Testing the simplifying assumption in high-dimensional vine copulas

Malte S. Kurz

TUM School of Management, Technical University of Munich,
Arcisstr. 21, 80333 Munich, Germany
e-mail: malte.kurz@tum.de

Fabian Spanhel

Department of Statistics, Ludwig-Maximilians-Universität München,
Akademiestr. 1, 80799 Munich, Germany
e-mail: spanhel@stat.uni-muenchen.de

Abstract: Testing the simplifying assumption in high-dimensional vine copulas is a difficult task. Tests must be based on estimated observations and check constraints on high-dimensional distributions. So far, corresponding tests have been limited to single conditional copulas with a low-dimensional set of conditioning variables. We propose a novel testing procedure that is computationally feasible for high-dimensional data sets and that exhibits a power that decreases only slightly with the dimension. By discretizing the support of the conditioning variables and incorporating a penalty in the test statistic, we mitigate the curse of dimensionality by looking for the possibly strongest deviation from the simplifying assumption. The use of a decision tree renders the test computationally feasible for large dimensions. We derive the asymptotic distribution of the test and analyze its finite sample performance in an extensive simulation study. An application of the test to four real data sets is provided.

Keywords and phrases: Conditional copula, pair-copula construction, partial vine copula, simplifying assumption, test for constant conditional correlation, vine copula.

Contents

1. Introduction

Vine copulas [24, 4, 1] are a popular tool to model multivariate dependence. An extensive literature concerning vine copulas has been based on the simplifying assumption [11, 19, 25, 28, 34]. This assumption states that every conditional copula in the vine copula does not vary in its conditioning arguments [21]. Multivariate distributions which can be represented as simplified vine copulas have been identified in early papers [21, 44]. More recently, the simplifying assumption has again attracted a lot of attention [10, 17, 18, 27, 32, 34, 43]. In the context of (bivariate) conditional copulas non- and semiparametric tests for the simplifying assumption have been developed [2, 17, 18]. See Derumigny and Fermanian [10] for a survey. In these studies the simplifying assumption is tested for one single conditional copula with a low-dimensional conditioning vector. The assumption is not tested for a vine copula where several conditional copulas with a possibly high-dimensional conditioning vector need to be checked.

We propose a framework for testing the simplifying assumption in high-dimensional vine copulas. By means of the partial vine copula, we introduce a novel stochastic interpretation of the simplifying assumption which is useful for testing it in high dimensions. We test the null hypothesis that the conditional correlation of the partial probability integral transforms associated with an edge of a vine is constant w.r.t. the conditioning variables. A rejection of this

hypothesis implies that the simplifying assumption can be rejected as well. To obtain a test which is still powerful in high dimensions, we discretize the support of the conditioning variables into a finite number of subsets and incorporate a penalty in the test statistic. To render the test computationally feasible in high dimensions, we apply a decision tree to find the possibly largest difference in the set of conditional correlations. The test can be applied in high dimensions which is demonstrated in real data applications and in simulation studies with up to 12-dimensional data sets. An accompanying R-package `pacotest` [29] is publicly available and has been applied to even higher-dimensional data sets [27]. The test can be used to detect building blocks of a vine copula where the simplifying assumption does not seem to be adequate and the estimation of a conditional copula that is varying in its conditioning variables can improve the modeling [40]. It can also be applied to construct new methods for the structure selection of vine copulas [27].

The organization of the paper is as follows. The partial vine copula and stochastic interpretations of the simplifying assumption are discussed in Section 2. In Section 3 we present the test for constant conditional correlations and derive its asymptotic distribution. A decision tree algorithm for searching for the largest deviation from the simplifying assumption is proposed in Section 4. An extensive analysis of the finite sample performance of the test is provided in Section 5. In Section 6 a hierarchical procedure to test the simplifying assumption in vine copulas is presented and applied to simulated and real data. Concluding remarks are given in Section 7.

Throughout the paper we use the following notation and assumptions. The cdf of a d-dimensional random vector $X_{1:d}$ is denoted by $F_{X_{1:d}} := \mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d)$. The distribution function or copula of a random vector $U_{1:d}$ with uniform margins is denoted by $F_{1:d} = C_{1:d}$. For simplicity, we assume that all random variables are real-valued and absolutely continuous. If X and Y are stochastically independent we write $X \perp Y$. For the indicator function we use $\mathbb{1}_{\{A\}} = 1$ if A is true, and $\mathbb{1}_{\{A\}} = 0$ otherwise. $\partial_\theta g(\theta)$ denotes the gradient w.r.t. θ and if $h(\gamma)$ is a d-dimensional function then $\partial_i h(\gamma)$ is the partial derivative w.r.t. the i-th element. All proofs are deferred to the Appendix.

2. The partial vine copula and stochastic interpretations of the simplifying assumption

In this section, we first introduce concepts that are required for the remainder of this paper. Thereafter, we establish a new stochastic interpretation of the simplifying assumption to check its validity.

Let $(A, B, \boldsymbol{Z}) \sim F_{A,B,\boldsymbol{Z}}$ be a random vector where A, B are random variables and $\boldsymbol{Z}$ is a vector of dimension at least 1. For $\boldsymbol{z} \in \text{supp}(\boldsymbol{Z})$ let $F_{A|\boldsymbol{Z}}(\cdot|\boldsymbol{z})$, $F_{B|\boldsymbol{Z}}(\cdot|\boldsymbol{z})$, and $F_{A,B|\boldsymbol{Z}}(\cdot, \cdot|\boldsymbol{z})$ be the corresponding conditional distributions of A, B, and (A, B) given $\boldsymbol{Z} = \boldsymbol{z}$, respectively.

Definition 1. (CPIT, bivariate conditional and partial copula)

(i) $U_{A|\boldsymbol{Z}} := F_{A|\boldsymbol{Z}}(A|\boldsymbol{Z})$ is the conditional probability integral transform (CPIT) of A w.r.t. $\boldsymbol{Z}$.

(ii) For $\boldsymbol{z} \in \text{supp}(\boldsymbol{Z})$, the bivariate conditional copula $C_{A,B;\boldsymbol{Z}}(\cdot,\cdot;\boldsymbol{z})$ of the bivariate conditional distribution $F_{A,B|\boldsymbol{Z}}(\cdot,\cdot|\boldsymbol{z})$ (Patton [37]) is defined as

$$C_{A,B;\boldsymbol{Z}}(a,b;\boldsymbol{z}) := \mathbb{P}(U_{A|\boldsymbol{Z}} \le a, U_{B|\boldsymbol{Z}} \le b | \boldsymbol{Z} = \boldsymbol{z}).$$

(iii) The bivariate partial copula $C^{\text{p}}_{A,B;\boldsymbol{Z}}$ of $F_{A,B|\boldsymbol{Z}}$ [5, 16, 41] is defined as

$$C^{\text{p}}_{A,B;\boldsymbol{Z}}(a,b) := \mathbb{P}(U_{A|\boldsymbol{Z}} \le a, U_{B|\boldsymbol{Z}} \le b).$$

While the bivariate conditional copula is the conditional distribution of a pair of uniform CPITs, the partial copula is the bivariate unconditional distribution of a pair of uniform CPITs. Sometimes, the function $C_{A,B;\boldsymbol{Z}}(\cdot,\cdot;\boldsymbol{z})$ is constant for all $\boldsymbol{z} \in \text{supp}(\boldsymbol{Z})$, e.g., if the distribution of $(A,B,\boldsymbol{Z})$ is multivariate normal. In this case, all conditional copulas $C_{A,B;\boldsymbol{Z}}(\cdot,\cdot;\boldsymbol{z}), \boldsymbol{z} \in \text{supp}(\boldsymbol{Z})$, are equal to the partial copula $C^{\text{p}}_{A,B;\boldsymbol{Z}}$. The bivariate conditional copula $C_{A,B;\boldsymbol{Z}}(\cdot,\cdot;\boldsymbol{z}), \boldsymbol{z} \in \text{supp}(\boldsymbol{Z})$, of the bivariate conditional distribution $F_{A,B|\boldsymbol{Z}}(\cdot,\cdot|\boldsymbol{z})$ arises if one expresses the cdf $F_{A,B,\boldsymbol{Z}}$ as follows

$$\begin{aligned} F_{A,B,\boldsymbol{Z}}(a,b,\boldsymbol{z}) &= \int_{\boldsymbol{t}\le\boldsymbol{z}} F_{A,B|\boldsymbol{Z}}(a,b|\boldsymbol{t})dF_{\boldsymbol{Z}}(\boldsymbol{t}) \\ &= \int_{\boldsymbol{t}\le\boldsymbol{z}} C_{A,B;\boldsymbol{Z}}\big(F_{A|\boldsymbol{Z}}(a|\boldsymbol{t}), F_{B|\boldsymbol{Z}}(b|\boldsymbol{t}); \boldsymbol{t}\big)dF_{\boldsymbol{Z}}(\boldsymbol{t}). \end{aligned} \tag{2.1}$$

A complete decomposition of the corresponding multivariate copula $C_{A,B,\boldsymbol{Z}}$ into bivariate conditional copulas of bivariate conditional distributions is given by an R-vine copula. The underlying graphical structure of an R-vine copula is an R-vine.

Definition 2. (R-vine – Bedford and Cooke [3])
The sequence of trees $\mathcal{V} := (T_1,\ldots,T_{d-1}) := ((N_1,E_1),\ldots,(N_{d-1},E_{d-1}))$ is an R-vine on d elements if

(i) T_1 is a tree with nodes $N_1 = \{1,\ldots,d\}$ and set of edges E_1.

(ii) For $j = 2,\ldots,d-1$: T_j is a tree with nodes $N_j = E_{j-1}$ and set of edges E_j.

(iii) Proximity condition for $j = 2,\ldots,d-1$: If two nodes in T_j are joined by an edge, the nodes (being edges in T_{j-1}) must share a common node in T_{j-1}.

The complete union associated with the edge $e = \{a,b\} \in E_j$ in tree T_j is defined as

$$\mathcal{U}_e := \{n \in N_1 : \exists e_1 \in E_1,\ldots,\exists e_{j-1} \in E_{j-1}, \text{ with } n \in e_1 \in \ldots \in e_{j-1} \in e\}.$$

The conditioning set D_e of the edge $e = \{a,b\} \in E_j$ in tree T_j is given by $D_e := \mathcal{U}_a \cap \mathcal{U}_b$ and always consists of $(j-1)$ elements. The conditioned sets

associated with the edge $e = \{a, b\} \in E_j$ in tree T_j are defined as $h_e := \mathcal{U}_a \setminus D_e$ and $i_e := \mathcal{U}_b \setminus D_e$ and are by construction singleton indices. We denote the set of all edges of the R-vine $\mathcal{V}$ from tree T_l on as $\mathcal{E}_l^d := \bigcup_{j=l}^{d-1} E_j$ so that the constraint set of the R-vine $\mathcal{V}$ is given as $\mathcal{CV} := \{(h_e, i_e; D_e) : e \in \mathcal{E}_1^d\}$.

The representation of a multivariate copula $C_{1:d}$ in terms of an R-vine copula arises if one assigns the corresponding bivariate copulas $C_{h_e, i_e}, e \in E_1$, of the bivariate distributions F_{h_e, i_e} to the edges of the first tree and the corresponding bivariate conditional copulas $C_{h_e, i_e; D_e}, e \in E_j, j = 2, \ldots, d-1$, of the bivariate conditional distributions $F_{h_e, i_e | D_e}$ to the edges of higher trees [3]. Using (2.1) for uniform margins and multiple times for different dimensions of $\boldsymbol{Z}$ and taking all partial derivatives to get the density, the following Proposition 1 can be shown.

Proposition 1. (R-vine copula representation – Bedford and Cooke [3])
Let $d \geq 3$ and $U_{1:d}$ be a uniform random vector with cdf $F_{1:d} = C_{1:d}$. Consider an R-vine structure $\mathcal{V}$ (Definition 2). Define $u_{l_e | D_e} := F_{l_e | D_e}(u_{l_e} | u_{D_e})$ for $e \in \mathcal{E}_2^d$, $l_e = h_e, i_e$ and denote the conditional copula of $F_{h_e, i_e | D_e}(\cdot, \cdot | u_{D_e})$ by $C_{h_e, i_e; D_e}(\cdot, \cdot; u_{D_e})$ (Definition 1). For $e \in E_1$ we set $u_{l_e | D_e} = u_{l_e}$ with $l_e = h_e, i_e$ and $C_{h_e, i_e; D_e} = C_{h_e, i_e}$. The density of $C_{1:d}$ can be expressed as

$$c_{1:d}(u_{1:d}) = \prod_{e \in \mathcal{E}_1^d} c_{h_e, i_e; D_e}\big(u_{h_e | D_e}, u_{i_e | D_e}; u_{D_e}\big).$$

In general, the estimation of an R-vine copula density is a difficult task if the dimension is not low. In order to simplify the modeling process and to overcome the curse of dimensionality, it is commonly assumed that the simplifying assumption holds for the data generating vine copula.

Assumption 1. (The simplifying assumption – Hobæk Haff, Aas and Frigessi [21])
The R-vine copula in Proposition 1 satisfies the simplifying assumption if $c_{h_e, i_e; D_e}(\cdot, \cdot; u_{D_e})$ does not depend on u_{D_e} for all $e \in \mathcal{E}_2^d$.

Assumption 1 characterizes the simplifying assumption in terms of restrictions that are placed on the functional form of bivariate conditional copulas of bivariate conditional distributions. That is, the simplifying assumption holds if each $(j + 1)$-dimensional[1] function $c_{h_e, i_e; D_e}(\cdot, \cdot; u_{D_e})$ only depends on its first two arguments, but the other $(j - 1)$ arguments u_{D_e} have no effect for all edges $e \in E_j, j = 2, \ldots, d-1$.

Note that the building blocks of the R-vine copula in Proposition 1 are determined by $C_{1:d}$. The other way round, we can assign arbitrary bivariate conditional copulas to the edges of an R-vine to construct a multivariate copula. It is also possible to assign bivariate copulas to the edges of an R-vine to construct a

---

[1]For $e \in E_j$, $c_{h_e, i_e; D_e}$ is a $(j + 1)$-dimensional function because it maps $(u_{h_e | D_e}, u_{i_e | D_e}, u_{D_e})$ to a value of the density of the conditional distribution of $(U_{h_e | D_e}, U_{i_e | D_e})$ given $U_{D_e} = u_{d_e}$.

multivariate copula. In this case, we call the resulting construction a simplified vine copula which is given as follows.

Definition 3. (Simplified vine copula (SVC) or pair-copula construction – Joe [23], Aas et al. [1])
Let $d \geq 3$ and consider an R-vine structure $\mathcal{V}$. For $e \in \mathcal{E}_1^d$, let $C_{h_e,i_e;D_e}^{\mathrm{SVC}}$ be a bivariate copula which we call a pair-copula.[2] Let $u_{l_e|D_e}^{\mathrm{SVC}} := u_{l_e}$ for $e \in E_1, l_e = h_e, i_e$. For $e \in E_j$, $j = 2, \ldots, d-1$, define $u_{l_e|D_e}^{\mathrm{SVC}}$ recursively as

$$u_{l_e|D_e}^{\mathrm{SVC}} := \partial_2 C_{l_{\bar{e}},m_{\bar{e}};D_{\bar{e}}}^{\mathrm{SVC}}(u_{l_{\bar{e}}|D_{\bar{e}}}^{\mathrm{SVC}}, u_{m_{\bar{e}}|D_{\bar{e}}}^{\mathrm{SVC}}), \tag{2.2}$$

with $\bar{e} \in E_{j-1}$ and where $m_{\bar{e}} \in D_e$ is selected such that $(l_{\bar{e}}, m_{\bar{e}}; D_{\bar{e}}) = (l_e, m_{\bar{e}}; D_e \setminus m_{\bar{e}}) \in \mathcal{CV}$.[3] The density of the resulting simplified vine copula is given by

$$c_{1:d}^{\mathrm{SVC}}(u_{1:d}) := \prod_{e \in \mathcal{E}_1^d} c_{h_e,i_e;D_e}^{\mathrm{SVC}}(u_{h_e|D_e}^{\mathrm{SVC}}, u_{i_e|D_e}^{\mathrm{SVC}}).$$

Note that $u_{l_e|D_e}^{\mathrm{SVC}}$ for $e \in E_j$, $j = 2, \ldots, d-1$, in Definition 3 is a function of (u_{l_e}, u_{D_e})[4] and that the density is specified by $j(j-1)/2$ pair-copulas $C_{h_e,i_e;D_e}^{\mathrm{SVC}}, e \in E_j, j = 1, \ldots, d-1$. From the second tree on, each of these pair-copulas, i.e., $C_{h_e,i_e;D_e}^{\mathrm{SVC}}$ for $e \in E_j, j = 2, \ldots, d-1$, together with $(u_{h_e|D_e}^{\mathrm{SVC}}, u_{i_e|D_e}^{\mathrm{SVC}})$, specifies the conditional distribution $F_{h_e,i_e|D_e}$ of $U_{1:d} \sim C_{1:d}^{\mathrm{SVC}}$. Moreover, $C_{h_e,i_e;D_e}^{\mathrm{SVC}}$ determines the bivariate conditional copula $C_{h_e,i_e;D_e}$ of the bivariate conditional distribution $F_{h_e,i_e|D_e}$ which is given by

$$\mathbb{P}(U_{h_e|D_e} \leq a, U_{i_e|D_e} \leq b | U_{D_e} = u_{D_e}) =: C_{h_e,i_e;D_e}(a,b;u_{D_e}) = C_{h_e,i_e;D_e}^{\mathrm{SVC}}(a,b).$$

If the data generating R-vine copula does not satisfy the simplifying assumption, a simplified vine copula can be used as an approximation. The partial vine copula is a special simplified vine copula which minimizes the Kullback-Leibler divergence from the data generating copula in a tree-wise fashion (Spanhel and Kurz [43]). Moreover, it gives rise to a stochastic interpretation of the simplifying assumption which is useful for testing the assumption in high dimensions. In order to define the partial vine copula (PVC) we have to construct partial probability integral transforms and higher-order partial copulas.

---

[2]We denote the building blocks $C_{h_e,i_e;D_e}^{\mathrm{SVC}}, e \in E_j, j = 1, \ldots, d-1$, of a simplified vine copula as pair-copulas to avoid any confusion with the bivariate copulas $C_{h_e,i_e}, e \in E_1$, in the first tree of the vine, or the bivariate conditional copulas $C_{h_e,i_e;D_e}, e \in E_j, j = 2, \ldots, d-1$, of bivariate conditional distributions $F_{h_e,i_e|D_e}$ in higher trees, where $F_{h_e,i_e|D_e}$ is a conditional distribution of $U_{1:d} \sim C_{1:d}^{\mathrm{SVC}}$.

[3]By the proximity condition of the R-vine $\mathcal{V}$ (Definition 2) there exists a $m_{\bar{e}} \in D_e$ such that either $(l_e, m_{\bar{e}}; D_e \setminus m_{\bar{e}}) \in \mathcal{CV}$ or $(m_{\bar{e}}, l_e; D_e \setminus m_{\bar{e}}) \in \mathcal{CV}$ with $\bar{e} \in E_{j-1} = N_j$ and w.l.o.g. the definition in equation (2.2) is restricted to the first case – otherwise one can use the identity $C_{i_e,h_e;D_e}^{\mathrm{SVC}}(a,b) = C_{h_e,i_e;D_e}^{\mathrm{SVC}}(b,a)$.

[4]It would be more explicit to write $u_{l_e|D_e}^{\mathrm{SVC}}(u_{l_e}, u_{D_e})$ instead of $u_{l_e|D_e}^{\mathrm{SVC}}$. However, the shortened notation is commonly used in the vine copula literature and simplifies the notation.

Definition 4. (Partial probability integral transforms and higher-order partial copulas – Spanhel and Kurz [43])
Let $d \geq 3$ and $U_{1:d}$ be a uniform random vector with cdf $F_{1:d} = C_{1:d}$. Consider an R-vine structure $\mathcal{V}$ (Definition 2). For $e \in E_2$ we define the **first-order partial copula** as

$$C^{\mathrm{PVC}}_{h_e, i_e; D_e} := C^{\mathrm{P}}_{h_e, i_e; D_e}. \tag{2.3}$$

Define the **partial probability integral transforms (PPITs)** as $U^{\mathrm{PVC}}_{l_e | D_e} := U_{l_e | D_e}$ for $e \in E_2$ with $l_e = h_e, i_e$ and for $e \in E_j$, $j = 3, \ldots, d-1$, as

$$U^{\mathrm{PVC}}_{l_e | D_e} := \partial_2 C^{\mathrm{PVC}}_{l_{\bar{e}}, m_{\bar{e}}; D_{\bar{e}}}(U^{\mathrm{PVC}}_{l_{\bar{e}} | D_{\bar{e}}}, U^{\mathrm{PVC}}_{m_{\bar{e}} | D_{\bar{e}}}), \tag{2.4}$$

with $\bar{e} \in E_{j-1}$ and where $m_{\bar{e}} \in D_e$ is selected such that $(l_{\bar{e}}, m_{\bar{e}}; D_{\bar{e}}) = (l_e, m_{\bar{e}}; D_e \setminus m_{\bar{e}}) \in \mathcal{CV}$.
For $e \in E_j$, $j = 3, \ldots, d-1$, the **$(j-1)$-th order partial copula** is defined as

$$C^{\mathrm{PVC}}_{h_e, i_e; D_e}(a, b) := \mathbb{P}(U^{\mathrm{PVC}}_{h_e | D_e} \leq a, U^{\mathrm{PVC}}_{i_e | D_e} \leq b). \tag{2.5}$$

Note that a first-order partial copula $C^{\mathrm{PVC}}_{h_e, i_e; D_e}$ in (2.3) equals the corresponding partial copula $C^{\mathrm{P}}_{h_e, i_e; D_e}$ for $e \in E_2$. However, a higher-order partial copula in (2.5) is, in general, not equal to the corresponding partial copula for $e \in E_j$, $j = 3, \ldots, d-1$. The partial vine copula is now constructed by assigning the specific first-order and higher-order partial copulas to the edges of an R-vine.

Definition 5. (Partial vine copula (PVC) – Spanhel and Kurz [43])
Let $d \geq 3$ and $U_{1:d}$ be a uniform random vector with cdf $F_{1:d} = C_{1:d}$. Consider an R-vine structure $\mathcal{V}$ (Definition 2). $C^{\mathrm{PVC}}_{1:d}$ denotes the **partial vine copula (PVC)** of $(C_{1:d}, \mathcal{V})$ and its density is given by

$$c^{\mathrm{PVC}}_{1:d}(u_{1:d}) := \prod_{e \in \mathcal{E}^d_1} c^{\mathrm{PVC}}_{h_e, i_e; D_e}(u^{\mathrm{PVC}}_{h_e | D_e}, u^{\mathrm{PVC}}_{i_e | D_e}),$$

where $c^{\mathrm{PVC}}_{h_e, i_e; D_e}$ is the density of $C^{\mathrm{PVC}}_{h_e, i_e; D_e}$ defined in (2.3) for $e \in E_2$ and (2.5) for $e \in E_j$, $j = 3, \ldots, d-1$. For $e \in E_1$ we set $C^{\mathrm{PVC}}_{h_e, i_e; D_e} = C_{h_e, i_e}$.

