

X-VINE MODELS FOR MULTIVARIATE EXTREMES

Anna Kiriliouk, Jeongjin Lee, Johan Segers

LIDAM Discussion Paper ISBA
2023 / 38

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

<https://uclouvain.be/en/research-institutes/lidam/isba/publication.html>

X-Vine Models for Multivariate Extremes

Anna Kiriliouk*

Jeongjin Lee*

Johan Segers†

December 22, 2023

Abstract

Regular vine sequences permit the organisation of variables in a random vector along a sequence of trees. Regular vine models have become greatly popular in dependence modelling as a way to combine arbitrary bivariate copulas into higher-dimensional ones, offering flexibility, parsimony, and tractability. In this project, we use regular vine structures to decompose and construct the exponent measure density of a multivariate extreme value distribution, or, equivalently, the tail copula density. Although these densities pose theoretical challenges due to their infinite mass, their homogeneity property offers simplifications. The theory sheds new light on existing parametric families and facilitates the construction of new ones, called X-vines. Computations proceed via recursive formulas in terms of bivariate model components. We develop simulation algorithms for X-vine multivariate Pareto distributions as well as methods for parameter estimation and model selection on the basis of threshold exceedances. The methods are illustrated by Monte Carlo experiments and a case study on US flight delay data.

Key words: exponent measure, graphical model, multivariate Pareto distribution, pair copula construction, regular vine, tail copula

1 Introduction

For multivariate extremes, margin-free tail dependence models based on max-stable distributions arise from classical limit theory for sample extremes (de Haan and Resnick, 1977). A question of high interest is the construction of such models that are flexible, parsimonious, and computationally tractable, and scale well as the dimension grows (Engelke and Ivanovs, 2021). To do so, we propose a novel approach based on regular vine tree sequences (Bedford and Cooke, 2001, 2002), called X-vines. The models can easily be built in arbitrary dimension by combining bivariate components only. The latter can be chosen independently from one another, giving great flexibility. The pairs are grouped in trees, the number of which determines the complexity of the model. Computations proceed by recursive algorithms.

For copula-based dependence modelling, outside the context of extreme value analysis, vine constructions have grown into a versatile and widely applied approach (Czado, 2019; Czado and Nagler, 2022; Nagler and Vatter, 2023). Our contribution is to make vine-based methods operational, in theory and practice, for densities of exponent measures of multivariate extreme value distributions. The challenge to overcome is that these densities do not integrate to one but to infinity. A change of margin transforms exponent measure densities into tail copula densities, which still have infinite mass, but with structural properties that resemble copula densities more closely.

Figure 1 shows a regular vine sequence $\mathcal{V}$ in dimension $d = 5$. The structure consists of four trees $T_1, \dots, T_4$, in which the edges of one tree become nodes in the next one. Each of the $d(d-1)/2 = 10$ pairs of variables appears as the leading pair before the semicolon on exactly one edge. The numbers behind the semicolons refer to conditioning variables. The first tree represents a Markov tree, while the subsequent trees add higher-order dependence relations.

An X-vine specification consists of a regular vine sequence $\mathcal{V}$, together with, for each edge in the first tree, a bivariate exponent measure or tail copula density, and, for each edge of the subsequent trees, a bivariate copula density, not necessarily stemming from extreme value theory. For the example in Fig. 1, four bivariate exponent measure densities (T_1) and six bivariate copula densities (T_2, T_3, T_4) would be required. These bivariate components can be chosen independently from another, without any constraints. Our main result shows how to combine bivariate building blocks into a single multivariate exponent measure or tail copula density. Further, we show that commonly used parametric models in extreme value analysis are actually examples of such X-vine models. We leverage the recursive nature of regular vine sequences both in theory and computations.

*Namur Institute for Complex Systems, University of Namur, Rue Grafé 2, 5000 Namur, Belgium. E-mail: anna.kiriliouk@unamur.be, jeongjin.lee@unamur.be

†LIDAM/ISBA, UCLouvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium. E-mail: johan.segers@uclouvain.be

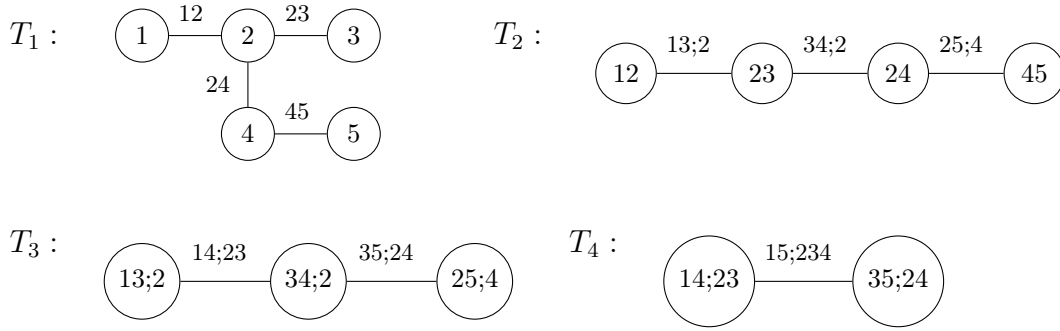


Figure 1: A regular vine sequence $\mathcal{V}$ in dimension $d = 5$ consisting of trees $T_1, \dots, T_4$, where T_j has $d - j + 1$ nodes and $d - j$ edges. The nodes of T_1 are $\{1, \dots, 5\}$, while the edges of T_j become the nodes of T_{j+1} .

The models constructed in this way turn out to be the tail limits of regular vine copulas. For the special case of D-vine copulas, such tail limits were already computed in Joe (1996) and Joe et al. (2010), focusing on tail dependence functions rather than densities. In Simpson et al. (2021), tails of D-vine and C-vine copulas were investigated from the perspective of the limiting shapes of sample clouds.

X-vine constructions are related to but different from recently proposed graphical models for extremes, either for directed or undirected graphs (Gissibl and Klüppelberg, 2018; Engelke and Hitz, 2020; Engelke et al., 2022). Markov trees (Segers, 2020; Hu et al., 2023) are a special case in the intersection of graphical and X-vine models.

After reviewing background on multivariate extremes and regular vine sequences in Section 2, we state a version of Sklar’s theorem for tail copula densities in Section 3. Some well-known parametric examples are worked out in Section 4, with a focus on the simplifying assumption that the conditional copula densities only depend on the values of the conditioning variables through their indices, but not their actual values. Section 5 contains the paper’s main results, showing on the one hand how a general tail copula density can be decomposed along any regular vine sequence, and, on the other hand, how to construct a tail copula density from a regular vine sequence and bivariate building blocks. The so-called X-vine models arising in this way are put to work in the subsequent sections, covering simulation algorithms (Section 6), parameter estimation and model selection (Section 7), simulation studies (Section 8), and a case study on US flight delay data (Section 9) taken from Hentschel et al. (2022). Section 10 concludes. The appendices contain a detailed example to illustrate the theory of regular vine sequences and X-vine models, the proofs of the paper’s results, additional theory on Sklar’s theorem, exponent measure densities and the connection with regular vine copulas, as well as additional numerical results. The methods are implemented in R (R Core Team, 2023), relying in particular on packages `graphicalExtremes` (Engelke et al., 2022) and `VineCopula` (Nagler et al., 2023); a dedicated software package is under construction.

2 Background

We write $[d] = \{1, \dots, d\}$ for the index set of the variables. Bold symbols refer to multivariate quantities. For a point $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ and a subset $J \subseteq [d]$, write subvectors as $\mathbf{x}_J = (x_j)_{j \in J}$ and $\mathbf{x}_{\setminus J} = (x_j)_{j \in [d] \setminus J}$. Mathematical operations on vectors such as addition, multiplication and comparison are considered component-wise.

2.1 Multivariate extreme value theory: tail copulas and their densities

Tail copulas

Classical extreme value theory starts from the assumption that the distribution function F of a random vector $\mathbf{X} = (X_1, \dots, X_d)$ is in the max-domain of attraction of a multivariate extreme value distribution. This assumption comprises two parts: (i) all univariate marginal distributions $F_1, \dots, F_d$ are in the domains of attraction of some univariate extreme value distributions; (ii) after reduction of the univariate margins to a common scale, the joint distribution of the random vector is multivariate regularly varying. We refer to Beirlant et al. (2004) and de Haan and Ferreira (2007) for textbook treatments of multivariate extreme value theory.

The focus in this paper is on part (ii), which concerns the tail dependence of $\mathbf{X}$, that is, the probabilities of high values occurring jointly among the variables X_j . To describe the mathematical theory behind tail dependence, let C be the survival copula of $\mathbf{X}$: under the standing assumption that the marginal distribution

functions $F_1, \dots, F_d$ are continuous, C is the distribution function of the random vector $\mathbf{U} = (U_1, \dots, U_d)$ with uniformly distributed components $U_j = 1 - F_j(X_j)$ for $j \in [d]$. Our interest is in high values of X_j and thus in low outcomes of U_j .

Definition 2.1 (Tail copula and tail copula measure). The (lower) *tail copula* R of C is the function on $\mathbb{E} = (0, \infty]^d \setminus \{\infty\}$ defined by

$$R(\mathbf{x}) = \lim_{t \searrow 0} t^{-1} C(t\mathbf{x}), \quad \mathbf{x} \in \mathbb{E}, \quad (2.1)$$

provided the limit exists. If it does, then the *tail copula measure*, denoted by the same symbol R , is the Borel measure on $\mathbb{E}$ determined by $R((\mathbf{0}, \mathbf{x}]) = R(\mathbf{x})$ for $\mathbf{x} \in \mathbb{E}$.

The term ‘tail copula’ stems from Schmidt and Stadtmüller (2006) in the bivariate case; Joe et al. (2010) call it ‘tail dependence function’. The tail copula measure already appears in Einmahl et al. (2001), where it is denoted by Λ . It is closely linked to the exponent measure ν in Eq. (2.9) below, introduced in de Haan and Resnick (1977) and which constitutes a fundamental notion in multivariate extreme value theory.

The total mass of the tail copula measure R equals infinity but $R(B)$ is finite for Borel sets B contained in $\{\mathbf{x} \in \mathbb{E} : \min \mathbf{x} \leq M\}$ for some $M > 0$. Its univariate margins are equal to the one-dimensional Lebesgue measure:

$$\forall j \in [d], \forall x_j \in (0, \infty), \quad R(\{\mathbf{y} \in \mathbb{E} : y_j \leq x_j\}) = x_j. \quad (2.2)$$

Equation (2.1) implies that R is homogeneous of order one, both as a function and as a measure: for $\mathbf{x} \in \mathbb{E}$, Borel sets $B \subseteq \mathbb{E}$, and $s \in (0, \infty)$,

$$R(s\mathbf{x}) = s R(\mathbf{x}) \quad \text{and} \quad R(sB) = s R(B). \quad (2.3)$$

Any Borel measure R on $\mathbb{E}$ satisfying (2.2) and (2.3) is a tail copula measure, i.e., is the limit in (2.1) for some copula C .

Tail copula densities

Throughout, we assume that the tail copula measure R has no mass on the hyperplanes through infinity, that is, $R(\{\mathbf{x} \in \mathbb{E} : x_j = \infty\}) = 0$ for all $j \in [d]$, so that R is concentrated on $(0, \infty)^d$. Furthermore, we assume that R is absolutely continuous with respect to the d -dimensional Lebesgue measure with continuous density $r : (0, \infty)^d \rightarrow [0, \infty)$, that is, $R(B) = \int_B r(\mathbf{x}) d\mathbf{x}$ for all Borel sets $B \subseteq (0, \infty)^d$. Choosing $B = (\mathbf{0}, \mathbf{x}]$ for $\mathbf{x} \in (0, \infty)^d$, we recover r from R by $r(\mathbf{x}) = \frac{\partial^d}{\partial x_1 \dots \partial x_d} R(\mathbf{x})$ for all $\mathbf{x} \in (0, \infty)^d$. The marginal constraint (2.2) implies

$$\forall j \in [d], \forall x_j \in (0, \infty), \quad \int_{(0, \infty)^{d-1}} r(\mathbf{x}) d\mathbf{x}_{\setminus j} = 1. \quad (2.4)$$

In view of (2.3), r is homogeneous of order $1 - d$:

$$\forall s \in (0, \infty), \forall \mathbf{x} \in (0, \infty)^d, \quad r(s\mathbf{x}) = s^{1-d} r(\mathbf{x}). \quad (2.5)$$

Properties (2.4) and (2.5) characterise the set of d -variate *tail copula densities*; see Section 4 for some well-known parametric families.

By Scheffé’s lemma from measure theory, if the copula C has density c and the tail copula measure R has density r , then (2.1) is implied by $\lim_{t \searrow 0} t^{1-d} c(t\mathbf{x}) = r(\mathbf{x})$ for all $\mathbf{x} \in (0, \infty)^d$. A word of caution: for a given copula density c , the limit $\lim_{t \searrow 0} t^{1-d} c(t\mathbf{x})$ may exist for all $\mathbf{x} \in (0, \infty)^d$ but not be a tail copula density, as the marginal constraint (2.4) may fail; a case in point is the independence copula, with density $c \equiv 1$, in which case the said limit is zero.

Margins of tail copulas

For a non-empty set $J \subseteq [d]$, let $\pi_J : (0, \infty)^d \rightarrow (0, \infty)^J$ denote the coordinate projection $\mathbf{x} \mapsto \pi_J(\mathbf{x}) = \mathbf{x}_J$ and let $R_J = R \circ \pi_J^{-1}$ denote the J -th marginal measure of R ,

$$R_J(B) = R(\pi_J^{-1}(B)) = R(\{\mathbf{x} \in (0, \infty)^d : \mathbf{x}_J \in B\}), \quad (2.6)$$

for Borel sets $B \subseteq (0, \infty)^J$. The choice $B = \prod_{j \in J} (0, x_j]$ for $x_j \in (0, \infty)$ shows that R_J is the tail copula measure of the copula C_J of $\mathbf{U}_J = \pi_J(\mathbf{U})$, as in Definition 2.1. For $\mathbf{x}_J \in (0, \infty]^J \setminus \{\infty\}$, we have

$$R_J(\mathbf{x}_J) = R(\tilde{\mathbf{x}}_J) \quad \text{with} \quad \tilde{x}_j = \begin{cases} x_j & \text{if } j \in J, \\ \infty & \text{if } j \in [d] \setminus J. \end{cases}$$

The density r_J of the measure R_J is

$$r_J(\mathbf{x}_J) = \frac{\partial^{|J|}}{\prod_{j \in J} \partial x_j} R_J \left(\prod_{j \in J} (0, x_j] \right) = \int_{\mathbf{x}_{\setminus J} \in (0, \infty)^{d-|J|}} r(\mathbf{x}) d\mathbf{x}_{\setminus J}, \quad (2.7)$$

for $\mathbf{x}_J \in (0, \infty)^J$ and, provided $J \neq [d]$, is obtained from r by integrating out the variables x_j with indices $j \notin J$.

Relations between tail copulas and other concepts in multivariate extremes

For later reference, we review a number of related terms from multivariate extreme value theory. This paragraph can be skipped at first reading, as the contributions in Sections 3 and 5 do not depend on it.

Equation (2.1) is equivalent to the assumption that the random vector $\mathbf{V} = (V_1, \dots, V_d)$ with $V_j = 1/U_j$ for $j \in [d]$ satisfies

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathbf{V} \leq n\mathbf{z})^n = G(\mathbf{z}) = \exp\{-\ell(1/\mathbf{z})\}, \quad \mathbf{z} \in (0, \infty)^d, \quad (2.8)$$

that is, $\mathbf{V}$ is in the max-domain of attraction of a *multivariate extreme value* (or *max-stable*) distribution G defined in terms of the *stable tail dependence function* $\ell(\mathbf{x}) = R(\mathbb{E} \setminus [\mathbf{x}, \infty])$ for $\mathbf{x} \in [0, \infty)^d$. The margins of $\mathbf{V}$ are unit-Pareto, $\mathbb{P}(V_j > x) = 1/x$ for $x \geq 1$, while those of G are unit-Fréchet, $G_j(x) = \exp(-1/x)$ for $x > 0$. Equivalently, the distribution of $\mathbf{V}$ is *multivariate regularly varying* (de Haan and Resnick, 1977; Resnick, 2007) with limit measure

$$\Lambda(\cdot) = R(\{\mathbf{x} \in \mathbb{E} : (1/x_1, \dots, 1/x_d) \in \cdot\}), \quad (2.9)$$

on $[0, \infty)^d \setminus \{\mathbf{0}\}$, that is, $t\mathbb{P}(\mathbf{V} \in tB) \rightarrow \Lambda(B)$ as $t \rightarrow \infty$ for Borel sets B contained in $[0, \infty)^d \setminus [0, \varepsilon]^d$ for some $\varepsilon > 0$ and satisfying $\Lambda(\partial B) = 0$, with ∂B the boundary of B . The measure Λ is called *exponent measure* because of the identity $G(\mathbf{z}) = \exp\{-\Lambda([0, \infty) \setminus [\mathbf{0}, \mathbf{z}])\}$ for $\mathbf{z} \in (0, \infty)^d$.

Another statement equivalent to (2.1) is that the conditional distribution of $\mathbf{V}/t$ given $\max \mathbf{V} > t$ is asymptotically *multivariate Pareto* for high threshold t , that is, we have the weak convergence relation

$$(\mathbf{V}/t \mid \max \mathbf{V} > t) \rightsquigarrow \mathbf{Y}, \quad t \rightarrow \infty, \quad (2.10)$$

where $\mathbf{Y}$ is a random vector supported on $\mathbb{L}_{>1} := \{\mathbf{y} \in [0, \infty)^d : \max \mathbf{y} > 1\}$ with distribution equal to the restriction of Λ in (2.9) to $\mathbb{L}_{>1}$ and normalized to a probability measure,

$$\mathbb{P}(\mathbf{Y} \in B) = \Lambda(B \cap \mathbb{L}_{>1}) / \Lambda(\mathbb{L}_{>1}) \quad (2.11)$$

for Borel sets $B \subseteq \mathbb{R}^d$ and normalizing constant $\Lambda(\mathbb{L}_{>1}) = \ell(\mathbf{1})$. For $\mathbf{y} \in \mathbb{L}_{>1}$, we also have $\mathbb{P}(\mathbf{Y} \geq \mathbf{y}) = R(1/\mathbf{y})/\ell(\mathbf{1})$.

Up to a location shift, multivariate Pareto distributions are a special case of *multivariate generalised Pareto distributions* introduced in Rootzén and Tajvidi (2006). They play a prominent role in the theory of graphical models for extremes (Engelke and Hitz, 2020). In the context of tail copula measures, it is more convenient to take component-wise reciprocals in (2.10) leading to *inverted multivariate Pareto distributions*, introduced in Section 6.

If the tail copula measure R is concentrated on $(0, \infty)^d$ and has density r , the exponent measure and the multivariate Pareto distribution have densities too. The exponent measure ν is concentrated on $(0, \infty)^d$ and has density

$$\lambda(\mathbf{y}) = r(1/\mathbf{y}) \prod_{j=1}^d y_j^{-2}, \quad \mathbf{y} \in (0, \infty)^d, \quad (2.12)$$

while the multivariate Pareto vector $\mathbf{Y}$ has probability density $\lambda(\mathbf{y})/\ell(\mathbf{1})$ for $\mathbf{y} \in \mathbb{L}_{>1}$.

By (2.3) and (2.9), the exponent measure ν is homogeneous of order -1 and thus determined by the *angular measure* H on the unit simplex $\Delta_{d-1} = \{\mathbf{x} \in [0, 1]^d : \|\mathbf{x}\|_1 = 1\}$ via

$$\Lambda(\{\mathbf{x} \in [0, \infty)^d : \|\mathbf{x}\|_1 > t, \mathbf{x}/\|\mathbf{x}\|_1 \in B\}) = t^{-1}H(B),$$

for $t > 0$ and Borel sets $B \subseteq \Delta_{d-1}$, with $\|\mathbf{x}\|_1 = \sum_{j=1}^d |x_j|$ for $\mathbf{x} \in \mathbb{R}^d$. Under the above smoothness assumption, the angular measure H is concentrated on the relative interior of the unit simplex Δ_{d-1} and its density h satisfies $h(\mathbf{w}) = \lambda(\mathbf{w})$ for $\mathbf{w} \in \Delta_{d-1}$ (Dombry et al., 2016), see also Lemma B.1 in Appendix B.

2.2 Regular vine sequences and the vine telescoping product formula

Informally, a regular vine sequence on d elements consists of a linked set of $d - 1$ trees, where the edges in tree $j - 1$ become nodes in tree j , joined by an edge only if they share a common node as edges in tree $j - 1$, for $j \in \{2, \dots, d - 1\}$ (Bedford and Cooke, 2002). *Regular vine copulas*, also called *pair copula constructions*, allow for flexible high-dimensional modelling using parametric families of bivariate copulas as building blocks for the edges of each tree. For recent overviews, see, for example, the monograph by Czado (2019) or the survey article by Czado and Nagler (2022). Our objective is to do the same for tail copula densities.

Recall that a tree $T = (N, E)$ is a connected, acyclic graph comprising a node set N and an edge set $E \subseteq \{\{x, y\} : x, y \in N, x \neq y\}$. Any two distinct nodes in a tree are connected by a unique path.

Definition 2.2 (Regular vine tree sequence). An ordered set of trees $\mathcal{V} = (T_1, \dots, T_{d-1})$ is a *regular vine tree sequence* on $d \geq 3$ elements if $T_1 = (N_1, E_1)$ is a tree with node set $N_1 = [d]$, while for $j \in \{2, \dots, d - 1\}$, the tree $T_j = (N_j, E_j)$ has node set $N_j = E_{j-1}$ and the *proximity condition* holds: for any $\{a, b\} \in E_j$, we have $|a \cap b| = 1$, that is, two nodes in T_j can be connected only if, as edges in T_{j-1} , they share a common node.

For brevity, we also use the term *regular vine sequence*. The proximity condition and other concepts defined below are illustrated via a five-dimensional example in Appendix A.

Definition 2.3 (Complete union, conditioning set, conditioned set). Let $\mathcal{V} = (T_1, \dots, T_{d-1})$ be a regular vine sequence. For any edge $e \in E_j$, the *complete union* of e is $A_e = e$ if $j = 1$ and $A_e = A_a \cup A_b$ if $e = \{a, b\}$ and $j \in \{2, \dots, d - 1\}$; informally, A_e is the subset of nodes in $[d]$ reachable from e by the membership relation. The *conditioning set* of an edge $e = \{a, b\} \in E_j$ with $j \geq 2$ is $D_e = A_a \cap A_b$ and the *conditioned set* of e is $\mathcal{C}_e = \mathcal{C}_{e,a} \cup \mathcal{C}_{e,b}$ with $\mathcal{C}_{e,a} = A_a \setminus D_e = A_a \setminus A_b$ and $\mathcal{C}_{e,b} = A_b \setminus D_e = A_b \setminus A_a$; if $e = \{a, b\} \in E_1$, we have $\mathcal{C}_{e,a} = \{a\}$, $\mathcal{C}_{e,b} = \{b\}$, $\mathcal{C}_e = e$ and $D_e = \emptyset$.

Clearly, $D_e = A_a \setminus \mathcal{C}_e = A_b \setminus \mathcal{C}_e$. Further, $\mathcal{C}_{e,a}$, $\mathcal{C}_{e,b}$ and D_e are disjoint subsets of N_1 ; $\mathcal{C}_{e,a}$ and $\mathcal{C}_{e,b}$ are singletons while D_e has $j - 1$ elements; and $\mathcal{C}_{e,a} \cup \mathcal{C}_{e,b} \cup D_e = A_e$, a node set with $j + 1$ elements. If $j \geq 2$, then $D_a \cup D_b \subseteq D_e$; indeed, by proximity, $a \cap b = \{k\}$ with $k \in N_{j-1}$, and $D_a \subset A_k \subset A_a \cap A_b = D_e$; similarly $D_b \subset D_e$. Further, the conditioned sets are related by $\mathcal{C}_{e,a} = \mathcal{C}_e \cap \mathcal{C}_a$ and $\mathcal{C}_{e,b} = \mathcal{C}_e \cap \mathcal{C}_b$. To see this, note that, by proximity, we have $a = \{a_1, k\}$ and $b = \{b_1, k\}$ for three distinct elements $a_1, b_1, k \in N_{j-1}$, and $\mathcal{C}_{e,a} = A_a \setminus A_b = (A_{a_1} \cup A_k) \setminus (A_{b_1} \cup A_k) \subseteq A_{a_1} \setminus A_k = \mathcal{C}_{a,a_1}$. Since both the left- and right-hand side are singletons, we must have $\mathcal{C}_{e,a} = \mathcal{C}_{a,a_1} = \mathcal{C}_e \cap \mathcal{C}_a$; similarly $\mathcal{C}_{e,b} = \mathcal{C}_{b,b_1} = \mathcal{C}_e \cap \mathcal{C}_b$.

