}(\boldsymbol{X}_1, \dots, \boldsymbol{X}_{1+p})\|)\right\} < \infty.$$

(iii) $\beta(t) = O(t^{-a})$ with $a > 1/(1 - 2\gamma)$ and there is $q > \max\{2, 2/(a - 1)\}$ such that

$$\mathbb{E}\left\{\|\boldsymbol{\phi}(\boldsymbol{X}_1, \dots, \boldsymbol{X}_{1+p})\|^q\right\} < \infty.$$

If the mixing decay is fast, we can therefore use weaker moment conditions. The latter two conditions already appeared in similar form in Chen and Fan (2006b). For $p = 1$, $d = 1$, a sizeable literature (including Chen and Fan, 2006b; Chen et al., 2009; Beare, 2010, 2012; 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.

5. Prediction

Vine copula models are quite complex and rarely allow closed-form expressions of conditional means, quantiles, or the predictive distribution. One may instead simulate (conditionally) from the estimated model and approximate such quantities by Monte Carlo methods. The standard simulation algorithm (e.g., Czado, 2019, Chapter 6) poses unnecessary computational demands, however. An efficient algorithm exploiting the Markov property is given in Section S3.2 of the supplementary material.

With the ability to simulate conditionally on the past, it is easy to compute predictions for all sorts of quantities, like conditional means or quantiles. More specifically, suppose we are interested in a functional $\mu = \psi(F_{k,p})$ of the conditional distribution $F_{k,p}(\cdot \mid \boldsymbol{x}_{t-1}, \dots, \boldsymbol{x}_{t-p}) = F_{\boldsymbol{X}_t,\dots,\boldsymbol{X}_{t+k}|\boldsymbol{X}_{t-1},\dots,\boldsymbol{X}_{t-p}}(\cdot \mid \boldsymbol{x}_{t-1}, \dots, \boldsymbol{x}_{t-p})$ of the next k time points given the past. We construct an estimator of this functional as follows:

1. Simulate N *iid* replicates $(\boldsymbol{X}^{(i)}_t, \dots, \boldsymbol{X}^{(i)}_{t+k})_{i=1}^N$ from the (estimated) conditional distribution of $(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+k})$ given $\boldsymbol{X}_{t-1} = \boldsymbol{x}_{t-1}, \dots, \boldsymbol{X}_{t-p} = \boldsymbol{x}_{t-p}$ using the estimated model (either parametric or semiparametric; see Section 4).
2. Compute $\psi(\widehat{F}_{k,p})$ where

$$\widehat{F}_{k,p}(\boldsymbol{x}_t, \dots, \boldsymbol{x}_{t+k}) = \frac{1}{n}\sum_{i=1}^N \mathbb{1}(\boldsymbol{X}^{(i)}_t \le \boldsymbol{x}_t, \dots, \boldsymbol{X}^{(i)}_{t+k} \le \boldsymbol{x}_{t+k}).$$

When simulating from an estimated parametric model $F_{k,p}(\cdot \mid \boldsymbol{x}_{t-1}, \dots, \boldsymbol{x}_{t-p}; \widehat{\boldsymbol{\eta}}^{(P)}, \widehat{\boldsymbol{\theta}}^{(P)})$, we call the resulting estimator $\widehat{\mu}^{(P)}$; when simulating from a semiparametric model $F_{k,p}(\cdot \mid \boldsymbol{x}_{t-1}, \dots, \boldsymbol{x}_{t-p}; \widehat{\boldsymbol{F}}, \widehat{\boldsymbol{\theta}}^{(SP)})$, we call the resulting estimator $\widehat{\mu}^{(SP)}$. The corresponding pseudo-true value μ^* is defined as $\psi\{F_{k,p}(\cdot \mid \boldsymbol{x}_{t-1}, \dots, \boldsymbol{x}_{t-p}; \boldsymbol{\eta}^*, \boldsymbol{\theta}^*)\}$ or $\psi\{F_{k,p}(\cdot \mid \boldsymbol{x}_{t-1}, \dots, \boldsymbol{x}_{t-p}; \boldsymbol{F}, \boldsymbol{\theta}^*)\}$ for the parametric and semiparametric cases respectively.

The following results account for the fact that we simulate from an estimated model. They are an immediate consequence of Theorem A.3 in Appendix A. In general, we assume that the map $F \mapsto \psi(F)$ is Frechet differentiable.

Theorem 8. *Suppose the map* $(\boldsymbol{\theta}, \boldsymbol{\eta}) \mapsto \psi\{F_{k,p}(\cdot \mid \boldsymbol{x}_{t-1}, \dots, \boldsymbol{x}_{t-p}; \boldsymbol{\eta}, \boldsymbol{\theta})\}$ *is continuously differentiable at* $(\boldsymbol{\theta}^*, \boldsymbol{\eta}^*)$ *with gradient* $\boldsymbol{\Psi}_{\boldsymbol{\theta}^*,\boldsymbol{\eta}^*}$.

1. *If* $N \to \infty$ *and conditions (P1)–(P2) hold, then* $\widehat{\mu}^{(P)} \to_p \mu^*$.
2. *If additionally* $T = o(N)$ *and conditions (P4)–(P3) hold, then* $\widehat{\mu}^{(P)} - \mu^* = O_p(T^{-1/2})$ *and*

$$\sqrt{T}(\widehat{\mu}^{(P)} - \mu^*) \to_d \mathcal{N}\big(0,\ \boldsymbol{\Psi}'_{\boldsymbol{\theta}^*,\boldsymbol{\eta}^*}\boldsymbol{J}^{-1}_{\boldsymbol{\eta}^*,\boldsymbol{\theta}^*}\boldsymbol{I}_{\boldsymbol{\eta}^*,\boldsymbol{\theta}^*}(\boldsymbol{J}^{-1}_{\boldsymbol{\eta}^*,\boldsymbol{\theta}^*})'\boldsymbol{\Psi}_{\boldsymbol{\theta}^*,\boldsymbol{\eta}^*}\big),$$

where $\boldsymbol{I}_{\boldsymbol{\eta}^*,\boldsymbol{\theta}^*}, \boldsymbol{J}_{\boldsymbol{\eta}^*,\boldsymbol{\theta}^*}$ *are defined in Section 4.3.1.*

For any function G, let $W(G) = G(\cdot)/w\{G(\cdot)\}$ and denote $W(\boldsymbol{G}) = (W(G_1), \dots, W(G_d))$.