To illustrate non-simplified vine copulas, simplified vine copulas and the partial vine copula, we consider a special case of Example 4.2 in Spanhel and Kurz [43]. For this purpose, let us introduce the following notation. If D_θ denotes a parametric copula family with dependence parameter $\theta \in \Theta$ and cdf $D_\theta(a, b)$ we write $C_{A,B; \boldsymbol{Z}} = D_{g(\boldsymbol{z})}$, where $g(\boldsymbol{z}) \in \Theta$, if $C_{A,B; \boldsymbol{Z}}(a, b; \boldsymbol{z}) = D_{g(\boldsymbol{z})}(a, b)$ for all $(a, b, \boldsymbol{z})$. Moreover, $C_{A,B; \boldsymbol{Z}} = D$, where D is a bivariate copula with cdf $D(a, b)$, if $C_{A,B; \boldsymbol{Z}}(a, b; \boldsymbol{z}) = D(a, b)$ for all $(a, b, \boldsymbol{z})$.

Example 1.
Let $C^\perp$ be the product copula with cdf $C^\perp(a, b) = ab$ and C^{FGM}_θ be the FGM copula with parameter $\theta \in [-1, 1]$ and cdf $C^{\mathrm{FGM}}_\theta(a, b) := ab[1 + \theta(1-a)(1-b)]$. The building-blocks of the four-dimensional vine copula are chosen to be

$$C_{12} = C_{23} = C_{34} = C^\perp, \quad C_{13; 2} = C^{\mathrm{FGM}}_{1-u_2}, \quad C_{24; 3} = C^{\mathrm{FGM}}_{1-u_3}, \quad C_{14; 23} = C^\perp.$$

(a) Vine copula Example 1.

(b) PVC of Example 1.
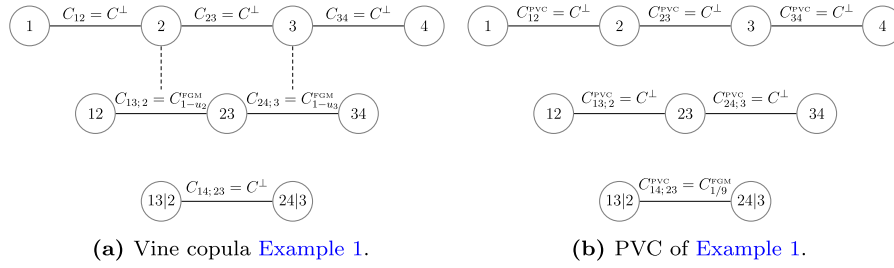
Fig 1: The non-simplified vine copula in Example 1 and its PVC. The influence of conditioning variables on the conditional copulas is indicated by dashed lines.

The density of the data generating R-vine copula defined in Example 1 is given by

$$
\begin{aligned}
c_{1:4}(u_{1:4}) &= \underbrace{c_{12}(u_1, u_2)\, c_{23}(u_2, u_3)\, c_{34}(u_3, u_4)}_{\text{first tree } T_1} \\
&\quad \times \underbrace{c_{13;2}(u_{1|2}, u_{3|2}; u_2)\, c_{24;3}(u_{2|3}, u_{4|3}; u_3)}_{\text{second tree } T_2} \times \underbrace{c_{14;23}(u_{1|23}, u_{4|23}; u_{2:3})}_{\text{third tree } T_3} \\
&= \underbrace{c_{13;2}(u_1, u_3; u_2)\, c_{24;3}(u_2, u_4; u_3)}_{\text{second tree } T_2} \\
&= c_{1-u_2}^{\text{FGM}}(u_1, u_3)\, c_{1-u_3}^{\text{FGM}}(u_2, u_4),
\end{aligned}
$$

where the second equality holds because the copulas in the first and third tree are product copulas. For this data generating process, all building blocks of the vine copula are product copulas except for the second tree T_2. In tree T_2, both $c_{13;2}(\cdot, \cdot; u_2)$ and $c_{24;3}(\cdot, \cdot; u_3)$ depend on the conditioning variables u_2 and u_3, respectively. Thus, the simplifying assumption does not hold. The building blocks of the non-simplified vine copula defined in Example 1 are illustrated in the left panel of Figure 1.

In order to approximate the non-simplified copula by a simplified vine copula one could assign arbitrary pair-copulas to the edges of the vine. For example the corresponding partial copulas. One can easily check that the partial copulas in the second tree are given by product copulas. Moreover, the partial copulas in the other trees are also given by product copulas. Thus, the resulting simplified vine copula model is the four-dimensional product copula with density given by

$$
c_{1:4}^{\text{SVC}}(u_{1:4}) = \underbrace{c_{13;2}^{\text{P}}(u_{1|2}, u_{3|2})\, c_{24;3}^{\text{P}}(u_{2|3}, u_{4|3})}_{\text{second tree } T_2} = 1.
$$

A better approximation, in terms of Kullback-Leibler divergence minimization, is given by the PVC (see Example 4.2 in Spanhel and Kurz [43]) with

density

$$c_{1:4}^{\mathrm{PVC}}(u_{1:4}) = \underbrace{c_{14;23}^{\mathrm{PVC}}(u_{1|23}^{\mathrm{PVC}}, u_{4|23}^{\mathrm{PVC}})}_{\text{third tree } T_3} = c_{14;23}^{\mathrm{PVC}}(u_1, u_4) = c_{1/9}^{\mathrm{FGM}}(u_1, u_4),$$

where the second equality holds because the copulas in the first and second tree of the PVC are product copulas. Note that the PVC approximates $C_{13;2}$ and $C_{24;3}$ in the second tree by the first-order partial copula which, by definition, always equals the partial copula. Thus, just like the previous approximation, the second tree is modeled by product copulas. However, the second-order partial copula $C_{14;23}^{\mathrm{PVC}}$ in the third tree T_3 is a FGM copula with positive dependence and not equal to the corresponding partial copula $C_{14;23}^{\mathrm{P}}$, which is the product copula. Consequently, the four-dimensional approximation given by the PVC is not the four-dimensional product copula. The building blocks of the PVC corresponding to Example 1 are illustrated in the right panel of Figure 1.

The partial copula as well as the PVC give rise to the following stochastic interpretations of the simplifying assumption.

Theorem 1. (Stochastic interpretations of the simplifying assumption)

Let $d \geq 3$ and $U_{1:d}$ be a uniform random vector. Consider a fixed R-vine structure $\mathcal{V}$ and the corresponding R-vine copula decomposition stated in Proposition 1. The following statements are equivalent:

(i) The R-vine copula satisfies the simplifying assumption (Assumption 1).
(ii) $\forall e \in \mathcal{E}_2^d :$ $(U_{h_e|D_e}, U_{i_e|D_e}) \perp U_{D_e}$
(iii) $\forall e \in \mathcal{E}_2^d :$ $(U_{h_e|D_e}^{PVC}, U_{i_e|D_e}^{PVC}) \perp U_{D_e}$

Theorem 1 highlights that the simplifying assumption is equivalent to $(d-1)(d-2)/2$ vectorial independence assumptions. Note that $(U_{h_e|D_e}, U_{i_e|D_e}) \perp U_{D_e}$ in (ii) can be replaced by $C_{h_e,i_e;D_e} = C_{h_e,i_e;D_e}^{\mathrm{P}}$ and that $(U_{h_e|D_e}^{\mathrm{PVC}}, U_{i_e|D_e}^{\mathrm{PVC}}) \perp U_{D_e}$ in (iii) can be replaced by $C_{h_e,i_e;D_e} = C_{h_e,i_e;D_e}^{\mathrm{PVC}}$. While the different stochastic interpretations (ii) and (iii) are equivalent in theory, (iii) is much more useful for testing the simplifying assumption. In practice, observations from the pair of CPITs $(U_{h_e|D_e}, U_{i_e|D_e})$ or the pair of PPITs $(U_{h_e|D_e}^{\mathrm{PVC}}, U_{i_e|D_e}^{\mathrm{PVC}})$ are not observable and have to be estimated from data. Observations from the CPIT $U_{l_e|D_e}$ can be obtained by estimating the j-dimensional conditional distribution function $F_{l_e|D_e}$ of U_{l_e} given U_{D_e}, which is not an easy task. For $e \in E_j$, $j = 2, \ldots, d-1$, the PPIT $U_{l_e|D_e}^{\mathrm{PVC}}$ is a composition of $j(j-1)/2$ pair-copulas which belong to the building blocks of the corresponding PVC given in equation (2.4) of Definition 4. Thus, for observations from a PPIT one can sequentially estimate a sequence of pair-copulas which is much simpler. Therefore, we use (iii) to construct a test which is based on pseudo-observations from the PPITs.

To illustrate the practical advantage of testing the simplifying assumption with PPITs ((iii) in Theorem 1) instead of CPITs ((ii) in Theorem 1) let us reconsider Example 1. In this case, the simplifying assumption can be formulated

either as

$$(ii): \quad \begin{aligned} &\text{Tree } T_2: \quad (U_{1|2}, U_{3|2}) \perp U_2 \quad \text{and} \quad (U_{2|3}, U_{4|3}) \perp U_3 \\ &\text{Tree } T_3: \quad (U_{1|23}, U_{4|23}) \perp U_{2:3}, \end{aligned}$$

or

$$(iii): \quad \begin{aligned} &\text{Tree } T_2: \quad (U_{1|2}, U_{3|2}) \perp U_2 \quad \text{and} \quad (U_{2|3}, U_{4|3}) \perp U_3 \\ &\text{Tree } T_3: \quad (U_{1|23}^{\mathrm{PVC}}, U_{4|23}^{\mathrm{PVC}}) \perp U_{2:3}. \end{aligned}$$

For the four-dimensional Example 1, both formulations (ii) and (iii) consist of three vectorial independencies, all of which must be true to be equivalent to the simplifying assumption, see Theorem 1. In tree T_2, the required conditions are identical for both (ii) and (iii). Since the copulas in tree T_2 are given by bivariate FGM copulas with a varying parameter, these vectorial independencies do not hold. Thus, the simplifying assumption does not hold according to both (ii) and (iii). Note that the difference between (ii) and (iii) becomes only apparent after the second tree. In the third tree, the vectorial independencies stated in (ii) and (iii) are quite different. For the data generating vine copula (Example 1) the statement in tree T_3 is obviously true for (ii). However, it is false for (iii) because $U_{1|23}^{\mathrm{PVC}} = U_1$ and, as can be readily verified, $U_1 \perp U_{2:3}$ is false.

If the data generating vine copula is unknown, it is more practical to base a test on (iii) than on (ii). In order to check whether the pair of CPITs $(U_{1|23}, U_{4|23})$ are jointly independent from $U_{2:3}$ in (ii), one has to estimate the unknown conditional distribution functions $F_{1|23}$ and $F_{4|23}$. The specification of a flexible parametric model for such conditional distributions is difficult. Non-parametric estimation might be sensible if the conditioning vector is very low-dimensional but suffers from the curse of dimensionality if the conditioning vector is high-dimensional. On the other side, (iii) just requires the estimation of pair-copulas, irrespective of the dimension of the conditioning vector. That is because both PPITs can be written as a composition of pair-copulas. For instance,

$$U_{1|23}^{\mathrm{PVC}} = \partial_2 C_{13;2}^{\mathrm{PVC}} \big(\partial_2 C_{12}(U_1, U_2), \partial_1 C_{23}(U_2, U_3) \big),$$

so that only the estimation of C_{12}, C_{23} and $C_{13;2}^{\mathrm{PVC}}$ are required to estimate $U_{1|23}^{\mathrm{PVC}}$. In many applications, it might be feasible to find good parametric models for these pair-copulas in order to obtain an appropriate estimate of $U_{1|23}^{\mathrm{PVC}}$.

3. Constant conditional correlation (CCC) test for $H_0: (U_{h_e|D_e}^{\mathrm{PVC}}, U_{i_e|D_e}^{\mathrm{PVC}}) \perp U_{D_e}$

Before we consider testing the simplifying assumption, we first develop in this and the next section tests for the null hypothesis $H_0: (U_{h_e|D_e}^{\mathrm{PVC}}, U_{i_e|D_e}^{\mathrm{PVC}}) \perp U_{D_e}$ for some $e \in E_j, j = 2, \ldots, d-1$. The main challenge is that the dimension $(j-1)$ of U_{D_e} can be rather large so that the power of consistent tests is not satisfying in higher trees. For instance, a consistent test could be obtained using a Cramér-von Mises test for vectorial independence (Kojadinovic and Holmes

[26] and Quessy [39]). However, as it is pointed out by Gijbels, Omelka and Veraverbeke [17] and shown in our simulation, the power of such a consistent test rapidly approaches the significance level if the dimension increases. Our focus is on a test that exhibits a power that is high for alternatives that one encounters in practical applications and that is quite robust to the dimension of the data set. For this reason, we consider the null hypothesis that the conditional correlation of the PPITs $(U_{h_e|D_e}^{\mathrm{PVC}}, U_{i_e|D_e}^{\mathrm{PVC}})$ associated with one edge of a vine is constant w.r.t. the conditioning variables U_{D_e}. A rejection of this hypothesis implies that the simplifying assumption can be rejected as well. To obtain a test whose power does not collapse substantially with the dimension of the conditioning vector, we now discretize the support of the conditioning vector into a finite number of subsets.

3.1. CCC test with known observations

For the ease of exposition, we introduce the main idea of the test without a reference to vine copulas. Let A, B be uniform random variables and $\boldsymbol{Z}$ be a random vector with uniform margins. We want to check the null hypothesis $H_0\colon (A, B) \perp \boldsymbol{Z}$. Let $\Lambda_0 := \mathrm{supp}(\boldsymbol{Z})$, $\Lambda_1, \Lambda_2 \subset \Lambda_0$ with $\Lambda_1 \cap \Lambda_2 = \emptyset$, and $\mathbb{P}(\boldsymbol{Z} \in \Lambda_1 \cup \Lambda_2) = 1$ with $\mathbb{P}(\boldsymbol{Z} \in \Lambda_1), \mathbb{P}(\boldsymbol{Z} \in \Lambda_2) > 0$. We call $\Gamma := \{\Lambda_1, \Lambda_2\}$ a partition of the support Λ_0 of $\boldsymbol{Z}$ into two disjoint subsets.[5] We are interested in the correlation between A and B in the two subgroups determined by Γ, i.e.,

$$r_l := \mathrm{Corr}(A, B | \boldsymbol{Z} \in \Lambda_l) = \frac{\mathrm{Cov}(A, B | \boldsymbol{Z} \in \Lambda_l)}{\sqrt{\mathrm{Var}(A | \boldsymbol{Z} \in \Lambda_l)\mathrm{Var}(B | \boldsymbol{Z} \in \Lambda_l)}},$$

for $l = 1, 2$. Under $H_0 : (A, B) \perp \boldsymbol{Z}$, it follows that

$$\mathrm{Corr}(A, B) = r_1 = r_2,$$

i.e., the conditional correlations are constant w.r.t. the conditioning event. As estimate for the correlation in the l-th group we use the sample version of the conditional correlation which we denote by $\hat{r}_l$.

A statistic for testing the equality of the correlations in the two samples is given by

$$T_n^\star(\Gamma) = n\frac{(\hat{r}_1 - \hat{r}_2)^2}{\hat{\sigma}^2(\hat{r}_1) + \hat{\sigma}^2(\hat{r}_2)},$$

where $\hat{\sigma}^2(\hat{r}_l), l = 1, 2$, is a consistent estimator (see Appendix A.2) for the asymptotic variance of $\sqrt{n}(\hat{r}_l - r_l)$. By construction of the estimators, the asymptotic covariance between $\hat{r}_1$ and $\hat{r}_2$ is zero. Thus, under regularity conditions and H_0 it can be readily verified that $T_n^\star(\Gamma) \xrightarrow{d} \chi^2(1)$.

---

[5]The idea of discretization has some similarity to the boxes approach of Derumigny and Fermanian [10] but differs substantially. We only discretize the support of the conditioning vector and the rejection of our null hypothesis is still a rejection of the simplifying assumption which is not always true for the approach of Derumigny and Fermanian [10]. Moreover, we present a data-driven approach to select the partition so that the idea of discretization can also be applied in high-dimensional settings without the need to impose strong a priori assumptions on the form of the partition.

In a more general setting, one can also use a partition of the support Λ_0 into $L \in \mathbb{N}$ pairwise disjoint subsets $\Gamma := \{\Lambda_1, \ldots, \Lambda_L\}$ and test whether

$$H_0 : r_1 = \ldots = r_L \quad \text{vs.} \quad H_A : \exists l_1, l_2 \in \{1, \ldots, L\}, l_1 \neq l_2 : r_{l_1} \neq r_{l_2}.$$

For this purpose, denote the vector of sample correlations in the groups by $\hat{R}_\Gamma^\star = (\hat{r}_1, \ldots, \hat{r}_L)^T$. Further, define the diagonal $L \times L$ matrix $\hat{\Sigma}_{R_\Gamma}^\star$, with diagonal elements $\hat{\Sigma}_{R_\Gamma, l, l}^\star = \hat{\sigma}^2(\hat{r}_l)$ and a $(L-1) \times L$ first-order difference matrix A so that $(A\hat{R}_\Gamma^\star)^T A\hat{R}_\Gamma^\star = \sum_{l=1}^{L-1}(\hat{r}_l - \hat{r}_{l+1})^2$. A statistic to test the equality of correlations in L groups is then defined by the quadratic form[6]

$$T_n^\star(\Gamma) = n(A\hat{R}_\Gamma^\star)^T(A\hat{\Sigma}_{R_\Gamma}^\star A^T)^{-1}A\hat{R}_\Gamma^\star,$$

with asymptotic distribution $T_n^\star(\Gamma) \overset{d}{\to} \chi^2(L-1)$ under $H_0 : (A, B) \perp \boldsymbol{Z}$.

3.2. CCC test with estimated pseudo-observations from the PPITs

In order to test the simplifying assumption, we have to test null hypotheses of the form $H_0 \colon (U_{h_e|D_e}^{\text{PVC}}, U_{i_e|D_e}^{\text{PVC}}) \perp U_{D_e}$. In practice, if $X_{1:d} \sim F_{X_{1:d}}$ is the data generating process, we cannot directly observe a sample of $(U_{h_e|D_e}^{\text{PVC}}, U_{i_e|D_e}^{\text{PVC}}, U_{D_e})$ but have to estimate such observations on the basis of a sample of $(X_{h_e}, X_{i_e}, X_{D_e})$. To obtain these pseudo-observations, we use a semi-parametric approach. First, we use normalized ranks to obtain pseudo-observations from $U_i = F_{X_i}(X_i)$ for $i = 1, \ldots, d$. We then consider a fixed R-vine structure $\mathcal{V}$ and assume that there is a parametric simplified vine copula model $\{C_{1:d; \theta_{1:d-1}}^{\text{SVC}} \colon \theta_{1:d-1} \in \Upsilon\}$ for the PVC $C_{1:d}^{\text{PVC}}$ of $(C_{1:d}, \mathcal{V})$ with $\theta_{1:d-1; 0} \in \Upsilon$ so that $C_{1:d; \theta_{1:d-1; 0}}^{\text{SVC}} = C_{1:d}^{\text{PVC}}$. Finally, we use the common stepwise ML estimator [20] to construct pseudo-observations from $(U_{h_e|D_e}^{\text{PVC}}, U_{i_e|D_e}^{\text{PVC}}, U_{D_e})$. The asymptotic distribution of the resulting test statistic with these estimated pseudo-observations is stated in Proposition 2.

Proposition 2.
Let $(X_{1:d}^k)_{k=1,\ldots,n}$ be n independent copies of $X_{1:d}$ and $C_{1:d}$ be the copula of $X_{1:d}$. Consider a fixed R-vine structure $\mathcal{V}$ (Definition 2) and assume there is a parametric simplified vine copula model $\{C_{1:d; \theta_{1:d-1}}^{SVC} \colon \theta_{1:d-1} \in \Upsilon\}$ for the PVC $C_{1:d}^{PVC}$ of $(C_{1:d}, \mathcal{V})$ with $\theta_{1:d-1; 0} \in \Upsilon$ so that $C_{1:d; \theta_{1:d-1; 0}}^{SVC} = C_{1:d}^{PVC}$. Let $e \in \mathcal{E}_2^d$ be fixed. Assume that the regularity conditions stated in Theorem 1 in Hobæk Haff [20] hold[7] and that the partition $\Gamma := \{\Lambda_1, \ldots, \Lambda_L\}$, where $\Lambda_1, \ldots, \Lambda_L \subset \Lambda_0 = \text{supp}(U_{D_e})$, satisfies

(i) $\Lambda_{l_1} \cap \Lambda_{l_2} = \emptyset$, for $l_1 \neq l_2$ with $1 \leq l_1, l_2 \leq L$,

---

[6]The statistic $T_n^\star(\Gamma)$ also follows from $(B\hat{R}_\Gamma^\star)^T B\hat{R}_\Gamma^\star = \sum_{l=1}^{L-1} \hat{\pi}_l(\hat{r}_l - \bar{r})^2$, where $\bar{r}$ is the average correlation, $\hat{\pi}_l$ is defined in Appendix A.2 as the fraction of data corresponding to the subset Λ_l and with an appropriately defined matrix B.

[7]Note that the results in Hobæk Haff [20] are written up for D-vine copulas but can be generalized to R-vine copulas [20]. The entries of the matrices needed to compute the standard errors of the sequentially estimated parameters of R-vine copulas can for example be found in Stöber and Schepsmeier [45].

(ii) $\mathbb{P}(U_{D_e} \in \bigcup_{l=1}^{L} \Lambda_l) = 1$ *with* $\mathbb{P}(U_{D_e} \in \Lambda_l) > 0$ *for all* $1 \leq l \leq L$.

Let $\hat{R}_\Gamma$ *be the vector of sample correlations that are computed using the pseudo-observations from the PPITs and* $\hat{\Sigma}_{R_\Gamma} := \hat{\Sigma}^\star_{R_\Gamma} + \hat{\Sigma}_{PVC} + \hat{\Sigma}_r$, *where* $\hat{\Sigma}_{R_\Gamma}$ *is defined in equation* $(A.8)$ *in Appendix A.2. Construct the test statistic*

$$T_n(\Gamma) = n(A\hat{R}_\Gamma)^T (A\hat{\Sigma}_{R_\Gamma} A^T)^{-1} A\hat{R}_\Gamma. \tag{3.1}$$

Under $H_0 : (U^{PVC}_{h_e|D_e}, U^{PVC}_{i_e|D_e}) \perp U_{D_e}$ *it holds that*

$$T_n(\Gamma) \xrightarrow{d} \chi^2(L-1).$$

The matrices $\Sigma_{\mathrm{PVC}} = \mathrm{plim}_{n \to \infty} \hat{\Sigma}_{\mathrm{PVC}}$ and $\Sigma_r = \mathrm{plim}_{n \to \infty} \hat{\Sigma}_r$ quantify the change in the asymptotic covariance due to the estimation of pseudo-observations from the PPITs. If the marginal distributions are known and we don't have to estimate ranks to obtain pseudo-observations from $C^{\mathrm{PVC}}_{1:d}$ it follows that $\Sigma_r = 0$. If the PVC $C^{\mathrm{PVC}}_{1:d}$ of $(C_{1:d}, \mathcal{V})$ is known it follows that $\Sigma_{\mathrm{PVC}} = 0$. Note that, in general, the off-diagonal elements of $\Sigma_{R_\Gamma} = \mathrm{plim}_{n \to \infty} \hat{\Sigma}_{R_\Gamma}$, i.e., the asymptotic covariances between estimated correlations in different groups, are no longer zero if observations from the PPITs are estimated.