The edge e is also written as $e = (\mathcal{C}_{e,a}, \mathcal{C}_{e,b}; D_e)$ and, since $\mathcal{C}_{e,a}$ and $\mathcal{C}_{e,b}$ are singletons, $\{a_e\}$ and $\{b_e\}$, say, we abbreviate $e = (a_e, b_e; D_e)$, with $a_e, b_e \in N_1$ and the convention that $a_e < b_e$, that is, $a_e = \min \mathcal{C}_e$ and $b_e = \max \mathcal{C}_e$. Similarly, for $e = \{a, b\} \in E_1$, the convention is that $a < b$. In Fig. 1, for instance, the edge e labelled (14; 23) in T_3 has $a_e = 1$, $b_e = 4$, and $D_e = \{2, 3\}$. Appendix A contains a fully worked out example to illustrate these concepts.

For positive scalars $\gamma_1, \dots, \gamma_d$ with $\gamma_1 = 1$, the identity $\prod_{j=2}^d \gamma_j / \gamma_{j-1} = \gamma_d$ is known as a *telescoping product*. Regular vine sequences enjoy a similar property that will be key to Theorem 5.1 and which, to the best of our knowledge, has not yet been stated in the literature.

Lemma 2.4 (Vine telescoping product). Let $d \geq 3$ be an integer and let $\gamma_J \in (0, \infty)$ for every non-empty $J \subseteq [d]$ with $\gamma_j := \gamma_{\{j\}} = 1$ for every $j \in [d]$. Given a regular vine sequence $\mathcal{V} = (T_j = (N_j, E_j))_{j=1}^{d-1}$ on $[d]$, we have

$$\gamma_{[d]} = \prod_{e \in E_1} \gamma_e \cdot \prod_{j=2}^{d-1} \prod_{e=\{a,b\} \in E_j} \frac{\gamma_{D_e} \cdot \gamma_{A_e}}{\gamma_{A_a} \cdot \gamma_{A_b}}. \quad (2.13)$$

With the notation $\gamma_{I|J} = \gamma_{I \cup J} / \gamma_J$, the last factor in (2.13) is

$$\frac{\gamma_{D_e} \cdot \gamma_{A_e}}{\gamma_{A_a} \cdot \gamma_{A_b}} = \frac{\gamma_{\{a_e, b_e\}|D_e}}{\gamma_{a_e|D_e} \cdot \gamma_{b_e|D_e}}. \quad (2.14)$$

The formula (2.13) is spelled out in Eq. (A.1) in Appendix A for a five-dimensional regular vine sequence.

Remark 2.5 (Edge-wise vine telescoping product). Given an edge $f \in E_j$ for some $j \in \{1, \dots, d - 1\}$ in a regular vine sequence $\mathcal{V}$ on $[d]$, we can consider the sub-vine $\mathcal{V}_f$ induced by $\mathcal{V}$ on the node set A_f in the natural way: we have $\mathcal{V}_f = (T_{f,i} = (N_{f,i}, E_{f,i}))_{i=1}^j$ where $N_{f,1} = A_f$ and $E_{f,i} = \{e \in E_i : A_e \subseteq A_f\}$ for $i \in \{1, \dots, j\}$. Applying Lemma 2.4 to $\mathcal{V}_f$, we obtain the seemingly more general formula

$$\gamma_{A_f} = \prod_{e \in E_1 : A_e \subseteq A_f} \gamma_e \cdot \prod_{i=2}^j \prod_{e \in E_i : A_e \subseteq A_f} \frac{\gamma_{\{a_e, b_e\}|D_e}}{\gamma_{a_e|D_e} \cdot \gamma_{b_e|D_e}}. \quad (2.15)$$

3 Sklar's theorem for tail copula densities

Let $r : (0, \infty)^d \rightarrow [0, \infty)$ be a d -variate tail copula density, that is, properties (2.4) and (2.5) hold. Recall its multivariate margins in (2.7). For non-empty, disjoint subsets $I, J \subset [d]$ and for $\mathbf{x}_I \in (0, \infty)^I$ and $\mathbf{x}_J \in (0, \infty)^J$ such that $r_J(\mathbf{x}_J) > 0$, define

$$\forall \mathbf{x}_{I \cup J} \in (0, \infty)^{I \cup J}, \quad r_{I|J}(\mathbf{x}_I | \mathbf{x}_J) := \frac{r_{I \cup J}(\mathbf{x}_{I \cup J})}{r_J(\mathbf{x}_J)}.$$

Viewing the “conditional” tail copula density $r_{I|J}$ as an $|I|$ -variate probability density, we can decompose it into $|I|$ univariate probability densities and a copula density.

Proposition 3.1 (Sklar's theorem: direct part). *Let r be a d -variate tail copula density and let $I, J \subset [d]$ be non-empty and disjoint. For $\mathbf{x}_J \in (0, \infty)^J$ such that $r_J(\mathbf{x}_J) > 0$, the function*

$$r_{I|J}(\cdot | \mathbf{x}_J) : (0, \infty)^I \rightarrow [0, \infty), \quad \mathbf{x}_I \mapsto r_{I|J}(\mathbf{x}_I | \mathbf{x}_J),$$

is a probability density on $(0, \infty)^I$. For $i \in I$, it has marginal probability density $x_i \mapsto r_{i|J}(x_i | \mathbf{x}_J)$ and cumulative distribution function

$$x_i \mapsto R_{i|J}(x_i | \mathbf{x}_J) := \int_0^{x_i} r_{i|J}(t | \mathbf{x}_J) dt,$$

with quantile function $u_i \mapsto R_{i|J}^{-1}(u_i | \mathbf{x}_J)$. If $|I| \geq 2$, then the copula density of $r_{I|J}(\cdot | \mathbf{x}_J)$ is

$$\forall \mathbf{u}_I \in (0, \infty)^I, \quad c_{I,J}(\mathbf{u}_I; \mathbf{x}_J) := \frac{r_{I|J}(\mathbf{x}_I | \mathbf{x}_J)}{\prod_{i \in I} r_{i|J}(x_i | \mathbf{x}_J)} \quad \text{with } x_i = R_{i|J}^{-1}(u_i | \mathbf{x}_J). \quad (3.1)$$

The homogeneity property (2.5) induces certain scaling properties.

Lemma 3.2 (Scale equi- and invariance). *In Proposition 3.1, for $\mathbf{x}_J \in (0, \infty)^J$ such that $r_J(\mathbf{x}_J) > 0$, the family $\{r_{I|J}(\cdot | t\mathbf{x}_J) : t \in (0, \infty)\}$ of probability densities is a scale family: if the random vector $\boldsymbol{\xi}_I$ has probability density $r_{I|J}(\cdot | \mathbf{x}_J)$ then the random vector $t\boldsymbol{\xi}_I$ has probability density $r_{I|J}(\cdot | t\mathbf{x}_J)$. As a consequence, for all $t \in (0, \infty)$, $\mathbf{x}_I \in (0, \infty)^I$, $\mathbf{u}_I \in (0, \infty)^I$ and $i \in I$, we have*

$$\begin{aligned} r_{i|J}(tx_i | t\mathbf{x}_J) &= r_{i|J}(x_i | \mathbf{x}_J), & R_{i|J}^{-1}(u_i | t\mathbf{x}_J) &= t \cdot R_{i|J}^{-1}(u_i | \mathbf{x}_J), & \text{and} \\ c_{I,J}(\mathbf{u}_I; t\mathbf{x}_J) &= c_{I,J}(\mathbf{u}_I; \mathbf{x}_J). \end{aligned} \quad (3.2)$$

In particular, if $J = \{j\}$ for some $j \in [d]$, then $c_{I,j}(\cdot; x_j) = c_{I,j}(\cdot; 1)$ does not depend on the value of $x_j \in (0, \infty)$.

Remark 3.3 (Conditional independence). Let $\mathbf{Y}$ be the multivariate Pareto vector in (2.11), suppose $d \geq 3$, and let i, j be distinct elements in $[d]$, with $J = [d] \setminus \{i, j\}$. Provided r is positive and continuous, Y_i and Y_j are conditionally independent given $\mathbf{Y}_J$ in the sense of Definition 1 in Engelke and Hitz (2020) if and only if $c_{i,j,J}(u_i, u_j; \mathbf{x}_J) = 1$ for all $(u_i, u_j) \in (0, 1)^2$ and $\mathbf{x}_J \in (0, \infty)^J$, the density of the independence copula. This statement follows by combining the factorization property in Proposition 1 in Engelke and Hitz (2020) with equations (2.12) and (3.1); see also (D.1) in Appendix D for an expression of $c_{I,J}$ directly in terms of the exponent measure density λ .

Definition 3.4 (Simplifying assumption). The family of copula densities $c_{I,J}(\cdot; \mathbf{x}_J)$ in (3.1) satisfies the *simplifying assumption* if the copula densities do not depend on $\mathbf{x}_J \in (0, \infty)^J$, that is, there exists a single copula density $c_{I,J}$ such that $c_{I,J}(\mathbf{u}_I; \mathbf{x}_J) = c_{I,J}(\mathbf{u}_I)$ for all $\mathbf{u}_I \in (0, 1)^I$ and $\mathbf{x}_J \in (0, \infty)^J$.

If J is a singleton, $J = \{j\}$, the scale invariance (3.2) implies that $\{c_{I,j}(\cdot; x_j) : x_j \in (0, \infty)\}$ always satisfies the simplifying assumption. In Section 4, we will see that the copula families induced by many commonly used parametric exponent measure models satisfy the simplifying assumption for all I and J .

Sklar's theorem for tail copula densities (Proposition 3.1) admits a converse part, stating how to combine several tail copula densities along a scale-invariant family of copula densities to form a tail copula density in a higher dimension. Proposition C.1 in Appendix C gives a precise statement. In the remainder of the paper, we will not use this result as such, although the X-vine construction in Section 5 is in the same spirit.

4 Parametric families of tail copula densities

We will compute the copula densities $c_{I,J}(\cdot; \mathbf{x}_J)$ in (3.1) for tail copula densities r arising from two parametric families: the scaled extremal Dirichlet model (Section 4.1) due to Belzile and Nešlehová (2017), encompassing the logistic, negative logistic and Dirichlet max-stable models, and the Hüsler–Reiss model (Section 4.2). For both families, the simplifying assumption (Definition 3.4) is satisfied for all choices of I and J , and the copula densities $c_{I,J}$ take on known parametric forms. This section can be skipped at first reading.

4.1 Scaled extremal Dirichlet model

In dimension $d \geq 2$, consider parameters $\alpha_1, \dots, \alpha_d > 0$ and $\rho > -\min(\alpha_1, \dots, \alpha_d)$ with $\rho \neq 0$. The angular measure density of the *scaled extremal Dirichlet* model is introduced in Belzile and Nešlehová (2017, Proposition 5). By Eq. (2.12) and Lemma B.1, the corresponding tail copula density is

$$r(\mathbf{x}) = \frac{\Gamma(\bar{\alpha} + \rho)}{d|\rho|^{d-1} \prod_{j=1}^d \Gamma(\alpha_j)} \cdot \left\{ \sum_{j=1}^d c(\alpha_j, \rho)^{1/\rho} x_j^{-1/\rho} \right\}^{-\rho - \bar{\alpha}} \cdot \prod_{j=1}^d c(\alpha_j, \rho)^{\alpha_j/\rho} x_j^{-\alpha_j/\rho - 1}, \quad (4.1)$$

for $\mathbf{x} \in (0, \infty)^d$, where $\bar{\alpha} = \alpha_1 + \dots + \alpha_d$ and $c(\alpha, \rho) = \Gamma(\alpha + \rho)/\Gamma(\alpha)$. The model is closed under marginalisation: for $J \subseteq [d]$ with $|J| \geq 2$, the marginal tail copula density r_J in (2.7) is of the same form in dimension $|J|$ and with parameters $(\alpha_j)_{j \in J}$ and ρ . The model unites and extends several well-known parametric families in multivariate extreme value analysis:

- if $\alpha_1 = \dots = \alpha_d = 1$ and $\rho > 0$, we get the negative logistic model (Joe, 1990);
- if $\alpha_1 = \dots = \alpha_d = 1$ and $-1 < \rho < 0$, we obtain the logistic model (Gumbel, 1960);
- if $\rho = 1$, we get the Coles–Tawn extremal Dirichlet model (Coles and Tawn, 1991).

In McNeil and Nešlehová (2010, Definition 3), the family of *Liouville* copulas is defined as the collection of survival copulas of random vectors of the form $\mathbf{Y} = S\mathbf{W}$, where $\mathbf{W}$ has a Dirichlet distribution on the unit simplex and is independent of the positive random variable S , whose distribution is referred to as the radial distribution. The special case where $\mathbf{W}$ is uniformly distributed on the unit simplex yields the family of Archimedean copulas (McNeil and Nešlehová, 2009).

Proposition 4.1. *Let r be the scaled extremal Dirichlet tail copula density in (4.1) in dimension $d \geq 3$. For every pair of disjoint, non-empty subsets $I, J \subset [d]$ with $|I| \geq 2$, the copula densities $c_{I,J}(\cdot; \mathbf{x}_J)$ in (3.1) satisfy the simplifying assumption (Definition 3.4). If $\rho > 0$, then $c_{I,J}$ is equal to density the $|I|$ -variate Liouville copula with Dirichlet parameters $(\alpha_i)_{i \in I}$ and with radial density on $(0, \infty)$ proportional to*

$$s \mapsto s^{\sum_{i \in I} \alpha_i - 1} \cdot (s + 1)^{-\rho - \sum_{k \in I \cup J} \alpha_k}.$$

If $\rho < 0$, then $c_{I,J}$ is equal to the density of the survival copula of the above Liouville copula.

In special cases, the copula densities $c_{I,J}$ can be identified more explicitly (calculations in Appendix B.3):

- if $\alpha_1 = \dots = \alpha_d = 1$ and $\rho > 0$, then $c_{I,J}$ is the density of the $|I|$ -variate Clayton copula with parameter $\theta/(1 + |J|\theta)$ with $\theta = 1/\rho > 0$;
- if $\alpha_1 = \dots = \alpha_d = 1$ and $-1 < \rho < 0$, then $c_{I,J}$ is the density of the $|I|$ -variate Clayton survival copula with parameter $\theta/(|J|\theta - 1)$ with $\theta = -1/\rho > 1$.

In both cases, the dependence *decreases* as the number $|J|$ of conditioning variables increases.

4.2 Hüsler–Reiss model

Let $\mathcal{D}_d$ be the set of symmetric, strictly conditionally negative definite matrices, that is, all matrices $\Gamma = (\Gamma_{ij})_{i,j=1}^d \in \mathbb{R}^{d \times d}$ of the form $\Gamma_{ij} = \text{Var}(A_i - A_j)$ for a d -variate random vector $\mathbf{A} = (A_1, \dots, A_d)$ with positive definite covariance matrix, i.e., Γ is a variogram matrix. Fix $k \in [d]$ and consider the $(d-1) \times (d-1)$ -dimensional positive definite covariance matrix

$$\Sigma^{(k)} = \frac{1}{2} (\Gamma_{ik} + \Gamma_{jk} - \Gamma_{ij})_{i,j \neq k}; \quad (4.2)$$

for Γ as in the previous sentence, we have $\Sigma_{ij}^{(k)} = \text{Cov}(A_i - A_k, A_j - A_k)$. The d -variate *Hüsler–Reiss* tail copula density is

$$r(\mathbf{x}; \Sigma^{(k)}) = \left(\prod_{i:i \neq k} x_i^{-1} \right) \phi_{d-1}(\bar{\mathbf{x}}_{\setminus k}; \Sigma^{(k)}), \quad \mathbf{x} \in (0, \infty)^d, \quad (4.3)$$

where ϕ_{d-1} is the centered $(d-1)$ -dimensional Gaussian density with the stated covariance matrix and $\bar{\mathbf{x}}_{\setminus k} = (\log(x_i/x_k) - \frac{1}{2}\Gamma_{ik})_{i:i \neq k}$. The exponent measure density λ associated to r via (2.12) appears in Engelke et al. (2015), where it is shown that it does not depend on the choice of $k \in [d]$. The corresponding max-stable distribution G in (2.8) goes back to Hüsler and Reiss (1989), who studied maxima of triangular arrays of Gaussian random vectors.

For a matrix M and for index sets K and L , write $M_{KL} = (M_{ij})_{i \in K, j \in L}$. Let Φ denote the standard normal distribution function and recall that the m -variate Gaussian copula density with $(m \times m)$ -dimensional positive definite correlation matrix $\mathbf{R}$ evaluated in $\mathbf{u} \in (0, 1)^m$ is

$$c_{\mathbf{R}}(\mathbf{u}) = |\mathbf{R}|^{-1/2} \cdot \exp \left\{ -\frac{1}{2} \mathbf{z}^\top (\mathbf{R}^{-1} - I_p) \mathbf{z} \right\} \quad \text{where} \quad z_i = \Phi^{-1}(u_i). \quad (4.4)$$

Proposition 4.2. *Let Γ be a $d \times d$ variogram with $d \geq 3$ and let r be the Hüsler–Reiss tail copula density in (4.3). For every pair of disjoint, non-empty subsets $I, J \subset [d]$ with $|I| \geq 2$, the copula densities $c_{I,J}(\cdot; \mathbf{x}_J)$ in (3.1) satisfy the simplifying assumption (Definition 3.4). The common copula density $c_{I,J}$ is equal to the $|I|$ -variate Gaussian copula density (4.4) with correlation matrix $\mathbf{R}_{I|J} = (\Delta_{I|J}^{(k)})^{-1/2} \Sigma_{I|J}^{(k)} (\Delta_{I|J}^{(k)})^{-1/2}$ for any $k \in J$, where $\Delta_{I|J}^{(k)}$ is the diagonal matrix with the same diagonal as*

$$\Sigma_{I|J}^{(k)} = \Sigma_{II}^{(k)} - \Sigma_{IJ}^{(k)} \left(\Sigma_{JJ}^{(k)} \right)^{-1} \Sigma_{JI}^{(k)}.$$

The expression of the correlation matrix $\mathbf{R}_{I|J}$ in Proposition 4.2 depends on the choice of $k \in J$, although the actual matrix does not. This is because the proof is based on identity (4.3) for the tail copula density. In Hentschel et al. (2022), a formula for the exponent measure density is derived that does not require the choice of such an index k . The formula involves what is called the Hüsler–Reiss precision matrix, denoted by Θ . It remains to be investigated how to express the correlation matrix $\mathbf{R}_{I|J}$ in terms of Θ .

5 Vine decompositions of tail copula densities

In this section, we show that any tail copula density r in dimension $d \geq 3$ can be decomposed along any regular vine sequence (Definition 2.2) on d elements into $d - 1$ bivariate tail copula densities and $\sum_{i=1}^{d-2} i = \binom{d-1}{2}$ bivariate copula densities. Moreover, the copula densities depending on a single conditioning variable satisfy the simplifying assumption (Definition 3.4). Conversely, starting from any regular vine sequence on $d \geq 3$ elements, any collection of $d - 1$ bivariate tail copula densities and any collection of $\binom{d-1}{2}$ bivariate copula densities, a d -variate tail copula density can be assembled. We coin tail copula densities constructed in this way as *X-vines*. By exploiting connection (2.12) between tail copula densities and exponent measure densities, all results carry over to the latter too. For the reader's convenience, the main formulas are spelled out in terms of exponent measure densities in Appendix D.

Theorem 5.1 (Tail copula density decomposition along a regular vine). *Let r be a d -variate ($d \geq 3$) tail copula density and let $\mathcal{V} = (T_j)_{j=1}^{d-1}$ with $T_j = (N_j, E_j)$ be a regular vine sequence on $[d]$. For $\mathbf{x} \in (0, \infty)^d$, we have*

$$r(\mathbf{x}) = \prod_{e \in E_1} r_{a_e, b_e}(x_{a_e}, x_{b_e}) \cdot \prod_{j=2}^{d-1} \prod_{e \in E_j} c_{a_e, b_e; D_e}(R_{a_e|D_e}(x_{a_e} | \mathbf{x}_{D_e}), R_{b_e|D_e}(x_{b_e} | \mathbf{x}_{D_e}); \mathbf{x}_{D_e}), \quad (5.1)$$

where, for $e = (a_e, b_e; D_e) \in E_2 \cup \dots \cup E_{d-1}$, the pair-copula density $c_{a_e, b_e; D_e}$ is

$$c_{a_e, b_e; D_e}(u_{a_e}, u_{b_e}; \mathbf{x}_{D_e}) = \frac{r_{a_e, b_e|D_e}(x_{a_e}, x_{b_e} | \mathbf{x}_{D_e})}{r_{a_e|D_e}(x_{a_e} | \mathbf{x}_{D_e}) \cdot r_{b_e|D_e}(x_{b_e} | \mathbf{x}_{D_e})},$$

with $x_{a_e} = R_{a_e|D_e}^{-1}(u_{a_e} | \mathbf{x}_{D_e})$ and $x_{b_e} = R_{b_e|D_e}^{-1}(u_{b_e} | \mathbf{x}_{D_e})$ for $u_{a_e}, u_{b_e} \in (0, 1)$.

Remark 5.2. The decomposition (5.1) is based on the vine telescoping product formulas (2.13)–(2.14) in Lemma 2.4. By applying the more general formula (2.15) in Remark 2.5, we find, for every edge $f \in E_j$ with $j \in \{1, \dots, d-1\}$, a decomposition of the marginal tail copula density on the complete union $J = A_f \subseteq [d]$, with cardinality $|J| = j + 1$: for $\mathbf{x}_J \in (0, \infty)^J$, we have

$$r_J(\mathbf{x}_J) = \prod_{\substack{e \in E_1 \\ A_e \subseteq J}} r_{a_e, b_e}(x_{a_e}, x_{b_e}) \cdot \prod_{i=2}^{|J|-1} \prod_{\substack{e \in E_i \\ A_e \subseteq J}} c_{a_e, b_e; D_e}(R_{a_e|D_e}(x_{a_e} | \mathbf{x}_{D_e}), R_{b_e|D_e}(x_{b_e} | \mathbf{x}_{D_e}); \mathbf{x}_{D_e}). \quad (5.2)$$

The decomposition (5.2) is true only for subsets $J \subseteq [d]$ of the form $J = A_f$ for some edge f in the regular vine sequence.

For the pair copula $C_{a_e, b_e; D_e}(\cdot, \cdot; \mathbf{x}_{D_e})$ associated to the edge $e = (a_e, b_e; D_e)$ in the regular vine sequence, consider the first-order partial derivatives

$$\begin{aligned} C_{a_e|b_e; D_e}(u_{a_e} | u_{b_e}; \mathbf{x}_{D_e}) &:= \frac{\partial}{\partial u_{b_e}} C_{a_e, b_e; D_e}(u_{a_e}, u_{b_e}; \mathbf{x}_{D_e}) = \int_{v=0}^{u_{a_e}} c_{a_e, b_e; D_e}(v, u_{b_e}; \mathbf{x}_{D_e}) dv, \\ C_{b_e|a_e; D_e}(u_{b_e} | u_{a_e}; \mathbf{x}_{D_e}) &:= \frac{\partial}{\partial u_{a_e}} C_{a_e, b_e; D_e}(u_{a_e}, u_{b_e}; \mathbf{x}_{D_e}) = \int_{v=0}^{u_{b_e}} c_{a_e, b_e; D_e}(u_{a_e}, v; \mathbf{x}_{D_e}) dv. \end{aligned} \quad (5.3)$$

Theorem 5.3 (Recursion and uniqueness). *In the setting of Theorem 5.1, for any $e = (a_e, b_e; D_e) \in E_2 \cup \dots \cup E_{d-1}$, we have*

$$\begin{aligned} R_{a_e|D_e \cup b_e}(x_{a_e} | \mathbf{x}_{D_e \cup b_e}) &= C_{a_e|b_e; D_e}(R_{a_e|D_e}(x_{a_e} | \mathbf{x}_{D_e}) | R_{b_e|D_e}(x_{b_e} | \mathbf{x}_{D_e}); \mathbf{x}_{D_e}), \\ R_{b_e|D_e \cup a_e}(x_{b_e} | \mathbf{x}_{D_e \cup a_e}) &= C_{b_e|a_e; D_e}(R_{b_e|D_e}(x_{b_e} | \mathbf{x}_{D_e}) | R_{a_e|D_e}(x_{a_e} | \mathbf{x}_{D_e}); \mathbf{x}_{D_e}). \end{aligned} \quad (5.4)$$

As a consequence, r is determined uniquely by the bivariate tail copula densities $r_{a,b}$ for $e = \{a, b\} \in E_1$ and the bivariate copula densities $c_{a_e, b_e; D_e}(\cdot; \mathbf{x}_{D_e})$ for $e \in E_2 \cup \dots \cup E_{d-1}$ and $\mathbf{x}_{D_e} \in (0, \infty)^{D_e}$.

Equation (5.4) is effectively a recursive relation, allowing to reduce the number of conditioning variables until there is only one conditioning variable left. The reason is that each of the indices $a_e|D_e$ and $b_e|D_e$ in (5.4) is itself of the form $a_f|D_f \cup b_f$ or $b_f|D_f \cup a_f$ for an edge $f \in e$ in E_{j-1} , i.e., one level lower than e .

Definition 5.4 (X-vine tail copula density). A d -variate tail copula density r is an *X-vine* along a regular vine sequence $\mathcal{V}$ if for each edge $e \in E_2 \cup \dots \cup E_{d-1}$, the pair copula densities $c_{a_e, b_e; D_e}(\cdot, \cdot; \mathbf{x}_{D_e})$ do not depend on the value of $\mathbf{x}_{D_e} \in (0, \infty)^{D_e}$. The d -variate exponent measure density λ in (2.12) is called an X-vine if the tail copula density r is an X-vine.