Theorem 9. *Suppose the map* $(W(\boldsymbol{G}), \boldsymbol{\theta}) \mapsto \psi\{F_{k,p}(\cdot \mid \boldsymbol{x}_{t-1}, \dots, \boldsymbol{x}_{t-p}; \boldsymbol{G}, \boldsymbol{\theta})\}$ *is Frechet differentiable at* $(W(\boldsymbol{F}), \boldsymbol{\theta}^*)$ *with derivative* $(W(\boldsymbol{h}), \boldsymbol{\theta}) \mapsto \sum_{j=1}^d \Psi_j(h_j) + \boldsymbol{\Psi}'_{\boldsymbol{\theta}}\boldsymbol{\theta}$.

1. *If* $N \to \infty$ *and conditions (SP1)–(SP2) hold, then* $\widehat{\mu}^{(SP)} \to_p \mu^*$.
2. *If additionally* $T = o(N)$ *and conditions (SP3)–(SP4) hold, then* $\widehat{\mu}^{(SP)} - \mu^* = O_p(T^{-1/2})$ *and*

$$\sqrt{T}(\widehat{\mu}^{(SP)} - \mu^*) \to_d \mathcal{N}(0, \sigma^2),$$

where $\sigma^2 = \mathrm{Var}(Z_1) + 2\sum_{t=2}^{\infty}\mathrm{Cov}(Z_1, Z_t)$ *with*

$$Z_t = \sum_{k=0}^{p}\sum_{j=1}^{d}\Psi_j\{\mathbb{1}(X_{t+k,j} \le \cdot) - F_j(\cdot)\} + \boldsymbol{\Psi}'_{\boldsymbol{\theta}}[\boldsymbol{\phi}^{(SP)}_{\boldsymbol{\theta}^*}\{\boldsymbol{F}(\boldsymbol{X}_t), \dots, \boldsymbol{F}(\boldsymbol{X}_{t+p})\} + \boldsymbol{D}(\boldsymbol{X}_t)]$$

and $\boldsymbol{D}(\boldsymbol{X}_t)$ *given in Theorem 7.*

Simulation-based prediction from vine copula models has been 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 allow for both parametric and semiparametric models, time series data, and a generic prediction target. In addition, Theorems 8 and 9 characterize a distributional limit for such predictions. This is of practical importance because it allows to properly assess estimation/prediction uncertainty. Of course, the results specialize to the *iid* case similarly to Corollaries 2 and 3. The asymptotic covariances are generally unknown and must be estimated. We propose computationally efficient methods in the following section.

6. Uncertainty quantification

In principle, the asymptotic covariances in the preceding theorems can be estimated by HAC methods (e.g., Andrews, 1991). In the prediction context such methods become numerically demanding, especially for the semiparametric estimator. (Semi-)parametric or block bootstrap methods (Künsch, 1989; Chen and Fan, 2006b; Genest and Rémillard, 2008) are general alternatives, but similarly demanding because the model has to be fit many times. We propose a more efficient bootstrap method based on an asymptotic approximation of the parameter estimates. In essence, we avoid refitting the entire model by performing only a single Newton–Raphson update on the bootstrapped likelihood.

We employ a dependent multiplier bootstrap scheme similar to Bücher and Kojadinovic (2016). Its idea is as follows. Let ℓ_T be a sequence with $\ell_T \to \infty$. We simulate a stationary time series of bootstrap weights $\xi_1, \dots, \xi_T$ that is ℓ_T-dependent, independent of the data, and satisfies $\mathbb{E}(\xi_1) = \mathrm{Var}(\xi_1) = 1$ and $\mathrm{Cov}(\xi_1, \xi_{1+t}) = 1 - o(t/\ell_T)$. Given a sufficiently regular stationary time series $Z_1, \dots, Z_T$, one can then show that $T^{1/2}\{T^{-1}\sum_{t=1}^T Z_t - \mathbb{E}(Z_1)\}$ and $T^{-1/2}\sum_{t=1}^T(\xi_t - 1)Z_t$ converge to independent copies of the same random variable (see, Bühlmann, 1993; Bücher and Kojadinovic, 2019). We shall apply this principle to bootstrap the step-wise log-likelihood and (if necessary) empirical marginal distributions.