3.3. CCC test with a combination of partitions

The selected partition Γ has an influence on how well a varying conditional correlation is detected by the test. To illustrate this and to motivate a test based on the combination of several partitions, we use the following Example 2.
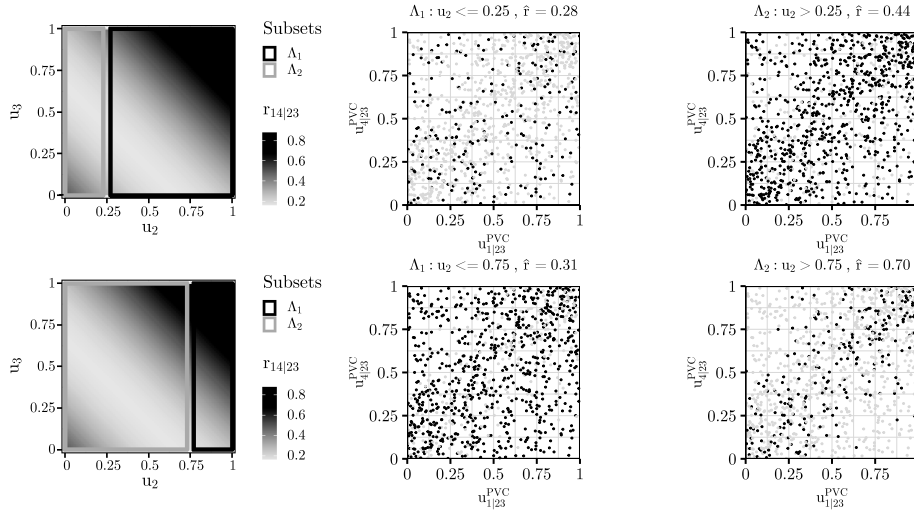
Example 2.
Let C^{Cl}_θ and C^{Fr}_θ be the Clayton and Frank copula with parameter θ, respectively. The building-blocks of the four-dimensional D-vine copula are chosen to be

$$\begin{aligned}
C_{12} = C_{23} = C_{34} &= C^{\mathrm{Cl}}_{\theta_1}, \\
C_{13;2} = C^{\mathrm{PVC}}_{13;2} = C_{24;3} = C^{\mathrm{PVC}}_{24;3} &= C^{\mathrm{Cl}}_{\theta_2}, \\
C_{14;23} &= C^{\mathrm{Fr}}_{\alpha(u_{2:3};\lambda)}, \\
\alpha(u_{2:3};\lambda) &= 1 + 2.5\lambda(1 - 1.5(u_2 + u_3))^2,
\end{aligned}$$

with $\theta_1 := \frac{2\tau}{1-\tau}$ and $\theta_2 := \frac{\theta_1}{1+\theta_1}$, where τ is the value of Kendall's τ of the bivariate margins in the first tree.[8]

For the illustration we set in Example 2 $\tau = 0.4$ and $\lambda = 1$ and simulate a sample of size $n = 1000$. For instance, if we choose $\Lambda_1 = \{(u_2, u_3) \in [0,1]^2 : u_3 \leq u_2\}$

---

[8]The parameter function is a generalization of the function $\theta(X) = 1 + 2.5(3 - X)^2$ used by Acar, Craiu and Yao [2] for a Frank copula with a one-dimensional conditioning set, where the conditioning variable X is assumed to be uniformly distributed on the interval $[2, 5]$. The parameters of the bivariate Clayton copulas in the first and second tree are specified in a way that Example 2 can be considered as a four-dimensional Clayton copula where the copula in the last tree is replaced by a Frank copula with varying parameter.

Fig 2: The plots show 1000 random draws from Example 2. On the left hand side the shaded background visualizes the conditional correlation $r_{14|23}$ of $C_{14;23}$ as a function of U_2 and U_3. In the middle and on the right hand side, realizations from the PPITs $(U_{1|23}^{\mathrm{PVC}}, U_{4|23}^{\mathrm{PVC}})$ grouped according to Γ_1 in the upper row and Γ_2 in the lower row are shown. The black points are observations being assigned to the corresponding subset and the light-gray points are observations which have been assigned to the other subset.

and $\Lambda_2 = \Lambda_0 \setminus \Lambda_1$, then $r_1 = r_2$ and the power of the test is asymptotically equal to the level of the test. Instead, we could use partitions such as $\Gamma_1 := \{\Lambda_1 = [0, 0.25] \times [0, 1], \Lambda_2 = (0.25, 1] \times [0, 1]\}$ or $\Gamma_2 := \{\Lambda_1 = [0, 0.75] \times [0, 1], \Lambda_2 = (0.75, 1] \times [0, 1]\}$. In Figure 2, we illustrate the resulting tests $T_n(\Gamma_1)$ and $T_n(\Gamma_2)$. The upper row corresponds to the first partition Γ_1 where the difference of the correlations in the two groups is $\hat{r}_2 - \hat{r}_1 = 0.16$, yielding a test statistic value of $T_n(\Gamma_1) = 5.41$. In contrast, if we consider the second partition Γ_2 shown in the lower row of Figure 2, we get $\hat{r}_2 - \hat{r}_1 = 0.39$ and $T_n(\Gamma_2) = 65.72$.

In order to increase the probability that the test detects a varying conditional correlation, it seems naturally to consider several partitions $\Gamma_0, \ldots, \Gamma_M$, $M \geq 1$, where each partition Γ_m is a collection of $L_m \geq 2$ subsets of the support Λ_0. A test statistic using a combination of partitions is given by

$$\Theta_n = \max \{T_n(\Gamma_0) + n\lambda_n, T_n(\Gamma_1), \ldots, T_n(\Gamma_M)\} - n\lambda_n, \qquad (3.2)$$

where λ_n is a penalty function and Γ_0 is the base partition whose corresponding test statistic $T_n(\Gamma_0)$ is the only one that is not penalized. The construction of such a test has some similarity to the approach of Lavergne and Patilea [30]. The idea is that by choosing an appropriate penalty function, under H_0, the asymptotic distribution of Θ_n and $T_n(\Gamma_0)$ should be equivalent. Moreover, if H_0 is not true, then Θ_n should have more power than $T_n(\Gamma_0)$ because, by construction, $\Theta_n \geq T_n(\Gamma_0)$. Precise conditions are given in the following proposition.

Proposition 3.

Assume that the conditions stated in [Proposition 2](#) hold and that the partitions $\Gamma_0, \ldots, \Gamma_M$ fulfill the conditions stated for Γ in [Proposition 2](#). Additionally, let $\lambda_n : \mathbb{N} \to \mathbb{R}^+$ be a penalty function such that

(i) $n\lambda_n \to \infty$ for $n \to \infty$,
(ii) $\lambda_n \to 0$ for $n \to \infty$.

Set $\Theta_n = \max\{T_n(\Gamma_0) + n\lambda_n, T_n(\Gamma_1), \ldots, T_n(\Gamma_M)\} - n\lambda_n$, where $T_n(\Gamma_m), 0 \leq m \leq M$, is defined like $T_n(\Gamma)$ in equation [(3.1)](#) in [Proposition 2](#) with Γ replaced by Γ_m. Under $H_0 : (U^{PVC}_{h_e|D_e}, U^{PVC}_{i_e|D_e}) \perp U_{D_e}$ it holds that

$$\Theta_n \xrightarrow{d} \chi^2(L_0 - 1),$$

where L_0 is the number of subsets of Γ_0. If there is a partition $\Gamma_{m^\star}, m^\star \in \{0, \ldots, M\}$, such that $\text{plim}_{n\to\infty} \frac{1}{n} T_n(\Gamma_{m^\star}) =: c > 0$ it follows that

$$\Theta_n \xrightarrow{p} \infty.$$

Thus, the critical value of Θ_n under H_0 asymptotically (as $n \to \infty$, for fixed M) only depends on the number of elements L_0 in Γ_0 but not on $\Gamma_1, \ldots, \Gamma_M$. In finite samples, however, the distribution of Θ_n under H_0 also depends on $\Gamma_1, \ldots, \Gamma_M$ and the usage of asymptotic critical values derived from the limiting distribution $\chi^2(L_0 - 1)$ may result in an empirical size that is not close to the theoretical level.[9] Moreover, if there is a partition $\Gamma_{m^\star}$ such that the corresponding correlations are not identical, i.e., $c > 0$, the power of the test approaches 1 if the sample size goes to infinity. Possible choices of the partitions $\Gamma_0, \ldots, \Gamma_M$ and their selection will be discussed in the next section.

4. A data-driven algorithm for testing with a high-dimensional conditioning vector

Both constant conditional correlation (CCC) tests, $T_n(\Gamma)$ and Θ_n, are based on partitions of $\Lambda_0 := \text{supp}(U_{D_e})$. We will first illustrate why an a priori determination of such partitions is problematic and then show how such partitions can be selected in a data-driven fashion.

4.1. A priori defined partitions

As a naive approach one could use the sample median $q^{(i)}_{0.5}$ of each conditioning variable u_i to split the observations into groups and then consider the partition

---

[9] In extensive simulation studies in [Section 5](#) and [Section 6.1](#) it is demonstrated that the empirical size is often comparable to the theoretical level and the choice of the penalty function is analyzed in a simulation study in [Appendix A.6](#)

that results from the combinations of these groups. For instance, for $C_{14;23}$ in the third tree, we obtain the following four subsets forming a partition

$$\Lambda_{(1,1)} = \{(u_2, u_3) \in \Lambda_0 \colon u_2 \le q_{0.5}^{(2)}, u_3 \le q_{0.5}^{(3)}\},$$
$$\Lambda_{(1,2)} = \{(u_2, u_3) \in \Lambda_0 \colon u_2 > q_{0.5}^{(2)}, u_3 \le q_{0.5}^{(3)}\},$$
$$\Lambda_{(2,1)} = \{(u_2, u_3) \in \Lambda_0 \colon u_2 \le q_{0.5}^{(2)}, u_3 > q_{0.5}^{(3)}\},$$
$$\Lambda_{(2,2)} = \{(u_2, u_3) \in \Lambda_0 \colon u_2 > q_{0.5}^{(2)}, u_3 > q_{0.5}^{(3)}\}.$$

Such an approach is however only feasible for a low-dimensional conditioning vector, since the number of subsets increases exponentially with the dimension of the conditioning vector. Moreover, the number of observations that are contained in a single subset might get too small. Alternatively, one could use the mean $\bar{u}_{D_e} = \frac{1}{j-1} \sum_{k=1}^{j-1} u_{D_e^k}$ of the conditioning variables u_{D_e} for $e \in E_j$, $j = 2, \ldots, d-1$, and consider the sample median $q_{0.5}^{\bar{u}_{D_e}}$ of $\bar{u}_{D_e}$ as split point to obtain the following partition $\Gamma_{\mathrm{med}} := \{\Lambda_1, \Lambda_2\}$ of the support $\Lambda_0 = \mathrm{supp}(U_{D_e})$:

$$\Lambda_1 := \left\{ u_{D_e} \in \Lambda_0 : \bar{u}_{D_e} \le q_{0.5}^{\bar{u}_{D_e}} \right\} \text{ and } \Lambda_2 := \Lambda_0 \setminus \Lambda_1. \tag{4.1}$$

However, in practice, such an a priori definition of the partition is arbitrary and might not detect a conditional variation. Therefore, we introduce now a decision tree algorithm which selects the partitions in a more data-driven way and is computationally feasible in high dimensions.

## 4.2. Data-driven choice of the partitions by means of a decision tree algorithm

The test statistic Θ_n can be rewritten in the following way

$$\Theta_n = \max\left\{T_n(\Gamma_0) + n\lambda_n, T_n(\Gamma_{\max})\right\} - n\lambda_n,$$

with $\Gamma_{\max} := \mathrm{argmax}_{\Gamma_m \in \{\Gamma_1, \ldots, \Gamma_M\}} T_n(\Gamma_m)$. The set $\Gamma_{\max}$ denotes the partition, excluding the base partition Γ_0, for which a possible violation of the H_0 is most pronounced. As base partition we set $\Gamma_0 = \Gamma_{\mathrm{med}}$ (see equation (4.1)). To find $\Gamma_{\max}$ in a data-driven and computationally efficient way we use decision trees with depth $J_{\max} = 2$ like the one shown in Figure 3. The decision tree recursively uses binary splits to partition the support $\Lambda_0 = \mathrm{supp}(U_{D_e})$ into disjoint subsets to obtain $\Gamma_{\max} := \{\Lambda_{(0,l,l)}, \Lambda_{(0,l,r)}, \Lambda_{(0,r,l)}, \Lambda_{(0,r,r)}\}$. The possible split points in the first level are given by the empirical quartiles of each conditioning variable and by the empirical quartiles of the mean of the conditioning vector. In the second level, the split points for each leaf are chosen similarly but conditional on the chosen subsets of the first level. Among these possible splits the split is chosen that maximizes the statistic of the CCC test. For algorithmic details of the decision tree with a general depth of $J_{\max}$ we refer to Appendix A.4.

We revisit Example 2 to illustrate how the decision tree adapts to the variation in the conditional correlation. For the same simulated sample that is used
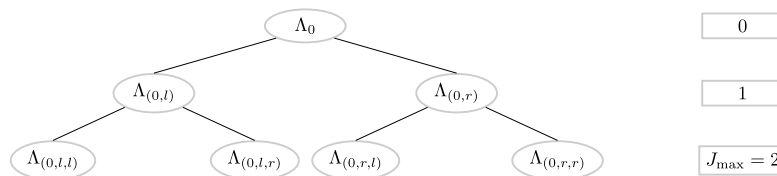
Fig 3: Partitioning of the support $\Lambda_0 = \text{supp}(U_{D_e})$ of the random vector U_{D_e} into disjoint subsets $\{\Lambda_{(0,l,l)}, \Lambda_{(0,l,r)}, \Lambda_{(0,r,l)}, \Lambda_{(0,r,r)}\}$ using a decision tree algorithm with depth $J_{\max} = 2$.
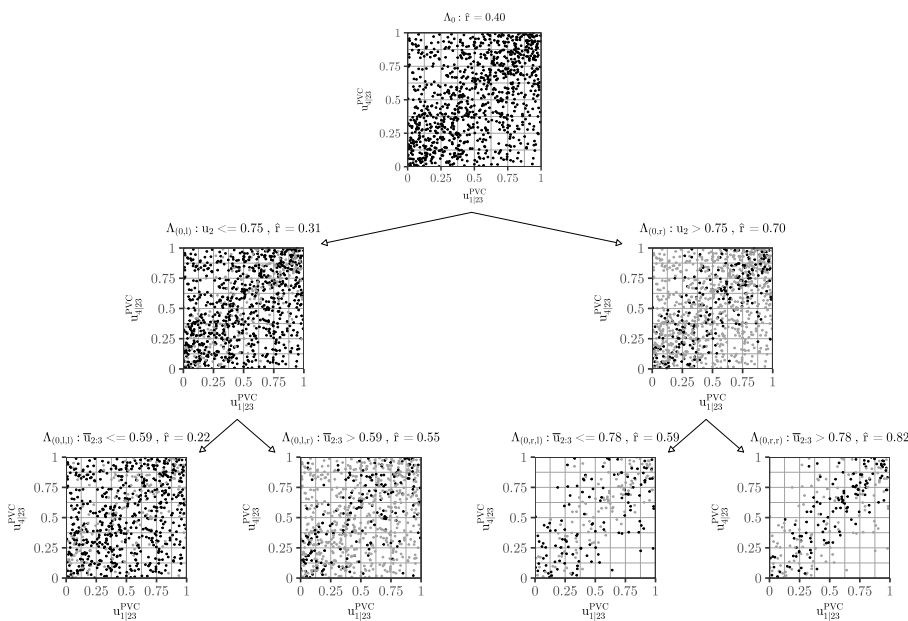
Fig 4: The plots show the same 1000 observations from Example 2 as Figure 2. Shown is the decision tree-based grouping of the PPITs $(U_{1|23}^{\text{PVC}}, U_{4|23}^{\text{PVC}})$ into four different groups according to the partition $\Gamma_{\max} := \{\Lambda_{(0,l,l)}, \Lambda_{(0,l,r)}, \Lambda_{(0,r,l)}, \Lambda_{(0,r,r)}\}$. The black points show observations being assigned to the corresponding subset of the support Λ_0 and the light-gray points correspond to the observations which have been assigned to the other subset due to the binary split. The mean of the conditioning variables u_2 and u_3 is denoted by $\bar{u}_{2:3}$.

for Figure 2, the decision tree is applied to test $H_0 : (U_{1|23}^{\text{PVC}}, U_{4|23}^{\text{PVC}}) \perp U_{2:3}$. The partitioning of Λ_0 into $\Gamma_{\max}$ is visualized in Figure 4 which shows the grouping of the observations from the PPITs $(U_{1|23}^{\text{PVC}}, U_{4|23}^{\text{PVC}})$ according to $\Gamma_{\max}$. In each scatter plot the black observations have been assigned to this leaf while the observations in light gray have been assigned to the other leaf. Furthermore, the estimated correlations in each group, which are used for the

CCC test, are shown. We see that the decision tree chooses a partition with estimated correlations that are quite different and a maximal difference of $\widehat{\text{Corr}}(U_{1|23}^{\text{PVC}}, U_{4|23}^{\text{PVC}} | U_{2:3} \in \Lambda_{(0,l,l)}) - \widehat{\text{Corr}}(U_{1|23}^{\text{PVC}}, U_{4|23}^{\text{PVC}} | U_{2:3} \in \Lambda_{(0,r,r)}) = 0.60$.

TABLE 1

The table refers to the same 1000 *observations from Example 2 as Figure 2. It lists the subsets formed with a decision tree to find* $\Gamma_{max} := \{\Lambda_{(0,l,l)}, \Lambda_{(0,l,r)}, \Lambda_{(0,r,l)}, \Lambda_{(0,r,r)}\}$. *The first column contains the decision tree level J, the second column the subsets and the third column the estimated correlations for the subsets.*

J	Subset Λ_γ	$\widehat{\text{Corr}}(U_{1|23}^{\text{PVC}}, U_{4|23}^{\text{PVC}} | U_{2:3} \in \Lambda_\gamma)$
0	$\Lambda_0 = [0,1]^2$	0.40
1	$\Lambda_{(0,l)} = [0, 0.75] \times [0,1]$	0.31
1	$\Lambda_{(0,r)} = [0.75, 1] \times [0,1]$	0.70
2	$\Lambda_{(0,l,l)} = \Lambda_{(0,l)} \cap \{(u_2, u_3) : \frac{1}{2}(u_2 + u_3) \leq 0.59\}$	0.22
2	$\Lambda_{(0,l,r)} = \Lambda_{(0,l)} \cap \{(u_2, u_3) : \frac{1}{2}(u_2 + u_3) > 0.59\}$	0.55
2	$\Lambda_{(0,r,l)} = \Lambda_{(0,r)} \cap \{(u_2, u_3) : \frac{1}{2}(u_2 + u_3) \leq 0.78\}$	0.59
2	$\Lambda_{(0,r,r)} = \Lambda_{(0,r)} \cap \{(u_2, u_3) : \frac{1}{2}(u_2 + u_3) > 0.78\}$	0.82

The subsets that are constructed by the decision tree to obtain the selected $\Gamma_{\max}$ partition are stated in Table 1 together with the corresponding estimated correlations. The decision tree generates partitions which are no longer simple one-directional splits as in Figure 2. Instead we now get more complex polygons as subsets which are visualized as colored frames in Figure 5. On the left hand side of Figure 5, the shaded area shows the variation in the correlation of $C_{14;23}$ as a function of U_2 and U_3. Areas with darker gray correspond to higher values of the conditional correlation $r_{14|23}$. The right hand side of Figure 5 illustrates how the dependence within the conditioning set, determined by the Clayton copula $C_{23} = C_{\theta_1}^{\text{Cl}}$, influences the variation of $C_{14;23}$. Shown are the realized values of (U_2, U_3) and their grouping into the subsets. The two plots on top correspond to the first binary split, which is done according to the 75% quartile of U_2. The two plots at the bottom show the splits in the second level of the decision tree, which is done according to quartiles of the mean of the conditioning variables U_2 and U_3. The adaption of the decision tree based partition to the variation in the conditional correlation becomes clear when looking at the shaded background which shows the conditional correlation $r_{14|23}$ of $C_{14;23}$ as a function of U_2 and U_3. We see how Λ_0 is partitioned into areas with relatively low $(\Lambda_{(0,l,l)})$, medium $(\Lambda_{(0,l,r)}$ and $\Lambda_{(0,r,l)})$ and high correlation $(\Lambda_{(0,r,r)})$.

In all simulations in Section 5 and Section 6.1, and the real data applications in Section 6.2, we use decision trees with depth $J_{\max} = 2$ and $\lambda_n = \frac{1}{\sqrt{n}}$ for the penalty. The choice of the penalty function is analyzed and explained in Appendix A.6. We further set $\Gamma_0 = \Gamma_{\text{med}}$ (see equation (4.1)) as base partition, because we assume to have no a priori information about the relative importance of each conditioning variable and because the median of the mean of the
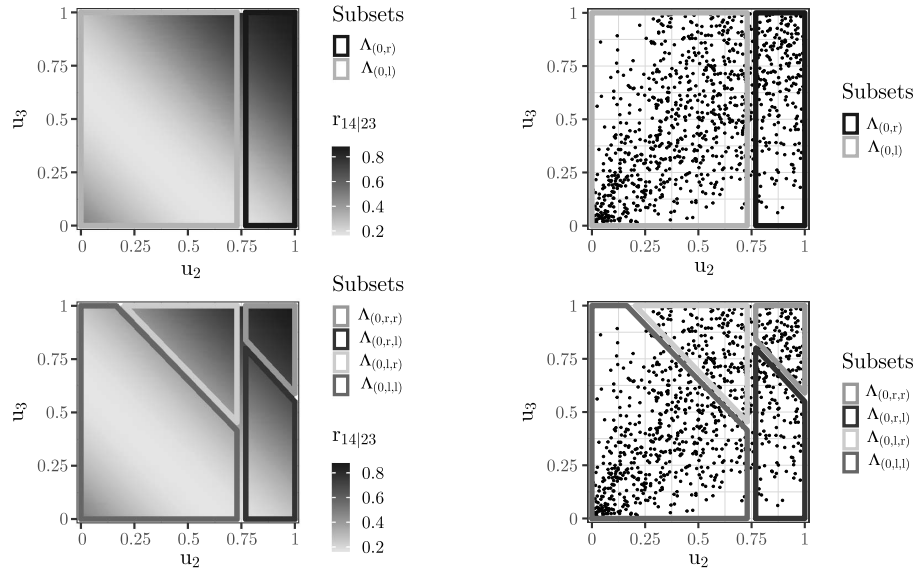
Fig 5: The plots show the same 1000 observations from Example 2 as Figure 2. On the left hand side the shaded background shows the conditional correlation $r_{14|23}$ of $C_{14;23}$ as a function of U_2 and U_3. On the right hand side the realizations of the conditioning variables U_2 and U_3 are shown. The upper plots correspond to the first binary split of the decision tree and the lower plots to the second and third binary split. The colored frames show the different subsets $\{\Lambda_{(0,l)}, \Lambda_{(0,r)}, \Lambda_{(0,l,l)}, \Lambda_{(0,l,r)}, \Lambda_{(0,r,l)}, \Lambda_{(0,r,r)}\}$ of the support $\Lambda_0 := [0,1]^2$.

conditioning vector as split point guarantees well-balanced sample sizes in the groups.[10]

5. CCC test: Simulation study

In the following, the finite-sample performance of the CCC test Θ_n is analyzed and compared to the performance of the vectorial independence (VI) test of Kojadinovic and Holmes [26]. The simulation study is build around Example 2 using a ceteris paribus setup in order to analyze the different key drivers of the empirical power. We will analyze the power of the CCC and the VI test w.r.t. different variations of the conditional copula in Section 5.1, illustrate the power gain of the CCC test due to the decision tree algorithm in Section 5.2, and investigate the performance of both tests w.r.t. the dimensionality of the

---

[10]If the application at hand suggests a priori that H_0 is likely to be false and a particular partition $\Gamma_\star$ would result in a more pronounced difference between the resulting conditional correlations, i.e., $T_n(\Gamma_\star) > T_n(\Gamma_{\mathrm{med}})$, then setting $\Gamma_0 = \Gamma_\star$ instead of $\Gamma_0 = \Gamma_{\mathrm{med}}$ would not deteriorate but possibly improve the power of the test provided H_0 is actually false. However, as it is shown in Section 5.2 the choice of Γ_0 is not crucial if the alternative partitions are selected in a data-driven way using the decision tree.

testing problem in Section 5.3. The effect of misspecified copula families in lower trees is discussed in Section 5.4.