Example 5.5 (Trivariate case). By Lemma 3.2, a trivariate tail copula density r is always an X-vine, and this along any of the three possible regular vine sequences on $\{1, 2, 3\}$. Indeed, for the vine sequence determined by $E_1 = \{\{1, 2\}, \{2, 3\}\}$, we can for instance write

$$r(x_1, x_2, x_3) = r_{12}(x_1, x_2) r_{23}(x_2, x_3) \cdot c_{13;2}(R_{1|2}(x_1|x_2), R_{3|2}(x_3|x_2)),$$

featuring the copula density

$$c_{13;2}(u_1, u_3) = \frac{r(x_1, 1, x_3)}{r_{12}(x_1, 1) r_{23}(1, x_3)} \quad \text{with} \quad x_j = R_{j|2}^{-1}(u_j|1), \quad j \in \{1, 3\}.$$

The function r is thus completely specified by the two bivariate tail copula densities r_{12} and r_{23} and the bivariate copula density $c_{13;2}$. In Theorem 5.7 below, this construction principle is lifted to higher dimensions.

By Propositions 4.1 and 4.2, the scaled extremal Dirichlet model (including the logistic, negative logistic and extremal Dirichlet models) and the Hüsler–Reiss family have conditional copula densities $c_{I,J}(\cdot; \mathbf{x}_J)$ that always satisfy the simplifying assumption. As a consequence, they are examples of X-vine tail copula densities too, and this along any regular vine sequence.

Definition 5.6 (X-vine specification). The triplet $(\mathcal{V}, \mathcal{R}, \mathcal{C})$ is an *X-vine specification* on d elements ($d \geq 3$) if:

1. $\mathcal{V} = (T_j)_{j=1}^{d-1}$ with $T_j = (N_j, E_j)$ is a regular vine sequence on $[d]$;
2. $\mathcal{R} = \{r_{a_e, b_e} : e = \{a_e, b_e\} \in E_1\}$ is a family of bivariate tail copula densities;
3. $\mathcal{C} = \{c_{a_e, b_e; D_e} : e = (a_e, b_e; D_e) \in \bigcup_{j \geq 2} E_j\}$ is a family of bivariate copula densities.

Fig. 2 in Section 8 shows an example of an X-vine specification involving a mix of parametric models for the bivariate (tail) copula densities.

Theorem 5.7 (X-vine tail copula density construction). *Let $(\mathcal{V}, \mathcal{R}, \mathcal{C})$ be an X-vine specification on $d \geq 3$ elements. Then the function r defined by*

$$r(\mathbf{x}) = \prod_{e \in E_1} r_{a_e, b_e}(x_{a_e}, x_{b_e}) \cdot \prod_{j=2}^{d-1} \prod_{e \in E_j} c_{a_e, b_e; D_e}(R_{a_e|D_e}(x_{a_e} | \mathbf{x}_{D_e}), R_{b_e|D_e}(x_{b_e} | \mathbf{x}_{D_e})) \quad (5.5)$$

with the functions $R_{\cdot|\cdot}$ defined recursively by

$$\begin{aligned} R_{a_e|D_e \cup b_e}(x_{a_e} | \mathbf{x}_{D_e \cup b_e}) &= C_{a_e|b_e; D_e}(R_{a_e|D_e}(x_{a_e} | \mathbf{x}_{D_e}) | R_{b_e|D_e}(x_{b_e} | \mathbf{x}_{D_e})) \\ R_{b_e|D_e \cup a_e}(x_{b_e} | \mathbf{x}_{D_e \cup a_e}) &= C_{b_e|a_e; D_e}(R_{b_e|D_e}(x_{b_e} | \mathbf{x}_{D_e}) | R_{a_e|D_e}(x_{a_e} | \mathbf{x}_{D_e})) \end{aligned} \quad (5.6)$$

is a d -variate tail copula density and its multivariate margins are the ones implied by the notation. In particular, r is an X-vine.

Remark 5.8 (Link with regular vine copulas). X-vine tail copula densities arise as the lower tail dependence limits of regular vine copula densities provided the pair copula densities at the edges of the first tree have the corresponding bivariate tail copula densities as lower tail dependence limits. In the passage to the limit, the regular vine sequence is preserved and so are the pair copulas at all trees starting from the second one. Max-stable dependence structures based on X-vines thus arise naturally from the max-domain of attraction condition of regular vine copulas, and the X-vine limit is particularly easy to calculate. These properties are worked out in Appendix E.

If the bivariate copula density $c_{a_e, b_e; D_e}$ associated with an edge $e \in E_2 \cup \dots \cup E_{d-1}$ is equal to the independence one, $c_{a_e, b_e; D_e} \equiv 1$, the corresponding factor drops out in (5.1) and the recursive formulas in (5.4) and (5.6) simplify to

$$\begin{aligned} R_{a_e|D_e \cup b_e}(x_{a_e} | \mathbf{x}_{D_e \cup b_e}) &= R_{a_e|D_e}(x_{a_e} | \mathbf{x}_{D_e}), \\ R_{b_e|D_e \cup a_e}(x_{b_e} | \mathbf{x}_{D_e \cup a_e}) &= R_{b_e|D_e}(x_{b_e} | \mathbf{x}_{D_e}). \end{aligned}$$

Sparse X-vine specifications arise when $c_{a_e, b_e; D_e} \equiv 1$ for many edges $e \in E_2 \cup \dots \cup E_{d-1}$. Model selection of sparse vine copulas in high dimensions has been investigated in Müller and Czado (2019) and Nagler et al. (2019). The case where all pair copulas are equal to the independence one for all edges starting from a given tree is of practical importance (Brechmann et al., 2012; Brechmann and Joe, 2015).

Definition 5.9. A *truncated regular vine tree sequence* $\mathcal{V} = (T_1, \dots, T_q)$ on $d \geq 3$ elements with truncation level $q \in \{1, \dots, d-2\}$ is an ordered set of trees for which there exists a regular vine tree sequence $\mathcal{V}' = (T'_1, \dots, T'_{d-1})$ on d elements such that $T_j = T'_j$ for all $j \in \{1, \dots, q\}$.

Definition 5.10. The triplet $(\mathcal{V}, \mathcal{R}, \mathcal{C})$ is a *truncated X-vine specification* on d elements ($d \geq 3$) with truncation level $q \in \{1, \dots, d-2\}$ if:

1. $\mathcal{V} = (T_j)_{j=1}^q$ is a truncated regular vine tree sequence on $[d]$;
2. $\mathcal{R} = \{r_{a,b} : e = \{a, b\} \in E_1\}$ is a family of bivariate tail copula densities;
3. $\mathcal{C} = \{c_{a_e, b_e; D_e} : e = (a_e, b_e; D_e) \in \bigcup_{j=2}^q E_j\}$ is a family of bivariate copula densities; in case $q = 1$, we have $\mathcal{C} = \emptyset$.

By definition, any truncated X-vine specification can be completed to a full X-vine specification by completing the truncated regular vine tree sequence $\mathcal{V}$ to a full one $\mathcal{V}'$ as in Definition 5.9 and by setting $c_{a_e, b_e; D_e} \equiv 1$ for all $e \in E'_{q+1} \cup \dots \cup E'_{d-1}$. The resulting X-vine copula density r does not depend on the way in which the truncated regular vine tree sequence $\mathcal{V}$ is completed, since the factorisation in (5.1) simplifies anyway to

$$r(\mathbf{x}) = \prod_{e \in E_1} r_{a_e, b_e}(x_{a_e}, x_{b_e}) \cdot \prod_{j=2}^q \prod_{e \in E_j} c_{a_e, b_e; D_e}(R_{a_e|D_e}(x_{a_e} | \mathbf{x}_{D_e}), R_{b_e|D_e}(x_{b_e} | \mathbf{x}_{D_e})),$$

with an empty second product in case $q = 1$. A truncated X-vine specification involves $d-1$ bivariate tail copula densities and $\sum_{j=2}^q (d-j)$ bivariate copula densities, yielding $qd - q(q+1)/2$ bivariate components in total. The truncation level q allows to tune the trade-off between sparsity and flexibility. If $q = 1$, the model is a Markov tree as in Engelke and Hitz (2020) and Segers (2020). Increasing q and adding trees ('layers') yield more complex dependence structures. Truncated X-vine specifications will be shown at work in the case study in Section 9.

6 Sampling from X-vine Pareto distributions

6.1 Inverted multivariate Pareto distributions

In Section 2, we introduced the multivariate Pareto distribution (2.11) as a limit model for high threshold excesses in (2.10). In the context of tail copula measures, it is more convenient to work with their reciprocals.

Definition 6.1. The d -variate random vector $\mathbf{Z}$ has an *inverted multivariate Pareto distribution* if there exists a d -variate tail copula measure R such that

$$\mathbb{P}(\mathbf{Z} \in B) = \frac{R(B \cap \mathbb{L})}{R(\mathbb{L})},$$

for Borel sets B , with $\mathbb{L} = \{\mathbf{x} \in (0, \infty]^d : \min \mathbf{x} < 1\}$; equivalently, if there exists a multivariate Pareto random vector $\mathbf{Y}$ such that $\mathbf{Z} = 1/\mathbf{Y}$.

If R is concentrated on $(0, \infty)^d$ and has density r , then $\mathbf{Z}$ has probability density $\mathbf{x} \mapsto r(\mathbf{x}) \mathbb{1}(\mathbf{x} \in \mathbb{L})/R(\mathbb{L})$. In general, for $j \in [d]$, the conditional distribution of Z_j given $Z_j < 1$ is uniform on $(0, 1)$; this is a consequence of the marginal constraint (2.2). More generally, for Borel sets B , we have

$$\mathbb{P}(\mathbf{Z} \in B \mid Z_j < 1) = R(B \cap \mathbb{L}^{(j)}), \quad (6.1)$$

where the set $\mathbb{L}^{(j)} = \{\mathbf{x} \in (0, \infty)^d : x_j < 1\}$ has R -measure one. Hence, on $\mathbb{L}^{(j)}$, the probability density $\mathbf{x} \mapsto r(\mathbf{x}) \mathbb{1}\{x_j < 1\}$ of $(\mathbf{Z} \mid Z_j < 1)$ coincides with the tail copula density r .

For the random vector $\mathbf{U}$ in Definition 2.1, the inverted multivariate Pareto vector $\mathbf{Z}$ is the weak limit in

$$(t\mathbf{U} \mid \min \mathbf{U} < 1/t) \rightsquigarrow \mathbf{Z}, \quad t \rightarrow \infty. \quad (6.2)$$

For non-empty $J \subset [d]$, we have

$$(t\mathbf{U}_J \mid \min \mathbf{U}_J < 1/t) \rightsquigarrow \mathbf{Z}_{|J} \stackrel{d}{=} (\mathbf{Z}_J \mid \min \mathbf{Z}_J < 1), \quad t \rightarrow \infty.$$

The conditional marginal $\mathbf{Z}_{|J}$ has a $|J|$ -variate inverted multivariate Pareto distribution associated with the marginal tail copula measure R_J in (2.6). In contrast, the random vector $\mathbf{Z}_J$ does not necessarily have an inverted multivariate Pareto distribution, as $\min \mathbf{Z}_J < 1$ is not guaranteed.

6.2 Random sampling from (inverted) multivariate Pareto distributions

Through equations (2.10) and (6.2), the (inverted) multivariate Pareto distribution serves as a model for a random vector conditionally on the event that at least one variable takes a value far in the tail of its respective marginal distribution. The L-shaped support of the (inverted) multivariate Pareto distribution makes direct random sampling from it a little awkward. Lemma 2 in Engelke and Hitz (2020) provides an ingenious algorithm that reduces the task of simulating from a multivariate Pareto random vector $\mathbf{Y}$ to random sampling from the conditional distributions $(\mathbf{Y} \mid Y_j > 1)$ for every $j \in [d]$. Below, we will study the equivalent problem of simulating from the conditional distribution of $(\mathbf{Z} \mid Z_j < 1)$ in case the tail copula density is an X-vine as in Definition 5.4.

We will do so by inverting the Rosenblatt transformation (Rosenblatt, 1952), applying conditional quantile functions with an increasing number of conditioning variables successively to independent uniform random variables. A judicious choice of the ordering of the variables permits to compute the required conditional quantile functions recursively in terms of the bivariate ingredients of the X-vine specification. This ordering is encoded in the permutation constructed in the next lemma.

Lemma 6.2. *Let $\mathcal{V}$ be a regular vine sequence on d elements. For every $j \in [d]$, there exists a permutation σ_j of $\{1, \dots, d\}$ such that (i) $\sigma_j(1) = j$ and (ii) there exist edges $e_{j,1} \in E_1, \dots, e_{j,d-1} \in E_{d-1}$ such that $\sigma_j(k) \in \mathcal{C}_{e_{j,k-1}}$ and $\{\sigma_j(i) : i = 1, \dots, k\} = A_{e_{j,k-1}}$ for each $k \in \{2, \dots, d\}$.*

Let r be an X-vine tail copula density. For each edge $e \in E_2 \cup \dots \cup E_{d-1}$ and for fixed $0 < u_{b_e} < 1$, let $u_{a_e} \mapsto C_{a_e|b_e;D_e}^{-1}(u_{a_e}|u_{b_e})$ be the inverse of the distribution function $u_{a_e} \mapsto C_{a_e|b_e;D_e}(u_{a_e}|u_{b_e})$, with $C_{a_e|b_e;D_e}$ defined in (5.3). Similarly for $C_{b_e|a_e;D_e}^{-1}$. Recall from Proposition 3.1 that for $i \in [d]$, non-empty $J \subseteq [d] \setminus \{i\}$, and $\mathbf{x}_J \in (0, \infty)^d$ such that $r_J(\mathbf{x}_J) > 0$, the quantile function $u_i \mapsto R_{i|J}^{-1}(u_i|\mathbf{x}_J)$ is the inverse of the distribution function $x_i \mapsto R_{i|J}(x_i|\mathbf{x}_J)$. Inverting (5.4) yields the recursive relations

$$\begin{aligned} R_{a_e|D_e \cup b_e}^{-1}(u_{a_e}|\mathbf{x}_{D_e \cup b_e}) &= R_{a_e|D_e}^{-1}\left(C_{a_e|b_e;D_e}^{-1}(u_{a_e}|R_{b_e|D_e}(x_{b_e}|\mathbf{x}_{D_e})) \mid \mathbf{x}_{D_e}\right), \\ R_{b_e|D_e \cup a_e}^{-1}(u_{b_e}|\mathbf{x}_{D_e \cup a_e}) &= R_{b_e|D_e}^{-1}\left(C_{b_e|a_e;D_e}^{-1}(u_{b_e}|R_{a_e|D_e}(x_{a_e}|\mathbf{x}_{D_e})) \mid \mathbf{x}_{D_e}\right). \end{aligned} \quad (6.3)$$

Let $W_1, \dots, W_d$ be independent random variables, all uniformly distributed on $(0, 1)$. For $j \in [d]$, let σ_j be a permutation of $[d]$ satisfying the two requirements in Lemma 6.2. Define a random vector $\mathbf{Z}^{(j)} = (Z_1^{(j)}, \dots, Z_d^{(j)})$ recursively as follows:

$$\begin{aligned} Z_j^{(j)} &= W_j, \\ Z_{\sigma_j(k)}^{(j)} &= R_{\sigma_j(k)|\sigma_j(\{1, \dots, k-1\})}^{-1}\left(W_{\sigma_j(k)} \mid \mathbf{Z}_{\sigma_j(\{1, \dots, k-1\})}^{(j)}\right), \quad k \in \{2, \dots, d\}. \end{aligned} \quad (6.4)$$

Proposition 6.3. *Let $\mathbf{Z}$ be an inverted multivariate Pareto random vector associated with the X-vine tail copula density r . For $j \in [d]$, the distribution of $\mathbf{Z}^{(j)}$ in (6.4) is equal to the one of $\mathbf{Z}$ conditionally on $Z_j < 1$. For every $k \in \{2, \dots, d\}$, the conditional quantile function $R_{\sigma_j(k)|\sigma_j(\{1, \dots, k-1\})}^{-1}$ is of the form $R_{a_e|D_e \cup b_e}^{-1}$ or $R_{b_e|D_e \cup a_e}^{-1}$ for some edge $e = e_{j,k-1} \in E_{k-1}$ and can thus be computed recursively via (6.3).*

The actual simulation algorithm is based on structure matrices encoding regular vine sequences as explained in Appendix A. In Section 8, we will apply the simulation technique based on Proposition 6.3 to assess the estimation methods we will introduce in Section 7 through Monte Carlo experiments.

7 Estimation and model selection

Let $\mathbf{X}_i = (X_{i,1}, \dots, X_{i,d})$ for $i \in \{1, \dots, n\}$ be an independent random sample from a distribution function F with continuous but unspecified margins $F_1, \dots, F_d$ and whose survival copula C has lower tail copula R (Definition 2.1). Suppose that the tail copula density r is an X-vine with specification $(\mathcal{V}, \mathcal{R}, \mathcal{C})$ (Definitions 5.4 and 5.6 and Theorem 5.7). We propose a procedure to estimate r from the excesses over a high multivariate threshold.

The regular vine sequence $\mathcal{V} = (T_j)_{j=1}^{d-1}$ with trees $T_j = (N_j, E_j)$ may be known or not. The bivariate tail copulas $r_{a,b}$ for edges $e = \{a, b\} \in E_1$ in the first tree and the bivariate copula densities $c_{a_e, b_e; D_e}$ for edges $e = (a_e, b_e; D_e) \in E_j$ in trees $j \geq 2$ are assumed to belong to prespecified (lists of) parametric families. The X-vine specification may be truncated (Definition 5.10), leading to a simpler model.

The basis of the method is a link between the conditional copula densities $c_{I,J}$ in Sklar's theorem (Proposition 3.1) on the one hand and the inverted multivariate Pareto distribution (Definition 6.1) on the other hand (Section 7.1). In Section 7.2, we propose parameter estimates, supposing that $\mathcal{V}$ is given and that parametric families of (tail) copula densities have been specified for all edges. In Section 7.3, finally, we treat model selection, which comprises the selection of the parametric families of the bivariate model components, the selection of the regular vine sequence $\mathcal{V}$, and the selection of the truncation level q .

7.1 Copulas and inverted multivariate Pareto distributions

Let r be a d -variate copula tail copula density, not necessarily an X-vine. In Sklar's theorem (Proposition 3.1), suppose that the copula density $c_{I,J}$ satisfies the simplifying assumption (Definition 3.4). The following proposition shows how to transform an inverted multivariate Pareto random vector $\mathbf{Z}$ associated to r into a random vector with density $c_{I,J}$.

Proposition 7.1. *Let r be a d -variate tail copula density and let $\mathbf{Z}$ be an inverted multivariate Pareto random vector associated to r . Let $I, J \subset [d]$ be non-empty and disjoint.*

- (i) *For $\mathbf{z}_J \in (0, \infty)^J$ such that $\min \mathbf{z}_J < 1$ and $r_J(\mathbf{z}_J) > 0$, the conditional density of $\mathbf{Z}_I$ given $\mathbf{Z}_J = \mathbf{z}_J$ is $r_{I|J}(\cdot | \mathbf{z}_J)$.*
- (ii) *Suppose $|I| \geq 2$. If $c_{I,J}(\cdot; \cdot)$ satisfies the simplifying assumption (Definition 3.4), then, conditionally on the event $\min \mathbf{Z}_J < 1$, the random vector $(R_{i|J}(Z_i | \mathbf{Z}_J))_{i \in I}$ is independent of $\mathbf{Z}_J$ and its density is $c_{I,J}$.*

By statement (ii), the density of $(R_{i|J}(Z_i | \mathbf{Z}_J))_{i \in I}$ given $\mathbf{Z}_J \in A$ is equal to $c_{I,J}$ for any non-empty set $A \subseteq \{\mathbf{z} \in (0, \infty)^J : \min \mathbf{z} < 1\}$. For estimation, we will use this property for $A = (0, 1)^J$, requiring in effect that *all* variables (rather than *at least one*) in J exceed a high threshold. We do so in order to avoid a potential bias stemming from including too many non-extreme values in the procedure. An alternative would be to opt for a censored likelihood approach (Ledford and Tawn, 1996), but this seems difficult to implement beyond dimensions two or three.

7.2 Sequential maximum likelihood estimation of X-vines

Let the d -variate tail copula density r be an X-vine specified by $(\mathcal{V}, \mathcal{R}, \mathcal{C})$ as in Theorem 5.7. Assume that the bivariate tail copula densities $(\mathcal{R})$ and bivariate copula densities $(\mathcal{C})$ belong to specified parametric families. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_{\mathcal{R}}, \boldsymbol{\theta}_{\mathcal{C}})$ denote the parameter vector associated with r . Here, $\boldsymbol{\theta}_{\mathcal{R}} = (\theta_e)_{e \in E_1}$ contains the parameters (or parameter vectors) $\theta_e \in \Theta_e$ associated with each pairwise tail copula density r_{a_e, b_e} for $e \in E_1$, while $\boldsymbol{\theta}_{\mathcal{C}} = (\boldsymbol{\theta}_{\mathcal{C},j})_{j=2}^{d-1}$ for $\boldsymbol{\theta}_{\mathcal{C},j} = (\theta_e)_{e \in E_j}$ denotes the vector of parameters (or parameter vectors) $\theta_e \in \Theta_e$ associated with each bivariate copula density $c_{a_e, b_e; D_e}$ for $e \in \bigcup_{j=2}^{d-1} E_j$. While it is possible to derive the full-likelihood of an X-vine (inverted) multivariate Pareto distribution, performing parameter estimation with the full model in high dimensions becomes challenging in extremes. Instead, using the X-vine decomposition into *bivariate* components and recursively defined quantities (Theorem 5.7), we outline a *sequential* procedure for parameter estimation, starting from T_1 to T_{d-1} tree by tree. This approach is inspired by the one for regular vine copulas (see, e.g., Czado, 2019), but with suitable adaptations to the extreme value context.

(1) Standardising the margins and selecting sub-samples

Recall from the beginning of the section that $\mathbf{X}_1, \dots, \mathbf{X}_n$, with $\mathbf{X}_i = (X_{i,1}, \dots, X_{i,d})$, is an independent random sample from the unknown d -variate distribution F , whose survival copula C is assumed to have tail copula R with tail copula density r . Set

$$\widehat{U}_{i,j} = 1 - \widehat{F}_j(X_{i,j}), \quad i = 1, \dots, n, \quad j \in [d], \quad (7.1)$$

where $\hat{F}_j$ denotes any estimator of the marginal distribution function $F_j(x) = \mathbb{P}(X_{i,j} \leq x)$. One possibility is the empirical distribution function, and to avoid boundary effects, we set

$$\hat{U}_{i,j} = 1 - (\text{rk}_{i,j} - 0.5)/n, \quad (7.2)$$

where $\text{rk}_{i,j} = \sum_{s=1}^n \mathbb{1}(X_{s,j} \leq X_{i,j})$ is the (maximal) rank of $X_{i,j}$ among $X_{1,j}, \dots, X_{n,j}$. We view the points $\hat{U}_i = (\hat{U}_{i,1}, \dots, \hat{U}_{i,d})$ as pseudo-observations from the survival copula C .

By Eq. (6.2), for large $t > 0$, the rescaled points $t\hat{U}_i$ for $i \in \{1, \dots, n\}$ such that $\min \hat{U}_i < 1/t$ constitute pseudo-observations from a distribution that approximates the inverted multivariate Pareto distribution associated with r . We set $t = n/k$ where $k \in \{1, \dots, n\}$ is such that k is ‘large’ and k/n is ‘small’; in the usual large-sample setting in extreme value theory, we have $k = k_n$ with $k \rightarrow \infty$ and $k/n \rightarrow 0$. For $j \in [d]$, let

$$K_j = \{i = 1, \dots, n : \hat{U}_{i,j} < k/n\} \quad (7.3)$$

be the set of indices i corresponding to observations whose the j th component $X_{i,j}$ is ‘large’; for the rank-based definition of $\hat{U}_{i,j}$ in the previous paragraph, this set of indices corresponds to the k largest observations for the j th variable. Further, put

$$K_J = \bigcap_{j \in J} K_j \quad \text{and} \quad K = \bigcup_{j \in [d]} K_j,$$

which are, respectively the set of indices i corresponding to large observations in *all* variables in $J \subseteq [d]$ *simultaneously* and the set of indices with large observations in *at least one* variable. Write $\hat{\mathbf{Z}}_i = (\hat{Z}_{i,j})_{j \in [d]}$ where $\hat{Z}_{i,j} = (n/k) \hat{U}_{i,j}$. In view of Eq. (6.2), we treat $\{\hat{\mathbf{Z}}_i\}_{i \in K}$ as a sample of $|K|$ pseudo-observations of the inverted multivariate Pareto distribution associated with r . The sample size $|K|$ is random, and in the rank-based definition above, we have $k \leq |K| \leq dk$. For non-empty $J \subseteq [d]$ and $i \in K$, we write $\hat{\mathbf{Z}}_{i,J} = (\hat{Z}_{i,j})_{j \in J}$.