Consider the bootstrapped estimating equation

$$\frac{1}{T}\sum_{t=1}^{T}\xi_t\boldsymbol{\phi}^{(P)}_{\boldsymbol{\eta},\boldsymbol{\theta}}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+p}) = 0.$$

We define our bootstrap replicates as an approximate one-step Newton–Raphson update from $(\widehat{\boldsymbol{\eta}}^{(P)}, \widehat{\boldsymbol{\theta}}^{(P)})$, i.e.,

$$\begin{pmatrix}\widetilde{\boldsymbol{\eta}}^{(P)}\\ \widetilde{\boldsymbol{\theta}}^{(P)}\end{pmatrix} = \begin{pmatrix}\widehat{\boldsymbol{\eta}}^{(P)}\\ \widehat{\boldsymbol{\theta}}^{(P)}\end{pmatrix} - \left(\sum_{t=1}^{T}\nabla_{(\boldsymbol{\eta},\boldsymbol{\theta})}\boldsymbol{\phi}^{(P)}_{(\widehat{\boldsymbol{\eta}}^{(P)},\widehat{\boldsymbol{\theta}}^{(P)})}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+p})\right)^{-1}\sum_{t=1}^{T}\xi_t\boldsymbol{\phi}^{(P)}_{(\widehat{\boldsymbol{\eta}}^{(P)},\widehat{\boldsymbol{\theta}}^{(P)})}(\boldsymbol{X}_t, \dots, \boldsymbol{X}_{t+p}).$$

Because $\nabla_{(\eta,\theta)}\phi^{(P)}_{(\widehat{\eta}^{(P)},\widehat{\theta}^{(P)})}$ and $\phi^{(P)}_{(\widehat{\eta}^{(P)},\widehat{\theta}^{(P)})}$ have already been evaluated when computing $(\widehat{\eta}^{(P)},\widehat{\theta}^{(P)})$, this update has negligible computational cost. One may then show that

$$\frac{1}{T}\sum_{t=1}^{T}\nabla_{(\eta,\theta)}\phi^{(P)}_{\widehat{\eta}^{(P)},\widehat{\theta}^{(P)}}(\boldsymbol{X}_t,\ldots,\boldsymbol{X}_{t+p}) \to_p \boldsymbol{J}_{\eta^*,\theta^*},$$

$$\frac{1}{T}\sum_{t=1}^{T}\xi_t\phi^{(P)}_{(\widehat{\eta}^{(P)},\widehat{\theta}^{(P)})}(\boldsymbol{X}_t,\ldots,\boldsymbol{X}_{t+p}) \to_d \mathcal{N}(0,\boldsymbol{I}_{\eta^*,\theta^*}),$$

with limiting variable independent of $(\widehat{\eta}^{(P)},\widehat{\theta}^{(P)})$.

In semiparametric models, we also have to bootstrap the empirical marginal distribution:

$$\widetilde{F}_j(x) = \frac{1}{T}\sum_{t=1}^{T}\xi_t\mathbb{1}(X_t \le x), \quad j = 1,\ldots,d.$$

Now consider the bootstrapped estimating equation

$$\frac{1}{T}\sum_{t=1}^{T}\xi_t\phi^{(SP)}_{\theta}\{\widetilde{\boldsymbol{F}}(\boldsymbol{X}_t),\ldots,\widetilde{\boldsymbol{F}}(\boldsymbol{X}_{t+p})\} = 0$$

and the approximate one-step update

$$\widetilde{\theta}^{(SP)} = \widehat{\theta}^{(SP)} - \left(\frac{1}{T}\sum_{t=1}^{T}\nabla_\theta\phi^{(SP)}_{\theta}\{\widehat{\boldsymbol{F}}(\boldsymbol{X}_t),\ldots,\widehat{\boldsymbol{F}}(\boldsymbol{X}_{t+p})\}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}\xi_t\phi^{(SP)}_{\widehat{\theta}^{(SP)}}\{\widetilde{\boldsymbol{F}}(\boldsymbol{X}_t),\ldots,\widetilde{\boldsymbol{F}}(\boldsymbol{X}_{t+p})\}.$$

Note that the function $\phi^{(SP)}_{\widehat{\theta}^{(SP)}}$ on the far right is evaluated at the bootstrapped margins. This is necessary to account for the estimation uncertainty in the margins. It also makes the update slightly more demanding, since we have to evaluate the function $\phi^{(SP)}_{\widehat{\theta}^{(SP)}}$ for every bootstrap replication. This cost is manageable, however. One may again show that $\widetilde{\theta}^{(SP)}$ and $\widehat{\theta}^{(SP)}$ converge in distribution to two *iid* variables.

To get bootstrap replicates for a prediction $\widehat{\mu}$, we simply simulate from bootstrapped models: for $r = 1,\ldots R$,

1. Simulate multipliers $\xi_1,\ldots,\xi_T$ independently from previous steps.
2. Compute a bootstrapped model (parametric or semiparametric) as outlined above.
3. Compute $\widetilde{\mu}_r$ as in Section 5, where $\boldsymbol{X}^{(i)}_t,\ldots,\boldsymbol{X}^{(i)}_{t+k}$, $i = 1,\ldots,N$ are simulated conditionally on the past from the bootstrapped model.

Validity of the above procedure can be established along the lines of Theorems 4–9 and arguments similar to Bühlmann (1993, Chapter 3). A formal proof is beyond the scope of this paper, but our simulation experiments in Section S4.2 of the supplementary materials indicate approximately correct coverage in a range of scenarios.

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 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). Vine structures are selected by a modification of the algorithm by Dissmann et al. (2013), the pair-copula families by the AIC criterion; we refer to Section S3 of the supplementary materials for details. We allow for all parametric families implemented in the `rvinecopulib` R package (Nagler and Vatter, 2020). This includes families without tail

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

Table 2
Companies, their coding, and industry branches.

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 3
Akaike's information criterion for the three vine copula time series models.