All results for the CCC test are computed with estimated pseudo-observations from the PPITs. Since the asymptotic distribution of the VI test with estimated pseudo-observations is unknown, we use the true observations from the PPITs for the VI test to compute p-values on the basis of 1000 bootstrap samples [39]. Therefore, in simulations with possible misspecifications we do not show results for the VI test.

5.1. Power study: The functional form of the conditional copula

The conditional copula $C_{14;23} = C^{\mathrm{Fr}}_{\alpha(u_{2:3};\lambda)}$ in Example 2 varies in $u_{2:3}$. To alter the variation, we choose values between zero and one for the parameter λ in the function

$$\alpha(u_{2:3};\lambda) = 1 + 2.5\lambda(1 - 1.5(u_2 + u_3))^2. \tag{5.1}$$

For $\lambda = 0$, $C_{14;23}$ does not vary in $u_{2:3}$. Recall that the simplifying assumption is satisfied in the second tree in Example 2, so that $H_0 : (U^{\mathrm{PVC}}_{1|23}, U^{\mathrm{PVC}}_{4|23}) \perp U_{2:3}$ is true if $\lambda = 0$. For $\lambda > 0$, $C_{14;23}$ varies in $u_{2:3}$ and $H_0 : (U^{\mathrm{PVC}}_{1|23}, U^{\mathrm{PVC}}_{4|23}) \perp U_{2:3}$ is false. For $\lambda = 1$ the variation is most pronounced. In Figure 6, the variation of the conditional correlation $r_{14|23}$ of $C_{14;23}$ as a function of the mean $\bar{u}_{2:3} = \frac{1}{2}(u_2 + u_3)$ is shown on the left hand side.[11]

For the sample sizes $n = 500, 1000$, and $\lambda = 0, 0.2, 0.4, 0.6, 0.8, 1$, we apply the CCC test Θ_n and the VI test [26] for the hypothesis $H_0 : (U^{\mathrm{PVC}}_{1|23}, U^{\mathrm{PVC}}_{4|23}) \perp U_{2:3}$. On the right hand side of Figure 6, empirical power values are plotted for different values of the parameter λ. The numbers are based on 1000 samples for each combination of λ and n. The level of the tests is chosen to be 5%. For both tests and sample sizes the empirical size (i.e. the case $\lambda = 0$) is close to the theoretical level of the test. The empirical power of both tests is clearly increasing for all values of λ if one doubles the sample size from 500 to 1000 observations. Furthermore, both tests are more powerful the more the variation in $C_{14;23}$ is pronounced, i.e., the larger the parameter λ is. In terms of empirical power, the CCC test outperforms the VI test in all settings with a relative improvement that often exceeds 300%. Both tests are implemented in an accompanying R-package pacotest [29] and are computationally feasible. For $\lambda = 1$ and a sample size of $n = 1000$, the computational time for the CCC test is 2.25 seconds and for the VI test 0.88 seconds.[12]

---

[11]Note that we have already seen $r_{14|23}$ for $\lambda = 1$ as a function of u_2 and u_3 as shaded background in the plots on the left hand side of Figure 2 and Figure 5.

[12]The reported computational times are the median from 20 repetitions on a single core of an AMD Ryzen 7 PRO 4750U processor. Note that a major part of the computational time for the CCC test is required for the computation of the asymptotic covariances that account for estimated pseudo-observations from the PPITs. In a fair comparison of computational times, the CCC test can, like the VI test, also be computed with known observations from the PPITs. The computational time of the CCC test then drops to 0.01 seconds and is much shorter than the 0.88 seconds of the VI test.
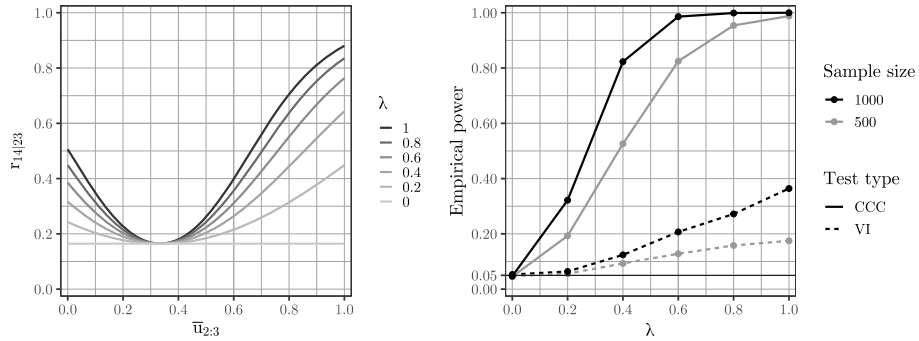
Fig 6: On the left hand side the conditional correlation $r_{14|23}$ of the Frank copula $C_{14;23}$ as a function of the mean $\bar{u}_{2:3}$ of the conditioning variables u_2 and u_3 is shown for different values of the parameter λ. The parameter function is stated in (5.1) and the data generating process is defined in Example 2. The plot on the right hand side shows the empirical power for different values of λ and the theoretical 5% level of the tests. Empirical power values are based on 1000 samples. The solid lines correspond to the CCC test Θ_n and the dashed lines to the VI test. Different sample sizes are shown color coded.

5.2. Power study: Improved power due to the decision tree algorithm

We now compare the CCC test based on the decision tree approach Θ_n with the CCC test $T_n(\Gamma_0)$ which only considers the base partition Γ_0. By construction, $\Theta_n \geq T_n(\Gamma_0)$ always holds, meaning that if we reject on the basis of $T_n(\Gamma_0)$, we also reject on the basis of Θ_n. As a consequence, the empirical power of Θ_n is never smaller than the empirical power of $T_n(\Gamma_0)$. The improvement in power due to the use of Θ_n instead of $T_n(\Gamma_0)$ depends on the data generating process and will be investigated in the following.

For the base partition we choose $\Gamma_0 = \Gamma_{\mathrm{med}}$ as in Section 4. As data generating processes we consider Example 2 and the resulting vine copulas that arise if the parameter $\alpha(u_{2:3}; \lambda)$ of $C^{\mathrm{Fr}}_{\alpha(u_{2:3};\lambda)}$ in the edge of tree T_3 of Example 2 is given by

$$\alpha_I(u_{2:3}; \lambda) = 1 + 2.5\lambda(1 - 2u_2(u_2 + u_3))^2, \tag{5.2}$$

or

$$\alpha_D(u_{2:3}; \lambda) = 1 + 2.5\lambda(1 - 2(u_2 - u_3))^2. \tag{5.3}$$

Figure 7 shows the empirical power of the CCC tests Θ_n and $T_n(\Gamma_0)$ for the hypothesis $H_0 : (U^{\mathrm{pvc}}_{1|23}, U^{\mathrm{pvc}}_{4|23}) \perp U_{2:3}$. For the case of Example 2 (left panel in Figure 7), the test with the fixed partition $\Gamma_0 = \Gamma_{\mathrm{med}}$ delivers a test $T_n(\Gamma_0)$ which performs almost as good as Θ_n. That is because the parameter $\alpha(u_{2:3}; \lambda)$ of $C^{\mathrm{Fr}}_{\alpha(u_{2:3};\lambda)}$ in Example 2 can be written as a function of the mean of the
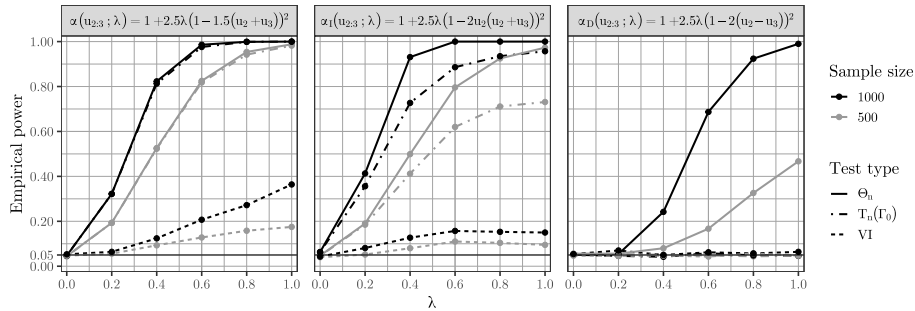
Fig 7: Empirical power of the CCC tests Θ_n and $T_n(\Gamma_0)$ and the VI test for the hypothesis $H_0 : (U_{1|23}^{\mathrm{PVC}}, U_{4|23}^{\mathrm{PVC}}) \perp U_{2:3}$ in Example 2. The three panels correspond to the parameter functions $\alpha(\cdot)$, $\alpha_I(\cdot)$ and $\alpha_D(\cdot)$, respectively. Empirical power values are based on 1000 samples and plotted against the parameter λ. The solid lines correspond to the CCC test Θ_n, the dashdotted lines to the CCC test $T_n(\Gamma_0)$ and the dashed lines to the VI test. Different sample sizes are shown color coded.

conditioning variables $\bar{u}_{2:3} = \frac{1}{2}(u_2 + u_3)$. Furthermore, the conditioning variables are positively associated due to the Clayton copula with $\tau_{23} = 0.4$. As a result, the decision tree rarely finds a partition $\Gamma_{\max}$ which in terms of the test statistic $T_n(\Gamma_{\max})$ improves over the base partition $\Gamma_0 = \Gamma_{\mathrm{med}}$ by an amount larger than the penalty λ_n. Therefore, the test statistic Θ_n often coincides with the test statistic $T_n(\Gamma_0)$ of the base partition.

For the other two cases, the partition Γ_{med} is not a good choice and the decision tree algorithm finds substantially better partitions in a data-driven way. The varying parameter $\alpha_I(u_{2:3}; \lambda)$ of $C_{\alpha_I(u_{2:3};\lambda)}^{\mathrm{Fr}}$ (middle panel in Figure 7) introduces an interaction effect between the two conditioning variables. Although the test with the fixed Γ_{med} partition can detect some variation in $C_{\alpha_I(u_{2:3};\lambda)}^{\mathrm{Fr}}$, the decision tree finds better partitions which can increase the empirical power by more than 20 percentage points. The gain of power is even more pronounced if the parameter $\alpha_D(u_{2:3}; \lambda)$ of $C_{\alpha_D(u_{2:3};\lambda)}^{\mathrm{Fr}}$ (right panel in Figure 7) is a function of the difference of the conditioning variables. Even if $\lambda = 1$ and $C_{\alpha_D(u_{2:3};\lambda)}^{\mathrm{Fr}}$ is strongly varying in $u_{2:3}$, the test with the fixed partition Γ_{med}, that is based on the mean of the conditioning variables, cannot recognize the variation. As a result, the empirical power is identical to the level of the test. On the contrary, the data-driven selection of the partition results in a substantial power increase. For $\lambda = 1$ and $n = 1000$, the power increases from 5% to 99%.

In summary, the choice of the base partition Γ_0 determines a lower bound for the empirical power of the test Θ_n and the decision tree can increase its power. The magnitude of the improvement depends on the data generating process and ranges from negligible, e.g., $\alpha(\cdot)$, to huge, e.g., $\alpha_D(\cdot)$. For all data generating processes, the power of the data-driven test Θ_n is much better than the power of the VI test. The difference is most pronounced for $\alpha_D(\cdot)$ where the empirical power of the VI test is always approximately 5% while the empirical power of the CCC test Θ_n can be 99%. In Appendix A.7 we investigate what kind of

partitions are typically selected by the decision trees and how they adapt to the different variations in the conditional correlation determined by $\alpha(\cdot), \alpha_I(\cdot)$, and $\alpha_D(\cdot)$.

5.3. Power study: The dimension of the conditioning set

For high-dimensional vine copulas, the dimension of the conditioning set of a conditional copula increases rapidly. Therefore, it is substantial that a test still has power if the dimension of the conditioning set is not small. To investigate the performance of the CCC test w.r.t. the dimension of the conditioning set, we consider a up to twelve-dimensional Clayton copula where the Clayton copula in the edge of the last tree is replaced by a Frank copula with varying parameter.
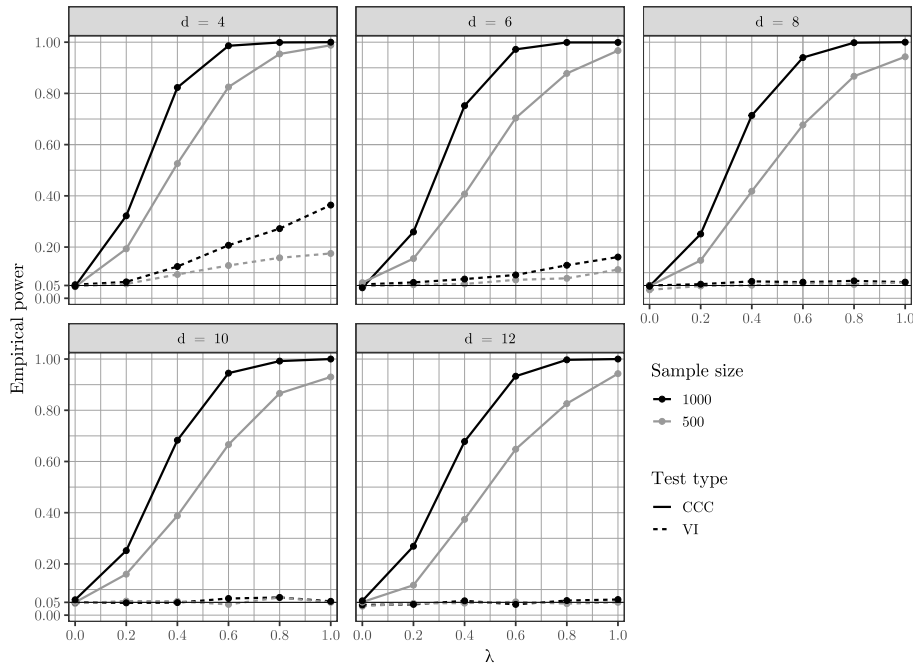
Example 3.
For $d \geq 4$, the building-blocks of the d-dimensional D-vine copula are chosen to be

$$C_{i,i+j;i+1:i+j-1} = C_{\theta_j}^{\text{Cl}}, \qquad 1 \leq j \leq d-2, 1 \leq i \leq d-j,$$
$$C_{1,d;2:d-1} = C_{\alpha(u_{2:3};\lambda)}^{\text{Fr}},$$
$$\alpha(u_{2:3}; \lambda) = 1 + 2.5\lambda(1 - 1.5(u_2 + u_3))^2,$$
$$\theta_j = \frac{\theta_1}{1 + (j-1)\theta_1}, \qquad 2 \leq j \leq d-2,$$
$$\theta_1 = \frac{2\tau}{1 - \tau}.$$

For $d = 4$ Example 3 coincides with Example 2 and as before we set $\tau = 0.4$ and consider different values for λ. For all dimensions, the functional form of the parameter only depends on the conditioning variables U_2 and U_3. Therefore, the variation of $C_{1,d;2:d-1} = C_{\alpha(u_{2:3};\lambda)}^{\text{Fr}}$ in $u_{2:d-1}$ is always the same but the dimension of the testing problem increases with d. Grouped by the dimension d, the empirical power and size of the VI and the CCC test Θ_n for the hypothesis $H_0 : (U_{1|2:d-1}^{\text{PVC}}, U_{d|2:d-1}^{\text{PVC}}) \perp U_{2:d-1}$ are shown in Figure 8.

While the empirical power of the VI test decreases a lot if the dimension of the conditioning set is increased, the empirical power of the CCC test decreases only slightly for higher values of d. In particular, for the setup $\lambda = 1$ and $n = 1000$, the power of the VI test drops from 36% to 5% if the dimension is increased from $d = 4$ to $d = 12$. On the contrary, the power of the CCC test for this setup is always 100%. Moreover, even when the power of the CCC test is not 100% for $d = 4$, the decrease in its power is still marginal. For instance, for $\lambda = 0.6$ and $n = 500$, the power of the CCC test only decreases from 83% to 67% while the power of the VI test quickly drops to approximately 5% if the dimension is increased from $d = 4$ to $d = 12$. Thus, the introduction of a penalty in the CCC test statistic Θ_n and the data-driven selection of the partition $\Gamma_{\max}$ by means of a decision tree yields a test whose power decreases only slightly with the dimension of the conditioning vector.
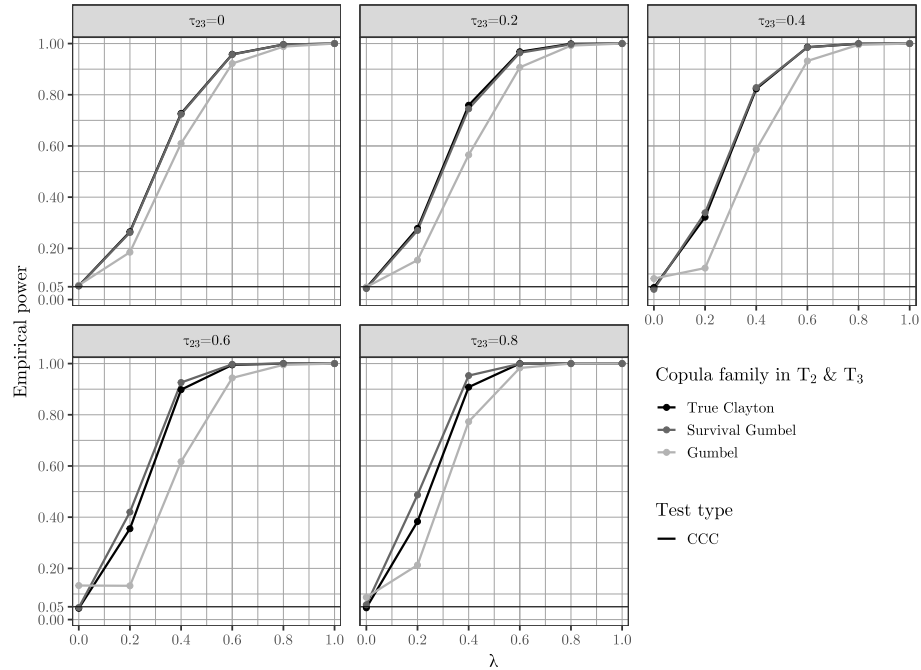
Fig 8: Empirical power of the CCC and VI test for the hypothesis H_0 : $(U_{1|2:d-1}^{\mathrm{PVC}}, U_{d|2:d-1}^{\mathrm{PVC}}) \perp U_{2:d-1}$ in Example 3. Empirical power values are based on 1000 samples and plotted against the parameter λ. For $\lambda = 0$ the plots show the empirical size of the tests. Each plot corresponds to a specific dimension d of the D-vine copula (Example 3). The solid lines correspond to the CCC test Θ_n and the dashed lines to the VI test. Different sample sizes are shown color coded.

5.4. Power study: Misspecification of the copulas in the lower trees

The true family of the five copulas in the first and second tree $C_{12}, C_{23}, C_{34}, C_{13;2}, C_{24;3}$ in Example 2 is the Clayton copula. To analyze the effect of misspecified copula families, we now vary the pairwise value of Kendall's tau $\tau_{23} = \tau_{12} = \tau_{34} = \tau_{13} = \tau_{24}$ between 0 and 0.8 and estimate either survival Gumbel or Gumbel copulas for all five copulas in the lower trees. The black lines in Figure 9 show the results for correctly specified Clayton copulas in the lower trees as a benchmark. Since the strength of the variation of $C_{14;23}$ is more pronounced for higher values of τ_{23}, the empirical power of the tests is also increasing in τ_{23}. The empirical size of the tests ($\lambda = 0$) is not influenced by τ_{23} and always close to the theoretical level of 5%.

The dark grey line in Figure 9 corresponds to a rather mild misspecification where we estimate survival Gumbel copulas in the first and second tree. We see that the empirical size is still very close to the theoretical level of 5% ($\lambda = 0$). Moreover, the power of the test with misspecified survival Gumbel copulas is almost indistinguishable from the power of the test with correctly specified

Fig 9: Empirical power of the CCC test Θ_n for the hypothesis $H_0 : (U_{1|23}^{\mathrm{PVC}}, U_{4|23}^{\mathrm{PVC}}) \perp U_{2:3}$ in Example 2. Empirical power values are based on 1000 samples of size $n = 1000$ and plotted against the parameter λ. For $\lambda = 0$ the plots show the empirical size of the test. Each plot corresponds to a specific value of Kendall's τ_{23} for the true copulas in the first and second tree. The black lines show the results where the PPITs are estimated with the true Clayton copula family in the first and second tree. The two grey lines correspond to the case where the PPITs are estimated with misspecified copula families (survival Gumbel and Gumbel, respectively).

Clayton copulas. If the degree of misspecification is severe and we fit Gumbel copulas (with upper tail dependence) to data generated from Clayton copulas (with lower tail dependence), differences in the empirical power of the CCC test become visible when comparing the light grey lines with the black lines in Figure 9. In the majority of the considered scenarios the empirical power is now a little smaller. In cases with high dependence, i.e., $\tau_{23} = 0.4, 0.6, 0.8$, the empirical size is increased. This shows that the test might not control the size if the copula families in the lower trees are severely misspecified. Note that we misspecify five copula families and that the misspecification in the second tree might be even worse because the data in the edges of the second tree is no longer generated by Clayton copulas if the copulas in the first tree are misspecified. Thus, the performance of the CCC test appears to be relatively robust w.r.t. such a severe misspecification.