(2) Estimating the tail copula parameters $\theta_{\mathcal{R}}$ in T_1

For each edge $e = \{a_e, b_e\} \in E_1$, we estimate the parameter (vector) θ_e associated with the bivariate tail copula density $r_{a_e, b_e}(\cdot; \theta_e)$. By Eq. (6.1), for $j \in \{a_e, b_e\}$, the conditional density of (Z_{a_e}, Z_{b_e}) given $Z_j < 1$ is $r_{a_e, b_e}(z_{a_e}, z_{b_e}; \theta_e) \mathbb{1}(z_j < 1)$. As one out of many possibilities, we use maximum pseudo-likelihood estimation to fit this density to the sub-samples $(\hat{Z}_{i, a_e}, \hat{Z}_{i, b_e})$ for both $i \in K_{a_e}$ and $i \in K_{b_e}$, with K_j as in Eq. (7.3). More precisely, we maximise each of the two pseudo-likelihoods

$$\mathcal{L}_{\mathcal{R}, e}(\theta_e^{(j)}; (\hat{Z}_{i, a_e}, \hat{Z}_{i, b_e}), i \in K_j) = \prod_{i \in K_j} r_{a_e, b_e}(\hat{Z}_{i, a_e}, \hat{Z}_{i, b_e}; \theta_e^{(j)}), \quad j \in \{a_e, b_e\}, \quad (7.4)$$

over $\theta_e^{(j)} \in \Theta_e$, yielding estimates $\hat{\theta}_e^{(a_e)}$ and $\hat{\theta}_e^{(b_e)}$, respectively. The final estimate is $\hat{\theta}_{a_e, b_e} = \{\hat{\theta}_e^{(a_e)} + \hat{\theta}_e^{(b_e)}\}/2$. The idea of averaging maximum pseudo-likelihood estimators on product spaces has already been suggested for the Hüsler–Reiss model in (Engelke et al., 2015), and implemented for multivariate Pareto distributions in the `graphicalExtremes` package (Engelke et al., 2022). Other approaches for parameter estimation can be considered: censored likelihoods (Ledford and Tawn, 1996; de Haan et al., 2008) or the empirical stable tail dependence functions (Einmahl et al., 2008, 2018).

(3) Estimating the copula parameters θ_C in $T_2, \dots, T_{d-1}$

Start with an edge $e = (a_e, b_e; D_e) \in E_2$, i.e., such that D_e is a singleton, $D_e = \{d_e\}$, say. Then $f = \{a_e, d_e\}$ and $g = \{b_e, d_e\}$ are edges in E_1 . By Proposition 7.1(ii), the density of the random pair

$$(R_{a_e|d_e}(Z_{a_e}|Z_{d_e}; \theta_f), R_{b_e|d_e}(Z_{b_e}|Z_{d_e}; \theta_g))$$

is $c_{a_e, b_e; D_e}(\cdot; \theta_e)$. The tail copula density parameters θ_f and θ_g have been estimated in the previous step, yielding estimates $\hat{\theta}_f$ and $\hat{\theta}_g$. Define pseudo-observations $(\hat{U}_{i, a_e; D_e}, \hat{U}_{i, b_e; D_e})$ for $i \in K_{D_e} = K_{d_e}$ from $c_{a_e, b_e; D_e}$ by

$$\hat{U}_{i, a_e; D_e} = R_{a_e|D_e}(\hat{Z}_{i, a_e} | \hat{Z}_{i, d_e}; \hat{\theta}_f), \quad \hat{U}_{i, b_e; D_e} = R_{b_e|D_e}(\hat{Z}_{i, b_e} | \hat{Z}_{i, d_e}; \hat{\theta}_g).$$

We estimate the parameter (vector) $\theta_e \in \Theta_e$ by maximising the pseudo-likelihood

$$\mathcal{L}_{C, e}(\theta_e; (\hat{U}_{i, a_e; D_e}, \hat{U}_{i, b_e; D_e}), i \in K_{D_e}) = \prod_{i \in K_{D_e}} c_{a_e, b_e; D_e}(\hat{U}_{i, a_e; D_e}, \hat{U}_{i, b_e; D_e}; \theta_e). \quad (7.5)$$

Other estimators are possible based on parametric relationships between the dependence coefficients such as Kendall's tau and Spearman's rho in relation to the parameters of corresponding bivariate copula densities (Czado, 2019).

Parameters (or parameter vectors) associated with edges $e = (a_e, b_e; D_e)$ in E_j for layers $j \in \{3, \dots, d-1\}$ can be estimated similarly, proceeding recursively. Suppose that all parameters associated with edges in $E_{1:(j-1)} = E_1 \cup \dots \cup E_{j-1}$ have already been estimated; let $\theta(E_{1:(j-1)})$ denote the parameters and $\hat{\theta}(E_{1:(j-1)})$ their estimate. By Proposition 7.1(i), the density of the random pair

$$(R_{a_e|D_e}(Z_{a_e} | \mathbf{Z}_{D_e}; \theta(E_{1:j-1})), R_{b_e|D_e}(Z_{b_e} | \mathbf{Z}_{D_e}; \theta(E_{1:j-1})))$$

is $c_{a_e, b_e; D_e}$. Define pseudo-observations $(\hat{U}_{i, a_e; D_e}, \hat{U}_{i, b_e; D_e})$ for $i \in K_{D_e}$ from $c_{a_e, b_e; D_e}$ by

$$\hat{U}_{i, a_e; D_e} = R_{a_e|D_e}(\hat{Z}_{i, a_e} | \hat{\mathbf{Z}}_{i, D_e}; \hat{\theta}(E_{1:j-1})), \quad \hat{U}_{i, b_e; D_e} = R_{b_e|D_e}(\hat{Z}_{i, b_e} | \hat{\mathbf{Z}}_{i, D_e}; \hat{\theta}(E_{1:j-1})). \quad (7.6)$$

In this definition, the conditional distribution functions $R_{a_e|D_e}$ and $R_{b_e|D_e}$ are calculated recursively via Eq. (5.6) with e in that formula replaced by the edges $f, g \in E_{j-1}$ that are joined by e , that is, such that $e = \{f, g\}$. The parameter (vector) θ_e is then obtained by maximising the pseudo-likelihood in (7.5).

7.3 Model selection for X-vines

Model selection for X-vines based on a random sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ from F as in Section 7.2 involves two steps:

- (1) selecting a regular vine sequence $\mathcal{V} = (T_1, \dots, T_{d-1})$;
- (2) given the regular vine sequence obtained in (1), choosing adequate bivariate parametric (tail) copula families, $\mathcal{R}$ and $\mathcal{C}$.

In fact, the two procedures are executed together sequentially, progressing from one tree to the next tree. First, we describe step (2) given the regular vine sequence $\mathcal{V}$.

Selecting parametric (tail) copula families given a truncated regular vine sequence

We first consider the specification of $\mathcal{R}$ in Definition 5.6 in tree T_1 . Let $\mathcal{B}_{\mathcal{R}, 1:T} = (\mathcal{B}_{\mathcal{R}, 1}, \dots, \mathcal{B}_{\mathcal{R}, T})$ be the list of candidate bivariate tail copula families. For each edge $e = (e_a, e_b) \in E_1$, specifying $\mathcal{R}$ involves choosing the bivariate tail copula family among $\mathcal{B}_{\mathcal{R}, 1:T}$. Similar to the idea of averaged maximum pseudo-likelihood estimates in Section 7.2, we use the averaged-AIC value for selecting bivariate tail copula families. For each edge e and each $\mathcal{B}_{\mathcal{R}, t}$, $t = 1, \dots, T$, we obtain two maximum pseudo-likelihood estimates $\hat{\theta}_e^{(t, a_e)}$ and $\hat{\theta}_e^{(t, b_e)}$, derived through the maximisation of log-likelihood functions on product spaces; see (7.4). The averaged-AIC value is

$$\text{AIC}_e(\mathcal{B}_{\mathcal{R}, t}) = 2\nu^{(t)} - \frac{1}{2} \left\{ \log \mathcal{L}_{\mathcal{R}, e}(\hat{\theta}_e^{(t, a_e)}) + \log \mathcal{L}_{\mathcal{R}, e}(\hat{\theta}_e^{(t, b_e)}) \right\},$$

where $\nu^{(t)}$ is the number of parameters in $\mathcal{B}_{\mathcal{R}, t}$. We select the bivariate tail copula family with the lowest averaged-AIC among $\text{AIC}_e(\mathcal{B}_{\mathcal{R}, 1}), \dots, \text{AIC}_e(\mathcal{B}_{\mathcal{R}, T})$.

Similarly, we specify $\mathcal{C}$ in Definition 5.6 in T_j for $j \in \{2, \dots, d-1\}$. Let $\mathcal{B}_{\mathcal{C}, 1:S} = (\mathcal{B}_{\mathcal{C}, 1}, \dots, \mathcal{B}_{\mathcal{C}, S})$ denote the list of candidate bivariate parametric copula families, including the independence copula (without parameters). Specifying $\mathcal{C}$ involves selecting the bivariate parametric copula family among $\mathcal{B}_{\mathcal{C}, 1:S}$ for each edge $e = (e_a, e_b; D_e) \in \bigcup_{j=2}^{d-1} E_j$. We do so by maximising the AIC, following the common practice in vine copula modelling; see Brechmann (2010) and (Czado, 2019, Section 8.1). For each edge e , we estimate the associated parameter (vector) $\theta_e^{(s)}$ by maximising the log-likelihood (7.5), and we calculate the corresponding AIC over the list of the bivariate parametric copula families $\mathcal{B}_{\mathcal{C}, s}$, $s = 1, \dots, S$,

$$\text{AIC}_e(\mathcal{B}_{\mathcal{C}, s}) = 2\nu^{(s)} - 2 \log \mathcal{L}_{\mathcal{C}, e}(\hat{\theta}_e^{(s)}),$$

where $\nu^{(s)}$ is the number of parameters in $\mathcal{B}_{\mathcal{C}, s}$, and $\nu^{(s)} = 0$ if $\mathcal{B}_{\mathcal{C}, s}$ contains only the independence copula. We then choose the bivariate parametric copula family with the lowest value among $\text{AIC}_e(\mathcal{B}_{\mathcal{C}, 1}), \dots, \text{AIC}_e(\mathcal{B}_{\mathcal{C}, S})$.

It is worthwhile to note that when implementing the model selection step for an edge e in E_j , only the trees $T_1, \dots, T_j$ need to have been selected, but not the trees $T_{j+1}, \dots, T_{d-1}$. This sequential approach aligns well with the vine learning procedure in the next paragraph.

Selecting the regular vine sequence

Morales-Napoles (2010) showed that the number of different regular vine sequences is equal to $d! 2^{\binom{d-2}{2}-1}$, making it impossible to go through all possible vine structures. We adopt the model selection approach similar to the one in Dissmann et al. (2013), choosing trees sequentially from T_1 to T_{d-1} .

To select the first tree, $T_1 = (N_1, E_1)$, on the node set $N_1 = [d]$, we follow a procedure as in Engelke and Volgushev (2022) and Hu et al. (2023). For every pair $\{a, b\}$ of distinct elements in $[d]$, let $w_{a,b}$ be a nonnegative weight derived from the data. We will be using an empirical version of the tail dependence coefficient $\chi_{a,b} = R_{a,b}(1, 1)$, namely we will set $w_{a,b}$ equal to

$$\hat{\chi}_{a,b} = \frac{1}{k} \sum_{i=1}^n \mathbb{1} \left\{ \hat{U}_{i,a} \leq k/n, \hat{U}_{i,b} \leq k/n \right\},$$

where $\hat{U}_{i,a}$ and $\hat{U}_{i,b}$ are defined as in (7.2). Another possible edge weight would be the empirical extremal variogram as in Engelke and Volgushev (2022). The first tree is then selected as the maximum-spanning tree based on the edge weights $w_{a,b}$, that is,

$$T_1 = \underset{\text{spanning tree } T=(N_1, E)}{\arg \max} \sum_{e=\{a,b\} \in E} w_{a,b}.$$

For each edge e in T_1 , we then select parametric tail copula densities and estimate their parameters as previously explained.

Next we describe, for $j \in \{2, \dots, d-1\}$, how to select the tree $T_j = (N_j, E_j)$ given the already selected trees $T_1, \dots, T_{j-1}$ and the parametric families and parameter estimates for edges $e \in E_1 \cup \dots \cup E_{j-1}$. The edge set of T_j is necessarily $N_j = E_{j-1}$. Let $E_{j,P}$ be the set of pairs of elements of N_j that satisfy the proximity condition (Definition 2.2); that is, every element e of $E_{j,P}$ is a pair $\{f, g\}$ of distinct elements in N_j so that $f \cap g$ is a singleton. For such $e \in E_{j,P}$, we can then consider the conditioning set D_e and the conditioned set $\mathcal{C}_e = \{a_e, b_e\}$ as before and, by recursion, compute $\hat{U}_{i,a_e;D_e}$ and $\hat{U}_{i,b_e;D_e}$ for $i \in N_{D_e}$ as in (7.6). Based on these, we compute an edge weight w_e ; we will take the absolute value of the empirical Kendall's tau, $|\hat{\tau}_e|$. We then select T_j as the maximum spanning tree on the node set N_j with edges restricted to $E_{j,P}$, that is,

$$T_j = \underset{\text{spanning tree } T=(N_j, E) \text{ with } E \subseteq E_{j,P}}{\arg \max} \sum_{e \in E} w_e.$$

For each edge $e \in E_j$, we further select the parametric copula families and estimate their parameters as before.

Selecting truncated regular vine tree sequences

The model selection procedure described above may set all pair-copulas to the independence copula in the subsequent trees from T_{q+1} to T_{d-1} for a truncation level $q \in \{1, \dots, d-2\}$. In this case, the resulting model corresponds to the truncated X-vine model in Definition 5.10. Besides an information criterion, we consider two additional criteria for selecting the independence copula at an edge e : when the effective sample size $|K_{D_e}| = n_{D_e}$ falls below a certain low value, and when the absolute value of the empirical Kendall's tau is close to zero. Sparsity, induced by a smaller effective sample size, is more likely when the total sample size n is relatively smaller than the dimension d ; recall that we condition on $(j-1)$ variables being extreme in tree T_j as explained in Section 7.2.

Even when the above criteria are not met, when d is large, it is natural to limit the number of model parameters by considering truncated X-vines, since Dissmann's algorithm captures as much dependence as possible in the first few trees. We use a modified Bayesian information (mBIC) to determine the truncation level, inspired by the one for regular vine copulas in Nagler et al. (2019). This modified version adjusts the prior probability in the BIC to penalise dependence copulas more severely in trees at higher levels. More specifically, assuming that for any edge $e \in E_j$ for $j \geq 2$, the parametric family has a single parameter θ_e and that a value of $\theta_e = 0$ corresponds to the independence copula, the mBIC includes independent Bernoulli variables $\mathbb{1}(\theta_e \neq 0)$ with mean $\psi_e = \psi_0^{j-1}$ for $e \in E_j$ and a hyperparameter $\psi_0 \in (0, 1)$: set $\text{mBIC}(1) = 0$ and, for $q \in \{2, \dots, d-1\}$, put

$$\text{mBIC}(q) = \sum_{j=2}^q \sum_{e \in E_j} \left[\mathbb{1}(\theta_e \neq 0) \left\{ \log n_{D_e} - 2 \log \left(\frac{\psi_0^{j-1}}{1 - \psi_0^{j-1}} \right) \right\} - 2 \log \mathcal{L}_{\mathcal{C},e}(\hat{\theta}_e) - 2 \log (1 - \psi_0^{j-1}) \right], \quad (7.7)$$

with $\mathcal{L}_{\mathcal{C},e}$ as in (7.5). The selected truncation level q^* is the value of q in $\{1, \dots, d-1\}$ that minimises $\text{mBIC}(q)$. In practice, we will set $\psi_0 = 0.9$ as in Nagler et al. (2019).

8 Simulation study

We conduct three simulation studies, evaluating the proposed procedures for parameter estimation, selection of bivariate parametric (tail) copula families, and vine truncation.

8.1 Parameter estimation

We first consider the 5-dimensional X-vine model r in Fig. 2. The bivariate tail copula densities r_{a_e, b_e} in the first tree $T_1 = (N_1, E_1)$ are chosen from the Hüsler–Reiss model, the negative logistic, logistic, and Dirichlet families (Section 4), while the bivariate copula densities $c_{a_e, b_e; D_e}$ in the subsequent trees $T_j = (N_j, E_j)$ for $j \in \{2, 3, 4\}$ are taken from the Clayton, Gumbel, and Gaussian copula families. We specify the parameter values for the bivariate parametric families as follows, including the tail dependence coefficient $\chi = R(1, 1)$ or Kendall’s τ for comparison: $\theta_{12} = 1.5$ ($\chi = 0.54$), $\theta_{23} = 2$ ($\chi = 0.71$), $\theta_{24} = 2.5$ ($\chi = 0.68$), $\theta_{45} = 2$ ($\chi = 0.62$); $\theta_{13;2} = 2$ ($\tau = 0.50$), $\theta_{34;2} = 2.5$ ($\tau = 0.60$), $\theta_{25;4} = 0.7$ ($\tau = 0.49$); $\theta_{14;23} = 0.4$ ($\tau = 0.17$), $\theta_{35;24} = -0.3$ ($\tau = -0.19$); and $\theta_{15;234} = 0.1$ ($\tau = 0.06$). The formulas connecting the parameters of the families of tail copula densities with the value of χ are given in Appendix F.2.

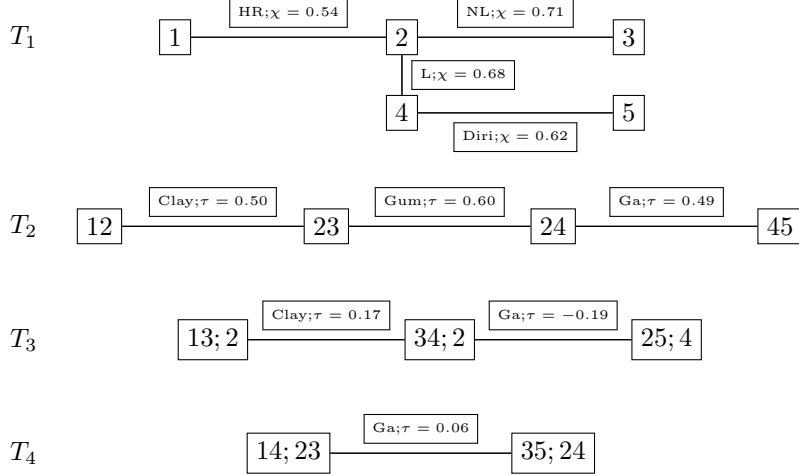


Figure 2: An X-vine specification $(\mathcal{V}, \mathcal{R}, \mathcal{C})$ on $d = 5$ variables. The bivariate tail copula densities r_{a_e, b_e} in the first tree T_1 are selected from the Hüsler–Reiss (HR), negative logistic (NL), logistic (L), and Dirichlet (Diri) parametric families, each with specified tail dependence coefficients χ . The bivariate copula densities $c_{a_e, b_e; D_e}$ in the subsequent trees T_2, T_3, T_4 are chosen from the Clayton (Clay), Gumbel (Gum), and Gaussian (Ga) copula families, each with specified Kendall’s tau τ .

Implementing simulation algorithms as detailed in Section 6 and Appendix A, we generate multivariate inverted Pareto random samples $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ associated with r in Section 6. Rather than using these directly, we perform a rank transformation $\hat{\mathbf{U}}_1, \dots, \hat{\mathbf{U}}_n$ as defined in (7.1) with $\hat{F}_j(Z_{i,j}) = (\text{rnk}_{i,j} - 0.5)/n$, where $\text{rnk}_{i,j}$ denotes the rank of $Z_{i,j}$ among $Z_{1,j}, \dots, Z_{n,j}$ for $j \in [d]$. We then take sub-samples $\{\hat{\mathbf{Z}}_i\}_{i \in K}$ with $K = \bigcup_{j \in [d]} K_j$ and K_j as in (7.3) to investigate simulation results with respect to threshold exceedances.

We set $(n, k) = (4000, 200)$ and perform sequential parameter estimation as in Section 7.2 over 200 repetitions. We obtain maximum pseudo-likelihood estimates and the model-based dependence measures, $\hat{\chi}$ and $\hat{\tau}$, using the parametric relations in Appendix F.2 for each of the ten edges in the vine. In Fig. 3a, box-plots present dependence measure estimates from specified parametric families based on the X-vine specification in Fig. 2. The four left-most box-plots (■) show the tail dependence coefficient estimates $\hat{\chi}$ for the four edges $e \in E_1$, and the remaining plots display Kendall’s tau estimates $\hat{\tau}$ for the six edges $e \in E_2 \cup E_3 \cup E_4$. Appendix F.2 includes additional box-plots: first, of dependence measures with varying sample sizes and threshold exceedances, and second, of maximum likelihood estimates of tail copula densities for a specific edge $e \in E_1$ with varying (n, k) .

8.2 Selecting parametric (tail) copula families

We assess algorithm effectiveness in selecting bivariate parametric families for each edge in each tree, using the same X-vine specification as in Fig. 2. While the flexibility of X-vine models allows us to consider any bivariate parametric family, we simplify the process by considering a catalogue of four candidate tail copula models for T_1 : the Hüsler–Reiss, negative logistic, logistic model, and Dirichlet model, along with a catalogue of nine candidate pair-copula families for T_2, T_3, T_4 : Independence, Gaussian, Clayton, Survival Clayton, Gumbel, Survival Gumbel, Frank, Joe, and Survival Joe copulas, as implemented in the VineCopula software package (Nagler et al., 2023).

Using again 200 repetitions with $(n, k) = (4000, 200)$, we assess the accuracy of the (averaged) AIC in Section 7.3 in selecting bivariate parametric families. The overall proportion of correctly selected families across all tree levels is 60%. For individual trees, the proportions are as follows: 62% for T_1 , 78% for T_2 ,

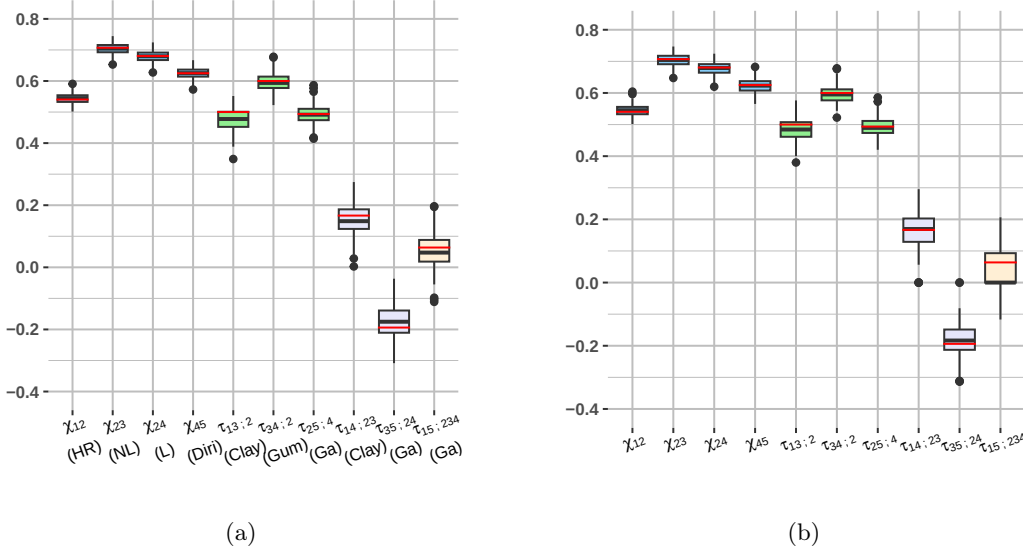


Figure 3: (a) Box-plots (■ for T_1 ■ for T_2 ■ for T_3 ■ for T_4) of dependence measure estimates from specified bivariate parametric families for each edge across the four trees in the X-vine specification in Fig. 2. (b) Similar as the plot on the left, but now bivariate parametric families are selected sequentially from the data, yielding larger estimation uncertainty. The red lines indicate the specified parameter values.

55% for T_3 , and 10% for T_4 . The specific proportions for each edge in each tree are presented in Table 1a. The lower proportions observed for e_{23} and e_{45} in E_1 result from the challenge in distinguishing between the negative logistic and logistic models. Overall, the proportion of accurately selected families decreases with declining effective sample size and Kendall’s tau estimate across tree levels. We observe the lowest proportion for the deepest edge $e_{15;234}$ where the corresponding pair-copula exhibits weak dependence; the algorithm selects the independence copula for this edge 110 times out of 200. Additionally, box-plots of dependence measure estimates from sequentially selected bivariate parametric families over 200 repetitions are presented in Fig. 3b.

We also investigate how effective sample sizes change over tree levels shown in Table 1b. We calculate averaged-effective sample size over 200 repetitions for each edge in each tree and express it as a percentage of the sample size n . The overall averaged-effective sample size decreases in the first level tree when considering stronger extremal structures (result is omitted). This reduction is due to the fact that, for a high threshold u , the probability $P[X_a > u \text{ or } X_b > u]$ decreases as the tail dependence between X_a and X_b increases. For the second level tree T_2 , the averaged-effective sample is $k/n = 5\%$, since the use of the rank transform implies (in the absence of ties) the identity $K_j = k$ for all $j \in [d]$.