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

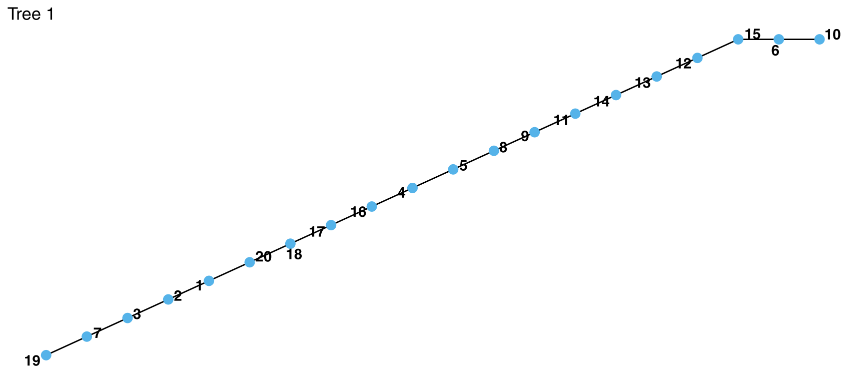

Fig. 6. First tree of M- and D-vines fitted on the whole data set. Trees across time-steps are connected at $(i_1, j_1) = (19, 19)$ (with $\widehat{\tau} \approx 0.02$) for the M-vine, and $(i_1, j_1) = (19, 10)$ (with $\widehat{\tau} \approx 0.05$) for the D-vine.

dependence (e.g., Gaussian copula), and with tail dependence in either one (e.g., Clayton copula), two (e.g., BB7 copula), or all four tails (e.g., the two-parameter t copula).

In Fig. 6 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 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 $(i_1, j_1) = (19, 19)$ (Airbus→Airbus) 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 (10) to Airbus (19) ($\widehat{\tau} = 0.05$).

The corresponding tree of the S-vine can be seen in Fig. 7. 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 Fig. 6, 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 flexibility 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 two 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 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 only consider parametric vine models, but also include three popular competitor models:

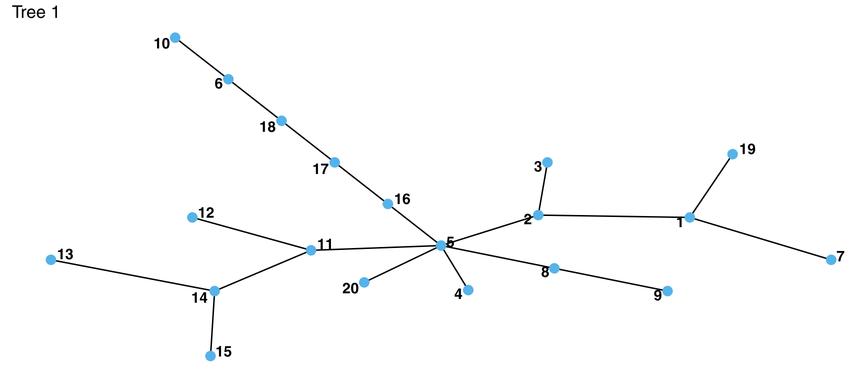

Fig. 7. First tree of an S-vine fitted on the whole data set. Trees across time-steps are connected at $(i_1, j_1) = (15, 6)$ (with $\hat{\tau} \approx 0.16$).

- *VAR*: A *vector autoregressive* model of order 1 (using the `vars` R package Pfaff, 2008).
- *GARCH-vine*: A combination of ARMA-GARCH marginal models (with skew-t residuals) and a vine copula for the residuals (using `rugarch` and `rvinecopulib`, Ghalanos, 2020; Nagler and Vatter, 2020). The ARMA-GARCH orders are selected for each marginal series individually by AIC. This model is an instance of the general residuals method proposed by Nasri and Rémillard (2019). Other choices for marginal models would also be possible.
- *DCC-GARCH*: the DCC-GARCH model of Engle and Sheppard (2001) based on a multivariate t distribution (using `rmgarch`, Ghalanos, 2019).

The VAR model clearly performs worst, since it cannot account for heteroscedasticity. We further see that the vine models outperform the GARCH-vine and DCC-GARCH 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).
- `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 Fig. 8. The dots are the average measure over the full period, the error bars 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.
- Overall all methods seem to provide reasonable predictions, including the residuals method (GARCH-vine).
- 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 predictive log-likelihood (`logS`). For 1-week-ahead forecasts the stationary vine models are outperformed by the DCC-GARCH for `CRPS` and `VaR95`. It compares favorably for the other measures.

We conclude that stationary vine models provide good forecasts for financial time series. This is somewhat remarkable since, in contrast to the pure 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

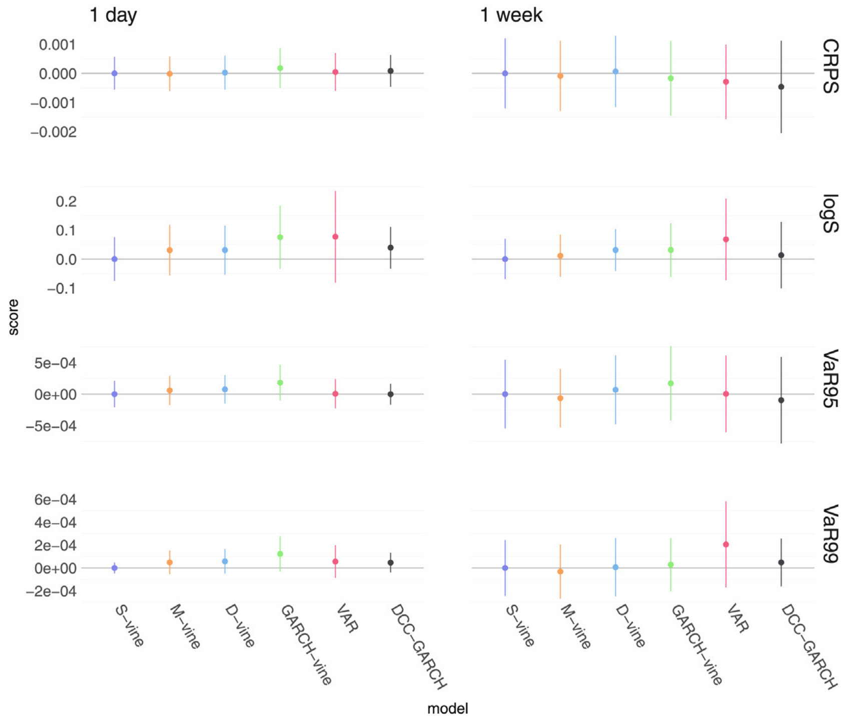

Fig. 8. Forecast performance of various time series models. Dots are mean performance, error bars indicate 90%-confidence intervals (accounting for 30 lags of autocorrelation). The left panel corresponds to 1-day-ahead, the right to 1-week-ahead forecasts. Scores are centered such that S-vine has score 0.