6. A hierarchical procedure for testing the simplifying assumption in vine copulas

Up to now, we have tested single building blocks of vine copulas, i.e., the hypothesis $H_0 \colon (U^{\mathrm{PVC}}_{h_e|D_e}, U^{\mathrm{PVC}}_{i_e|D_e}) \perp U_{D_e}$ for a fixed edge $e \in \mathcal{E}^d_2$. To check the simplifying assumption for a d-dimensional R-vine copula with fixed structure $\mathcal{V}$, one has to check this constraint for $M = (d-1)(d-2)/2$ edges. In order to control the family-wise error rate α, we apply the Bonferroni correction and test the set of hypotheses[13]

$$\forall e \in \mathcal{E}^d_2 : H_0 : (U^{\mathrm{PVC}}_{h_e|D_e}, U^{\mathrm{PVC}}_{i_e|D_e}) \perp U_{D_e}.$$

We use a hierarchical procedure and begin with testing the hypotheses in the second tree. We then only check the hypotheses for the next tree T_{j+1} if the validity of the simplifying assumption could not be rejected for tree T_j, for $j = 2, \dots, d-2$. The hierarchical procedure is stopped whenever an individual hypothesis $H_0 \colon (U^{\mathrm{PVC}}_{h_e|D_e}, U^{\mathrm{PVC}}_{i_e|D_e}) \perp U_{D_e}$ is rejected at a level of α/M. As a result, the hierarchical procedure detects critical building blocks of a vine copula model where a pair-copula $C^{\mathrm{SVC}}_{h_e,i_e;D_e}$ does not seem to be an adequate model for the bivariate conditional copula $C_{h_e,i_e;D_e}$ of the bivariate conditional distribution $F_{h_e,i_e|D_e}$. The procedure is also in line with the common sequential specification and estimation of simplified vine copulas and can be integrated in model selection algorithms as demonstrated in Kraus and Czado [27].

6.1. Simulation study

In the following simulation study we test the simplifying assumption for the four-dimensional Clayton, Gaussian, Gumbel, and Frank copula with pair-wise values of Kendall's τ of $0.2, 0.4, 0.6, 0.8$. Since each copula is exchangeable an arbitrary structure can be fixed. We fit a D-vine copula and select the copula families of $C^{\mathrm{SVC}}_{12}, C^{\mathrm{SVC}}_{23}, C^{\mathrm{SVC}}_{34}, C^{\mathrm{SVC}}_{13;2}, C^{\mathrm{SVC}}_{24;3}, C^{\mathrm{SVC}}_{14;23}$ using the AIC. The hierarchical procedure is applied to test the simplifying assumption at a theoretical level of 5%. Thus, in tree T_2 two conditional copulas are tested with an individual level of 1.67%. If we do not reject the H_0 for both copulas in tree T_2, we continue in tree T_3 and test $H_0 : (U^{\mathrm{PVC}}_{1|23}, U^{\mathrm{PVC}}_{4|23}) \perp U_{2:3}$ at an individual level of 1.67%.

The two upper panels in Figure 10 report the results for the four-dimensional Clayton and Gaussian copula for which the simplifying assumption is satisfied [44]. The empirical size of the hierarchical test procedure with the CCC test Θ_n is close to the theoretical level even under consideration of possibly misspecified copula families. In the lower panels of Figure 10, the empirical power for the four-dimensional Frank and Gumbel copula is plotted. The Frank and Gumbel copula violate the simplifying assumption as long as $\tau \notin \{0, 1\}$ [44]. Although

---

[13]For each hypothesis corresponding to a fixed edge $e \in \mathcal{E}^d_2$ we use the same CCC test settings as before, i.e., decision trees with depth $J_{\max} = 2$, $\lambda_n = \frac{1}{\sqrt{n}}$ for the penalty and $\Gamma_0 = \Gamma_{\mathrm{med}}$ (see equation (4.1)) as base partition.
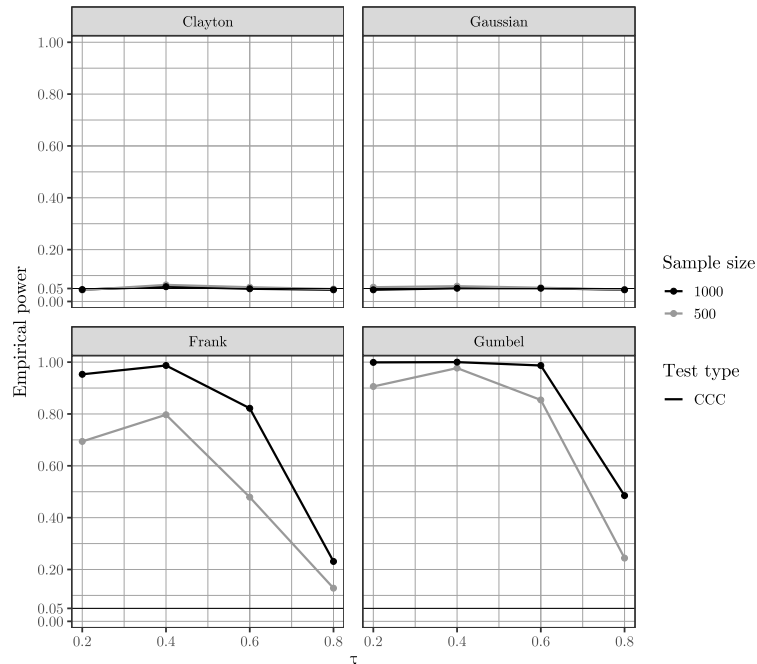
Fig 10: Empirical power of the hierarchical procedure with the CCC test Θ_n for the four-dimensional Clayton, Gaussian, Frank and Gumbel copula. Vine copula models are estimated using the AIC for selecting the copula families. Empirical power values are based on 1000 samples and plotted against the pairwise value of Kendall's τ. Different sample sizes are shown color coded.

the variation in the conditional copulas induced by the four-dimensional Frank and Gumbel copulas is rather mild (see Figure 11), the CCC test often rejects the simplifying assumption. That the power has a minimum at $\tau = 0.8$ can be explained by the fact that both copulas satisfy the simplifying assumption for $\tau \to 1$ and the variation of the conditional correlations (see Figure 11) is less pronounced than for $\tau = 0.2, 0.4, 0.6$.

How the decision trees adapt to the variation of the conditional correlations can be seen by looking at the most frequently selected decision trees and the resulting $\Gamma_{\max}$ partitions. For the Frank copula (with $n = 1000$ and $\tau = 0.2, 0.4, 0.6, 0.8$) the most frequently selected decision tree in tree T_2 of the vine (selected in 21.28% of all simulated cases) consists of a first split according to the 25% quartile and a second split at the 50% quartile in the left part and at the 25% quartile in the right part, i.e.,

$$\Gamma_{\max} = \Big\{ \Lambda_{(0,l,l)} = [0, 0.125], \, \Lambda_{(0,l,r)} = (0.125, 0.25], \, \Lambda_{(0,r,l)} = (0.25, 0.4375],$$
$$\Lambda_{(0,r,r)} = (0.4375, 1] \Big\}.$$
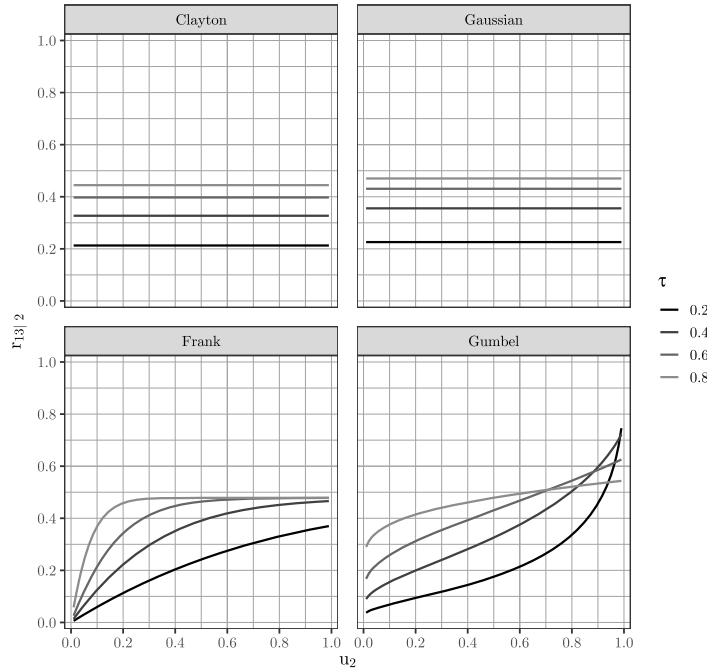
Fig 11: Conditional correlations $r_{13|2} = \mathrm{Corr}(U_{1|2}, U_{3|2}|U_2)$ of $C_{13;2}$ in tree T_2 of the Clayton, Gaussian, Frank and Gumbel copula.

Thus, the $\Gamma_{\max}$ partition is most granular for smaller values of the conditioning variable. This is also the area where the variation of the Frank copula is most pronounced as can be seen from Figure 11. In contrast, for the Gumbel copula (with $n = 1000$ and $\tau = 0.2, 0.4, 0.6, 0.8$) the most common selected $\Gamma_{\max}$ partition in tree T_2 (selected in 18.51% of all simulated cases) is

$$\Gamma_{\max} = \Big\{ \Lambda_{(0,l,l)} = [0, 0.375], \, \Lambda_{(0,l,r)} = (0.375, 0.75], \, \Lambda_{(0,r,l)} = (0.75, 0.875], \\ \Lambda_{(0,r,r)} = (0.875, 1] \Big\}.$$

Comparing $\Gamma_{\max}$ with Figure 11, which shows the conditional correlation for the Gumbel copula, we see that the decision tree adapts to the variation of the conditional correlation by selecting the most granular splits for higher values of the conditioning variable.

Two hypotheses are tested in tree T_2 and, provided there is no rejection in tree T_2, one additional hypothesis is tested in tree T_3 when the hierarchical procedure for testing the simplifying assumption is applied to four-dimensional vine copulas. In the simulation study, most of the rejections for the four-dimensional Frank and Gumbel copula happen in tree T_2. In Figure 12 we decompose the empirical power of the hierarchical procedure with the CCC test Θ_n into second tree and third tree rejections. Over all considered scenarios 94.09% and 98.06%
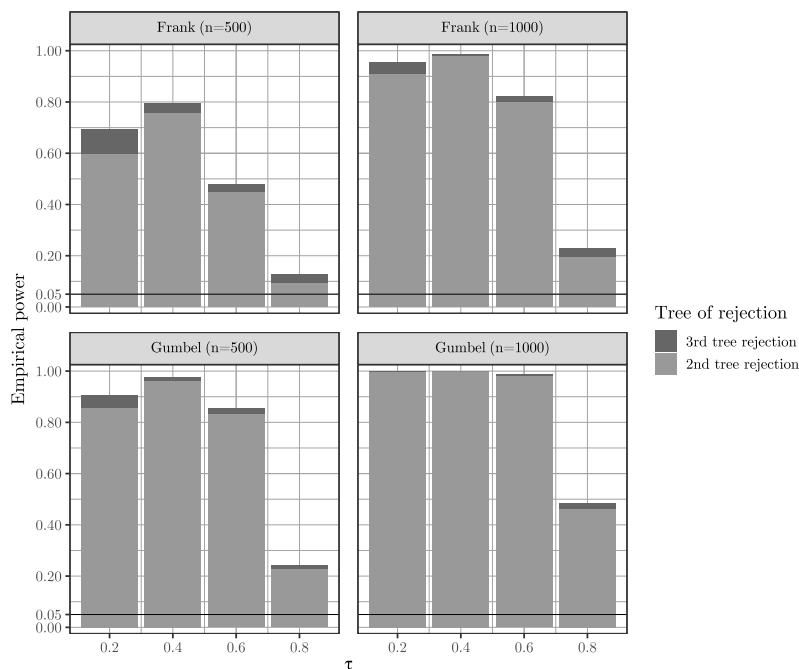
Fig 12: Empirical power of the hierarchical procedure with the CCC test Θ_n for the four-dimensional Frank and Gumbel copula. The tree of rejections is shown color coded.

of the rejections occur in the second tree T_2 for the Frank and Gumbel copula, respectively.

Finally, we revisit the up to twelve-dimensional Example 3 from Section 5.3. In Section 5.3 we only tested one hypothesis in the last tree and analyzed the effect of an increasing dimensionality of the conditioning set on the empirical power of the CCC test. Now we apply the hierarchical procedure to test the simplifying assumption for the entire D-vine copula with fixed structure defined in Example 3 for $d = 4, 6, 8, 10, 12$ and test the set of hypotheses

$$\forall j = 2, \ldots, d - 1 \, \forall i = 1, \ldots, d - j :$$
$$H_0 : (U_{i|i+1:i+j-1}^{\mathrm{PVC}}, U_{i+j|i+1:i+j-1}^{\mathrm{PVC}}) \perp U_{i+1:i+j-1}.$$

As mentioned before, the simplifying assumption for a d-dimensional vine copula is equivalent to $M = (d-1)(d-2)/2$ hypotheses. This means that for Example 3 we might test $M = 3, 10, 21, 36, 55$ hypotheses for $d = 4, 6, 8, 10, 12$ but only the last hypothesis actually violates the simplifying assumption for $\lambda \neq 0$.

The empirical size and power of the hierarchical procedure with the CCC test Θ_n to test the simplifying assumption for data generated from Example 3 is shown in Figure 13. For $\lambda = 0$ the simplifying assumption is satisfied and the empirical size is close to the theoretical level of 5%. Similar to Section 5.3, we
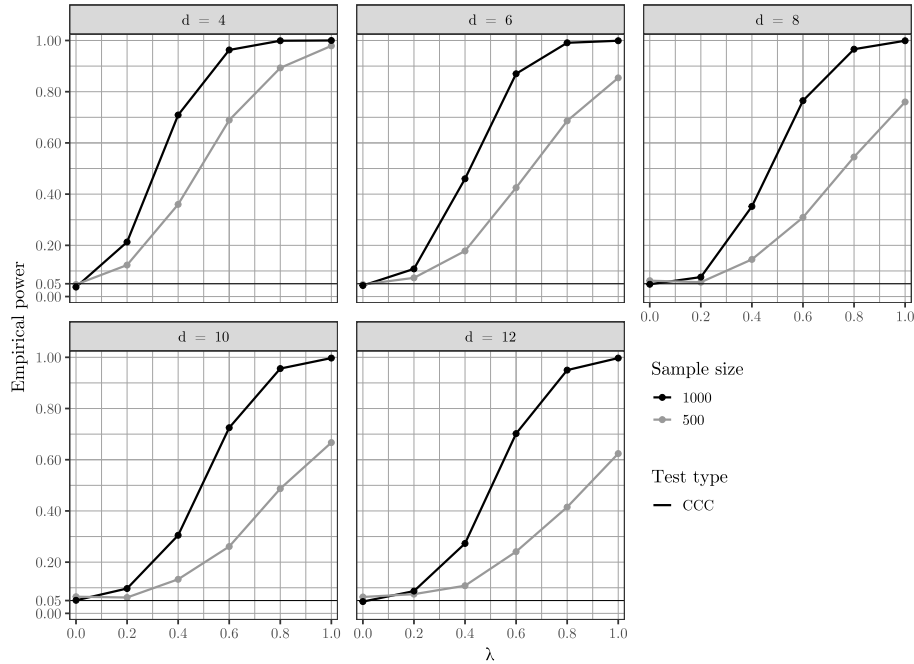
Fig 13: Empirical power of the hierarchical procedure with the CCC test Θ_n for the hypothesis that the simplifying assumption is satisfied for the vine copula defined in Example 3. Empirical power values are based on 1000 samples and plotted against the parameter λ. For $\lambda = 0$ the plots show the empirical size of the test. Each plot corresponds to a specific dimension d of the D-vine copula (Example 3). Different sample sizes are shown color coded.

observe that the empirical power of the hierarchical procedure with the CCC test Θ_n is slightly decreasing in the dimension d. The effect is more pronounced for the smaller sample size $n = 500$ as compared to $n = 1000$. Note that for $d = 12$ only one of 55 building blocks of the vine copula violates the simplifying assumption. This explains why for $n = 500$ and $\lambda = 1$ the empirical power of 62% is lower than the empirical power of 94% when only the last building block is tested (see Figure 8). However, a sample size of $n = 1000$ is sufficiently large so that the empirical power of the hierarchical procedure with the CCC test Θ_n is still almost 100% in this scenario.

6.2. Real data applications

We now use the hierarchical procedure with the CCC test Θ_n to test the simplifying assumption for R-vine copulas fitted to four different real data sets. The dimensionality of the data varies between 3 and 10 and the number of observations between 655 and 5032. On the one hand, we consider the data set `uranium`

TABLE 2

Results for real data sets: The simplifying assumption is tested by applying the hierarchical procedure with the CCC test Θ_n.

Name	uranium	FF3F	FF5F	Ind10
Description	Uranium Exploration Data Set	Fama/French 3 Factors	Fama/French 5 Factors	10 Industry Portfolios
Source	Cook and Johnson [9], R-package copula	Kenneth R. French – Data Library	Kenneth R. French – Data Library	Kenneth R. French – Data Library
Variables	log concentration of Uranium, Lithium, Cobalt, Potassium, Cesium, Scandium, Titanium	SMB, HML, $R_m - R_f$	SMB, HML, RMW, CMA, $R_m - R_f$	10 industry portfolios formed according to four-digit SIC codes.
Period	—	02-Jan-2001 to 31-Dec-2020 (daily)	02-Jan-2001 to 31-Dec-2020 (daily)	02-Jan-2001 to 31-Dec-2020 (daily)
Dimension	7 (21 pair-copulas)	3 (3 pair-copulas)	5 (10 pair-copulas)	10 (45 pair-copulas)
n	655	5032	5032	5032
Test decision of the hierarchical procedure (at a 5% significance level)	The simplifying assumption can be **rejected** in tree T_2.	The simplifying assumption **cannot** be rejected.	The simplifying assumption **cannot** be rejected.	The simplifying assumption **cannot** be rejected.
Smallest p-value [Bonferroni adjusted p-value] of one hypothesis H_0: $(U^{\mathrm{PVC}}_{h_e\|D_e}, U^{\mathrm{PVC}}_{i_e\|D_e}) \perp U_{D_e}$ of the hierarchical procedure	< 0.01 [< 0.01]	0.20 [0.20]	0.21 [1.00]	0.10 [1.00]

[9] and obtain normalized ranks as pseudo-observations from the copula by means of the rescaled ecdf. On the other hand, we consider three financial data sets from the *Kenneth R. French – Data Library* (available under: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). We apply ARMA(1,1)-GARCH(1,1)-filtering [12, 6] with t-distributed innovations and apply the rescaled ecdf to the residuals to obtain pseudo-observations from the copulas.[14]

In order to test the simplifying assumption with the CCC test, the researcher has to specify the vine structure $\mathcal{V}$, which also determines the H_0, and the copula families for the edges. The selection of the vine structure $\mathcal{V}$ and copula families is commonly based on the data, e.g., by using the algorithm of Dißmann et al. [11]. If model selection and statistical inference is done on the same data, statistical inference might no longer be valid (see Fithian, Sun and Taylor [13] or Lee et al. [31] for a discussion of post model selection inference). Therefore, we randomly split the samples into two parts so that the partial vine copula model is selected on one half of the data and the other half of the data is used to apply the CCC

---

[14]Copula modeling for GARCH-filtered return data has been studied in [8, 7] and [38] provides a review of copula models for economic time series.
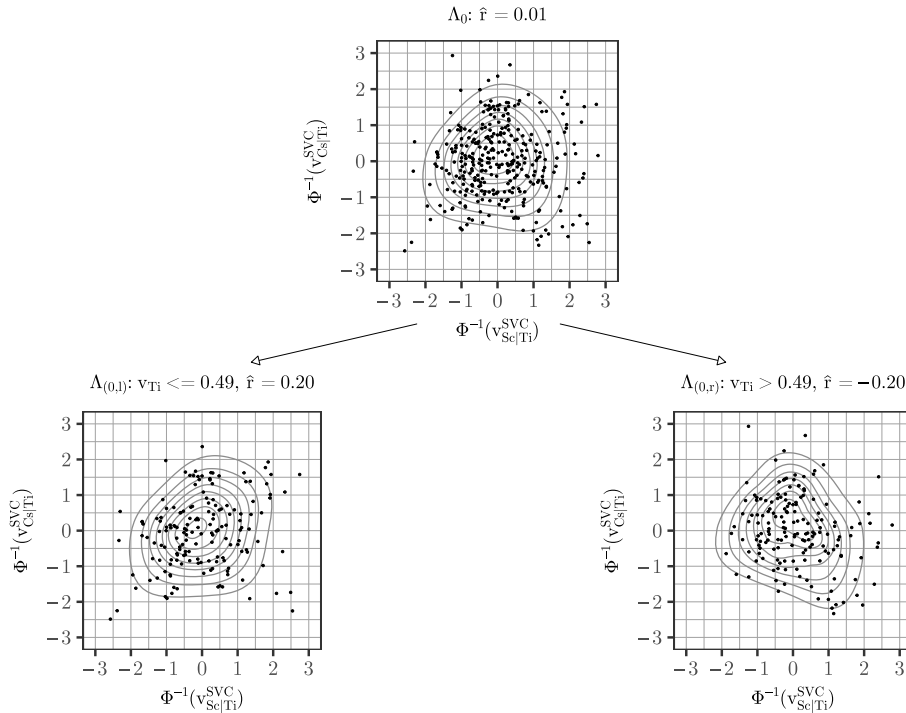
Fig 14: Scatterplots of pseudo-observations from the PPITs (see equation (A.7)) corresponding to the conditional copula of Scandium (`Sc`) and Cesium (`Cs`) given Titanium (`Ti`) for the `uranium` dat set. The grouped scatterplots are formed according to the decision tree and show the data points that are used for the CCC test Θ_n to test the simplifying assumption. The margins are transformed to be standard normal and the contours are derived from copula kernel density estimates.

test. To obtain parametric models for the PVC we apply the standard R-vine model selection algorithm proposed by Dißmann et al. [11] which is implemented in the R-package `VineCopula` [35]. The resulting vine structure $\mathcal{V}$ and copula families are then used to fit a model for the partial vine copula and conduct tests on the other half of the data.

In Table 2 we provide information about all four data sets and report the results of the hierarchical test procedure with the CCC test Θ_n.[15] For all cases where the validity of the simplifying assumption is rejected, we show the first tree in which we reject at least one null hypothesis of the form $H_0 : (U_{h_e|D_e}^{\text{PVC}}, U_{i_e|D_e}^{\text{PVC}}) \perp U_{D_e}$ and stop the hierarchical test procedure. Moreover, the smallest p-value of one hypothesis of the hierarchical procedure is also depicted.

---

[15]Note that the CCC test Θ_n is not a consistent test because the conditional correlation can be constant if the simplifying assumption is false. Moreover, if the conditional correlation is varying it may be possible that this variation is not detected by the proposed CCC test Θ_n with the decision tree based partitioning as described in Section 4.2.