T_1	99%	40%	60%	49%	T_1	7.30%	6.47%	6.60%	6.87%
T_2		58%	93%	82%	T_2		5.00%	5.00%	5.00%
T_3			42%	68%	T_3			3.52%	3.38%
T_4				11%	T_4				3.12%

(a)
(b)

Table 1: (a) Proportion of correctly selected families for each edge in each tree. (b) Averaged-effective sample sizes as percentages for the same edges. The order of the components in Table 1, from left to right and then from top to bottom, corresponds to the x-axis order of box-plots in Fig. 3.

8.3 Tree selection and truncation

We consider a higher dimensional X-vine model where residual dependence weakens with increasing tree level. Specifically, for $\mathcal{V}$ we consider a 50-dimensional C-vine, that is, in each tree T_j , there is a single node $a_j \in [d]$ such that a_j belongs to the conditioned set $\mathcal{C}_e$ of all edges $e \in E_j$. The first tree includes the Hüsler–Reiss models and negative logistic models with randomly assigned parameter values $\theta_e \in [1, 2]$ for $e \in E_1$. Subsequent trees contain bivariate Gaussian copulas with partial correlations $\rho_e = 1.1 - 0.1j$ for $e \in E_j$ and $j \in \{2, \dots, 9\}$, and $\rho_e = 0.1$ for $e \in E_j$ with $j \geq 10$. This X-vine specification allows us to explore truncated X-vine models by setting pair-copulas with weak dependence to independence copulas.

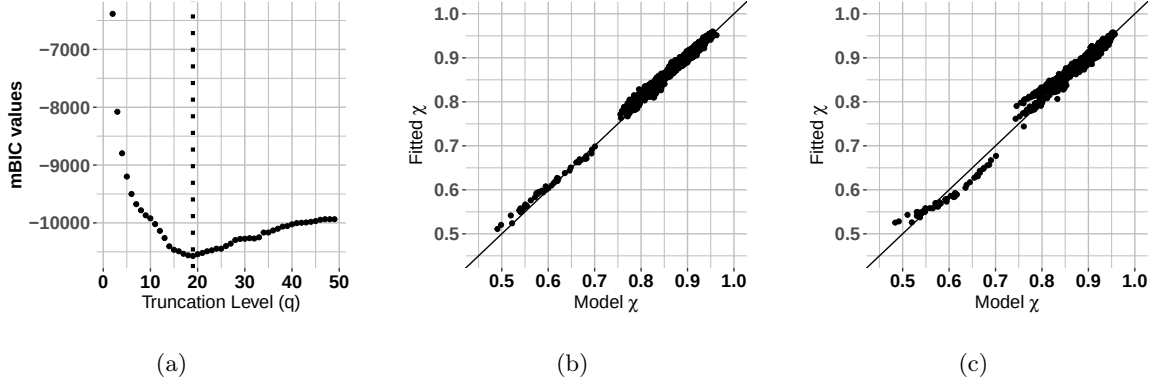


Figure 4: (a) The mBIC plotted across tree levels with a dotted line indicating the selected mBIC-optimal truncation level of $q^* = 19$. (b) χ -plot comparing pairwise χ values from the specified X-vine model to those obtained from the fitted X-vine model via Monte Carlo simulations. (c) Similar to (b), a χ -plot for the truncated X-vine model.

We directly use inverted multivariate Pareto samples of size $n = 1000$ from the X-vine model for a later comparison of model-based tail dependence coefficients. Using the procedures explained in Sections 7.2 and 7.3, we sequentially select trees $T_1, \dots, T_{d-1}$ using $\hat{\chi}_e$ for $e \in E_1$ and $\hat{\tau}_e$ for $e \in E_j, j \geq 2$, as edge weights. For each selected tree, we choose the bivariate parametric families with the lowest (averaged) AIC for each edge and estimate the associated parameters. Additionally, the independence copulas in subsequent trees are chosen if either $|\hat{\tau}_e| < 0.05$ or $n_{D_e} < 10$ for each edge.

To explore truncated X-vine models, we use the mBIC with $\psi_0 = 0.9$ as in Nagler et al. (2019). For a specific iteration, we select the mBIC-optimal truncation level of $q^* = 19$, indicated as a dotted line in Fig. 4a. We assess the goodness of fit by comparing pairwise χ -values from the specified 50-dimensional X-vine model with those obtained from the fitted truncated X-vine model via Monte Carlo simulations in Fig. 4b (for bivariate tail copula densities, closed-form expressions of χ are available for most well-known parametric tail copula density families; see Appendix F). In the case of the truncated X-vine model, the chi-plot in Fig. 4c resembles that of the full model but exhibits more variability, particularly for lower χ values.

9 Application: US flight delay data

We apply X-vine models to investigate extremal dependence among large flight delays in the US flight network. The raw data-set is accessible through the US Bureau of Transportation Statistics¹. These flight delay data were analysed by Hentschel et al. (2022) who first selected airports with a minimum of 1000 flights per year and applied a k -medoids clustering algorithm to identify homogeneous clusters in terms of extremal dependence. This clustering approach not only makes the analysis suitable for modelling extremal dependence but also reveals shared frequent flight connections between airports and similar geographical characteristics in each cluster. Focusing on the Hüsler–Reiss family, Hentschel et al. (2022) fitted an extremal graphical model to large flight delays of airports in the Texas cluster to investigate conditional independence.

For the purpose of model comparison, we also analyse daily total delays (in minutes) in the Texas cluster. The cluster comprises $d = 29$ airports and counts $n = 3603$ days from 2005 to 2020, during which all airports have recorded measurements. This pre-processed data is available through the R-package `graphicalExtremes` (Engelke et al., 2022). A graphical representation of the actual flight connections between airports is presented in Fig. 5a. While the graph derived from flight connections does not guarantee the optimal graph in terms of extremal behavior, it offers valuable insights into the flight network.

Let $\hat{\mathbf{x}}_i, i = 1, \dots, n$, denote the observed measurements. Following the standardisation process with the rank transformation as in Section 7.2, we take sub-samples $\hat{\mathbf{z}}_i = (n/k)\hat{\mathbf{u}}_i$ for $i \in K$ where k/n is the threshold of the uniform margin $\hat{u}_{i,j}$. We choose k such that $k/n = 0.13$ in order to have a large enough effective sample size in relation to the number of variables.

Assigning $\hat{\chi}_e$ for $e \in E_1$ and $\hat{\tau}_e$ for $e \in E_j, j \geq 2$, as edge weights, we select the vine tree structure sequentially as in Section 7.3. The first maximum spanning tree is obtained with a range of estimated tail dependence coefficients $\hat{\chi}_e$ between 0.42 and 0.65 and a mean of 0.55 in Fig. 9 in Appendix F.3. Out of the $d - 1 = 28$ edges in T_1 , selected bivariate tail copula families include the Hüsler–Reiss (8 edges), negative logistic (3 edges) and Dirichlet models (17 edges). Subsequent trees with a total of 378 edges are then

¹<https://www.bts.dot.gov>

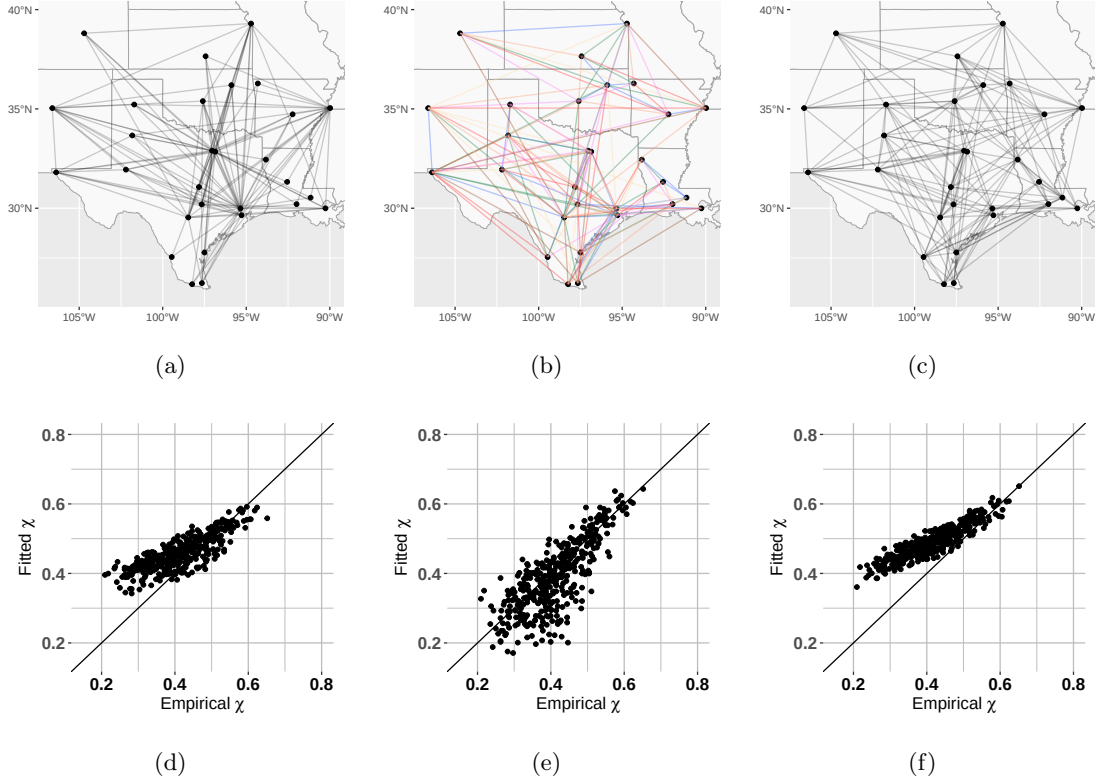


Figure 5: (a) Flight graph illustrating connection flights between airports joined by edges in the Texas cluster. (b) The X-vine graph with the first seven trees superimposed, showing for trees T_2 to T_7 only the 113 out of 147 edges for which a copula other than the independence copula is selected, for 141 edges in total. (c) The estimated extremal Hüsler–Reiss graph structure using the regularised method with a tuning parameter value of $\rho^* = 0.1$, totalling 148 edges. Corresponding χ -plots comparing empirical pairwise tail dependence coefficients with those from the fitted graphical models: (d) the flight graph, (e) the truncated X-vine graph, and (f) the extremal Hüsler–Reiss graph using the EGlearn algorithm.

chosen with the following pair-copula families (the number of edges inside parentheses): Independence (201), Gaussian (31), Clayton (8), Gumbel (26), Frank (34), Joe (36), Survival Clayton (31), Survival Gumbel (9), and Survival Joe (2). As the tree level rises, we observe a decline in residual dependence and an increase in the number of independence copulas. The X-vine algorithm sets all pair copulas to the independence copula from tree level 21, that is, the vine is truncated at level $q = 20$.

To evaluate the X-vine model’s efficacy in capturing extremal dependence for large flight delays between airports in the Texas cluster, we estimate bivariate and trivariate tail dependence coefficients. We compare empirical pairwise tail dependence coefficients $\hat{\chi}_{a,b}$ for $a, b \in [d]$ with those from the fitted X-vine model in Fig. 10a. Additionally, Fig. 10b in Appendix F.3 compares trivariate empirical tail dependence coefficients $\hat{\chi}_{abc}$ for $a, b, c \in [d]$ with those obtained from the fitted X-vine model. Both assessments indicate a satisfactory fit for the large flight delays data.

To further explore truncated X-vine models with a lower truncation level, we use the mBIC in Eq. (7.7). Fig. 11a illustrates the mBIC-optimal truncation level of $q^* = 7$, which corresponds to 113 dependence copulas [Gaussian (17), Clayton (3), Gumbel (23), Frank (16), Joe (25), Survival Clayton (22), Survival Gumbel (6), and Survival Joe (1)] out of the 147 total edges from T_2 to T_7 . The total number of bivariate model components is thus $28 + 113 = 141$. Additionally, we investigate how the truncation level changes over the threshold range of $k/n \in (0.05, 0.15)$ in Fig. 11b in Appendix F.3. It appears that the mBIC-optimal truncation levels are not overly sensitive to the choice of the threshold. The superimposed graph of the first seven vine trees is shown in Fig. 5b, showing for T_2 to T_7 only the 113 out of 147 edges for which a copula other than the independence one is selected. The level by level vine tree sequence from T_1 to T_7 is presented in Fig. 9 in Appendix F.3.

Returning attention to model comparison, we consider extremal graphical models with the Hüsler–Reiss distribution, another sparse model for extremes, detailed in Engelke et al. (2021) and Hentschel et al. (2022). Hentschel et al. (2022) obtained the sparse Hüsler–Reiss graphical model through a structure learning method, using the EGlearn algorithm of Engelke et al. (2021) for the sequential estimation of extremal graphical structure from data. The tuning parameter $\rho \geq 0$ controls sparsity, and the empirical extremal variogram matrix $\hat{\Gamma}$ is used as edge weight for the *minimum* spanning tree with smaller elements in $\hat{\Gamma}$ indicating stronger extremal dependence.

Following the vignette in Engelke et al. (2022), the data set is split in half: a training set for model fitting and a test set for tuning parameter selection. We determine the optimal tuning parameter, $\rho^* = 0.1$, by evaluating the Hüsler–Reiss log-likelihood values across tuning parameter values ρ from the test set. The resulting sparse extremal graph with 148 edges at $\rho^* = 0.1$ is shown in Fig. 5c.

We assess the goodness of fit using the entire sub-samples and compare empirical tail dependence coefficients to those obtained from the fitted graphical models: the Hüsler–Reiss extremal graphical model for the flight graph in Fig. 5d, the truncated X-vine model with $q^* = 7$ in Fig. 5e, and the extremal Hüsler–Reiss graphical model with $\rho^* = 0.1$ in Fig. 5f, respectively.

The flexibility of a regular vine structure allows the selection of various (tail) copula families with the lowest AIC. Recall from Section 4.2 that the Hüsler–Reiss model arising from the X-vine specification is constructed using bivariate Hüsler–Reiss tail copula densities in the first tree and bivariate Gaussian copulas in subsequent trees. In the truncated X-vine model, 8 out of 28 edges prefer Hüsler–Reiss models in T_1 , while in trees T_j for $j \geq 2$, 96 out of 113 edges favour copula families other than the Gaussian one. However, uncertainty in sequential parameter estimation tends to accumulate across tree levels. Consequently, the χ -plot of the truncated X-vine model shows a better fit but more variability. In contrast, the χ -plot of the extremal Hüsler–Reiss graphical model has less variability but seems to be biased towards higher tail dependence.

Similarly, focusing exclusively on the Hüsler–Reiss model, Fig. 12 presents a model comparison between the extremal Hüsler–Reiss graphical model based on the flight graph, the Hüsler–Reiss model derived from the X-vine structure, and the extremal Hüsler–Reiss graphical model from the EGlearn algorithm. Additional details about the fitted models and model comparison are provided in Appendix F.3.

10 Discussion

We have opened the door to the construction and the use of extremal dependence models based on regular vine tree sequences. For (ordinary) copulas, the methodology has been extensively developed in the literature since its inception more than two decades ago. Our contribution is to deliver the theoretical and methodological advances required to apply vine machinery in the extreme value context too. The key consists of a version of Sklar’s theorem applied to tail copula densities, which stand in one-to-one relation to exponent measure densities, together with a telescoping product formula for regular vines. For many popular parametric families of tail dependence models, the resulting conditional copula densities satisfy a simplifying assumption. Tail copula densities with this property, called X-vines, are specified by a triple consisting of a regular vine sequence, a vector of bivariate tail copula densities and a vector of bivariate copula densities. These bivariate

building blocks can be chosen independently from another, yielding great flexibility. At the same time, the recursive properties of regular vines ensure that all computations can be reduced to operations involving the bivariate building blocks only. Further, we have studied the inverted multivariate Pareto distributions associated to these X-vines, as dependence models for excesses over high multivariate thresholds. A recursive algorithm serves to sample from such distributions, and we have developed a statistical methodology on them, in order to learn an X-vine from data. The method comprises parameter estimation, model selection, and vine structure learning, all proceeding sequentially, tree by tree. In large dimensions, truncating the vine sequence produces models with fewer parameters. We have illustrated the performance of the various methods on simulated and real data.

While the proposed methodology is fully operational, it is clearly open to improvement, while many open questions remain, just as for ordinary regular vine copulas. We name just a few. The vine structure learning method inspired by Dissmann et al. (2013) aims to capture dependence by the first several trees in the sequence, but there is no guarantee that it retrieves the true structure. The regular vine atlas of Morales-Nápoles et al. (2023) can serve as a test bed for evaluating vine learning approaches. While the parameter estimators were shown to perform well in simulations, their large-sample theory remains to be developed, in line with Hobæk Haff (2013). For the Hüsler–Reiss model, the precise connection between the variogram matrix and the correlation parameters of the bivariate Gaussian copulas remains to be elucidated and the effect of truncation on the Hüsler–Reiss precision matrix (Hentschel et al., 2022) to be uncovered. The relation between X-vines and graphical models for extremes as in Engelke and Hitz (2020) deserves further investigation, perhaps leveraging results of Zhu and Kurowicka (2022). Finally, as our approach is limited to dependence structures generated from max-stable distributions, a completely open question is whether it can be extended to other settings in multivariate extreme value analysis, such as the conditional extremal dependence model (Heffernan and Tawn, 2004) or the geometric sample-cloud approach (Nolde, 2014; Simpson et al., 2021).

Acknowledgments

This work is supported in part by funds from the *Fonds de la Recherche Scientifique – FNRS*, Belgium (grant number T.0203.21).

References

- Bedford, T. and R. M. Cooke (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial intelligence* 32, 245–268.
- Bedford, T. and R. M. Cooke (2002). Vines—a new graphical model for dependent random variables. *The Annals of Statistics* 30(4), 1031–1068.
- Beirlant, J., Y. Goegebeur, J. Segers, and J. L. Teugels (2004). *Statistics of Extremes: Theory and Applications*, Volume 558. John Wiley & Sons.
- Belzile, L. R. and J. G. Nešlehová (2017). Extremal attractors of Liouville copulas. *Journal of Multivariate Analysis* 160, 68–92.
- Brechmann, E. (2010). Truncated and simplified regular vines and their applications. Master’s thesis, Technische Universität München.
- Brechmann, E. C., C. Czado, and K. Aas (2012). Truncated regular vines in high dimensions with application to financial data. *Canadian journal of statistics* 40(1), 68–85.
- Brechmann, E. C. and H. Joe (2015). Truncation of vine copulas using fit indices. *Journal of Multivariate Analysis* 138, 19–33.
- Coles, S. G. and J. A. Tawn (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B (Methodological)* 53(2), 377–392.
- Czado, C. (2019). *Analyzing dependent data with vine copulas*, Volume 222 of *Lecture Notes in Statistics*. Springer, Cham. A practical guide with R.
- Czado, C. and T. Nagler (2022). Vine copula based modeling. *Annual Review of Statistics and Its Application* 9, 453–477.
- de Haan, L. and A. Ferreira (2007). *Extreme Value Theory: An Introduction*. Springer Science & Business Media.
- de Haan, L., C. Neves, and L. Peng (2008). Parametric tail copula estimation and model testing. *Journal of Multivariate Analysis* 99(6), 1260–1275.
- de Haan, L. and S. I. Resnick (1977). Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 40(4), 317–337.
- Dissmann, J., E. C. Brechmann, C. Czado, and D. Kurowicka (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis* 59, 52–69.

- Dombry, C., S. Engelke, and M. Oesting (2016). Asymptotic properties of the maximum likelihood estimator for multivariate extreme value distributions. *arXiv preprint arXiv:1612.05178*.
- Einmahl, J. H., A. Kiriliouk, and J. Segers (2018). A continuous updating weighted least squares estimator of tail dependence in high dimensions. *Extremes* 21, 205–233.
- Einmahl, J. H., A. Krajina, and J. Segers (2008). A method of moments estimator of tail dependence. *Bernoulli* 14(4), 1003–1026.
- Einmahl, J. H. J., V. I. Piterbarg, and L. de Haan (2001). Nonparametric estimation of the spectral measure of an extreme value distribution. *The Annals of Statistics* 29(5), 1401–1423.
- Engelke, S., A. Hitz, N. Gnecco, and M. Hentschel (2022). graphicalExtremes: Statistical methodology for graphical extreme value models. R package version 0.2.0.
- Engelke, S. and A. S. Hitz (2020). Graphical models for extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82(4), 871–932.
- Engelke, S. and J. Ivanovs (2021). Sparse structures for multivariate extremes. *Annual Review of Statistics and Its Application* 8, 241–270.
- Engelke, S., J. Ivanovs, and K. Strokorb (2022). Graphical models for infinite measures with applications to extremes and Lévy processes. *arXiv preprint arXiv:2211.15769*.
- Engelke, S., M. Lalancette, and S. Volgushev (2021). Learning extremal graphical structures in high dimensions. *arXiv preprint arXiv:2111.00840*.
- Engelke, S., A. Malinowski, Z. Kabluchko, and M. Schlather (2015). Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 77(1), 239–265.
- Engelke, S. and S. Volgushev (2022). Structure learning for extremal tree models. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84(5), 2055–2087.
- Gissibl, N. and C. Klüppelberg (2018). Max-linear models on directed acyclic graphs. *Bernoulli* 24(4A), 2693–2720.
- Gumbel, E. J. (1960). Distributions des valeurs extrêmes en plusieurs dimensions. *Publ. Inst. Statist. Univ. Paris* 9, 171–173.
- Heffernan, J. E. and J. A. Tawn (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society Series B: Statistical Methodology* 66(3), 497–546.
- Hentschel, M., S. Engelke, and J. Segers (2022). Statistical inference for Hüsler–Reiss graphical models through matrix completions. *arXiv preprint arXiv:2210.14292*.
- Hobæk Haff, I. (2013). Parameter estimation for pair-copula constructions. *Bernoulli* 19(2), 462–491.
- Hu, S., Z. Peng, and J. Segers (2023). Modelling multivariate extreme value distributions via markov trees. *Scandinavian Journal of Statistics* n/a(n/a), <https://doi.org/10.1111/sjos.12698>.
- Hüsler, J. and R.-D. Reiss (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statistics & Probability Letters* 7(4), 283–286.
- Joe, H. (1990). Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statistics & Probability Letters* 9(1), 75–81.
- Joe, H. (1996). Families of m -variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In *Distributions with fixed marginals and related topics (Seattle, WA, 1993)*, Volume 28 of *IMS Lecture Notes Monogr. Ser.*, pp. 120–141. Inst. Math. Statist., Hayward, CA.
- Joe, H., H. Li, and A. K. Nikoloulopoulos (2010). Tail dependence functions and vine copulas. *Journal of Multivariate Analysis* 101(1), 252–270.
- Ledford, A. W. and J. A. Tawn (1996). Statistics for near independence in multivariate extreme values. *Biometrika* 83(1), 169–187.
- McNeil, A. J. and J. Nešlehová (2010). From Archimedean to Liouville copulas. *Journal of Multivariate Analysis* 101(8), 1772–1790.
- McNeil, A. J. and J. Nešlehová (2009). Multivariate Archimedean copulas, d -monotone functions and ℓ_1 -norm symmetric distributions. *The Annals of Statistics* 37(5B), 3059–3097.
- Morales-Napoles, O. (2010). Counting vines. In *Dependence modeling: Vine copula handbook*, pp. 189–218. World Scientific.
- Morales-Nápoles, O., M. Rajabi-Bahaabadi, G. A. Torres-Alves, and C. M. P. ’t Hart (2023). Chimera: An atlas of regular vines on up to 8 nodes. *Scientific Data* 10(1), 337.
- Müller, D. and C. Czado (2019). Selection of sparse vine copulas in high dimensions with the lasso. *Statistics and Computing* 29(2), 269–287.
- Nagler, T., C. Bumann, and C. Czado (2019). Model selection in sparse high-dimensional vine copula models with an application to portfolio risk. *Journal of Multivariate Analysis* 172, 180–192.
- Nagler, T., U. Schepsmeier, J. Stoeber, E. C. Brechmann, B. Graeler, and T. Erhardt (2023). *VineCopula: Statistical Inference of Vine Copulas*. R package version 2.4.5.
- Nagler, T. and T. Vatter (2023). *rvinecopulib: High Performance Algorithms for Vine Copula Modeling*. R package version 0.6.3.1.1.

- Nolde, N. (2014). Geometric interpretation of the residual dependence coefficient. *Journal of Multivariate Analysis* 123, 85–95.
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Resnick, S. I. (2007). *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Science & Business Media.
- Rootzén, H. and N. Tajvidi (2006). Multivariate generalized Pareto distributions. *Bernoulli* 12(5), 917–930.
- Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics* 23(3), 470–472.
- Schmidt, R. and U. Stadtmüller (2006). Non-parametric estimation of tail dependence. *Scandinavian Journal of Statistics* 33(2), 307–335.
- Segers, J. (2020). One-versus multi-component regular variation and extremes of Markov trees. *Advances in Applied Probability* 52(3), 855–878.
- Simpson, E. S., J. L. Wadsworth, and J. A. Tawn (2021). A geometric investigation into the tail dependence of vine copulas. *Journal of Multivariate Analysis* 184, 104736.
- Zhu, K. and D. Kurowicka (2022). Regular vines with strongly chordal pattern of (conditional) independence. *Computational Statistics & Data Analysis* 172, 107461.