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 structure. But more importantly, our contribution gives a final answer in the search for vine copula models suitable for stationary time series.

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. The proposed bootstrap procedure avoids refitting the models through a one-step approximation. The method appears to be new and may prove useful beyond the current scope. The bootstrap technique also does not require explicit estimation of the rather complicated limiting variances in our theorems. It might be possible to achieve this even more efficiently using a blocking technique similar to Ibragimov and Müller (2010).

We further provide theoretical justifications 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. In particular, they provide post-hoc justification for what is already practiced widely: step-wise estimation and simulation-based inference in fully parametric, but usually misspecified R-vine models. Our main results are empowered by more abstract theorems given in Appendix A. They deal with general semiparametric method-of-moment type estimators with potentially non-negligible nuisance parameter. As this is a common setup, especially in copula models, these abstract results shall prove powerful beyond the present paper. For example, generalizations of the results in Tsukahara (2005) are obtained as easy corollaries. The results shall also help in other interesting extensions of our model, for example accounting for long-memory dependence or non-stationarity (see, e.g., Ibragimov and Lentzas, 2017; Chen et al., 2020).

Despite confirmatory numerical experiments, a limitation of the results is an assumption on the decay of mixing coefficients (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 – is therefore an urging problem for future research.

Acknowledgments

We thank the Editor and three anonymous referees for many helpful suggestions that substantially improved our paper.

Appendix A. General results for semiparametric method-of-moment estimation

The proofs for the parametric and semiparametric cases are largely similar. To avoid duplication, we first establish general results that cover both cases. The statements and proofs make extensive use of empirical process techniques (e.g., Van Der Vaart and Wellner, 1996; Dehling and Philipp, 2002) and the associated short notation $\mathbb{P}_T g = \frac{1}{T}\sum_{t=1}^T g(\boldsymbol{X}_t)$ and $Pg = \mathbb{E}\{g(\boldsymbol{X}_t)\}$ for the empirical measure and expectation over an arbitrary function g.

Suppose we want to estimate a Euclidean parameter $\boldsymbol{\alpha}^* \in \mathcal{A} \subseteq \mathbb{R}^p$ in the presence of a nuisance parameter $\boldsymbol{\nu}^* \in \mathfrak{N}$, possibly infinite-dimensional. Let $\boldsymbol{\phi}_{\boldsymbol{\alpha},\boldsymbol{\nu}} = (\phi_{\boldsymbol{\alpha},\boldsymbol{\nu},1}, \dots, \phi_{\boldsymbol{\alpha},\boldsymbol{\nu},r})$ be a map $\mathbb{R}^s \to \mathbb{R}^r$ such that $P\boldsymbol{\phi}_{\boldsymbol{\alpha}^*,\boldsymbol{\nu}^*} = 0$. Given an estimator $\widehat{\boldsymbol{\nu}}$ of $\boldsymbol{\nu}^*$ define $\widehat{\boldsymbol{\alpha}}$ as the solution to $\mathbb{P}_T\boldsymbol{\phi}_{\boldsymbol{\alpha},\widehat{\boldsymbol{\nu}}} = 0$. We shall assume that $\mathfrak{N}$ is a subset of a Banach space and define

$$\mathcal{A}(\delta) = \{\boldsymbol{\alpha} \in \mathcal{A} : \|\boldsymbol{\alpha} - \boldsymbol{\alpha}^*\| \le \delta\}, \qquad \mathfrak{N}(\delta) = \{\boldsymbol{\nu} \in \mathfrak{N} : \|\boldsymbol{\nu} - \boldsymbol{\nu}^*\| \le \delta\}.$$

We impose the following general conditions:

(C1) The series $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is strictly stationary and absolutely regular.
(C2) For every $\delta > 0$, $\mathrm{P}(\|\widehat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*\| \le \delta) \to 1$ as $T \to \infty$.
(C3) For every $\epsilon > 0$, it holds $\inf_{\|\boldsymbol{\alpha}-\boldsymbol{\alpha}^*\|>\epsilon} \|P\boldsymbol{\phi}_{\boldsymbol{\alpha},\boldsymbol{\nu}^*}\| > 0$.
(C4) For every $K > 0$, it holds $P\sup_{\boldsymbol{\alpha}\in\mathcal{A}(K)} \|\boldsymbol{\phi}_{\boldsymbol{\alpha},\boldsymbol{\nu}^*}\| < \infty$ and there is $\delta > 0$ such that

$$P\left\{\sup_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2\in\mathcal{A}(K)}\ \sup_{\boldsymbol{\nu}_1,\boldsymbol{\nu}_2\in\mathfrak{N}(\delta)} \frac{\|\boldsymbol{\phi}_{\boldsymbol{\alpha}_1,\boldsymbol{\nu}_1} - \boldsymbol{\phi}_{\boldsymbol{\alpha}_2,\boldsymbol{\nu}_2}\|}{\|\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2\| + \|\boldsymbol{\nu}_1 - \boldsymbol{\nu}_2\|}\right\} < \infty.$$