For the non-financial data example `uranium` we reject the simplifying assumption already in tree T_2. The rejection is in line with the results reported by Gijbels, Omelka and Veraverbeke [17] and Kraus and Czado [27]. In Figure 14 we provide a visualization of the rejected building block for the `uranium` data set. The grouped scatterplots are formed according to the decision tree used for the CCC test Θ_n. For easier visual inspection the margins are transformed to be standard normal and contours derived from copula kernel density estimates (`R`-package `kdecopula` [33]) are shown. The plots show that the conditional copula of Scandium (`Sc`) and Cesium (`Cs`) given Titanium (`Ti`) is varying with an estimated positive correlation of 0.20 for values of Titanium being smaller than its median and a negative estimated correlation of −0.20 for larger values of Titanium.

For the three filtered financial returns the simplifying assumption cannot be rejected on the basis of the CCC test Θ_n (Table 2). This indicates that the possible violation of the simplifying assumption for the vine copulas selected by Dißmann's algorithm might be less severe for this kind of data as compared to `uranium`. This is consistent with the findings of Kraus and Czado [27] who also use the CCC test and report that the simplifying assumption seems to be rather appropriate for filtered financial returns.

7. Conclusion

In practical applications, a test for the simplifying assumption in high-dimensional vine copulas must be computationally feasible and tackle the curse of dimensionality. The introduced hierarchical procedure with the CCC test Θ_n addresses these two issues.

The asymptotic distribution of the CCC test statistic is derived under the assumption of semi-parametrically estimated pseudo-observations from the partial probability integral transforms. Since the test has a known asymptotic distribution and is based on the stepwise maximum likelihood estimator, it is computationally feasible also in high dimensions. To prevent suffering from the curse of dimensionality, the CCC test utilizes a novel stochastic interpretation of the simplifying assumption based on the partial vine copula. Moreover, we propose a discretization of the support of the conditioning vector into a finite number of subsets and incorporate a penalty in the test statistic. A decision tree algorithm looks for the largest deviation from the simplifying assumption measured in terms of conditional correlations and also contributes to a computationally feasible test.

In a simulation study we provide a thorough analysis of the finite sample performance of the CCC test for various kinds of data generating processes. The CCC test outperforms the vectorial independence test by a large margin if the conditional correlation is varying. Even more important for high-dimensional applications, the simulation study demonstrates that the power of the test decreases only slightly with the dimension of the conditioning vector. A moderate misspecification of the parametric copula families does not affect the power

properties of the CCC test. We also investigate the performance of a hierarchical procedure that utilizes the CCC test to test the simplifying assumption. An application to four real data sets demonstrates the usefulness of the test and indicates that the validity of the simplifying assumption should be checked individually for each data set.

The CCC test can also be utilized to improve the modeling of data with vine copulas. Schellhase and Spanhel [40] make use of the CCC test to identify building blocks of vine copulas where the simplifying assumption does not seem to be adequate and the estimation of a conditional copula that is varying in its conditioning variables can improve the modeling. Additionally, Kraus and Czado [27] use the CCC test to find alternative vine copula structures which might be more in line with the simplifying assumption.

Appendix

A.1. Proof of Theorem 1

That (i) and (ii) are equivalent follows from the definition of the simplifying assumption in Assumption 1 and the definition of the conditional and partial copula in Definition 1.

For $e \in \mathcal{E}_2^d$, $(U_{h_e|D_e}^{\mathrm{PVC}}, U_{i_e|D_e}^{\mathrm{PVC}}) \perp U_{D_e}$ implies $(U_{h_e|D_e}^{\mathrm{PVC}}, U_{i_e|D_e}^{\mathrm{PVC}}) \overset{\text{a.s.}}{=} (U_{h_e|D_e}, U_{i_e|D_e})$ by Lemma 3.1 in Spanhel and Kurz [43]. Thus, $(U_{h_e|D_e}^{\mathrm{PVC}}, U_{i_e|D_e}^{\mathrm{PVC}}, U_{D_e})$ and $(U_{h_e|D_e}, U_{i_e|D_e}, U_{D_e})$ are equal in distribution and $(U_{h_e|D_e}, U_{i_e|D_e}) \perp U_{D_e}$. This shows (ii) $\Leftarrow$ (iii).

To show (ii) $\Rightarrow$ (iii), we use induction over the trees. By the definition of the PPITs in the second tree T_2 we have that

$$\forall e \in E_2 : \quad (U_{h_e|D_e}^{\mathrm{PVC}}, U_{i_e|D_e}^{\mathrm{PVC}}) \overset{\text{a.s.}}{=} (U_{h_e|D_e}, U_{i_e|D_e}).$$

Because (ii) implies that

$$\forall e \in E_2 : \quad (U_{h_e|D_e}, U_{i_e|D_e}) \perp U_{D_e},$$

it follows that

$$\forall e \in E_2 : \quad (U_{h_e|D_e}^{\mathrm{PVC}}, U_{i_e|D_e}^{\mathrm{PVC}}) \perp U_{D_e}$$

and the base case of the induction is proved.

We now assume the induction hypothesis that for tree T_{j-1}

$$\forall \bar{e} \in E_{j-1} : \quad (U_{h_{\bar{e}}|D_{\bar{e}}}^{\mathrm{PVC}}, U_{i_{\bar{e}}|D_{\bar{e}}}^{\mathrm{PVC}}) \perp U_{D_{\bar{e}}}$$

holds. Note that Lemma 3.1 in Spanhel and Kurz [43] then implies that

$$\forall \bar{e} \in E_{j-1} : \quad (U_{h_{\bar{e}}|D_{\bar{e}}}^{\mathrm{PVC}}, U_{i_{\bar{e}}|D_{\bar{e}}}^{\mathrm{PVC}}) \overset{\text{a.s.}}{=} (U_{h_{\bar{e}}|D_{\bar{e}}}, U_{i_{\bar{e}}|D_{\bar{e}}}). \tag{A.1}$$

By (ii), we get for each $e \in E_j$ in tree T_j and $l_e = h_e, i_e$ that

$$\begin{aligned} U_{l_e|D_e} &\overset{\text{a.s.}}{=} \partial_2 C_{l_{\bar{e}}, m_{\bar{e}}; D_{\bar{e}}}(U_{l_{\bar{e}}|D_{\bar{e}}}, U_{m_{\bar{e}}|D_{\bar{e}}}; U_{D_{\bar{e}}}) \\ &\overset{\text{a.s.}}{=} \partial_2 C_{l_{\bar{e}}, m_{\bar{e}}; D_{\bar{e}}}^{\mathrm{P}}(U_{l_{\bar{e}}|D_{\bar{e}}}, U_{m_{\bar{e}}|D_{\bar{e}}}), \end{aligned} \tag{A.2}$$

where $\bar{e} \in E_{j-1}$ and $m_{\bar{e}} \in D_e$ are chosen such that $(l_{\bar{e}}, m_{\bar{e}}; D_{\bar{e}}) = (l_e, m_{\bar{e}}; D_e \setminus m_{\bar{e}}) \in \mathcal{CV}$.[16] Note that $C^{\mathrm{P}}_{l_{\bar{e}}, m_{\bar{e}}; D_{\bar{e}}}$ is the distribution of $(U_{l_{\bar{e}}|D_{\bar{e}}}, U_{m_{\bar{e}}|D_{\bar{e}}})$ and $C^{\mathrm{PVC}}_{l_{\bar{e}}, m_{\bar{e}}; D_{\bar{e}}}$ the distribution of $(U^{\mathrm{PVC}}_{l_{\bar{e}}|D_{\bar{e}}}, U^{\mathrm{PVC}}_{m_{\bar{e}}|D_{\bar{e}}})$. By (A.1) it follows that $C^{\mathrm{P}}_{l_{\bar{e}}, m_{\bar{e}}; D_{\bar{e}}} = C^{\mathrm{PVC}}_{l_{\bar{e}}, m_{\bar{e}}; D_{\bar{e}}}$. Thus,

$$\begin{aligned} U_{l_e|D_e} &\stackrel{\mathrm{a.s.}}{=} \partial_2 C^{\mathrm{PVC}}_{l_{\bar{e}}, m_{\bar{e}}; D_{\bar{e}}}(U_{l_{\bar{e}}|D_{\bar{e}}}, U_{m_{\bar{e}}|D_{\bar{e}}}) \\ &\stackrel{\mathrm{a.s.}}{=} \partial_2 C^{\mathrm{PVC}}_{l_{\bar{e}}, m_{\bar{e}}; D_{\bar{e}}}(U^{\mathrm{PVC}}_{l_{\bar{e}}|D_{\bar{e}}}, U^{\mathrm{PVC}}_{m_{\bar{e}}|D_{\bar{e}}}) \qquad\qquad\quad (\text{A.3}) \\ &\stackrel{\mathrm{a.s.}}{=} U^{\mathrm{PVC}}_{l_e|D_e}, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (\text{A.4}) \end{aligned}$$

where (A.3) follows from (A.1) and (A.4) is the definition of $U^{\mathrm{PVC}}_{l_e|D_e}$. As a result, we have shown that

$$\forall e \in E_j: \quad (U^{\mathrm{PVC}}_{h_e|D_e}, U^{\mathrm{PVC}}_{i_e|D_e}) \stackrel{\mathrm{a.s.}}{=} (U_{h_e|D_e}, U_{i_e|D_e}).$$

Finally, (ii) implies that

$$\forall e \in E_j: \quad (U^{\mathrm{PVC}}_{h_e|D_e}, U^{\mathrm{PVC}}_{i_e|D_e}) \perp U_{D_e},$$

which completes the induction to show (ii) $\Rightarrow$ (iii).

A.2. Proof of Proposition 2

We first prove the following lemma stating the asymptotic distribution of the test statistic $T^\star_n(\Gamma) = n(A\hat{R}^\star_\Gamma)^T (A\hat{\Sigma}^\star_{R_\Gamma} A^T)^{-1} A\hat{R}^\star_\Gamma$ under $H_0: (U^{\mathrm{PVC}}_{h_e|D_e}, U^{\mathrm{PVC}}_{i_e|D_e}) \perp U_{D_e}$ and the assumption that observations from the PPITs are observable.

Lemma 1.
Let $(U^k_{1:d})_{k=1:n}$ be n independent copies of $U_{1:d} \sim C_{1:d}$ and $\Lambda_0 := \mathrm{supp}(U_{D_e})$. Consider a fixed R-vine structure $\mathcal{V}$ and let $e \in \mathcal{E}^d_2$ be a fixed edge. Assume that the partition $\Gamma := \{\Lambda_1, \ldots, \Lambda_L\}$ satisfies the conditions stated in Proposition 2. Under $H_0: (U^{PVC}_{h_e|D_e}, U^{PVC}_{i_e|D_e}) \perp U_{D_e}$ it holds that

$$T^\star_n(\Gamma) \xrightarrow{d} \chi^2(L-1).$$

Proof. We first derive the asymptotic distribution of $\hat{R}^\star_\Gamma = (\hat{r}_1, \ldots, \hat{r}_L)$ under H_0 before showing that $T^\star_n(\Gamma)$ has an asymptotic chi-square distribution under H_0. For this purpose, let $e_5 := (0,0,0,0,1)^T$, $\otimes$ denote the Kronecker product, 1_L be a $L \times 1$ column vector of ones and I_L be the $L \times L$ identity matrix, so that $(I_L \otimes e_5)^T$ is a $L \times 5L$ matrix that can be used to extract every fifth element from a $5L$-dimensional column vector. The correlations are then given by $\hat{R}^\star_\Gamma = (I_L \otimes e_5)^T \hat{\alpha}$, with $\hat{\alpha}$ being the unique solution of the estimating equation

$$\frac{1}{n} \sum_{k=1}^n g_\Gamma(U^k_{1:d}, \hat{\pi}, \alpha) \stackrel{!}{=} 0, \qquad\qquad\qquad\qquad (\text{A.5})$$

---

[16]By the proximity condition of the R-vine $\mathcal{V}$ (Definition 2) there exists a $m_{\bar{e}} \in D_e$ such that either $(l_e, m_{\bar{e}}; D_e \setminus m_{\bar{e}}) \in \mathcal{CV}$ or $(m_{\bar{e}}, l_e; D_e \setminus m_{\bar{e}}) \in \mathcal{CV}$ with $\bar{e} \in E_{j-1} = N_j$ and w.l.o.g. we use the first case in (A.2).

where the estimating function g_Γ will be stated in the following.

Define

$$g_\pi(U_{1:d}^k, \pi) = \left(\pi_1 - \mathbb{1}_{\{U_{D_e}^k \in \Lambda_1\}}, \dots, \pi_L - \mathbb{1}_{\{U_{D_e}^k \in \Lambda_L\}}\right)^T,$$

where $\pi := \pi_{1:L} \in (0,1)^L$. The solution $\hat{\pi}$ of $\frac{1}{n}\sum_{k=1}^n g_\pi(U_{1:d}^k, \pi) \stackrel{!}{=} 0$ denotes the random fraction of data corresponding to Λ_l, i.e.,

$$\hat{\pi}_l := \frac{1}{n}\sum_{k=1}^n \mathbb{1}_{\{U_{D_e}^k \in \Lambda_l\}}, \quad l = 1, \dots, L.$$

Define

$$h(U_{1:d}^k, \phi) = \begin{pmatrix} \phi_1 - U_{h_e|D_e}^{\mathrm{PVC},k} \\ \phi_2 - U_{i_e|D_e}^{\mathrm{PVC},k} \\ \phi_3 - (U_{h_e|D_e}^{\mathrm{PVC},k} - \phi_1)^2 \\ \phi_4 - (U_{i_e|D_e}^{\mathrm{PVC},k} - \phi_2)^2 \\ \phi_5 - (U_{h_e|D_e}^{\mathrm{PVC},k} - \phi_1)(U_{i_e|D_e}^{\mathrm{PVC},k} - \phi_2)(\phi_3\phi_4)^{-\frac{1}{2}} \end{pmatrix},$$

where $\phi := \phi_{1:5} \in \mathbb{R}^5$. For $\Lambda_l \in \Gamma = \{\Lambda_1, \dots, \Lambda_L\}$ we set

$$f(U_{1:d}^k, \pi_l, \phi) = \pi_l^{-1}\mathbb{1}_{\{U_{D_e}^k \in \Lambda_l\}}h(U_{1:d}^k, \phi).$$

The estimating function g_Γ in (A.5) is given by

$$g_\Gamma(U_{1:d}^k, \pi, \alpha) := \left(f(U_{1:d}^k, \pi_1, \alpha_1)^T, \dots, f(U_{1:d}^k, \pi_L, \alpha_L)^T\right)^T.$$

Let π_0 be the unique solution of $\mathbb{E}[g_\pi(U_{1:d}^k, \pi)] = 0$, ϕ_0 be the unique solution of $\mathbb{E}[h(U_{1:d}^k, \phi)] = 0$, and $\alpha_0 = (\phi_0^T, \dots, \phi_0^T)^T$ so that $\mathbb{E}[g_\Gamma(U_{1:d}^k, \pi, \alpha_0)] = 0$ for all $\pi \in (0,1)^L$ under H_0 because for each l-th block element of $\mathbb{E}[g_\Gamma(U_{1:d}^k, \pi, \alpha_0)]$ it holds that

$$\left(\mathbb{E}[g_\Gamma(U_{1:d}^k, \pi, \alpha_0)]\right)_l := \mathbb{E}[f(U_{1:d}^k, \pi_l, \phi_0)] = \mathbb{E}\left[\pi_l^{-1}\mathbb{1}_{\{U_{D_e}^k \in \Lambda_l\}}h(U_{1:d}^k, \phi_0)\right]$$

$$\stackrel{H_0}{=} \mathbb{E}\left[\pi_l^{-1}\mathbb{1}_{\{U_{D_e}^k \in \Lambda_l\}}\right]\underbrace{\mathbb{E}[h(U_{1:d}^k, \phi_0)]}_{=0} = 0.$$

Using the same steps it can be readily verified that $\mathbb{E}[\partial_{\pi^T} g_\Gamma(U_{1:d}^k, \pi, \alpha_0)] = 0$ for all $\pi \in (0,1)^L$ under H_0. Thus, under H_0, the standard theory of estimating equations for two-step estimators, e.g., Theorem 6.1 in Newey and McFadden [36], yields that

$$\sqrt{n}(\hat{\alpha} - \alpha_0) \stackrel{d}{\to} N_{5L}(0, G_\Gamma^{-1}\Omega_\Gamma(G_\Gamma^{-1})^T), \tag{A.6}$$

where $G_\Gamma := \mathbb{E}\big[\partial_{\alpha^T} g_\Gamma(U_{1:d}^k, \pi_0, \alpha_0)\big]$, $\Omega_\Gamma := \mathrm{Cov}\big[g_\Gamma(U_{1:d}^k, \pi_0, \alpha_0)\big]$ and $N_d(\mu, \Sigma)$ denotes a d-dimensional normal distribution with mean vector μ and covariance matrix Σ.

If we now extract every fifth element from $\hat\alpha$ using $\hat R_\Gamma^\star = (I_L \otimes e_5)^T \hat\alpha$, we obtain the joint asymptotic distribution of the estimated correlations under H_0 as

$$\sqrt{n}((\hat r_1, \ldots, \hat r_L)^T - (r_1, \ldots, r_L)^T) = \sqrt{n}(\hat R_\Gamma^\star - R_\Gamma) \overset{d}{\to} N_L(0, \Sigma_{R_\Gamma}^\star),$$

so that

$$\Sigma_{R_\Gamma}^\star = (I_L \otimes e_5)^T G_\Gamma^{-1} \Omega_\Gamma (G_\Gamma^{-1})^T (I_L \otimes e_5).$$

Under H_0 it holds that $r_1 = \ldots = r_L = r = \mathrm{Corr}(U_{h_e|D_e}^{\mathrm{PVC}}, U_{i_e|D_e}^{\mathrm{PVC}})$ and therefore it follows with the $(L-1) \times L$ first-order difference matrix A and the continuous mapping theorem, that

$$\sqrt{n} A \hat R_\Gamma^\star \overset{d}{\to} N_{L-1}(0, A \Sigma_{R_\Gamma}^\star A^T).$$

To obtain the statistic of the CCC test when a sample from the PPITs is observable, the covariance matrix

$$\Sigma_{R_\Gamma}^\star = \mathrm{Cov}\big[(I_L \otimes e_5)^T G_\Gamma^{-1} g_\Gamma(U_{1:d}^k, \pi_0, \alpha_0)\big]$$

has to be consistently estimated, e.g., by $\hat\Sigma_{R_\Gamma}^\star = \widehat{\mathrm{Cov}}[(I_L \otimes e_5)^T \hat G_\Gamma^{-1} g_\Gamma(U_{1:d}^k, \hat\pi, \hat\alpha)]$, where $\widehat{\mathrm{Cov}}[X]$ denotes the sample covariance of the random vector X. By applying once more the continuous mapping theorem and Slutsky's theorem, we get

$$T_n^\star(\Gamma) = n(A \hat R_\Gamma^\star)^T (A \hat\Sigma_{R_\Gamma}^\star A^T)^{-1} A \hat R_\Gamma^\star \overset{d}{\to} \chi^2(L-1)$$

and Lemma 1 is proven. ∎

The remaining part of the proof of Proposition 2 requires the definition of the pseudo stepwise maximum likelihood estimator of the vine copula parameters. This estimator can be obtained as the solution of estimating equations (Hobæk Haff [20], Spanhel and Kurz [42], Tsukahara [46]). By extending these estimating equations by the ones for the correlations defined in the proof of Lemma 1 we derive the asymptotic distribution of the CCC test when pseudo-observations from the PPITs are estimated. Consider a fixed R-vine structure $\mathcal{V}$ and let $\{C_{1:d;\,\theta_{1:d-1}}^{\mathrm{SVC}} : \theta_{1:d-1} \in \Upsilon\}$ be a parametric simplified vine copula such that $\exists \theta_{1:d-1;\,0} \in \Upsilon$ so that $C_{1:d;\,\theta_{1:d-1;\,0}}^{\mathrm{SVC}} = C_{1:d}^{\mathrm{PVC}}$, where $C_{1:d}^{\mathrm{PVC}}$ denotes the PVC of $(C_{1:d}, \mathcal{V})$. The density of $C_{1:d;\,\theta_{1:d-1;\,0}}^{\mathrm{SVC}}$ is given by

$$c_{1:d;\,\theta_{1:d-1}}^{\mathrm{SVC}}(u_{1:d}) = \prod_{e \in \mathcal{E}_1^d} c_{h_e, i_e;\, D_e;\, \theta_{j,e}}^{\mathrm{SVC}}(u_{h_e|D_e}^{\mathrm{SVC}}(\theta_{1:j-1}), u_{i_e|D_e}^{\mathrm{SVC}}(\theta_{1:j-1})),$$

where $C_{h_e, i_e;\, D_e;\, \theta_{j,e}}^{\mathrm{SVC}}$ is a bivariate copula for each $e \in \mathcal{E}_1^d$ with parameter $\theta_{j,e}$. For $1 \leq j \leq d-1$, the vector $\theta_j := (\theta_{j,e})_{e \in E_j}^T$ collects the parameters in tree T_j

and the vector $\theta_{1:j} := (\theta_1, \ldots, \theta_j)^T$ collects all parameters up to and including tree T_j. The individual stepwise pseudo score functions for the copulas in tree T_j are given by

$$\partial_{\theta_j} \ell_j(V_{1:d}^k; \theta_{1:j}) := \partial_{\theta_j} \sum_{e \in E_j} \ln\left(c_{h_e,i_e;D_e;\theta_{j,e}}^{\mathrm{SVC}}(V_{h_e|D_e}^{\mathrm{SVC},k}(\theta_{1:j-1}), V_{i_e|D_e}^{\mathrm{SVC},k}(\theta_{1:j-1}))\right),$$

$j = 1, \ldots, d-1$. Here, the pseudo-observations of the PPITs for $e \in E_1$, $l_e = h_e, i_e$ are defined by

$$V_{l_e|D_e}^{\mathrm{SVC},k}(\theta_{1:0}) := V_{l_e}^k := \frac{1}{n+1} \sum_{m=1}^n \mathbb{1}_{\{X_{l_e}^m \leq X_{l_e}^k\}},$$

and for $e \in \mathcal{E}_2^d$ as

$$V_{l_e|D_e}^{\mathrm{SVC},k}(\theta_{1:j-1}) := \partial_2 C_{l_{\bar{e}},m_{\bar{e}};D_{\bar{e}};\theta_{j-1,\bar{e}}}^{\mathrm{SVC}}(V_{l_{\bar{e}}|D_{\bar{e}}}^{\mathrm{SVC},k}(\theta_{1:j-2}), V_{m_{\bar{e}}|D_{\bar{e}}}^{\mathrm{SVC},k}(\theta_{1:j-2})), \quad \text{(A.7)}$$

where $l_e = h_e, i_e$ and $(l_{\bar{e}}, m_{\bar{e}}; D_{\bar{e}}) = (l_e, m_{\bar{e}}; D_e \setminus m_{\bar{e}}) \in \mathcal{CV}$ is selected as in Definition 3.

Set $\theta := \theta_{1:d-1}$ and define the estimating function

$$g_S(V_{1:d}^k, \theta) = (\partial_{\theta_1} \ell_1(V_{1:d}^k; \theta_1)^T, \ldots, \partial_{\theta_{d-1}} \ell_{d-1}(V_{1:d}^k; \theta_{1:d-1})^T)^T,$$

so that the solution $\hat{\theta}$ of $\frac{1}{n} \sum_{k=1}^n g_S(V_{1:d}^k, \theta) \overset{!}{=} 0$ is the pseudo stepwise maximum likelihood estimator.