SUPPLEMENTARY MATERIAL

In Appendix A, we provide detailed illustrations of a number of concepts, results and methods for a particular five-dimensional regular vine sequence. Appendix B contains the proofs of all results in the paper. A converse to Sklar's theorem in Proposition 3.1 is stated and proved in Appendix C. For readers more familiar with exponent measure densities than with tail copula densities, a number of theoretical results are reformulated in terms of exponent measure densities in Appendix D. In Appendix E, it is shown that X-vine models are the tail limits of regular vine copula models in the sense of Definition 2.1. Finally, Appendix F.2 contains additional numerical results, both for the Monte Carlo simulations in Section 8 and the US flight data case study in Section 9.

A Example in dimension five

Consider the regular vine sequence $\mathcal{V} = (T_1, \dots, T_4)$ on $d = 5$ nodes specified by the following trees $T_j = (N_j, E_j)$ for $j = 1, \dots, 4$:

$$\begin{aligned} T_1 : N_1 &= \{1, 2, 3, 4, 5\} \\ E_1 &= \{\{1, 2\}, \{2, 3\}, \{2, 4\}, \{4, 5\}\} \\ T_2 : N_2 &= E_1 = \{\{1, 2\}, \{2, 3\}, \{2, 4\}, \{4, 5\}\} \\ E_2 &= \{\{\{1, 2\}, \{2, 3\}\}, \{\{2, 3\}, \{2, 4\}\}, \{\{2, 4\}, \{4, 5\}\}\} \\ T_3 : N_3 &= E_2 = \{\{\{\{1, 2\}, \{2, 3\}\}, \{\{2, 3\}, \{2, 4\}\}\}, \{\{2, 4\}, \{4, 5\}\}\} \\ E_3 &= \{\{\{\{\{1, 2\}, \{2, 3\}\}, \{\{2, 3\}, \{2, 4\}\}\}, \{\{\{2, 3\}, \{2, 4\}\}, \{\{2, 4\}, \{4, 5\}\}\}\} \\ T_4 : N_4 &= E_3 = \{\{\{\{\{\{1, 2\}, \{2, 3\}\}, \{\{2, 3\}, \{2, 4\}\}\}, \{\{\{2, 3\}, \{2, 4\}\}, \{\{2, 4\}, \{4, 5\}\}\}\}\} \\ E_4 &= \{\{\{\{\{\{1, 2\}, \{2, 3\}\}, \{\{2, 3\}, \{2, 4\}\}\}, \{\{\{2, 3\}, \{2, 4\}\}, \{\{2, 4\}, \{4, 5\}\}\}\}\} \end{aligned}$$

In tree T_3 , consider the edge $e = \{a, b\} = \{\{\{1, 2\}, \{2, 3\}\}, \{\{2, 3\}, \{2, 4\}\}\} \in E_3$, which is also a node in tree T_4 . The proximity condition is satisfied since $|a \cap b| = |\{\{2, 3\}\}| = 1$: note that $a \cap b$ is indeed a singleton, the unique element of which is the set $\{2, 3\}$. The edge e has complete union $A_e = \{1, 2, 3, 4\}$, which is the union of the conditioning sets $A_a = \{1, 2, 3\}$ and $A_b = \{2, 3, 4\}$. Further, e has conditioning set $D_e = \{2, 3\} = A_a \cap A_b$, and conditioned set $C_e = A_a \triangle A_b = \{1, 4\}$, which is itself the union $C_e = C_{e,a} \cup C_{e,b}$ of $C_{e,a} = A_a \setminus A_b = \{1\}$ and $C_{e,b} = A_b \setminus A_a = \{4\}$. Writing this edge in its conditioned form gives better readability: $e = (C_{e,a}, C_{e,b}; D_e) = (14; 23)$. Fig. 6 shows the regular vine sequence, where the nodes and edges have been labeled using the conditioned forms.

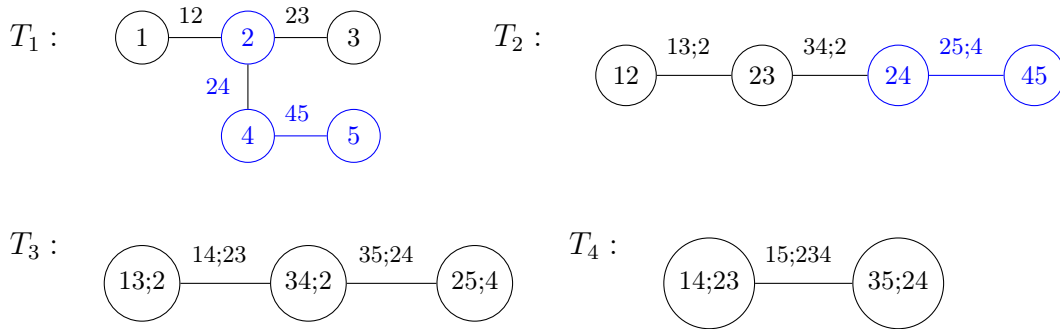


Figure 6: A graphical representation of the five-dimensional regular vine sequence $\mathcal{V}$. The sub-vine $\mathcal{V}_f$ induced by $\mathcal{V}$ on the node set $A_f = \{2, 4, 5\}$ for edge $f = (25; 4)$ in T_2 is shown in blue.

Vine telescoping product

Given are scalars $\gamma_J \in (0, \infty)$ for non-empty $J \subseteq \{1, \dots, 5\}$ such that $\gamma_j = 1$ for every $j \in \{1, \dots, 5\}$. Also, write $\gamma_{I|J} = \gamma_{I \cup J} / \gamma_J$ for disjoint and non-empty $I, J \subset [d]$. Developing the product formula in Lemma 2.4 along the given regular vine sequence $\mathcal{V}$ gives

$$\gamma_{12345} = \gamma_{12}\gamma_{23}\gamma_{24}\gamma_{45} \cdot \frac{\gamma_{13|2}\gamma_{34|2}\gamma_{25|4}}{\gamma_{1|2}\gamma_{3|2}\gamma_{2|3}\gamma_{2|4}\gamma_{2|4}\gamma_{5|2}} \cdot \frac{\gamma_{14|23}\gamma_{35|24}}{\gamma_{1|23}\gamma_{4|23}\gamma_{3|24}\gamma_{5|24}} \cdot \frac{\gamma_{15|234}}{\gamma_{1|234}\gamma_{5|234}}. \quad (\text{A.1})$$

As an illustration of Remark 2.5, consider the edge $f = \{\{2, 4\}, \{4, 5\}\} \in E_2$ (or $f = (25; 4)$ in its conditioned form). The vine $\mathcal{V}_f$ induced by $\mathcal{V}$ on the node set $A_f = \{2, 4, 5\}$ consists of two trees, $\mathcal{V}_f = (T_{f,1}, T_{f,2})$, with edge sets $E_{f,1} = \{\{2, 4\}, \{4, 5\}\}$ and $E_{f,2} = \{\{\{2, 4\}, \{4, 5\}\}\}$; see Fig. 6. The product formula (2.15) yields the identity

$$\gamma_{245} = \gamma_{24}\gamma_{25} \cdot \frac{\gamma_{25|4}}{\gamma_{2|4}\gamma_{5|4}}.$$

Tail copula density decomposition

By Theorem 5.1, any 5-variate tail copula density r can be decomposed along the given regular vine sequence $\mathcal{V}$ as

$$\begin{aligned} r(\mathbf{x}) = & r_{12}(x_1, x_2) r_{23}(x_2, x_3) r_{24}(x_2, x_4) r_{45}(x_4, x_5) \\ & \cdot c_{13;2}(R_{1|2}, R_{3|2}) c_{34;2}(R_{3|2}, R_{4|2}) c_{25;4}(R_{2|4}, R_{5|4}) \\ & \cdot c_{14;23}(R_{1|23}, R_{4|23}; \mathbf{x}_{23}) c_{35;24}(R_{3|24}, R_{5|24}; \mathbf{x}_{24}) \\ & \cdot c_{15;234}(R_{1|234}, R_{5|234}; \mathbf{x}_{234}), \end{aligned} \quad (\text{A.2})$$

where, for brevity, we write $R_{j|D} = R_{j|D}(x_j | \mathbf{x}_D)$ for $j \in [d]$ and $D \subseteq [d] \setminus \{j\}$. The bivariate copula densities in the decomposition are given by formula (3.1) in Sklar's theorem for tail copula densities (Proposition 3.1). The copulas associated with the edges in tree T_2 necessarily satisfy the simplifying assumption (see the paragraph right after Definition 3.4): for instance, the value of $c_{13;2}(u_1, u_3 | x_2)$ does not depend on x_2 , which is why we write $c_{13;2}(R_{1|2}, R_{3|2})$ rather than $c_{13;2}(R_{1|2}, R_{3|2}; x_2)$ on the second line of (A.2).

Recursion

To completely specify the tail copula density r in terms of bivariate components, we can calculate the arguments $R_{j|D}$ of the pair copulas through the recursive formula of Theorem 5.3, as explained also in the paragraph after that theorem. For example, in Eq. (A.2), starting from the first argument of the copula $c_{15;234}$ associated with the deepest edge, the conditional distribution function $R_{1|234}$ can be expressed in terms of the conditional distribution functions $R_{1|23}$ and $R_{4|23}$ and the bivariate copula $C_{14;23}$ via

$$R_{1|234}(x_1 | \mathbf{x}_{234}) = C_{14;23}(R_{1|23}(x_1 | \mathbf{x}_{23}) | R_{4|23}(x_4 | \mathbf{x}_{23}); \mathbf{x}_{23}), \quad (\text{A.3})$$

where $C_{14;23}(u_1 | u_4; \mathbf{x}_{23}) = \int_0^{u_1} c_{14;23}(v, u_4; \mathbf{x}_{23}) dv$. Eq. (A.3) is the first identity in Eq. (5.4) applied to $e = \{\{\{1, 2\}, \{2, 3\}\}, \{\{2, 3\}, \{3, 4\}\}\}$ with $a_e = 1$, $b_e = 4$, and $D_e = \{2, 3\}$. In turn, the conditional distribution function $R_{1|23}$ can be expressed as (similarly for $R_{4|23}$),

$$R_{1|23}(x_1 | \mathbf{x}_{23}) = C_{13;2}(R_{1|2}(x_1 | x_2) | R_{3|2}(x_3 | x_2)),$$

where $C_{13;2}(u_1 | u_3) = \int_0^{u_1} c_{13;2}(v, u_3) dv$, $R_{1|2}(x_1 | x_2) = \int_0^{x_1} r_{12}(t, x_2) dt$ and $R_{3|2}(x_3 | x_2) = \int_0^{x_3} r_{23}(x_2, t) dt$, i.e., in terms of *bivariate* (tail) copula densities r_{12} , r_{23} and $c_{13;2}$ only.

Unique-edge property

Proofs involving regular vine sequences often operate via a recursive argument in which a node $a \in [d]$ and the edges that contain it are peeled off, leaving a regular vine sequence on $d - 1$ elements; see for instance the proof of Theorem 5.3, which relies on Lemma B.2 below. Let us illustrate the latter for the regular vine sequence $\mathcal{V}$ in Figure 6. The conditioned set of the deepest edge $e = (15; 234)$ is $\mathcal{C}_e = \{1, 5\}$. For each of the latter two nodes and for each level $j \in \{1, \dots, d - 1\}$, there is a unique edge $e_j \in E_j$ that contains this node in its complete union. For instance, for node 1 these are the edges (12) , $(13; 2)$, $(14; 23)$ and $(15; 234)$; moreover, node 1 belongs to the conditioned set of each of these edges. Removing node 1 and these four edges leaves a regular vine sequence on the four remaining elements $\{2, \dots, 5\}$, which is the sub-vine induced by $\mathcal{V}$ on the node set $\{2, \dots, 5\}$.

Simulation

Let the random vector $\mathbf{Z} = (Z_1, \dots, Z_5)$ have the inverted multivariate Pareto distribution as in Definition 6.1 associated to an X-vine tail copula density r as in (A.2). By the rejection algorithm in Lemma 2 in Engelke and Hitz (2020), the task of simulating $\mathbf{Z}$ can be reduced to the one of simulating the random vector $\mathbf{Z}^{(j)}$, by definition equal in distribution to $(\mathbf{Z} | Z_j < 1)$, and this for all $j = 1, \dots, 5$. Let $\mathbf{W} = (W_1, \dots, W_5)$ be a vector of independent random variables uniformly distributed on $(0, 1)$. We generate $\mathbf{Z}^{(j)}$ from $\mathbf{W}$ by applying the inverse Rosenblatt transform (6.4) with the five variables $Z_1^{(j)}, \dots, Z_5^{(j)}$ arranged according

j	σ_j	$e_{j,1}$	$e_{j,2}$	$e_{j,3}$	$e_{j,4}$
1	(1, 2, 3, 4, 5)	12	13; 2	14; 23	15; 234
2	(2, 1, 3, 4, 5)	12	13; 2	14; 23	15; 234
3	(3, 2, 1, 4, 5)	23	13; 2	14; 23	15; 234
4	(4, 5, 2, 3, 1)	45	25; 4	35; 24	15; 234
5	(5, 4, 2, 3, 1)	45	25; 4	35; 24	15; 234

Table 2: Five possible permutations σ_j and the edge sequences $e_{j,1}, \dots, e_{j,4}$ for $j \in \{1, \dots, 5\}$ of Lemma 6.2 for the regular vine tree $\mathcal{V}$ in Figure 6.

to a permutation σ_j of $\{1, \dots, 5\}$ such that $\sigma_j(1) = j$ and such that for each $k = 2, \dots, 5$, the node set $\{\sigma_j(1), \dots, \sigma_j(k)\}$ is equal to the conditioning set of some edge $e_{j,k-1}$ in tree T_{k-1} (Lemma 6.2).

The permutations σ_j can be constructed via a simple algorithm. Given $j \in [d]$:

- (1) Put $\sigma_j(1) = j$.
- (2) Find $e_{j,1} \in E_1$ such that $j \in e_{j,1}$.
- (3) Let $\sigma_j(2)$ be the unique element in $e_{j,1} \setminus \{j\}$.
- (4) For $k = 3, \dots, d$, do
 - (a) Find $e_{j,k-1} \in E_{k-1}$ such that $e_{j,k-2} \in e_{j,k-1}$.
 - (b) Let $\sigma_j(k)$ be the unique element in $A_{e_{j,k-1}} \setminus A_{e_{j,k-2}}$.

Steps (2) and (4.a) leave freedom for choice, and this is why there may be multiple permutations σ_j that satisfy the requirements of Lemma 6.2. For the regular vine tree $\mathcal{V}$ in Fig. 6, five possible permutations σ_j for $j \in \{1, \dots, 5\}$ and the corresponding edge sequences $e_{j,1}, \dots, e_{j,4}$ are listed in Table 2. Another choice for σ_4 would have been (4, 2, 5, 3, 1), for instance.

Spelling out the inverse Rosenblatt transform (6.4) for $j = 4$ and $\sigma_4 = (4, 5, 2, 3, 1)$ gives $Z_4^{(4)} = W_4$ and

$$\begin{aligned} Z_5^{(4)} &= R_{5|4}^{-1} \left(W_5 | Z_4^{(4)} \right), & Z_2^{(4)} &= R_{2|45}^{-1} \left(W_2 | Z_{45}^{(4)} \right), \\ Z_3^{(4)} &= R_{3|245}^{-1} \left(W_3 | Z_{245}^{(4)} \right), & Z_1^{(4)} &= R_{1|2345}^{-1} \left(W_1 | Z_{2345}^{(4)} \right). \end{aligned}$$

Thanks to the stated properties of the permutations σ_j , the conditional quantile functions $R_{k|\sigma_j(\{1, \dots, k-1\})}^{-1}$ can be computed recursively via (6.3) in terms of the bivariate ingredients of the X-vine specification of r : for instance, starting from edge $e_{4,2} = (25; 4)$ in T_2 , we obtain

$$R_{2|45}^{-1}(u_2 | \mathbf{x}_{45}) = R_{2|4}^{-1} \left(C_{2|5;4}^{-1}(u_2 | R_{5|4}(x_5 | x_4)) | x_4 \right),$$

an expression in terms of the tail copula densities r_{24} and r_{45} of edges in T_1 and the bivariate copula density $c_{25;4}$ of an edge in T_2 .

Structure matrices

For programming purposes, it is convenient to encode a regular vine sequence in matrix form. Such a regular vine matrix or structure matrix in short is a $d \times d$ upper triangular matrix $M = (m_{i,j})_{i,j=1}^d$ with elements $m_{i,j}$ in $[d]$ and whose diagonal $(m_{1,1}, \dots, m_{d,d})$ is a permutation of $[d]$; see Czado (2019, Section 5.5). A recipe to construct such a matrix is explained in Example 5.11 of Czado (2019). A recursive algorithm to do so is as follows:

- (1) Let e_d be the single edge in E_{d-1} .
- (2) Choose $m_{d,d} \in \mathcal{C}_{e_d}$.
- (3) For $j \in \{1, \dots, d-1\}$:
 - (a) Let $e_j \in E_j$ be the unique edge that contains $m_{d,d}$ in its complete union and hence in its conditioned set (see Lemma B.2).
 - (b) Let $m_{j,d}$ be the unique element in $\mathcal{C}_{e_j} \setminus \{m_{d,d}\}$.
- (4) Remove node $m_{d,d}$ and the edges $e_1, \dots, e_{d-1}$ in Step (3) from the regular vine sequence.

- (5) Proceed recursively with the resulting regular vine sequence on the $d - 1$ elements $[d] \setminus \{m_{d,d}\}$ to fill up the $(d - 1)$ th column $m_{1,d-1}, \dots, m_{d-1,d-1}$; and so on.

Note that step (3a) involves the unique-edge property described in a previous paragraph. Step (2) of the algorithm involves an arbitrary choice, and this is why the same regular vine sequence can be encoded by several structure matrices. Given an element $j \in [d]$, it is always possible to choose the diagonal elements $m_{2,2}, \dots, m_{d,d}$ to be different from j , thus ensuring $m_{1,1} = j$. The diagonal of the resulting structure matrix M then forms a permutation σ_j meeting the requirements of Lemma 6.2.

For the regular vine sequence in Fig. 6, some possible structure matrices M are

$$\begin{bmatrix} 1 & 1 & 2 & 2 & 4 \\ & 2 & 1 & 3 & 2 \\ & & 3 & 1 & 3 \\ & & & 4 & 1 \\ & & & & 5 \end{bmatrix}, \begin{bmatrix} 2 & 2 & 2 & 2 & 4 \\ & 1 & 1 & 3 & 2 \\ & & 3 & 1 & 3 \\ & & & 4 & 1 \\ & & & & 5 \end{bmatrix}, \begin{bmatrix} 3 & 3 & 2 & 2 & 4 \\ & 2 & 3 & 3 & 2 \\ & & 1 & 1 & 3 \\ & & & 4 & 1 \\ & & & & 5 \end{bmatrix}, \begin{bmatrix} 4 & 4 & 4 & 2 & 2 \\ & 5 & 5 & 4 & 3 \\ & & 2 & 5 & 4 \\ & & & 3 & 5 \\ & & & & 1 \end{bmatrix}, \begin{bmatrix} 5 & 5 & 4 & 2 & 2 \\ & 4 & 5 & 4 & 3 \\ & & 2 & 5 & 4 \\ & & & 3 & 5 \\ & & & & 1 \end{bmatrix}.$$

The diagonals of these matrices are equal to the permutations σ_j in Table 2.

Truncated X-vine models

To encode a truncated regular vine sequence $\mathcal{V} = (T_1, \dots, T_q)$ on d elements with truncation level $q \in \{1, \dots, d - 2\}$ (Definition 5.9), one possible way would be to complete it to a (full) regular vine sequence $\mathcal{V}' = (T_1, \dots, T_{d-1})$ and use a structure matrix of the latter, imputing independence copulas for the resulting X-vine model. This would be wasteful, however, especially if d is large and q is small. Instead, we avoid such arbitrary and unnecessary completions and work with truncated structure matrices instead. The form of such a matrix is the same as of a structure matrix in the previous paragraph, but with zeroes on the unspecified elements above the diagonal.

Consider for instance the truncated regular vine sequence (T_1, T_2) on $d = 5$ elements for T_1 and T_2 as in Fig. 2. One possible structure matrix would be

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

which is the first of the five structure matrices in the previous paragraph but with zeroes on positions that correspond to edges in trees T_j with $j > q = 2$. Such a matrix can be directly obtained from the truncated regular vine sequence by an algorithm similar to the one in the previous paragraph:

- (1) Choose $m_{d,d} \in [d]$ such that there is only a single edge $e_q \in E_q$ with $m_{d,d} \in A_{e_q}$ and hence $m_{d,d} \in \mathcal{C}_{e_q}$.
- (2) For $j \in \{1, \dots, q\}$:
 - (a) Let $e_j \in E_j$ be the unique edge that contains $m_{d,d}$ in its complete union and hence in its conditioned set.
 - (b) Let $m_{j,d}$ be the unique element in $\mathcal{C}_{e_j} \setminus \{m_{d,d}\}$.
- (3) Remove node $m_{d,d}$ and the edges $e_1, \dots, e_q$ from Step (2) from the truncated regular vine sequence.
- (4) Proceed recursively with the resulting regular vine sequence, truncated or not.

Recursive estimation

We illustrate the recursive nature of the sequential estimation procedure in Section 7.2 for the regular vine sequence in Fig. 6. For the edge $e = (25; 4)$ in E_2 , we consider pseudo-observations

$$\hat{U}_{i,2;4} = R_{2|4}(\hat{Z}_{i,2} \mid \hat{Z}_{i,4}; \hat{\theta}_{24}), \quad \hat{U}_{i,5;4} = R_{5|4}(\hat{Z}_{i,5} \mid \hat{Z}_{i,4}; \hat{\theta}_{45})$$

for $i \in N_4$, i.e., for those observations with a large value in variable $j = 4$. Here, the tail copula parameters $\hat{\theta}_{24}$ and $\hat{\theta}_{45}$ have already been estimated in a previous step based on the pairs $(\hat{Z}_{i,2}, \hat{Z}_{i,4})$ for $i \in N_2 \cup N_4$ and the pairs $(\hat{Z}_{i,4}, \hat{Z}_{i,5})$ for $i \in N_4 \cup N_5$, respectively, using the averaging procedure based on the pseudo-likelihoods (7.4). For instance, we have $\hat{\theta}_{24} = (\hat{\theta}_{24}^{(2)} + \hat{\theta}_{24}^{(4)})/2$ where $\hat{\theta}_{24}^{(2)}$ is estimated based on the pairs $(\hat{Z}_{i,2}, \hat{Z}_{i,4})$ for $i \in N_2$, while $\hat{\theta}_{24}^{(4)}$ is based on the same pairs but for $i \in N_4$.

Next, consider the edge $e = (35; 24)$ in E_3 , which itself joins the edges $f = (34; 2)$ and $g = (25; 4)$ in E_2 . In view of Proposition 7.1(ii) with $I = \{3, 5\}$ and $J = \{2, 4\}$, we define pseudo-observations $(\widehat{U}_{i,3;24}, \widehat{U}_{i,5;24})$ for $i \in N_{24} = N_2 \cap N_4$ from $c_{3,5;24}$ by

$$\begin{aligned}\widehat{U}_{i,3;24} &= R_{3|24} \left(\widehat{Z}_{i,3} \mid \widehat{Z}_{i,24}; \hat{\theta}(E_{1:2}) \right) \\ &= C_{3|4;2} \left(R_{3|2} \left(\widehat{Z}_{i,3} \mid \widehat{Z}_{i,2}; \hat{\theta}_{23} \right) \mid R_{4|2} \left(\widehat{Z}_{i,4} \mid \widehat{Z}_{i,2}; \hat{\theta}_{24} \right); \hat{\theta}_{34;2} \right)\end{aligned}$$

and

$$\begin{aligned}\widehat{U}_{i,5;24} &= R_{5|24} \left(\widehat{Z}_{i,5} \mid \widehat{Z}_{i,24}; \hat{\theta}(E_{1:2}) \right) \\ &= C_{5|2;4} \left(R_{5|4} \left(\widehat{Z}_{i,5} \mid \widehat{Z}_{i,4}; \hat{\theta}_{45} \right) \mid R_{2|4} \left(\widehat{Z}_{i,2} \mid \widehat{Z}_{i,4}; \hat{\theta}_{24} \right); \hat{\theta}_{25;4} \right).\end{aligned}$$

Note that for the pair $\{2, 4\}$, both conditional distribution functions $R_{2|4}$ and $R_{4|2}$ are required.

B Proofs

B.1 Proofs for Section 2

First we state an auxiliary lemma linking the angular measure density h to the exponent measure density λ and the tail copula density r .