(C5) There is $\delta > 0$ such that

$$P\left\{\sup_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2\in\mathcal{A}(\delta)}\ \sup_{\boldsymbol{\nu}_1,\boldsymbol{\nu}_2\in\mathfrak{N}(\delta)} \frac{\|(\boldsymbol{\phi}_{\boldsymbol{\alpha}_1,\boldsymbol{\nu}_1} - \boldsymbol{\phi}_{\boldsymbol{\alpha}_2,\boldsymbol{\nu}_1}) - (\boldsymbol{\phi}_{\boldsymbol{\alpha}_1,\boldsymbol{\nu}_2} - \boldsymbol{\phi}_{\boldsymbol{\alpha}_2,\boldsymbol{\nu}_2})\|}{\|\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2\|\|\boldsymbol{\nu}_1 - \boldsymbol{\nu}_2\|}\right\} < \infty.$$

(C6) $T^{1/2}(\widehat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*)$ converges weakly to a tight, centered Gaussian limit $\boldsymbol{N}$.
(C7) The map $(\boldsymbol{\alpha}, \boldsymbol{\nu}) \mapsto P\boldsymbol{\phi}_{\boldsymbol{\alpha},\boldsymbol{\nu}}$ from $\mathcal{A} \times \mathfrak{N}$ to $\mathbb{R}^r$ is Fréchet differentiable at $(\boldsymbol{\alpha}^*, \boldsymbol{\nu}^*)$ with derivative $(\boldsymbol{a}, \boldsymbol{b}) \mapsto \Phi_{\boldsymbol{\alpha}^*,\boldsymbol{\nu}^*,1}(\boldsymbol{a}) + \Phi_{\boldsymbol{\alpha}^*,\boldsymbol{\nu}^*,2}(\boldsymbol{b})$. That is, $\Phi_{\boldsymbol{\alpha}^*,\boldsymbol{\nu}^*,1}$, $\Phi_{\boldsymbol{\alpha}^*,\boldsymbol{\nu}^*,2}$ are continuous, linear maps such that for every $\|\boldsymbol{a}\| \to 0$, $\|\boldsymbol{b}\| \to 0$,

$$\left\|P\boldsymbol{\phi}_{\boldsymbol{\alpha}^*+\boldsymbol{a},\boldsymbol{\nu}^*+\boldsymbol{b}} - P\boldsymbol{\phi}_{\boldsymbol{\alpha}^*,\boldsymbol{\nu}^*} - \Phi_{\boldsymbol{\alpha}^*,\boldsymbol{\nu}^*,1}(\boldsymbol{a}) - \Phi_{\boldsymbol{\alpha}^*,\boldsymbol{\nu}^*,2}(\boldsymbol{b})\right\| = o(\|\boldsymbol{a}\| + \|\boldsymbol{b}\|).$$

Further assume that $\Phi_{\boldsymbol{\alpha}^*,\boldsymbol{\nu}^*,1}$ is invertible.
(C8) The β-mixing coefficients of $(X_t)_{t\in\mathbb{Z}}$ satisfy $\sum_{t=0}^\infty \beta(t) < \infty$ and $\sum_{t=0}^\infty \int_0^{\beta(t)} Q^2(u)du < \infty$, where Q is the inverse survival function of $\|\boldsymbol{\phi}_{\boldsymbol{\alpha}^*,\boldsymbol{\nu}^*}\|$.

Conditions (C2) and (C6) make convergence of $\widehat{\boldsymbol{\nu}}$, the estimator of the nuisance parameter, a prerequisite. The other conditions concern the regularity of the time series and identifying functions $\boldsymbol{\phi}_{\boldsymbol{\alpha},\boldsymbol{\nu}}$. The following theorems establish consistency and asymptotic normality of $\widehat{\boldsymbol{\alpha}}$, our estimator for the parameter of interest.

Theorem A.1. *If (C1)–(C4) hold, then* $\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\| = o_p(1)$.

Theorem A.2. *Under conditions (C1)–(C8), it holds*

$$\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^* = -\Phi^{-1}_{\boldsymbol{\alpha}^*,\boldsymbol{\nu}^*,1}\left\{\mathbb{P}_T\boldsymbol{\phi}_{\boldsymbol{\alpha}^*,\boldsymbol{\nu}^*} + \Phi_{\boldsymbol{\alpha}^*,\boldsymbol{\nu}^*,2}(\widehat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*)\right\} + o_p(T^{-1/2}),$$

and

$$T^{1/2}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*) \to_d \mathcal{N}\{0, \Phi^{-1}_{\boldsymbol{\alpha}^*,\boldsymbol{\nu}^*,1}\boldsymbol{\Sigma}_{\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*}(\Phi^{-1}_{\boldsymbol{\alpha}^*,\boldsymbol{\nu}^*,1})'\},$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*}$ *is the limiting covariance of* $T^{1/2}\{\mathbb{P}_T\boldsymbol{\phi}_{\boldsymbol{\alpha}^*,\boldsymbol{\nu}^*} + \Phi_{\boldsymbol{\alpha}^*,\boldsymbol{\nu}^*,2}(\widehat{\boldsymbol{\nu}} - \boldsymbol{\nu}^*)\}$.

Now let $F_{\boldsymbol{\alpha},\boldsymbol{\nu}}$ be a cumulative distribution function parametrized by $(\boldsymbol{\alpha}, \boldsymbol{\nu})$. Suppose the parameter of interest is defined as $\mu^* = \psi(F_{\boldsymbol{\alpha}^*,\boldsymbol{\nu}^*})$ for some functional ψ. Denote $F_{N,\boldsymbol{\alpha},\boldsymbol{\nu}}$ the empirical measure over N *iid* realizations from $F_{\boldsymbol{\alpha},\boldsymbol{\nu}}$. For estimators $\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\nu}}$, define $\widehat{\mu} = \psi(F_{N,\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\nu}}})$.