Moreover, $g_\Gamma(V_{1:d}^k, \pi, \theta, \alpha)$ denotes the estimating function of the correlations when pseudo-observations from the PPITs are used, i.e.,

$$g_\Gamma(V_{1:d}^k, \pi, \theta, \alpha) := \left(f(V_{1:d}^k, \pi_1, \theta, \alpha_1)^T, \ldots, f(V_{1:d}^k, \pi_L, \theta, \alpha_L)^T\right)^T,$$

where

$$f(V_{1:d}^k, \pi_l, \theta, \phi) = \pi_l^{-1} \mathbb{1}_{\{U_{D_e}^k \in \Lambda_l\}} \begin{pmatrix} \phi_1 - V_{h_e|D_e}^{\mathrm{SVC},k}(\theta_{1:j-1}) \\ \phi_2 - V_{i_e|D_e}^{\mathrm{SVC},k}(\theta_{1:j-1}) \\ \phi_3 - (V_{h_e|D_e}^{\mathrm{SVC},k}(\theta_{1:j-1}) - \phi_1)^2 \\ \phi_4 - (V_{i_e|D_e}^{\mathrm{SVC},k}(\theta_{1:j-1}) - \phi_2)^2 \\ \phi_5 - \frac{(V_{h_e|D_e}^{\mathrm{SVC},k}(\theta_{1:j-1})-\phi_1)(V_{i_e|D_e}^{\mathrm{SVC},k}(\theta_{1:j-1})-\phi_2)}{(\phi_3\phi_4)^{\frac{1}{2}}} \end{pmatrix}.$$

Let $\beta := (\theta^T, \alpha^T)^T$ so that the estimating function of the vine copula parameters and the correlations is given by

$$g(V_{1:d}^k, \pi, \beta) = \begin{pmatrix} g_S(V_{1:d}^k, \theta) \\ g_\Gamma(V_{1:d}^k, \pi, \theta, \alpha) \end{pmatrix}.$$

The rank approximate estimator $\hat{\beta}$ is then given as the solution of $\frac{1}{n}\sum_{k=1}^{n} g(V_{1:d}^k, \hat{\pi}, \beta) \stackrel{!}{=} 0$ where $\hat{\pi}$ is given as in the proof of Lemma 1. To derive the asymptotic distribution of $\hat{\beta}$, introduce

$$W_i := \int \partial_{u_i} g(u_{1:d}, \pi_0, \beta_0) \mathbb{1}_{\{U_i \leq u_i\}} \mathrm{d}C_{1:d}(u_{1:d}), \qquad i = 1, \ldots, d,$$

$$\tilde{\Omega} := \mathrm{Cov}\left(g(U_{1:d}^k, \pi_0, \beta_0) + \sum_{i=1}^{d} W_i \right),$$

where $\beta_0 := (\theta_{1:d-1;0}^T, \alpha_0^T)^T$ is the unique solution of $\mathbb{E}[g(U_{1:d}^k, \pi, \beta_0)] = 0$ for all $\pi \in (0,1)^L$ under H_0. By the same reasoning as in the proof of Lemma 1 and because $g_S(U_{1:d}^k, \theta)$ does not depend on π it follows that $\mathbb{E}\left[\partial_{\pi^T} g(U_{1:d}^k, \pi, \beta_0)\right] = 0$ for all $\pi \in (0,1)^L$ under H_0. Thus, provided the regularity conditions in Theorem 1 in Hobæk Haff [20] are satisfied, it follows that

$$\sqrt{n}(\hat{\beta} - \beta_0) \stackrel{d}{\to} N_{n_\theta + 5L}(0, \tilde{\Sigma}),$$

where $\tilde{\Sigma} = G^{-1}\tilde{\Omega}(G^{-1})^T$ with $G := -\mathbb{E}[\partial_{\beta^T} g(U_{1:d}^n, \pi_0, \beta_0)]$ and n_θ is the number of vine copula parameters, i.e., the length of the vector θ.

To extract the estimated correlations $\hat{R}_\Gamma$ from $\hat{\beta}$ and to obtain the corresponding asymptotic covariance matrix, we can exploit the block-structure of G as follows

$$\begin{aligned}
G &= \begin{pmatrix} \mathbb{E}[\partial_{\theta^T} g_S(U_{1:d}^k, \theta_0)] & 0 \\ \mathbb{E}[\partial_{\theta^T} g_\Gamma(U_{1:d}^k, \pi_0, \theta_0, \alpha_0)] & \mathbb{E}[\partial_{\alpha^T} g_\Gamma(U_{1:d}^k, \pi_0, \theta_0, \alpha_0)] \end{pmatrix} \\
&= \begin{pmatrix} \mathbb{E}[\partial_{\theta^T} g_S(U_{1:d}^k, \theta_0)] & 0 \\ \mathbb{E}[\partial_{\theta^T} g_\Gamma(U_{1:d}^k, \pi_0, \theta_0, \alpha_0)] & G_\Gamma \end{pmatrix}.
\end{aligned}$$

Denote the $n_\theta \times L$ matrix consisting of zeros by $0_{n_\theta \times L}$ and define $\delta := (0_{n_\theta \times L}^T, (I_L \otimes e_5)^T)^T$ so that $\hat{R}_\Gamma = \delta^T \hat{\beta}$. The asymptotic covariance matrix of $\sqrt{n}(\hat{R}_\Gamma - R_\Gamma)$ is then

$$\begin{aligned}
\Sigma_{R_\Gamma} &= \mathrm{Cov}\left[\delta^T G^{-1}\left(g(U_{1:d}, \pi_0, \beta_0) + \sum_{i=1}^{d} W_i\right)\right] \\
&= \Sigma_{R_\Gamma}^\star + \left(\mathrm{Cov}\left[\delta^T G^{-1} g(U_{1:d}, \pi_0, \beta_0)\right] - \Sigma_{R_\Gamma}^\star\right) \\
&\quad + \left(\mathrm{Cov}\left[\delta^T G^{-1}\left(g(U_{1:d}, \pi_0, \beta_0) + \sum_{i=1}^{d} W_i\right)\right] - \mathrm{Cov}\left[\delta^T G^{-1} g(U_{1:d}, \pi_0, \beta_0)\right]\right) \\
&= \Sigma_{R_\Gamma}^\star + \Sigma_{\mathrm{PVC}} + \Sigma_r.
\end{aligned}$$

Thus, under H_0 it follows that

$$\sqrt{n}(\hat{R}_\Gamma - R_\Gamma) \stackrel{d}{\to} N_L(0, \Sigma_{R_\Gamma}).$$

With the same arguments as in the proof of Lemma 1 this implies under H_0

$$T_n(\Gamma) = n(A\hat{R}_\Gamma)^T(A\hat{\Sigma}_{R_\Gamma}A^T)^{-1}A\hat{R}_\Gamma \xrightarrow{d} \chi^2(L-1),$$

where

$$\hat{\Sigma}_{R_\Gamma} = \widehat{\text{Cov}}\left(\delta^T\hat{G}^{-1}(g(U_{1:d}^k, \hat{\pi}, \hat{\beta}) + \sum_{i=1}^d \hat{W}_i)\right) \tag{A.8}$$

is a consistent estimator of Σ_{R_Γ}.[17]

A.3. Proof of Proposition 3

To obtain the asymptotic distribution of the test statistic Θ_n, we need the following lemmas.

Lemma 2.
Let $Y \sim F_Y$, where F_Y is the cdf of a continuous probability distribution. Additionally, let λ_n be a penalty function satisfying the conditions stated in Proposition 3. If $Y_n \xrightarrow{d} Y$ it holds that $Y_n - n\lambda_n \xrightarrow{p} -\infty$, i.e.,

$$\forall \alpha \in \mathbb{R} : \lim_{n\to\infty} \mathbb{P}(Y_n - n\lambda_n \leq \alpha) = 1.$$

Proof. Let $\alpha \in \mathbb{R}$. Since $n\lambda_n \to \infty$ it holds that

$$\forall \varepsilon > 0 \, \exists n_1 \in \mathbb{N} \, \forall n \geq n_1 : F_Y(\alpha + n\lambda_n) > 1 - \frac{\varepsilon}{2}. \tag{A.9}$$

By assumption Y_n converges in distribution to $Y \sim F_Y$, therefore

$$\forall \varepsilon > 0 \, \forall n_1 \in \mathbb{N} \, \exists n_2 \in \mathbb{N} \, \forall n \geq n_2 : |F_{Y_n}(\alpha + n_1\lambda_{n_1}) - F_Y(\alpha + n_1\lambda_{n_1})| < \frac{\varepsilon}{2}. \tag{A.10}$$

Moreover, $\exists n_3 \in \mathbb{N} \, \forall n \geq n_3 : n\lambda_n \geq n_1\lambda_{n_1}$. Thus, $\forall \varepsilon > 0 \, \forall n \geq \max(n_1, n_2, n_3)$ it holds that

$$\begin{aligned}
\mathbb{P}(Y_n \leq \alpha + n\lambda_n) &\geq \mathbb{P}(Y_n \leq \alpha + n_1\lambda_{n_1}) = F_{Y_n}(\alpha + n_1\lambda_{n_1}) \\
&= F_{Y_n}(\alpha + n_1\lambda_{n_1}) - F_Y(\alpha + n_1\lambda_{n_1}) + F_Y(\alpha + n_1\lambda_{n_1}) \\
&\overset{(A.10)}{>} F_Y(\alpha + n_1\lambda_{n_1}) - \frac{\varepsilon}{2} \overset{(A.9)}{>} 1 - \varepsilon.
\end{aligned}$$

Thus,

$$\lim_{n\to\infty} \mathbb{P}(Y_n - n\lambda_n \leq \alpha) = 1. \qquad \blacksquare$$

---

[17]See Genest, Ghoudi and Rivest [15] for a consistent estimator of $W_i = \int \partial_{u_i} g(u_{1:d}, \pi_0, \beta_0) \mathbb{1}_{\{U_i \leq u_i\}} dC_{1:d}(u_{1:d})$.

In the following Lemma 3 the asymptotic behavior of $\delta_n := \max\{Y_0^n + n\lambda_n, Y_1^n, Y_2^n, \ldots, Y_M^n\} - n\lambda_n$ is analyzed.

Lemma 3.
Let $(Y_0^k)_{k=1:n}, (Y_1^k)_{k=1:n}, \ldots, (Y_M^k)_{k=1:n}$ be $(M+1)$ sequences of random variables and $Y_m \sim F_{Y_m}$, $0 \leq m \leq M$, random variables with continuous cumulative distribution functions. Further let $\lambda_n : \mathbb{N} \to \mathbb{R}^+$ be a penalty function satisfying the conditions stated in Proposition 3.
Define $\delta_n := \max\{Y_0^n + n\lambda_n, Y_1^n, Y_2^n, \ldots, Y_M^n\} - n\lambda_n$.

(i) If $Y_m^n \xrightarrow{d} Y_m$ for each $0 \leq m \leq M$, it holds that $\delta_n \xrightarrow{d} Y_0$.
(ii) If there is an $m^\star \in \{0, \ldots, M\}$ such that $\operatorname{plim}_{n\to\infty} \frac{1}{n} Y_{m^\star}^n = y_{m^\star} > 0$ then
$\delta_n \xrightarrow{p} \infty$.

Proof. Proof of (i). Let $\alpha \in \mathbb{R}$, then

$$
\begin{aligned}
\mathbb{P}(\delta_n \leq \alpha) &= \mathbb{P}\left(\max\{Y_0^n + n\lambda_n, Y_1^n, \ldots, Y_M^n\} - n\lambda_n \leq \alpha\right) \\
&= \mathbb{P}\left(\max\{Y_0^n, Y_1^n - n\lambda_n, \ldots, Y_M^n - n\lambda_n\} \leq \alpha\right) \\
&= \mathbb{P}\left(Y_0^n \leq \alpha, Y_1^n - n\lambda_n \leq \alpha, \ldots, Y_M^n - n\lambda_n \leq \alpha\right).
\end{aligned}
$$

Using the Fréchet-Hoeffding inequalities [14, 22] we have

$$
\mathbb{P}(\delta_n \leq \alpha) \geq \max\left\{0, \mathbb{P}\left(Y_0^n \leq \alpha\right) + \sum_{m=1}^M \mathbb{P}\left(Y_m^n - n\lambda_n \leq \alpha\right) - M\right\}
$$

and

$$
\mathbb{P}(\delta_n \leq \alpha) \leq \min\left\{\mathbb{P}\left(Y_0^n \leq \alpha\right), \min_{1 \leq m \leq M}\left\{\mathbb{P}\left(Y_m^n - n\lambda_n \leq \alpha\right)\right\}\right\}.
$$

Due to the continuity of the minimum and maximum as well as Lemma 2 it follows that

$$
\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(\delta_n \leq \alpha) &\geq \lim_{n\to\infty} \max\left\{0, \mathbb{P}\left(Y_0^n \leq \alpha\right) + \sum_{m=1}^M \mathbb{P}\left(Y_m^n - n\lambda_n \leq \alpha\right) - M\right\} \\
&= \max\left\{0, \lim_{n\to\infty} \mathbb{P}\left(Y_0^n \leq \alpha\right) + \sum_{m=1}^M \lim_{n\to\infty} \mathbb{P}\left(Y_m^n - n\lambda_n \leq \alpha\right) - M\right\} \\
&= \max\left\{0, F_{Y_0}(\alpha) + \sum_{m=1}^M 1 - M\right\} = F_{Y_0}(\alpha)
\end{aligned}
$$

and

$$\lim_{n\to\infty} \mathbb{P}(\delta_n \leq \alpha) \leq \lim_{n\to\infty} \min\left\{ \mathbb{P}\left(Y_0^n \leq \alpha\right), \min_{1\leq m\leq M}\left\{\mathbb{P}\left(Y_m^n - n\lambda_n \leq \alpha\right)\right\}\right\}$$

$$= \min\left\{ \lim_{n\to\infty} \mathbb{P}\left(Y_0^n \leq \alpha\right), \min_{1\leq m\leq M}\left\{ \lim_{n\to\infty} \mathbb{P}\left(Y_m^n - n\lambda_n \leq \alpha\right)\right\}\right\} \tag{A.11}$$

$$= \min\left\{ F_{Y_0}(\alpha), \min_{1\leq m\leq M}\{1\}\right\} = F_{Y_0}(\alpha).$$

Thus,

$$\delta_n \xrightarrow{d} Y_0.$$

Proof of (ii). For $m = 0, \ldots, M$, define $Z_m^n := Y_m^n - n\lambda_n \mathbb{1}_{\{m\neq 0\}}$. Because $\lambda_n \to 0$ it follows that $\operatorname{plim}_{n\to\infty} \frac{1}{n} Z_{m^\star}^n = y_{m^\star}$. Note that $\mathbb{P}(Z_{m^\star}^n \leq \alpha) \xrightarrow{n\to\infty} 0$ for all $\alpha \in \mathbb{R}$ because $\frac{1}{n} Z_{m^\star}^n \xrightarrow{P} y_{m^\star} > 0$.

Thus, the Fréchet-Hoeffding upper bound implies that for any $\alpha \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}(\delta_n \leq \alpha) \leq \lim_{n\to\infty} \min\left\{ \mathbb{P}(Z_{m^\star}^n \leq \alpha), \min_{m\in\{0,\ldots,M\}\backslash m^\star}\left\{\mathbb{P}\left(Z_m^n \leq \alpha\right)\right\}\right\}$$

$$= \min\left\{ \lim_{n\to\infty} \mathbb{P}(Z_{m^\star}^n \leq \alpha), \min_{m\in\{0,\ldots,M\}\backslash m^\star}\left\{ \lim_{n\to\infty} \mathbb{P}\left(Z_m^n \leq \alpha\right)\right\}\right\}$$

$$= \min\Big\{0, \underbrace{\min_{m\in\{0,\ldots,M\}\backslash m^\star}\left\{ \lim_{n\to\infty} \mathbb{P}\left(Z_m^n \leq \alpha\right)\right\}}_{\in[0,1]}\Big\} = 0,$$

and the proof is complete. ∎

Using Proposition 2 and setting $Y_m^n = T_n(\Gamma_m)$ in Lemma 3 (i) it follows that the statistic Θ_n converges under H_0 to a $\chi^2(L_0 - 1)$ distribution.

Now assume that the correlations conditional on the subsets of the partition $\Gamma_{m^\star}$ are not identical, i.e., $\operatorname{plim}_{n\to\infty} \frac{1}{n} T_n(\Gamma_{m^\star}) = c > 0$ for some $m^\star \in \{0, \ldots, M\}$. Setting $Y_m^n = T_n(\Gamma_m)$ in Lemma 3 (ii) shows that the test statistic Θ_n converges in probability to infinity.

A.4. The decision tree: Algorithmic details

Every leaf in the decision tree represents a subset of the support Λ_0 of the random vector U_{D_e}. The maximum depth of the decision tree is denoted by $J_{\max}$ and every leaf is assigned to a level J in the decision tree ($0 \leq J \leq J_{\max}$). The level of a leaf refers to the number of splits which have already been used to arrive at the leaf, starting from the root leaf Λ_0 (see Figure 15).

A leaf is denoted by $\Lambda_{\gamma_{0:J}}$, where the $(J+1)$-dimensional vector $\gamma_{0:J} := (\gamma_0, \gamma_1, \ldots, \gamma_J) \in \{0\} \times \{l, r\}^J$ is the unique identifier for a leaf in the J-th level of the decision tree. That is, the two leaves in the $(J+1)$-th level of the decision tree being connected via edges to the leaf $\gamma_{0:J} = (\gamma_0, \gamma_1, \ldots, \gamma_J)$ in the J-th
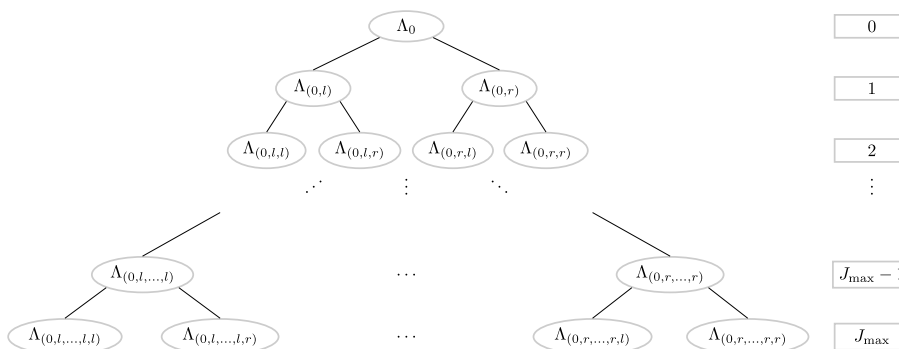
Fig 15: Partitioning of the support $\Lambda_0 = \mathrm{supp}(U_{D_e})$ of the random vector U_{D_e} into disjoint subsets $\Lambda_{\gamma_{0:J}}$, where $\gamma_{0:J} := (\gamma_0, \gamma_1, \ldots, \gamma_J) \in \{0\} \times \{l, r\}^J$, $0 \leq J \leq J_{\max}$, using a decision tree algorithm with maximum depth $J_{\max}$.

level are identified by $\gamma_{0:J+1} := (\gamma_{0:J}, k) := (\gamma_0, \gamma_1, \ldots, \gamma_J, k)$ with $k \in \{l, r\}$. The subsets assigned to the leafs in the $(J+1)$-th level by a binary split are given by

$$\Lambda_{(\gamma_{0:J},k)} \subset \Lambda_{\gamma_{0:J}}, \qquad J \geq 0, \qquad k \in \{l, r\}, \qquad \Lambda_{(\gamma_{0:J},l)} \cap \Lambda_{(\gamma_{0:J},r)} = \emptyset.$$

Every split is chosen out of a finite number $\mathcal{M}$ of possible splits. A possible split $S_{\gamma_{0:J}}^m$ in the leaf $\gamma_{0:J}$ is defined as a pair of disjoint subsets of $\Lambda_{\gamma_{0:J}}$, i.e., $S_{\gamma_{0:J}}^m := \left(\Lambda_{(\gamma_{0:J},l)}^m, \Lambda_{(\gamma_{0:J},r)}^m\right) \subset \Lambda_{\gamma_{0:J}} \times \Lambda_{\gamma_{0:J}}$ with $\Lambda_{(\gamma_{0:J},l)}^m \cap \Lambda_{(\gamma_{0:J},r)}^m = \emptyset$. From these possible splits, the split is selected that maximizes the statistic of the CCC test. Meaning that every split is defined as

$$S_{\gamma_{0:J}} = (\Lambda_{(\gamma_{0:J},l)}, \Lambda_{(\gamma_{0:J},r)}) = \operatorname*{argmax}_{S_{\gamma_{0:J}}^m \in \{S_{\gamma_{0:J}}^1, \ldots, S_{\gamma_{0:J}}^{\mathcal{M}}\}} T_n\left(S_{\gamma_{0:J}}^m\right).$$

Thus, the subsets that are transferred to leaf $(\gamma_{0:J}, k)$, $k \in \{l, r\}$, after using the optimal split $S_{\gamma_{0:J}}$, are given by $\Lambda_{(\gamma_{0:J},l)}$ and $\Lambda_{(\gamma_{0:J},r)}$. In the last level $J_{\max}$ we obtain a final partition of the support Λ_0 into $2^{J_{\max}}$ disjoint subsets given by $\Gamma_{\max} := \{\Lambda_{(0,l,\ldots,l,l)}, \Lambda_{(0,l,\ldots,l,r)}, \ldots, \Lambda_{(0,r,\ldots,r,l)}, \Lambda_{(0,r,\ldots,r,r)}\}$. For the final partition we compute the value of the test statistic

$$\Theta_n = \max\{T_n(\Gamma_0) + n\lambda_n, T_n(\Gamma_{\max})\} - n\lambda_n.$$

In all simulations in Section 5 and the real data applications in Section 6.2, we choose $\lambda_n = \frac{1}{\sqrt{n}}$ for the penalty and $\Gamma_0 = \Gamma_{\mathrm{med}}$ as base partition.[18] Further tuning parameters of the decision tree are the maximum depth $J_{\max}$ of the tree and the set of possible splits $\{S_{\gamma_{0:J}}^1, \ldots, S_{\gamma_{0:J}}^{\mathcal{M}}\}$. To keep the test computationally feasible and because it performs well in simulations, we consider a maximum depth of $J_{\max} = 2$ and the number of possible splits in each leaf $\gamma_{0:J}$ is restricted

---

[18]The partition Γ_{med} is defined in equation (4.1) in Section 4.1.

to be at most $\mathcal{M} = 3(j - 1 + \mathbb{1}_{\{j \geq 3\}})$ in tree T_j. The considered splits are as follows: To obtain the sets $\Lambda_{(0,l)}$ and $\Lambda_{(0,r)}$ for the two leaves in level 1, we consider the empirical 25%, 50% and 75% quartiles for each conditioning variable U_k, $k \in D_e$. If $j \geq 3$, we additionally take the empirical 25%, 50% and 75% quartiles of the mean aggregated conditioning variables into account, resulting in $3 \cdot j$ possible splits. A formal definition of the set of possible splits is given in Appendix A.5. The sets $\{\Lambda_{(0,l,l)}, \Lambda_{(0,l,r)}\}$ and $\{\Lambda_{(0,r,l)}, \Lambda_{(0,r,r)}\}$ for the four leaves in level 2 are obtained in the same fashion except that we now condition on $U_{D_e} \in \Lambda_{(0,l)}$ or $U_{D_e} \in \Lambda_{(0,r)}$, respectively. Furthermore, we use several restrictions in the decision tree algorithm to guarantee that the final data sets do not become too small.[19]

A.5. *Formal definition of the set of possible splits for the decision tree*

If $Q_q[(X_i)_{i \in \mathcal{I}}]$ denotes the empirical q-quantile of the vector $(X_i)_{i \in \mathcal{I}}$, the set of possible splits in the leaf $\gamma_{0:J}$, for $J = 0, 1$ and $e \in E_j$, $j \geq 3$ with $(j-1)$-dimensional conditioning set $D_e := D_e^{1:j-1}$ is given by

$$\mathcal{S}_{\gamma_{0:J}}^{1:3j} := \left\{ S_{\gamma_{0:J}}^1, \ldots, S_{\gamma_{0:J}}^{3j} \right\}$$
$$= \left\{ (\Lambda_{(\gamma_{0:J},l)}^1, \Lambda_{(\gamma_{0:J},r)}^1)), \ldots, (\Lambda_{(\gamma_{0:J},l)}^{3j}, \Lambda_{(\gamma_{0:J},r)}^{3j}) \right\},$$

with

$$\Lambda_{(\gamma_{0:J},l)}^m := \begin{cases} \left\{ u_{D_e} \in \Lambda_{\gamma_{0:J}} : u_{D_e^m} \leq Q_{0.25}[(V_{D_e^m}^k)_{k \in \mathcal{I}_{\gamma_{0:J}}}] \right\}, & 1 \leq m \leq j - 1 \\[2mm] \left\{ u_{D_e} \in \Lambda_{\gamma_{0:J}} : u_{D_e^{m_1}} \leq Q_{0.5}[(V_{D_e^{m_1}}^k)_{k \in \mathcal{I}_{\gamma_{0:J}}}] \right\}, & j \leq m \leq 2j - 2 \\[2mm] \left\{ u_{D_e} \in \Lambda_{\gamma_{0:J}} : u_{D_e^{m_2}} \leq Q_{0.75}[(V_{D_e^{m_2}}^k)_{k \in \mathcal{I}_{\gamma_{0:J}}}] \right\}, & 2j - 1 \leq m \leq 3j - 3 \\[2mm] \left\{ u_{D_e} \in \Lambda_{\gamma_{0:J}} : g(u_{D_e}) \leq Q_{0.25 m_3}[(g(V_{D_e}^k))_{k \in \mathcal{I}_{\gamma_{0:J}}}] \right\}, & 3j - 2 \leq m \leq 3j \end{cases}$$

and $\Lambda_{(\gamma_{0:J},r)}^m = \Lambda_{\gamma_{0:J}} \setminus \Lambda_{(\gamma_{0:J},l)}^m$. We further used the notation $m_\ell := m - \ell(j-1)$ for $1 \leq \ell \leq 3$ and the index set $\mathcal{I}_{\gamma_{0:J}}$ is defined as $\mathcal{I}_{\gamma_{0:J}} := \{k \in \{1, \ldots, n\} : V_{D_e}^k \in \Lambda_{\gamma_{0:J}}\}$.