Lemma B.1. *We have $h(\mathbf{w}) = \lambda(\mathbf{w})$ for $\mathbf{w} \in \Delta_{d-1}$ and thus also*

$$r(\mathbf{x}) = \lambda(1/\mathbf{x}) \prod_{j=1}^d x_j^{-2} = \|1/\mathbf{x}\|_1^{-d-1} \cdot h\left(\frac{1/\mathbf{x}}{\|1/\mathbf{x}\|_1}\right) \prod_{j=1}^d x_j^{-2}, \quad \mathbf{x} \in (0, \infty)^d.$$

Proof of Lemma B.1. Let $V(\mathbf{y}) = \Lambda([0, \infty) \setminus [\mathbf{0}, \mathbf{y}])$ for $\mathbf{y} \in (0, \infty)^d$, the exponent measure function. For $V_{1:d}(\mathbf{y}) := \partial^d V / (\partial y_1 \cdots \partial y_d)$, Coles and Tawn (1991) show that

$$V_{1:d}(\mathbf{y}) = -\|\mathbf{y}\|_1^{-d-1} h(\mathbf{y}/\|\mathbf{y}\|_1).$$

On the other hand, by the inclusion-exclusion formula we also have

$$\begin{aligned}V_{1:d}(\mathbf{y}) &= \sum_{\emptyset \neq J \subseteq [d]} (-1)^{|J|-1} \frac{\partial^d \nu(\{\mathbf{z} \geq \mathbf{0} : \forall j \in J, z_j > y_j\})}{\partial y_1 \cdots \partial y_d} \\ &= (-1)^{d-1} (-1)^d \lambda(\mathbf{y}) = -\lambda(\mathbf{y}).\end{aligned}$$

We find

$$\lambda(\mathbf{y}) = \|\mathbf{y}\|_1^{-d-1} h(\mathbf{y}/\|\mathbf{y}\|_1), \quad \mathbf{y} \in (0, \infty)^d.$$

If $\|\mathbf{y}\|_1 = 1$, this specializes to $\lambda(\mathbf{y}) = h(\mathbf{y})$. The connection with the tail copula density r follows from (2.12), which can be inverted to

$$r(\mathbf{x}) = \lambda(1/\mathbf{x}) \prod_{j=1}^d x_j^{-2}, \quad \mathbf{x} \in (0, \infty)^d. \quad \square$$

The following lemma provides the basis for recursive arguments concerning regular vines.

Lemma B.2. *Let $\mathcal{V} = (T_j)_{j=1}^{d-1}$ be a regular vine sequence on d elements, with $d \geq 2$. Let e be the unique edge in E_{d-1} . For every $a \in \mathcal{C}_e$ and every $j \in \{1, \dots, d-1\}$, there is a unique edge $e_j \in E_j$ such that $a \in A_{e_j}$, and then in fact $a \in \mathcal{C}_{e_j}$.*

Proof of Lemma B.2. The existence is clear: at every level $j \in \{1, \dots, d-1\}$, the union $\bigcup_{e' \in E_j} A_{e'}$ of the complete unions is equal to $[d]$. Indeed, this holds for E_1 since T_1 is a tree on $[d]$ and it holds for the other edge sets by the recursive definition of the complete unions.

The uniqueness is specific to the two nodes in the conditioned set of the single edge e in E_{d-1} . Suppose that for some $a \in \mathcal{C}_e$ there is an edge set E_j , with $j \leq d-2$, for which there are two different edges, say f and g , that both contain a in their complete unions. If f and g are linked up by some edge h in the next tree T_{j+1} , then $a \in D_h$ and thus, eventually $a \in D_e$ (see the paragraph following Definition 2.3), a contradiction. Otherwise, there exist two different edges f' and g' in E_{j+1} that contain a in their complete unions, and the argument can be repeated, if needed until the last tree has been reached. Finally, since the conditioning sets are nested and since a does not belong to the conditioning set of e , it does not belong to the conditioning set of another edge either. $\square$

Proof of Lemma 2.4. Identity (2.14) follows from

$$\frac{\gamma_{D_e} \cdot \gamma_{A_e}}{\gamma_{A_a} \cdot \gamma_{A_b}} = \frac{\frac{\gamma_{A_e}}{\gamma_{D_e}}}{\frac{\gamma_{A_a}}{\gamma_{D_e}} \cdot \frac{\gamma_{A_b}}{\gamma_{D_e}}}$$

together with $A_a = \{a_e\} \cup D_e$, $A_b = \{b_e\} \cup D_e$, and $A_e = \{a_e, b_e\} \cup D_e$.

The proof of (2.13) is by induction on d . For $d = 3$, we can relabel the indices so that the first edge set $E_1 = \{\{1, 2\}, \{2, 3\}\}$ while E_2 contains the single edge $(13; 2)$. The formula then states that

$$\gamma_{123} = \gamma_{12} \cdot \gamma_{23} \cdot \frac{\gamma_2 \cdot \gamma_{123}}{\gamma_{12} \cdot \gamma_{23}},$$

which is obviously true since $\gamma_2 = 1$ by assumption.

Let $d \geq 4$ and suppose the formula is true for regular vine sequences of dimension up to $d - 1$. Let $e = (a_e, b_e; D_e)$ be the single edge in the deepest tree $T_{d-1} = (N_{d-1}, E_{d-1})$ of the vine. By Lemma B.2, there exist unique edges $e_j \in E_j$ for $j \in \{1, \dots, d - 1\}$ with $e_{d-1} = e$ such that

$$a_e \in e_1 \in e_2 \in \dots \in e_{d-1}.$$

Moreover, the node a_e belongs to the conditioned set of each edge e_j , so we can write $\mathcal{C}_{e_j} = \{a_e, b_j\}$ for $j \in \{1, \dots, d - 1\}$ and some $b_j \in [d] \setminus \{a_e\}$. For $j \in \{2, \dots, d - 1\}$, we have $e_j = \{e_{j-1}, f_{j-1}\}$ for some edge $f_{j-1} \in E_{j-1}$.

The regular vine sequence $\mathcal{V}$ induces a regular vine sequence $\bar{\mathcal{V}}$ on $\bar{N}_1 = [d] \setminus \{a_e\}$ with edge sets $\bar{E}_j = E_j \setminus \{e_j\}$ for $j \in \{1, \dots, d - 2\}$; this is the sub-vine induced by $\mathcal{V}$ on $[d] \setminus \{a_e\}$. By the induction hypothesis, we have

$$\gamma_{[d] \setminus \{a_e\}} = \prod_{\bar{e} \in \bar{E}_1} \gamma_{\bar{e}} \cdot \prod_{j=2}^{d-2} \prod_{\bar{e} = \{\bar{f}, \bar{g}\} \in \bar{E}_j} \frac{\gamma_{D_{\bar{e}}} \cdot \gamma_{A_{\bar{e}}}}{\gamma_{A_{\bar{f}}} \cdot \gamma_{A_{\bar{g}}}}.$$

The product of the additional factors in (2.13) is

$$\gamma_{e_1} \cdot \prod_{j=2}^{d-1} \frac{\gamma_{D_{e_j}} \cdot \gamma_{A_{e_j}}}{\gamma_{A_{e_{j-1}}} \cdot \gamma_{A_{f_{j-1}}}} = \left(\gamma_{e_1} \cdot \prod_{j=2}^{d-1} \frac{\gamma_{A_{e_j}}}{\gamma_{A_{e_{j-1}}}} \right) \cdot \prod_{j=2}^{d-1} \frac{\gamma_{D_{e_j}}}{\gamma_{A_{f_{j-1}}}}.$$

It is sufficient to show that this product is equal to $\gamma_{[d]}/\gamma_{[d] \setminus \{a_e\}}$. On the one hand, since $A_{e_1} = e_1$ and $A_{e_{d-1}} = [d]$, we have the telescoping product

$$\gamma_{e_1} \cdot \prod_{j=2}^{d-1} \frac{\gamma_{A_{e_j}}}{\gamma_{A_{e_{j-1}}}} = \gamma_{A_{e_{d-1}}} = \gamma_{[d]}.$$

On the other hand, we have

$$\prod_{j=2}^{d-1} \frac{\gamma_{D_{e_j}}}{\gamma_{A_{f_{j-1}}}} = \gamma_{D_{e_2}} \cdot \frac{\prod_{j=3}^{d-1} \gamma_{D_{e_j}}}{\prod_{j=2}^{d-2} \gamma_{A_{f_{j-1}}}} \cdot \frac{1}{\gamma_{A_{f_{d-2}}}}.$$

The first factor is $\gamma_{D_{e_2}} = 1$ since D_{e_2} is a singleton, while the last factor is $1/\gamma_{A_{f_{d-2}}} = 1/\gamma_{[d] \setminus \{a_e\}}$. It remains to show that the middle factor is equal to 1 too. To do so, write it as

$$\prod_{j=3}^{d-1} \frac{\gamma_{D_{e_j}}}{\gamma_{A_{f_{j-2}}}},$$

which is indeed equal to 1, since $D_{e_j} = A_{f_{j-2}}$ for $j \in \{3, \dots, d - 1\}$. Indeed, to see the latter identity, write

$$e_j = \{e_{j-1}, f_{j-1}\} = \{\{e_{j-2}, f_{j-2}\}, f_{j-1}\}.$$

Recall $\mathcal{C}_{e_j} = \{a_e, b_j\}$. Then $D_{e_j} = A_{e_j} \setminus \{a_e, b_j\}$ as well as $A_{e_j} = A_{e_{j-1}} \cup \{b_j\}$ and $A_{j-1} = A_{f_{j-2}} \cup \{a\}$, the unions being disjoint. It follows that $A_{f_{j-2}} = A_{e_j} \setminus \{a, b_j\} = D_{e_j}$. The proof is complete. $\square$

B.2 Proofs for Section 3

Proof of Proposition 3.1. The fact that $r_{I|J}(\cdot | \mathbf{x}_J)$ is a probability density function on $(0, \infty)^I$ follows from the fact that integrating a tail copula density over certain coordinates while keeping the other coordinates fixed produces the marginal tail copula density in the remaining coordinates, see (2.7). The statements about the univariate margins follow in the same way. Finally, the expression of the copula density follows from Sklar's theorem for probability densities applied to $r_{I|J}(\cdot | \mathbf{x}_J)$. $\square$

Proof of Lemma 3.2. By the homogeneity of tail copula densities as stated in (2.5), we have

$$r_{I|J}(\mathbf{x}_I | t\mathbf{x}_J) = \frac{r_{I \cup J}(\mathbf{x}_I, t\mathbf{x}_J)}{r_J(t\mathbf{x}_J)} = \frac{t^{1-|I \cup J|} \cdot r_{I \cup J}(t^{-1}\mathbf{x}_I, \mathbf{x}_J)}{t^{1-|J|} \cdot r_J(\mathbf{x}_J)} = t^{-|I|} \cdot r_{I|J}(t^{-1}\mathbf{x}_I | \mathbf{x}_J).$$

If the random vector $\boldsymbol{\xi}_I$ has probability density function $\mathbf{x}_I \mapsto r_{I|J}(\mathbf{x}_I | \mathbf{x}_J)$, then the probability density function of the random vector $t\boldsymbol{\xi}_I$ is indeed $\mathbf{x}_I \mapsto t^{-|I|} \cdot r_{I|J}(t^{-1}\mathbf{x}_I | \mathbf{x}_J) = r_{I|J}(\mathbf{x}_I | t\mathbf{x}_J)$. The statements on the cumulative distribution and quantile functions follow easily. Finally, the copula of $t\boldsymbol{\xi}_I$ does not depend on $t > 0$, yielding (3.2). $\square$

B.3 Proofs for Section 4

Proof of Proposition 4.1. Since the lower dimensional margins of the scaled extremal Dirichlet model (4.1) are of the same form as r itself, we find for the conditional probability density in the variables $\mathbf{x}_I \in (0, \infty)^I$ the expressions

$$r_{I|J}(\mathbf{x}_I | \mathbf{x}_J) = K \cdot \frac{\left\{ \sum_{i \in I} c(\alpha_i, \rho)^{1/\rho} x_i^{-1/\rho} + s(\mathbf{x}_J) \right\}^{-\rho - \sum_{k \in I \cup J} \alpha_k}}{s(\mathbf{x}_J)^{-\rho - \sum_{j \in J} \alpha_j}} \cdot \prod_{i \in I} x_i^{-\alpha_i/\rho - 1},$$

where K is a normalising constant that depends on the parameters $\alpha_1, \dots, \alpha_d, \rho$ and the sets I and J but not on the arguments $\mathbf{x}_{I \cup J}$, while $s(\mathbf{x}_J) = \sum_{j \in J} c(\alpha_j, \rho)^{1/\rho} x_j^{-1/\rho}$. For a well-chosen scalar $t > 0$ depending on $\mathbf{x}_J$, we have $s(t\mathbf{x}_J) = 1$. By Lemma 3.2, a scalar multiplication of the vector of conditioning variables $\mathbf{x}_J$ does not affect the copula density $c_{I;J}(\cdot; \mathbf{x}_J)$, i.e., $c_{I;J}(\cdot; t\mathbf{x}_J) = c_{I;J}(\cdot; \mathbf{x}_J)$. We obtain that $c_{I;J}(\cdot; \mathbf{x}_J) \equiv c_{I;J}(\cdot)$ is equal to the copula density of the $|I|$ -variate probability density function proportional to

$$\mathbf{x}_I \mapsto \prod_{i \in I} x_i^{-\alpha_i/\rho - 1} \cdot \left\{ \sum_{i \in I} c(\alpha_i, \rho)^{1/\rho} x_i^{-1/\rho} + 1 \right\}^{-\rho - \sum_{k \in I \cup J} \alpha_k}.$$

This fact already implies that the simplifying assumption holds.

To identify $c_{I;J}$, we can incorporate the constants $c(\alpha_i, \rho)$ into the variables x_i by marginal scale transformations without changing the copula. Hence, $c_{I;J}$ is equal to the copula density of the $|I|$ -variate probability density function proportional to

$$\mathbf{x}_I \mapsto \prod_{i \in I} x_i^{-\alpha_i/\rho - 1} \cdot \left(\sum_{i \in I} x_i^{-1/\rho} + 1 \right)^{-\rho - \sum_{k \in I \cup J} \alpha_k}. \quad (\text{B.1})$$

Consider two cases: $\rho > 0$ or $\rho < 0$. Suppose first $\rho > 0$. Applying the component-wise *decreasing* transformations $x_i \mapsto y_i = x_i^{-1/\rho}$, we see that the copula density $c_{I;J}$ is equal to the *survival* copula density of the $|I|$ -variate probability density function proportional to

$$\mathbf{y}_I \mapsto \prod_{i \in I} y_i^{\alpha_i - 1} \cdot \left(\sum_{i \in I} y_i + 1 \right)^{-\rho - \sum_{k \in I \cup J} \alpha_k}.$$

Let $\mathbf{Y} = (Y_i)_{i \in I}$ be a random vector on $(0, \infty)^I$ with the latter density function and consider the random vector $t(\mathbf{Y}) = (\mathbf{W}, S)$ with $S = \sum_{i \in I} Y_i$ and $W_i = Y_i/S$ for $i \in I$; note that $\sum_{i \in I} W_i = 1$, so we consider $\mathbf{W}$ as a random vector on the $(|I| - 1)$ -dimensional unit simplex $\Delta_I = \{\mathbf{w} \in [0, 1]^I : \sum_{i \in I} w_i = 1\}$, where one of the coordinates, say w_{i_0} for some fixed $i_0 \in I$, is written as $w_{i_0} = 1 - \sum_{i \in I \setminus \{i_0\}} w_i$. The inverse transformation is $t^{-1}(\mathbf{w}; s) = (sw_i)_{i \in I}$. By the multivariate changes-of-variables formula, the probability density of $t(\mathbf{Y})$ is

$$\prod_{i \in I} w_i^{\alpha_i - 1} \cdot s^{\sum_{i \in I} \alpha_i - 1} (s + 1)^{-\rho - \sum_{k \in I \cup J} \alpha_k}$$

for $(\mathbf{w}, s) \in \Delta_I \times (0, \infty)$. The random vector $\mathbf{W}$ thus has a Dirichlet distribution with parameter vector $(\alpha_i)_{i \in I}$ and is independent of the positive random variable S . Therefore, $\mathbf{Y} = S\mathbf{W}$ has a Liouville distribution and its survival copula is a Liouville copula.

Second, suppose $\rho < 0$. We can then apply exactly the same component-wise transformations $x_i \mapsto y_i = x_i^{-1/\rho}$ as in the case $\rho > 0$, but now these transformations are *increasing*, so that $c_{I,J}$ is equal to the copula density of the random vector $\mathbf{Y}$. As the survival copula of $\mathbf{Y}$ is a Liouville copula, $c_{I,J}$ itself is a survival Liouville copula. $\square$

Calculating $c_{I,J}$ for the negative logistic model. If $\alpha_1 = \dots = \alpha_d = 1$ and $\rho > 0$, then, writing $\theta = 1/\rho > 0$, the probability density in (B.1) is proportional to the function

$$\mathbf{x}_I \mapsto \prod_{i \in I} x_i^{-\theta-1} \cdot \left(1 + \sum_{i \in I} x_i^{-\theta} \right)^{-1/\theta - |I| - |J|},$$

for $\mathbf{x}_I \in (0, \infty)^I$. We have $1/\theta + |J| = 1/\theta_J$ with $\theta_J = \theta/(|J|\theta + 1)$. The copula density does not change if we apply the component-wise increasing transformations $(0, \infty) \rightarrow (0, 1) : x_i \mapsto u_i = (x_i^{-\theta} + 1)^{-1/\theta_J}$ to the $|I|$ variables. Since $x_i^{-\theta} = u_i^{-\theta_J} - 1$ and $dx_i/du_i = (\theta_J/\theta)x_i^{1+\theta}u_i^{-\theta_J-1}$, the resulting probability density is proportional to

$$\mathbf{u}_I \mapsto \prod_{i \in I} u_i^{-\theta_J-1} \cdot \left(1 - |I| + \sum_{i \in I} u_i^{-\theta_J} \right)^{-1/\theta_J - |I|},$$

which is proportional to the density of the $|I|$ -variate Clayton copula with parameter θ_J . $\square$

Calculating $c_{I,J}$ for the logistic model. If $\alpha_1 = \dots = \alpha_d$ and $\rho = -1/\theta$ with $\theta > 1$, then we seek the copula density of the probability density proportional to

$$\mathbf{x}_I \mapsto \prod_{i \in I} x_i^{\theta-1} \cdot \left(1 + \sum_{i \in I} x_i^{\theta} \right)^{1/\theta - |I| - |J|}, \quad \mathbf{x}_I \in (0, \infty)^I.$$

This time, we have $1/\theta - |J| = -1/\vartheta_J$ with $\vartheta_J = \theta/(|J|\theta - 1)$. We apply the component-wise *decreasing* transformations $(0, \infty) \rightarrow (0, 1) : x_i \mapsto u_i = (x_i^{\theta} + 1)^{-1/\vartheta_J}$. Since $x_i^{\theta} = u_i^{-\vartheta_J} - 1$ and $dx_i/du_i = -(\vartheta_J/\theta)x_i^{1-\theta}u_i^{-\vartheta_J-1}$, the copula density we seek is equal to the one of the *survival* copula of the probability density proportional to

$$\mathbf{u}_I \mapsto \prod_{i \in I} u_i^{-\vartheta_J-1} \cdot \left(1 - |I| + \sum_{i \in I} u_i^{-\vartheta_J} \right)^{-1/\vartheta_J - d}, \quad \mathbf{u}_I \in (0, 1)^I.$$

Again, this function is proportional to the density of the $|I|$ -variate Clayton copula with parameter ϑ_J . $\square$

Proof of Proposition 4.2. Write $K = I \cup J$. The Hüsler–Reiss tail copula density $r(\cdot, \Sigma^{(k)})$ in (4.3) is marginally closed in the sense that the $|K|$ -variate tail copula density $r_K(\cdot; \Sigma^{(k)})$ resulting from marginalization as in (2.7) is of the same parametric form but now with respect to the $(|K| \times |K|)$ -dimensional variogram Γ_{KK} . The covariance matrix (4.2) with Γ replaced by Γ_{KK} is $\Sigma_{KK}^{(k)}$. We obtain

$$r_K(\mathbf{x}; \Sigma^{(k)}) = r(\mathbf{x}; \Sigma_{KK}^{(k)}) = \left(\prod_{i \in K \setminus \{k\}} x_i^{-1} \right) \phi_{|K|-1}(\bar{\mathbf{x}}_{\setminus k}; \Sigma_{KK}^{(k)}),$$

for $\mathbf{x} \in (0, \infty)^K$, with $\bar{\mathbf{x}}_{\setminus k} = (\log(x_i/x_k) - \frac{1}{2}\Gamma_{ik})_{i \in K \setminus \{k\}}$. For fixed $\mathbf{x}_J \in (0, \infty)^J$, the function $\mathbf{x}_I \mapsto r_{I \cup J}(\mathbf{x}_{I \cup J}; \Sigma^{(k)})$ is proportional to an $|I|$ -variate log-normal density with covariance matrix $\Sigma_{I|J}^{(k)}$ and some mean vector depending on $\mathbf{x}_J$; this follows from a well-known property of conditional distributions of the Gaussian distribution and in particular from the block-matrix inversion formula of

$$\Sigma_{KK}^{(k)} = \begin{bmatrix} \Sigma_{II}^{(k)} & \Sigma_{IJ}^{(k)} \\ \Sigma_{JI}^{(k)} & \Sigma_{JJ}^{(k)} \end{bmatrix}.$$

Since the copula of a random vector is invariant with respect to component-wise increasing transformations, the copula density of the probability density $r_{I|J}(\cdot | \mathbf{x}_J; \Sigma^{(k)})$ is equal to the stated Gaussian copula density. $\square$

B.4 Proofs for Section 5

Proof of Theorem 5.1 and Remark 5.2. Applying the regular vine telescoping equations (2.13) and (2.14) to $\gamma_J := r_J(\mathbf{x}_J)$, we find

$$r(\mathbf{x}) = \prod_{e \in E_1} r_{a_e, b_e}(x_{a_e}, x_{b_e}) \cdot \prod_{j=2}^{d-1} \prod_{e \in E_j} \frac{r_{a_e, b_e|D_e}(x_{a_e}, x_{b_e}|\mathbf{x}_{D_e})}{r_{a_e|D_e}(x_{a_e}|\mathbf{x}_{D_e}) \cdot r_{b_e|D_e}(x_{b_e}|\mathbf{x}_{D_e})}.$$

By Eq. (3.1) in Sklar's theorem for tail copula densities (Proposition 3.1) applied to $I = \{a_e, b_e\}$ and $J = D_e$, the last fraction is equal to $c_{a_e, b_e|D_e}(u_{a_e}, u_{b_e}; \mathbf{x}_{D_e})$ with $u_{a_e} = R_{a_e|D_e}(x_{a_e}|\mathbf{x}_{D_e})$ and $u_{b_e} = R_{b_e|D_e}(x_{b_e}|\mathbf{x}_{D_e})$. Formula (5.2) in Remark 5.2 follows in the same way from the more general product formula (2.15) in Remark 2.5. $\square$

Proof of Theorem 5.3. Let $j_0 \in \{2, \dots, d-1\}$ and $e_0 \in E_{j_0}$. The complete union $J_0 = A_{e_0} \subset [d]$ has cardinality $j_0 + 1$. We can restrict the regular vine sequence $\mathcal{V}$ to J_0 in the obvious way to obtain a regular vine sequence $\mathcal{V}_0 = (T_{0,j})_{j=1}^{j_0}$ on $j_0 + 1$ elements: the first tree $T_{0,1}$ has node set $N_{0,1} = J_0$ and edge set $E_{0,1} = \{e \in E_1 : e \subset J_0\}$, and so on. The edge e_0 is the single one in the deepest tree T_{0,j_0} of $\mathcal{V}_0$. By applying Theorem 5.1 to r_{J_0} , we obtain the same decomposition as in (5.1), with the same bivariate (tail) copula densities, but restricted to $\mathcal{V}_0$. This means that, without loss of generality, we can assume that e_0 is the deepest edge of the original regular vine sequence $\mathcal{V}$, that is, e_0 is the unique element, say e , in E_{d-1} .