Theorem A.3. *Suppose that the maps $F \mapsto \psi(F)$ and $(\boldsymbol{\alpha}, \boldsymbol{\nu}) \to \psi(F_{\boldsymbol{\alpha},\boldsymbol{\nu}})$ are Frechet differentiable at $F_{\boldsymbol{\alpha}^*,\boldsymbol{\nu}^*}$ and $(\boldsymbol{\alpha}^*, \boldsymbol{\nu}^*)$ respectively.*

1. *If $T \to \infty$ and $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\nu}}) \to_p (\boldsymbol{\alpha}^*, \boldsymbol{\nu}^*)$, then $\widehat{\mu} \to_p \mu^*$.*
2. *If $T = o(N)$ and $T^{1/2}\{(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\nu}}) - (\boldsymbol{\alpha}^*, \boldsymbol{\nu}^*)\}$ converges weakly to a tight process $(\boldsymbol{A}, \boldsymbol{N})$, then*
$$T^{1/2}(\widehat{\mu} - \mu^*) \to_d \Psi_{(\boldsymbol{\alpha}^*)}(\boldsymbol{A}) + \Psi_{(\boldsymbol{\nu}^*)}(\boldsymbol{N}),$$
where $(\boldsymbol{a}, \boldsymbol{b}) \mapsto \Psi_{(\boldsymbol{\alpha}^*)}(\boldsymbol{a}) + \Psi_{(\boldsymbol{\nu}^*)}(\boldsymbol{b})$ is the Frechet derivative of the map $(\boldsymbol{a}, \boldsymbol{b}) \mapsto \psi(F_{\boldsymbol{a},\boldsymbol{b}})$ at $(\boldsymbol{\alpha}^*, \boldsymbol{\nu}^*)$.

In the context of our paper, $\widehat{\boldsymbol{\nu}}$ is a vector of empirical distribution functions. Lemma 4.1 of Chen and Fan (2006b) establishes (C2) and (C6), but under conditions slightly stronger than our (C1) and (C8). The following lemma improves their result accordingly. For sake of completeness, we give a detailed proof in Section S6.6 of the supplementary material.

Lemma A.1. *Let $Z_1, \ldots, Z_T \in \mathbb{R}$ be a stationary time series with β-mixing coefficients $\beta(t), t \geq 0$. Define $F_T(z) = (T+1)^{-1}\sum_{t=1}^T \mathbb{1}(Z_t \leq z)$ and $W_T = (F_T - F_Z)(z)/w\{F_Z(z)\}$, where $w(u) = u^\gamma(1-u)^\gamma$, $\gamma \in [0, 1)$.*

1. *If $\beta(t) \to 0$, $\sup_{z\in\mathbb{R}} |W_T(z)| \to 0$ almost surely.*
2. *If $\gamma \in [0, 1/2)$ and $\beta(t) = O(t^{-a})$ with $a > 1/(1-2\gamma)$, the process W_T converges weakly in $\ell^\infty(\mathbb{R})$ to a tight Gaussian limit W with mean zero and covariance*
$$E\{W(z_1)W(z_2)\} = \frac{\mathrm{Var}\{\mathbb{1}(Z_1 \leq z_1), \mathbb{1}(Z_1 \leq z_2)\} + 2\sum_{t=2}^\infty \mathrm{Cov}\{\mathbb{1}(Z_1 \leq z_1), \mathbb{1}(Z_t \leq z_2)\}}{w(z_1)w(z_2)}.$$

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2021.11.015. It contains additional illustrations, simulation results, and all proofs.

References

Aas, K., 2016. Pair-copula constructions for financial applications: A review. Econometrics 4 (4), 43.
Aas, K., Czado, C., Frigessi, A., Bakken, H., 2009. Pair-copula constructions of multiple dependence. Insurance Math. Econom. 44 (2), 182–198.
Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59 (3), 817–858.
Beare, B.K., 2010. Copulas and temporal dependence. Econometrica 78 (1), 395–410.
Beare, B.K., 2012. Archimedean copulas and temporal dependence. Econom. Theory 28 (6), 1165–1185.
Beare, B.K., Seo, J., 2015. Vine copula specifications for stationary multivariate Markov chains. J. Time Series Anal. 36 (2), 228–246.
Bedford, T., Cooke, R.M., 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., Cooke, R.M., 2002. Vines—a new graphical model for dependent random variables. Ann. Stat. 30 (4), 1031–1068.
Bladt, M., McNeil, A.J., 2020. Time series copula models using d-vines and v-transforms: an alternative to GARCH modelling. arXiv:2006.11088.
Brechmann, E.C., Czado, C., 2015. COPAR—multivariate time series modeling using the copula autoregressive model. Appl. Stoch. Models Bus. Ind. 31 (4), 495–514.
Bücher, A., Kojadinovic, I., 2016. A dependent multiplier bootstrap for the sequential empirical copula process under strong mixing. Bernoulli 22 (2), 927–968.
Bücher, A., Kojadinovic, I., 2019. A note on conditional versus joint unconditional weak convergence in bootstrap consistency results. J. Theoret. Probab. 32 (3), 1145–1165.
Bühlmann, P.L., 1993. The Blockwise Bootstrap in Time Series and Empirical Processes (Ph.D. thesis). ETH Zurich.
Chen, X., Fan, Y., 2006a. Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. J. Econom. 135 (1), 125–154.
Chen, X., Fan, Y., 2006b. Estimation of copula-based semiparametric time series models. J. Econom. 130 (2), 307–335.
Chen, X., Huang, Z., Yi, Y., 2021. Efficient estimation of multivariate semi-nonparametric GARCH filtered copula models. J. Econom. 222 (1), 484–501, Annals Issue:Financial Econometrics in the Age of the Digital Economy.
Chen, X., Wu, W.B., Yi, Y., 2009. Efficient estimation of copula-based semiparametric Markov models. Ann. Stat. 37 (6B), 4214–4253.
Chen, X., Xiao, Z., Wang, B., 2020. Copula-based time series with filtered nonstationarity. J. Econom. in press.
Czado, C., 2019. Analyzing Dependent Data With Vine Copulas. In: Lecture Notes in Statistics, Springer, vol. 222, Springer.
Darsow, W.F., Nguyen, B., Olsen, E.T., 1992. Copulas and Markov processes. Illinois J. Math. 36 (4), 600–642.
Dehling, H., Philipp, W., 2002. Empirical process techniques for dependent data. In: Empirical Process Techniques For Dependent Data. Springer, pp. 3–113.
Dissmann, J., Brechmann, E.C., Czado, C., Kurowicka, D., 2013. Selecting and estimating regular vine copulae and application to financial returns. Comput. Statist. Data Anal. 59, 52–69.
Engle, R.F., Sheppard, K., 2001. Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Technical report, National Bureau of Economic Research.
Genest, C., Ghoudi, K., Rivest, L.-P., 1995. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika 82 (3), 543–552.