A.6. *Choosing the penalty function: A finite sample analysis*

To apply the test based on the statistic Θ_n, a penalty function λ_n has to be specified and any choice satisfying the conditions stated in Proposition 3 results in an asymptotically valid test. However, the size and power for finite sample

---

[19]A decision tree with two or three splits is only applied if we have a certain amount of data. This is implemented by introducing a tuning parameter which controls the minimum sample size per leaf in the decision tree (the default value is 100 observations). As a result we do not always use the 25%, 50% and 75% quartiles as thresholds but depending on the available sample size we may only use the 50% quartile or even don't apply any additional split at all.

sizes depends on the chosen penalty function λ_n. In the following, the choice of the penalty function in finite samples will be analyzed in a simulation study under H_0, i.e., with a focus on the empirical size.

In all simulations in Section 5 and the real data applications in Section 6.2, we choose $\lambda_n = \frac{1}{\sqrt{n}}$ for the penalty and $\Gamma_0 = \Gamma_{\text{med}}$ as base partition.[20] As mentioned in Section 5.2, the test statistic $T_n(\Gamma_0)$ (the CCC test with fixed partition Γ_0) is related to the test statistic Θ_n (the CCC test with a decision tree selected partition) in the following way. For $\Theta_n = \max\{T_n(\Gamma_0), T_n(\Gamma_{\max}) - n\lambda_n\}$, with $\Gamma_{\max} := \text{argmax}_{\Gamma_m \in \{\Gamma_1, \ldots, \Gamma_M\}} T_n(\Gamma_m)$, it holds

$$T_n(\Gamma_0) \leq \Theta_n,$$

meaning that if we reject based on $T_n(\Gamma_0)$, we also reject based on Θ_n. It follows that the empirical size of Θ_n is bounded from below by the empirical size of $T_n(\Gamma_0)$ when both tests are applied to the same collection of data sets in a monte carlo simulation to compute the empirical size.

We now derive a condition on λ_n such that Θ_n and $T_n(\Gamma_0)$ result in equivalent test decisions. This means that the test statistic Θ_n is analyzed relative to $T_n(\Gamma_0)$.[21] Let $\tau := F^{-1}_{\chi^2(L_0-1)}(1-\alpha)$, where L_0 is the cardinality of Γ_0, i.e., the number of subsets forming the partition Γ_0. If the penalty function λ_n satisfies

$$\lambda_n > \frac{1}{n}\Big(T_n(\Gamma_{\max}) - \tau\Big) =: b_n, \tag{A.12}$$

it follows that $\Theta_n = \max\{T_n(\Gamma_0), T_n(\Gamma_{\max}) - n\lambda_n\} \leq \max\{T_n(\Gamma_0), \tau\}$. Therefore, if we cannot reject at a α-level based on $T_n(\Gamma_0)$ and if λ_n satisfies (A.12), we also cannot reject based on Θ_n, i.e., if $\lambda_n > b_n$ it holds

$$T_n(\Gamma_0) < \tau \quad \Rightarrow \quad \Theta_n < \tau.$$

As a result, if $\lambda_n > b_n$, both tests result in the same α-level test decisions, i.e.,

$$T_n(\Gamma_0) \geq \tau \quad \Leftrightarrow \quad \Theta_n \geq \tau.$$

Note that $T_n(\Gamma)$ converges in distribution to a χ^2-distribution under H_0 and by Slutsky's theorem it follows that $b_n \xrightarrow{p} 0$. Therefore, the lower bound b_n is bounded in probability, i.e.,

$$\forall \varepsilon > 0 \ \exists B > 0 \ \forall n \geq 1: \ \mathbb{P}(|b_n| \geq B) < \varepsilon.$$

Meaning that for any $\varepsilon > 0$, we can choose λ_n such that $\mathbb{P}(|b_n| \geq \lambda_n) < \varepsilon$, which restricts the probability of different test decisions (i.e., rejecting the H_0 with Θ_n but not rejecting the H_0 with $T_n(\Gamma_0)$) at a α-level to ε because

$$\mathbb{P}(\Theta_n \geq \tau, T_n(\Gamma_0) < \tau) = \mathbb{P}(\max\{T_n(\Gamma_0), T_n(\Gamma_{\max}) - n\lambda_n\} \geq \tau, T_n(\Gamma_0) < \tau)$$
$$= \mathbb{P}(T_n(\Gamma_{\max}) - n\lambda_n \geq \tau, T_n(\Gamma_0) < \tau)$$
$$\leq \mathbb{P}(T_n(\Gamma_{\max}) - n\lambda_n \geq \tau) = \mathbb{P}(b_n \geq \lambda_n) < \varepsilon.$$

---

[20]The partition Γ_{med} is defined in equation (4.1) in Section 4.1.

[21]An extensive simulation study of the finite sample performance of the proposed test Θ_n is presented in Section 5 where the empirical size relative to the theoretical level of the test is studied.
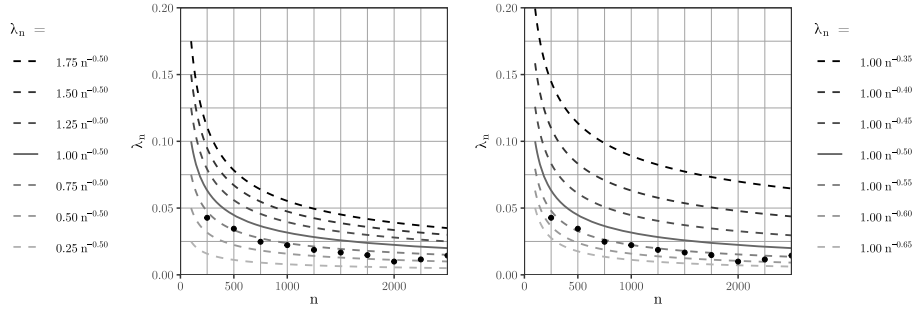
Fig 16: The dots correspond to the maximum of the lower bound for the penalty function in 1000 simulated samples each consisting of n observations. The data generating process is the four-dimensional D-vine copula defined in Example 2. The lines correspond to different choices of the penalty function $\lambda_n = cn^{-\beta}$. The solid line corresponds to the default penalty function $\lambda_n = \frac{1}{\sqrt{n}}$ used in all simulations and applications. On the left hand side the dashed lines correspond to different choices of the level c of the penalty function λ_n and on the right hand side to different choices for the power β of the penalty function.

This implies that for any $\varepsilon > 0$, we can choose λ_n such that $\mathbb{P}(|b_n| \geq \lambda_n) < \varepsilon$ and therefore

$$
\begin{aligned}
\mathbb{P}&\Big(T_n(\Gamma_0) \geq \tau \quad \Leftrightarrow \quad \Theta_n \geq \tau \Big) \\
&= \mathbb{P}(T_n(\Gamma_0) \geq \tau, \Theta_n \geq \tau) + \mathbb{P}(T_n(\Gamma_0) < \tau, \Theta_n < \tau) \\
&= 1 - \underbrace{\mathbb{P}(T_n(\Gamma_0) \geq \tau, \Theta_n < \tau)}_{=0} - \mathbb{P}(T_n(\Gamma_0) < \tau, \Theta_n \geq \tau) \geq 1 - \varepsilon.
\end{aligned}
$$

In practical applications, we are interested in the finite sample distribution of the lower bound b_n of the penalty function λ_n. Using resampling techniques, we can determine this lower bound for λ_n under H_0.

To illustrate how one can use resampling techniques to determine the parameters c and β of the penalty function $\lambda_n = cn^{-\beta}$, we again consider the data generating process given in Example 2. For $\lambda = 0$, the null hypothesis $H_0 : (U_{1|23}^{\mathrm{pvc}}, U_{4|23}^{\mathrm{pvc}}) \perp U_{2:3}$ is true. For each considered sample size n we generate 1000 random samples of size n from the four-dimensional D-vine copula and compute for each sample the lower bound b_n of the penalty function λ_n. In Figure 16, the maximum of all 1000 lower bounds in the different samples is plotted for different sample sizes n as dots. By taking the maximum over all resampled lower bounds we identify a lower bound for the penalty which would guarantee that in every of the 1000 samples the asymptotic α-level test decisions are the same, i.e., $T_n(\Gamma_0) \geq \tau \quad \Leftrightarrow \quad \Theta_n \geq \tau$. The level of the test is chosen to be 5% and the lines correspond to different choices of the penalty function $\lambda_n = cn^{-\beta}$. The level c of the penalty function is varied for a fixed power of $\beta = 0.5$ in the plot on the left hand side of Figure 16 and the power β of the penalty function is varied for a fixed level of $c = 1$ on the right hand side. The solid line corre-

sponds to the penalty function $\lambda_n = \frac{1}{\sqrt{n}}$ which we use in all the simulations and applications. One can see that the choice of the penalty function is reasonable in comparison to the lower bounds obtained via resampling techniques for all sample sizes between 250 and 2500 observations, as the penalty is for all sample sizes considerably larger than the lower bounds.

A.7. An analysis of the typical partitions selected by the decision trees

The following analysis complements the simulation study in Section 5.2 by an analysis of the partitions typically selected by the decision trees. To investigate what kind of partitions are selected by the decision trees and how they adapt to the different variations in the conditional correlation determined by $\alpha(\cdot), \alpha_I(\cdot)$, and $\alpha_D(\cdot)$ (see (5.1), (5.2) and (5.3)), we focus on the case $\lambda = 1$ and a sample size of $n = 1000$. If the variation of the copula parameter is given by $\alpha(\cdot)$, the first split is most often done with respect to the 75% quartile of $\bar{U}_{2:3} := \frac{1}{2}(U_2 + U_3)$ (in 93.6% of the simulated samples). The most frequently selected $\Gamma_{\max}$ partition (in 40.6% of the samples) consists of two more splits with respect to the quartiles of $\bar{U}_{2:3}$ in the next level. This partition is visualized in the left panel of Figure 17. The adaption of the decision tree based partition to the
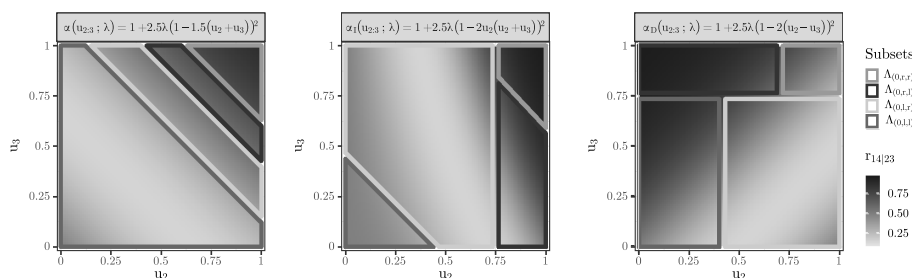
Fig 17: Visualization of the most frequently selected $\Gamma_{\max} := \{\Lambda_{(0,l,l)}, \Lambda_{(0,l,r)}, \Lambda_{(0,r,l)}, \Lambda_{(0,r,r)}\}$ partitions with the decision trees in the simulation study for $\lambda = 1$ and 1000 samples of size $n = 1000$. The three panels correspond to the parameter functions $\alpha(\cdot)$, $\alpha_I(\cdot)$ and $\alpha_D(\cdot)$, respectively. The shaded background shows the conditional correlation $r_{14|23}$ of $C_{14;23}$ as a function of U_2 and U_3.

variation in the conditional correlation becomes evident when looking at the shaded background which shows the conditional correlation $r_{14|23}$ of $C_{14;23}$ as a function of U_2 and U_3.

The middle panel in Figure 17 shows the most frequently selected $\Gamma_{\max}$ partition (in 14.8% of the samples) if the variation of the copula parameter is given by $\alpha_I(\cdot)$ which contains an interaction effect between the two conditioning variables. In this case, the first split is most often done with respect to the 75% quartile of U_2 (in 61.7% of the samples). The decision tree adapts to this variation in the conditional correlation, shown as shaded background, by choosing splits with respect to the quartiles of $\bar{U}_{2:3}$ in the second level.

The right panel in Figure 17 visualizes the most frequently selected $\Gamma_{\max}$ partition (in 22.7% of the samples) if the variation of the copula parameter is given by $\alpha_D(\cdot)$. Here, the two conditioning variables have an opposite sign in the definition of $\alpha_D(\cdot)$. In this case, the first split is most often done with respect to the 75% quartile of U_3 (in 52.8% of the samples). Once again, the decision tree adapts to the variation in the conditional correlation, shown as shaded background, by choosing splits with respect to the quartiles of the other variable U_2 in the second level.

Acknowledgments

References

[1] Aas, K., Czado, C., Frigessi, A. and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics* **44** 182–198. MR2517884

[2] Acar, E. F., Craiu, R. V. and Yao, F. (2013). Statistical testing of covariate effects in conditional copula models. *Electronic Journal of Statistics* **7** 2822–2850. MR3148369

[3] Bedford, T. and Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence* **32** 245–268. MR1859866

[4] Bedford, T. and Cooke, R. M. (2002). Vines – A new graphical model for dependent random variables. *The Annals of Statistics* **30** 1031–1068. MR1926167

[5] Bergsma, W. P. (2004). Testing conditional independence for continuous random variables. URL: https://www.eurandom.tue.nl/reports/2004/048-report.pdf.

[6] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31** 307–327. MR0853051

[7] Chan, N.-H., Chen, J., Chen, X., Fan, Y. and Peng, L. (2009). Statistical Inference for Multivariate Residual Copula of GARCH Models. *Statistica Sinica* **19** 53–70. MR2487877

[8] Chen, X. and Fan, Y. (2006). Estimation of copula-based semiparametric time series models. *Journal of Econometrics* **130** 307–335. MR2211797

[9] Cook, R. D. and Johnson, M. E. (1986). Generalized Burr-Pareto-Logistic Distributions With Applications to a Uranium Exploration Data Set. *Technometrics* **28** 123–131. MR0841485

[10] Derumigny, A. and Fermanian, J. D. (2017). About tests of the "simplifying" assumption for conditional copulas. *Dependence Modeling* **5** 154–197. MR3694366

[11] DISSMANN, J., BRECHMANN, E. C., CZADO, C. and KUROWICKA, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis* **59** 52–69. MR3000041

[12] ENGLE, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* **50** 987–1007. MR0666121

[13] FITHIAN, W., SUN, D. and TAYLOR, J. (2014). Optimal Inference After Model Selection. ArXiv e-prints arXiv:1410.2597.

[14] FRÉCHET, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Annales de l'Université de Lyon, Section A, Series 3* **14** 53–77. MR0049518

[15] GENEST, C., GHOUDI, K. and RIVEST, L. P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82** 543–552. MR1366280

[16] GIJBELS, I., OMELKA, M. and VERAVERBEKE, N. (2015). Partial and average copulas and association measures. *Electronic Journal of Statistics* **9** 2420–2474. MR3425363

[17] GIJBELS, I., OMELKA, M. and VERAVERBEKE, N. (2017). Nonparametric testing for no covariate effects in conditional copulas. *Statistics* **51** 475–509. MR3630461

[18] GIJBELS, I., OMELKA, M., PEŠTA, M. and VERAVERBEKE, N. (2017). Score tests for covariate effects in conditional copulas. *Journal of Multivariate Analysis* **159** 111–133. MR3668550

[19] GROTHE, O. and NICKLAS, S. (2013). Vine constructions of Lévy copulas. *Journal of Multivariate Analysis* **119** 1–15. MR3061411

[20] HOBÆK HAFF, I. (2013). Parameter estimation for pair-copula constructions. *Bernoulli* **19** 462–491. MR3037161

[21] HOBÆK HAFF, I., AAS, K. and FRIGESSI, A. (2010). On the simplified pair-copula construction – Simply useful or too simplistic? *Journal of Multivariate Analysis* **101** 1296–1310. MR2595309

[22] HOEFFDING, W. (1940). *Maßtabinvariante Korrelationstheorie. Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik an der Universität Berlin, 5.* MR0004426

[23] 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, IMS Lecture Notes – Monograph Series* (L. Rüschendorf, B. Schweizer and M. D. Taylor, eds.) 120–141. MR1485527

[24] JOE, H. (1997). *Multivariate Models and Dependence Concepts.* Chapman & Hall, London. MR1462613

[25] JOE, H., LI, H. and NIKOLOULOPOULOS, A. K. (2010). Tail dependence functions and vine copulas. *Journal of Multivariate Analysis* **101** 252–270. MR2557632

[26] KOJADINOVIC, I. and HOLMES, M. (2009). Tests of independence among continuous random vectors based on Cramér-von Mises functionals of the

empirical copula process. *Journal of Multivariate Analysis* **100** 1137–1154. MR2508377

[27] KRAUS, D. and CZADO, C. (2017). Growing simplified vine copula trees: improving Dißmann's algorithm. ArXiv e-prints `arXiv:1703.05203`.

[28] KUROWICKA, D. and JOE, H., eds. (2011). *Dependence Modeling*. World Scientific, Singapore. MR2856966

[29] KURZ, M. S. (2021). pacotest: Testing for Partial Copulas and the Simplifying Assumption in Vine Copulas URL: `https://CRAN.R-project.org/package=pacotest`, R-Package version 0.4.1.

[30] LAVERGNE, P. and PATILEA, V. (2008). Breaking the curse of dimensionality in nonparametric testing. *Journal of Econometrics* **143** 103–122. MR2384435

[31] LEE, J. D., SUN, D. L., SUN, Y. and TAYLOR, J. E. (2016). Exact post-selection inference, with application to the lasso. *The Annals of Statistics* **44** 907–927. MR3485948

[32] MROZ, T., FUCHS, S. and TRUTSCHNIG, W. (2021). How simplifying and flexible is the simplifying assumption in pair-copula constructions – analytic answers in dimension three and a glimpse beyond. *Electronic Journal of Statistics* **15** 1951–1992. MR4255318

[33] NAGLER, T. (2018). kdecopula: An R Package for the Kernel Estimation of Bivariate Copula Densities. *Journal of Statistical Software* **84** 1–22.

[34] NAGLER, T. and CZADO, C. (2016). Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas. *Journal of Multivariate Analysis* **151** 69–89. MR3545278

[35] NAGLER, T., SCHEPSMEIER, U., STOEBER, J., BRECHMANN, E. C., GRAELER, B. and ERHARDT, T. (2021). VineCopula: Statistical Inference of Vine Copulas URL: `https://CRAN.R-project.org/package=VineCopula`, R-Package version 2.4.2.

[36] NEWEY, W. K. and MCFADDEN, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics* **4** 2111–2245. MR1315971

[37] PATTON, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review* **47** 527–556. MR2216591

[38] PATTON, A. J. (2012). A review of copula models for economic time series. *Journal of Multivariate Analysis* **110** 4–18. MR2927506

[39] QUESSY, J. F. (2010). Applications and asymptotic power of marginal-free tests of stochastic vectorial independence. *Journal of Statistical Planning and Inference* **140** 3058–3075. MR2659837

[40] SCHELLHASE, C. and SPANHEL, F. (2018). Estimating non-simplified vine copulas using penalized splines. *Statistics and Computing* **28** 387–409. MR3747570

[41] SPANHEL, F. and KURZ, M. S. (2016a). The partial copula: Properties and associated dependence measures. *Statistics & Probability Letters* **119** 76–83. MR3555271

[42] SPANHEL, F. and KURZ, M. S. (2016b). Estimating parametric simplified vine copulas under general conditions. unpublished working paper.

[43] SPANHEL, F. and KURZ, M. S. (2019). Simplified vine copula models:

Approximations based on the simplifying assumption. *Electronic Journal of Statistics* **13** 1254–1291. MR3935849

[44] Stöber, J., Joe, H. and Czado, C. (2013). Simplified pair copula constructions – Limitations and extensions. *Journal of Multivariate Analysis* **119** 101–118. MR3061418

[45] Stöber, J. and Schepsmeier, U. (2013). Estimating standard errors in regular vine copula models. *Computational Statistics* **28** 2679–2707. MR3141358

[46] Tsukahara, H. (2005). Semiparametric estimation in copula models. *Canadian Journal of Statistics* **33** 357–375. MR2193980