Genest, C., Rémillard, B., 2008. Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. Annales De L'Institut Henri PoincarÉ, ProbabilitÉs Et Statistiques 44 (6), 1096–1127.

Ghalanos, A., 2019. rmgarch: Multivariate GARCH Models. CRAN, R package version 1.3-7.

Ghalanos, A., 2020. rugarch: Univariate GARCH Models. CRAN, R package version 1.4-4.

Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. J. Am. Stat. Assoc. 102 (477), 359–378.

Gong, J., Li, Y., Peng, L., Yao, Q., 2015. Estimation of extreme quantiles for functions of dependent random variables. J. R. Stat. Soc. Ser. B Stat. Methodol. 77 (5), 1001–1024.

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

Hu, L., 2006. Dependence patterns across financial markets: a mixed copula approach. Appl. Financial Econ. 16 (10), 717–729.

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

Ibragimov, R., Lentzas, G., 2017. Copulas and long memory. Probab. Surv. 14, 289–327.

Ibragimov, R., Müller, U.K., 2010. T-statistic based correlation and heterogeneity robust inference. J. Bus. Econ. Stat. 28 (4), 453–468.

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). In: IMS Lecture Notes Monograph Series, vol. 28, Inst. Math. Statist., Hayward, CA, pp. 120–141.

Joe, H., 1997. Multivariate Models And Dependence Concepts. In: Monographs on Statistics and Applied Probability, vol. 73, Chapman & Hall, London.

Joe, H., 2005. Asymptotic efficiency of the two-stage estimation method for copula-based models. J. Multivariate Anal. 94 (2), 401–419.

Joe, H., 2014. Dependence Modeling With Copulas. CRC Press.

Jordan, A., Krüger, F., Lerch, S., 2019. Evaluating probabilistic forecasts with scoringrules. J. Stat. Softw. 90 (12), 1–37.

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

Krüger, D., 2018. General Vine Copula Models for Stationary Multivariate Time Series. Technical University of Munich.

Künsch, H.R., 1989. The jackknife and the bootstrap for general stationary observations. Ann. Stat. 17 (3), 1217–1241.

Loaiza-Maya, R., Smith, M.S., Maneesoonthorn, W., 2018. Time series copulas for heteroskedastic data. J. Appl. Econom. 33 (3), 332–354.

Longla, M., Peligrad, M., 2012. Some aspects of modeling dependence in copula-based Markov chains. J. Multivariate Anal. 111, 234–240.

Nagler, T., Vatter, T., 2020. rvinecopulib: High Performance Algorithms for Vine Copula Modeling. CRAN, R package version 0.5.2.1.0.

Nasri, B.R., Rémillard, B.N., 2019. Copula-based dynamic models for multivariate time series. J. Multivariate Anal. 172, 107–121, Dependence Models.

Nasri, B.R., Rémillard, B.N., Bouezmarni, T., 2019. Semi-parametric copula-based models under non-stationarity. J. Multivariate Anal. 173, 347–365.

Oh, D.H., Patton, A.J., 2017. Modeling dependence in high dimensions with factor copulas. J. Bus. Econ. Stat. 35 (1), 139–154.

Patton, A.J., 2006. Modelling asymmetric exchange rate dependence. Internat. Econom. Rev. 47 (2), 527–556.

Patton, A.J., 2009. Copula–based models for financial time series. In: Mikosch, T., Kreiß, J.-P., Davis, R.A., Andersen, T.G. (Eds.), Handbook Of Financial Time Series. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 767–785.

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

Pfaff, B., 2008. ?Var, SVAR and SVEC models: Implementation within R Package vars. J. Stat. Softw. 27 (4), 1–32.

Rémillard, B., Papageorgiou, N., Soustra, F., 2012. Copula-based semiparametric models for multivariate time series. J. Multivariate Anal. 110, 30–42.

Rio, E., 2017. Asymptotic Theory Of Weakly Dependent Random Processes. vol. 80, Springer.

Simard, C., Rémillard, B., 2015. Forecasting time series with multivariate copulas. Depend. Model. 3 (1), 59–82.

Sklar, M., 1959. Fonctions de répartition à n dimensions et leurs marges. Publications De L'Institut Statistique De L'UniversitÉ De Paris 8, 229–231.

Smith, M.S., 2015. Copula modelling of dependence in multivariate time series. Int. J. Forecast. 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.

Tsukahara, H., 2005. Semiparametric estimation in copula models. Canad. J. Statist. 33 (3), 357–375.

Van Der Vaart, A.W., Wellner, J., 1996. Weak convergence and empirical processes: with applications to statistics. Springer Science & Business Media.