We have $e = (a_e, b_e; D_e) = \{f, g\}$ with $f, g \in N_{d-1} = E_{d-2}$. The complete union of e is $A_e = [d]$. Furthermore, upon switching the names of f and g if necessary, we have $A_f = D_e \cup \{a_e\} = [d] \setminus \{b_e\}$ and $A_g = D_e \cup \{b_e\} = [d] \setminus \{a_e\}$. We can apply the decomposition (5.1) to both r and its $(d-1)$ -dimensional marginal density r_{A_g} , as in the previous paragraph; see also Eq. (5.2). The two decompositions are the same, except that the one for r has some additional factors, namely the ones that involve node a_e . For each tree, there is only one such additional factor, since in each tree T_j , there is only a single edge, say e_j , that contains a_e in its complete union (Lemma B.2). For these edges e_j , we have in fact $a_e \in \mathcal{C}_{e_j}$, so we can write $e_j = (a_e, b_{e_j}; D_{e_j})$. Note that $e_{d-1} = e$ and $b_{e_{d-1}} = b_e$, while $e_{d-2} = f$. It follows that

$$\begin{aligned} r_{a_e|D_e \cup b_e}(x_{a_e}|\mathbf{x}_{D_e \cup b_e}) &= \frac{r(\mathbf{x})}{r_{A_g}(\mathbf{x}_{A_g})} \\ &= r_{a_e, b_{e_1}}(x_{a_e}, x_{b_{e_1}}) \cdot \prod_{j=2}^{d-1} c_{a_e, b_{e_j}|D_{e_j}} \left(R_{a_e|D_{e_j}}(x_{a_e}|\mathbf{x}_{D_{e_j}}), R_{b_{e_j}|D_{e_j}}(x_{b_{e_j}}|\mathbf{x}_{D_{e_j}}); \mathbf{x}_{D_{e_j}} \right). \end{aligned}$$

Recall that $A_f = D_e \cup \{a_e\}$. Applying the same argument to edge f via the regular vine sequence induced by $\mathcal{V}$ on A_f as in the first paragraph of the proof, we find

$$r_{a_e|D_e}(x_{a_e}|\mathbf{x}_{D_e}) = r_{a_e, b_{e_1}}(x_{a_e}, x_{b_{e_1}}) \cdot \prod_{j=2}^{d-2} c_{a_e, b_{e_j}|D_{e_j}} \left(R_{a_e|D_{e_j}}(x_{a_e}|\mathbf{x}_{D_{e_j}}), R_{b_{e_j}|D_{e_j}}(x_{b_{e_j}}|\mathbf{x}_{D_{e_j}}); \mathbf{x}_{D_{e_j}} \right),$$

where the product is empty if $d = 3$. Comparing the two identities, we get

$$r_{a_e|D_e \cup b_e}(x_{a_e}|\mathbf{x}_{D_e \cup b_e}) = r_{a_e|D_e}(x_{a_e}|\mathbf{x}_{D_e}) \cdot c_{a_e, b_e|D_e} \left(R_{a_e|D_e}(x_{a_e}|\mathbf{x}_{D_e}), R_{b_e|D_e}(x_{b_e}|\mathbf{x}_{D_e}); \mathbf{x}_{D_e} \right).$$

Since the partial derivative of $R_{a_e|D_e}(x_{a_e}|\mathbf{x}_{D_e})$ with respect to x_{a_e} is $r_{a_e|D_e}(x_{a_e}|\mathbf{x}_{D_e})$, the first of the two equalities in (5.4) follows by integration. The second equality in (5.4) is shown in the same way. $\square$

Proof of Theorem 5.7. The proof is by induction on d . For $d = 3$, we can, upon renumbering the nodes, assume the regular vine sequence has edge set $E_1 = \{\{1, 2\}, \{2, 3\}\}$, and (5.5) becomes

$$r(x_1, x_2, x_3) = r_{12}(x_1, x_2) r_{23}(x_2, x_3) \cdot c_{13;2} \left(R_{1|2}(x_1|x_2), R_{3|2}(x_3|x_2) \right).$$

By Proposition C.1, this function is a valid tail copula density with margins r_{12} and r_{23} . Furthermore,

$$r_{1|23}(x_1|x_2, x_3) = \frac{r(x_1, x_2, x_3)}{r_{23}(x_2, x_3)} = r_{12}(x_1, x_2) \cdot c_{13;2} \left(R_{1|2}(x_1|x_2), R_{3|2}(x_3|x_2) \right).$$

Integrating over x_1 yields, after the change of variables $u_1 = R_{1|2}(x_1|x_2)$, the identity

$$R_{1|23}(x_1|x_2, x_3) = C_{1|3;2} \left(R_{1|2}(x_1|x_2), R_{3|2}(x_3|x_2) \right).$$

The proof of the recursive formula for $R_{3|12}(x_3|x_1, x_2)$ is entirely similar.

For the induction step, let $d \geq 4$ and assume the stated properties are true for X-vine specifications on node sets of cardinality up to $d - 1$. To show that r defined in (5.5) is a d -variate tail copula density, we check the marginal constraint (2.4) and the homogeneity (2.5).

To show (2.5), let $\mathbf{x} \in (0, \infty)^d$ and $s \in (0, \infty)$. Since r_{a_e, b_e} is a bivariate tail copula density for each $e \in E_1$, we have, by (2.5),

$$\prod_{e \in E_1} r_{a_e, b_e}(sx_{a_e}, sx_{b_e}) = \prod_{e \in E_1} (s^{-1} r_{a_e, b_e}(x_{a_e}, x_{b_e})) = s^{-(d-1)} \prod_{e \in E_1} r_{a_e, b_e}(x_{a_e}, x_{b_e}),$$

as there are $d - 1$ edges in the first tree. Further, by the induction hypothesis and by scale equivariance (Lemma 3.2), we have $R_{a_e|D_e}(sx_{a_e}|s\mathbf{x}_{D_e}) = R_{a_e|D_e}(x_{a_e}|s\mathbf{x}_{D_e})$ and similarly for a_e replaced by b_e . It follows that r defined in (5.5) satisfies (2.5).

Next, we show that r defined in (5.5) satisfies (2.4). Let $e = \{f, g\} = (a_e, b_e; D_e)$ be the single edge in the deepest tree T_{d-1} , with $f, g \in E_{d-2}$ and where the labels are such that $a_e \in \mathcal{C}_f$ and $b_e \in \mathcal{C}_g$. The regular vine sequence $\mathcal{V}$ on $[d]$ induces regular vine sequences $\mathcal{V}_f$ and $\mathcal{V}_g$ on the node sets

$$A_f = [d] \setminus \{b_e\} = D_e \cup \{a_e\} \quad \text{and} \quad A_g = [d] \setminus \{a_e\} = D_e \cup \{b_e\},$$

respectively. The X-vine specification $(\mathcal{V}, \mathcal{R}, \mathcal{C})$ can be restricted in similar ways, by keeping only those bivariate (tail) copula densities associated to edges with complete unions contained in A_f and A_g , respectively, yielding X-vine specifications $(\mathcal{V}_f, \mathcal{R}_f, \mathcal{C}_f)$ and $(\mathcal{V}_g, \mathcal{R}_g, \mathcal{C}_g)$. By the induction hypothesis, these two X-vine specifications produce $(d - 1)$ -dimensional tail copula densities r_{A_f} and r_{A_g} as in (5.5) but with factors restricted to those edges whose complete unions are contained in A_f and A_g , respectively. Let b_f be the other node in $\mathcal{C}_f$ besides a_e , i.e., $\mathcal{C}_f = \{a_e, b_f\}$; similarly, let a_g be the other node in $\mathcal{C}_g$ besides b_e , i.e., $\mathcal{C}_g = \{a_g, b_e\}$. We have

$$D_e = D_f \cup \{b_f\} = D_g \cup \{a_g\}.$$

By the induction hypothesis applied to $(\mathcal{V}_f, \mathcal{R}_f, \mathcal{C}_f)$, we can obtain $R_{a_e|D_e} = R_{a_e|D_f \cup b_f}$ by the recursive formula (5.6); the same is true for $R_{b_e|D_e} = R_{b_e|D_g \cup a_g}$.

For each $j \in \{1, \dots, d - 1\}$, there exists a unique edge $e_j \in E_j$ such that $a_e \in A_{e_j}$ and then in fact $a_e \in \mathcal{C}_{e_j}$ (Lemma B.2); clearly $e_{d-1} = e$ and $e_{d-2} = f$. For $j \in \{1, \dots, d - 1\}$, let $b_j = \mathcal{C}_{e_j} \setminus \{a_e\}$ be the other node in the conditioned set of e_j besides a_e ; we have $b_{d-1} = b_e$ and $b_{d-2} = b_f$. Isolating in (5.5) the factors indexed by edges $e_1, \dots, e_{d-1}$, we obtain

$$r(\mathbf{x}) = r_{A_g}(\mathbf{x}_{A_g}) \cdot r_{a_e, b_1}(x_{a_e}, x_{b_1}) \cdot \prod_{j=2}^{d-1} c_{a_e, b_j; D_{e_j}} \left(R_{a_e; D_{e_j}}(x_{a_e} | \mathbf{x}_{D_{e_j}}), R_{b_j; D_{e_j}}(x_{b_j} | \mathbf{x}_{D_{e_j}}) \right).$$

By the chain rule, the derivative of the function

$$x_{a_e} \mapsto C_{a_e|b_e; D_e} \left(R_{a_e|D_e}(x_{a_e} | \mathbf{x}_{D_e}), R_{b_e|D_e}(x_{b_e} | \mathbf{x}_{D_e}) \right)$$

is equal to the function

$$x_{a_e} \mapsto c_{a_e, b_e; D_e} \left(R_{a_e|D_e}(x_{a_e} | \mathbf{x}_{D_e}), R_{b_e|D_e}(x_{b_e} | \mathbf{x}_{D_e}) \right) \cdot \frac{\partial}{\partial x_{a_e}} R_{a_e|D_e}(x_{a_e} | \mathbf{x}_{D_e}).$$

Recall $D_e \cup \{a_e\} = A_f$, $a_e \in \mathcal{C}_f$, and $f = e_{d-2}$. Applying the induction hypothesis to $(\mathcal{V}_f, \mathcal{R}_f, \mathcal{C}_f)$, we find

$$\begin{aligned} \frac{\partial}{\partial x_{a_e}} R_{a_e|D_e}(x_{a_e} | \mathbf{x}_{D_e}) &= \frac{r_{a_e \cup D_e}(\mathbf{x}_{a_e \cup D_e})}{r_{D_e}(\mathbf{x}_{D_e})} \\ &= r_{a_e, b_1}(x_{a_e}, x_{b_1}) \cdot \prod_{j=2}^{d-2} c_{a_e, b_j; D_{e_j}} \left(R_{a_e; D_{e_j}}(x_{a_e} | \mathbf{x}_{D_{e_j}}), R_{b_j; D_{e_j}}(x_{b_j} | \mathbf{x}_{D_{e_j}}) \right). \end{aligned}$$

Combining these identities, we get

$$r(\mathbf{x}) = r_{A_g}(\mathbf{x}_{A_g}) \cdot \frac{\partial}{\partial x_{a_e}} C_{a_e|b_e; D_e} \left(R_{a_e|D_e}(x_{a_e} | \mathbf{x}_{D_e}), R_{b_e|D_e}(x_{b_e} | \mathbf{x}_{D_e}) \right). \quad (\text{B.2})$$

It follows that

$$\int_0^\infty r(\mathbf{x}) dx_{a_e} = r_{A_g}(\mathbf{x}_{A_g}).$$

Since r_{A_g} is a $(d-1)$ -variate tail copula density, we can fix variable $x_i \in (0, \infty)$ with $i \in A_g$ and integrate further over all $(d-2)$ remaining variables in A_g , to find that $\int_{(0, \infty)^{d-1}} r(\mathbf{x}) d\mathbf{x}_{\setminus i} = 1$. Repeating the same argument with the roles of f and g interchanged leads to

$$r(\mathbf{x}) = r_{A_f}(\mathbf{x}_{A_f}) \cdot \frac{\partial}{\partial x_{b_e}} C_{b_e|a_e; D_e}(R_{b_e|D_e}(x_{b_e}|\mathbf{x}_{D_e}), R_{a_e|D_e}(x_{a_e}|\mathbf{x}_{D_e}))$$

and thus

$$\int_0^\infty r(\mathbf{x}) dx_{b_e} = r_{A_f}(\mathbf{x}_{A_f}),$$

from which also $\int_{(0, \infty)^{d-1}} r(\mathbf{x}) d\mathbf{x}_{\setminus i} = 1$ for $i \in A_f$ and $x_i \in (0, \infty)$. As $A_f \cup A_g = A_e = [d]$, we have shown (2.4).

Finally, we show the recursive formulas (5.6). From $D_e \cup \{b_e\} = A_g$, it follows that

$$\begin{aligned} \frac{\partial}{\partial x_{a_e}} R_{a_e|D_e \cup b_e}(x_{a_e}|\mathbf{x}_{D_e \cup b_e}) &= r_{a_e|D_e \cup b_e}(x_{a_e}|\mathbf{x}_{D_e \cup b_e}) = \frac{r(\mathbf{x})}{r_{A_g}(\mathbf{x}_{A_g})} \\ &= \frac{\partial}{\partial x_{a_e}} C_{a_e|b_e; D_e}(R_{a_e|D_e}(x_{a_e}|\mathbf{x}_{D_e}), R_{b_e|D_e}(x_{b_e}|\mathbf{x}_{D_e})), \end{aligned}$$

where we used (B.2) in the last step. Integrating over x_{a_e} yields the first identity in (5.6). The second identity follows in the same way. The proof of Theorem 5.7 is complete. $\square$

B.5 Proofs for Section 6

Proof of Lemma 6.2. The proof is by induction on d . If $d = 3$, we can, after relabeling the nodes if necessary, assume that $E_1 = \{\{1, 2\}, \{2, 3\}\}$. Then $E_2 = \{e_2\}$ with $e_2 = (13; 2)$. The three permutations are then as follows:

- For $j = 1$, we have $\sigma_1 = (1, 2, 3)$ with $e_{1,1} = \{1, 2\}$.
- For $j = 2$, we have $\sigma_2 = (2, 1, 3)$ with $e_{2,1} = \{1, 2\}$ or $\sigma_2 = (2, 3, 1)$ with $e_{2,1} = \{2, 3\}$.
- For $j = 3$, we have $\sigma_3 = (3, 2, 1)$ with $e_{3,1} = \{2, 3\}$.

Next, let $d \geq 4$ and assume the statement is true for regular vine sequence on $d-1$ elements. Let $e = (a_e, b_e; D_e) = \{f, g\}$ be the single edge in E_{d-1} , with $f, g \in E_{d-2}$ labeled in such a way that $a_e \in \mathcal{C}_f$ and $b_e \in \mathcal{C}_g$. The regular vine sequence $\mathcal{V}$ induces regular vine sequences $\mathcal{V}_f$ and $\mathcal{V}_g$ on the node sets $A_f = [d] \setminus \{b_e\}$ and $A_g = [d] \setminus \{a_e\}$, both of which have $d-1$ elements; the edges $\bar{e}$ of the two regular vine sequences $\mathcal{V}_f$ and $\mathcal{V}_g$ are those of $\mathcal{V}$ that satisfy $A_{\bar{e}} \subseteq A_f$ or $A_{\bar{e}} \subseteq A_g$, respectively.

Let $j \in [d]$. If $j \neq b_e$, then we can apply the induction hypothesis to $\mathcal{V}_f$ to find a bijection $\sigma_{f,j} : \{1, \dots, d-1\} \rightarrow A_f = [d] \setminus \{b_e\}$ such that (i) $\sigma_{f,j}(1) = j$ and (ii) $\sigma_{f,j}(k)$ belongs to the conditioned set of some edge e_{k-1} in the $(k-1)$ th tree of $\mathcal{V}_f$ as well as $\{\sigma_{f,j}(i) : i = 1, \dots, k\} = A_{e_{k-1}}$, for all $k \in \{2, \dots, d-1\}$. Extend $\sigma_{f,j}$ to a map σ_j on $[d]$ by $\sigma_j(d) = b_e$. Then σ_j meets all the requirements.

If $j = b_e$, then $j \neq a_e$, and we can apply the same argument to $\mathcal{V}_g$. $\square$

Proof of Proposition 6.3. The distribution of $Z_j^{(j)}$ is that of Z_j given that $Z_j < 1$, and this is the uniform distribution on $(0, 1)$. The fact that the distribution of $\mathbf{Z}^{(j)}$ defined in (6.4) is a consequence of the inverse Rosenblatt transform.

Further, the permutation σ_j in Lemma 6.2 is such that for every $k \in \{2, \dots, d\}$, there exists an edge $e_{j,k-1} \in E_{k-1}$ such that $\sigma_j(k) \in \mathcal{C}_{e_{j,k-1}} = \{a_{e_{j,k-1}}, b_{e_{j,k-1}}\}$ and $\{\sigma_j(i) : i = 1, \dots, k\} = A_{e_{j,k-1}} = \{a_{e_{j,k-1}}, b_{e_{j,k-1}}\} \cup D_{e_{j,k-1}}$. These properties imply the stated representation of the conditional quantile function. $\square$

B.6 Proofs for Section 7

Proof of Proposition 7.1. (i) By Definition 6.1, the density of $\mathbf{Z}$ is the function $\mathbf{z} \mapsto r(\mathbf{z}) \mathbb{1}_{\mathbb{L}}(\mathbf{z})/R(\mathbb{L})$ with $\mathbb{L} = \{\mathbf{x} : \min(\mathbf{z}) < 1\}$. Let $K \subset [d]$ be non-empty and let $\mathbf{z}_K \in (0, \infty)^K$ be such that $\min \mathbf{z} < 1$. The marginal density of $\mathbf{Z}_K = (Z_k)_{k \in K}$ evaluated in $\mathbf{z}_K$ is

$$\int_{(0, \infty)^{[d] \setminus K}} \frac{r(\mathbf{z}) \mathbb{1}_{\mathbb{L}}(\mathbf{z})}{R(\mathbb{L})} d\mathbf{z}_{[d] \setminus K}.$$

Since $\min \mathbf{z}_K < 1$, we also have $\min \mathbf{z} < 1$ for $\mathbf{z}_{[d] \setminus K} \in (0, \infty)^{[d] \setminus K}$. As a consequence, the indicator function in the integral is always equal to one and, by Eq. (2.7), the density of $\mathbf{Z}_K$ evaluated in $\mathbf{z}_K$ simplifies to $r_K(\mathbf{z}_K)/r(\mathbb{L})$.

Applying this identity to $K = J$ and $K = I \cup J$, we find, for $\mathbf{z}_I \in (0, \infty)^I$ and for $\mathbf{z}_J \in (0, \infty)^J$ such that $\min \mathbf{z}_J < 1$ and $r_J(\mathbf{z}_J) > 0$, the conditional density of $\mathbf{Z}_I$ given $\mathbf{Z}_J = \mathbf{z}_J$ evaluated in $\mathbf{z}_I$:

$$\frac{r_{I \cup J}(\mathbf{z}_{I \cup J})/R(\mathbb{L})}{r_J(\mathbf{z}_J)/R(\mathbb{L})} = \frac{r_{I \cup J}(\mathbf{z}_{I \cup J})}{r_J(\mathbf{z}_J)} = r_{I|J}(\mathbf{z}_I|\mathbf{z}_J).$$

(ii) Let $\mathbf{z}_J \in (0, \infty)^J$ be such that $\min \mathbf{z}_J < 1$ and $r_J(\mathbf{z}_J) > 0$. By statement (i) and by Eq. (3.1) in Sklar's theorem (Proposition 3.1), the conditional density of $(R_{i|J}(Z_i|\mathbf{Z}_J))_{i \in I}$ given $\mathbf{Z}_J = \mathbf{z}_J$ is $c_{I;J}$. Since, by the simplifying assumption, the latter copula density does not depend on $\mathbf{z}_J$, the random vector $(R_{i|J}(Z_i|\mathbf{Z}_J))_{i \in I}$ is independent of $\mathbf{Z}_J$, conditionally on $\min \mathbf{Z}_J < 1$. $\square$

C Converse of Sklar's theorem for tail copula densities

As a converse to Sklar's theorem for tail copula densities (Proposition 3.1), we can combine several tail copula densities along a scale-invariant family of copula densities to form a tail copula density in higher dimensions. With slight abuse of notation, write $i \cup J := \{i\} \cup J$ for $i \in I$. A family of k -variate copula densities $\{c(\cdot; \theta) : \theta \in \Theta\}$ indexed by a Borel set $\Theta \subseteq \mathbb{R}^m$ is *jointly measurable* if the map $(\mathbf{u}, \theta) \mapsto c(\mathbf{u}; \theta)$ is Borel measurable on $(0, 1)^k \times \Theta$. Joint measurability is trivial if $c(\cdot; \theta)$ does not depend on θ .

Proposition C.1 (Sklar's theorem: converse part). *Let $I \cup J$ be a partition of $[d]$ into two disjoint, non-empty sets, with $d \geq 3$ and $|I| \geq 2$. Let $(r_{i \cup J})_{i \in I}$ be an $|I|$ -tuple of $(|J| + 1)$ -variate tail copula densities with common margin r_J . Let $\{c_{I;J}(\cdot; \mathbf{x}_J) : \mathbf{x}_J \in (0, \infty)^J\}$ be a jointly measurable family of $|I|$ -variate copula densities such that (3.2) holds. Then the function $r : (0, \infty)^d \rightarrow [0, \infty)$ defined by*

$$r(\mathbf{x}) := r_J(\mathbf{x}_J) \cdot \prod_{i \in I} r_{i|J}(x_i|\mathbf{x}_J) \cdot c_{I;J}(R_{i|J}(x_i|\mathbf{x}_J), i \in I; \mathbf{x}_J), \quad (\text{C.1})$$

for $\mathbf{x} \in (0, \infty)^d$ such that $r_J(\mathbf{x}_J) > 0$ and zero otherwise is a d -variate tail copula density with margins $r_{i \cup J}$ for $i \in I$.

Proof of Proposition C.1. First, we show that r is a tail copula density. To this end, we need to show two properties: the normalization constraint (2.4) and the homogeneity relation (2.5).

To show the homogeneity relation (2.5), let $t \in (0, \infty)$, and let $\mathbf{x} \in (0, \infty)^d$ be such that $r_J(\mathbf{x}_J) > 0$. Then

$$\begin{aligned} r(t\mathbf{x}) &= r_J(t\mathbf{x}_J) \cdot \prod_{i \in I} r_{i|J}(tx_i|t\mathbf{x}_J) \cdot c_{I;J}(R_{i|J}(tx_i|t\mathbf{x}_J), i \in I; t\mathbf{x}_J) \\ &= (r_J(t\mathbf{x}_J))^{1-|I|} \cdot \prod_{i \in I} r_{i \cup J}(tx_i, t\mathbf{x}_J) \cdot c_{I;J}(R_{i|J}(tx_i|t\mathbf{x}_J), i \in I; t\mathbf{x}_J). \end{aligned}$$

By Lemma 3.2, the last factor does not depend on t at all. By (2.5), the scalar t can be taken out of the tail copula densities yielding a factor $(t^{1-|J|})^{1-|I|} \cdot (t^{1-(1+|J|)})^{|I|}$. The exponent of t is, as required,

$$(1 - |J|)(1 - |I|) - |I| \cdot |J| = 1 - |I| - |J| = 1 - d.$$

To show the normalization constraint (2.4), let $k \in \{1, \dots, d\}$ and $x_k \in (0, \infty)$. We need to show that $\int_{(0, \infty)^{d-1}} r(\mathbf{x}) d\mathbf{x}_{\setminus k} = 1$, where the integral is over all coordinates except for the k th one. There are two cases: $k \in I$ or $k \in J$.

First, if $k \in I$, the integral of interest is

$$\int_{(0, \infty)^J} r_J(\mathbf{x}_J) \left\{ \int_{(0, \infty)^{I \setminus k}} \prod_{i \in I} r_{i|J}(x_i|\mathbf{x}_J) \cdot c_{I;J}(R_{i|J}(x_i|\mathbf{x}_J), i \in I; \mathbf{x}_J) d\mathbf{x}_{I \setminus k} \right\} d\mathbf{x}_J.$$

From the inner integral, the factor $r_{k|J}(x_k|\mathbf{x}_J)$ can be taken out. The function that remains in the inner integral has integral equal to one: substitute $u_i = R_{i|J}(x_i|\mathbf{x}_J)$ for $i \in I \setminus k$ and use the fact that the copula density $c_{I;J}$ has uniform margins. The integral simplifies to

$$\int_{(0, \infty)^J} r_J(\mathbf{x}_J) \cdot r_{k|J}(x_k|\mathbf{x}_J) d\mathbf{x}_J = \int_{(0, \infty)^J} r_{k \cup J}(x_k, \mathbf{x}_J) d\mathbf{x}_J = 1,$$

as $r_{k \cup J}$ is a tail copula density.

Second, if $k \in J$, the integral of interest is

$$\int_{(0, \infty)^{J \setminus k}} r_J(\mathbf{x}_J) \left\{ \int_{(0, \infty)^I} \prod_{i \in I} r_{i|J}(x_i|\mathbf{x}_J) \cdot c_{I;J}(R_{i|J}(x_i|\mathbf{x}_J), i \in I; \mathbf{x}_J) d\mathbf{x}_I \right\} d\mathbf{x}_{J \setminus k}.$$

The inner integral is equal to one, as the integrand is a probability density function, by Sklar's theorem. The integral of interest thus simplifies to

$$\int_{(0,\infty)^{J \setminus k}} r_J(\mathbf{x}_J) d\mathbf{x}_{J \setminus k} = 1,$$

since r_J is a tail copula density.

The proof that r has the required margins $r_{i \cup J}$ for $i \in I$ is similar to the case $k \in I$ in the previous part of the proof.

Finally, statement (3.1) easily follows by applying Sklar's theorem to $r_{I|J}(\cdot | \mathbf{x}_J)$. $\square$

In the special case that J is a singleton, $J = \{j\}$, we have $I = [d] \setminus j$, the $(d-1)$ tail copula densities $(r_{i,j})_{i \in [d] \setminus j}$ are bivariate, and the common-margin assumption is automatic by (2.2); in particular, $r_{i|j}(x_i | x_j) = r_{i,j}(x_i, x_j)$. Further, the scale-invariance property (3.2) enforces the simplifying assumption, so that the construction (C.1) becomes

$$r(\mathbf{x}) := c(R_{i|j}(x_i | x_j), i \in [d] \setminus j) \cdot \prod_{i \in [d] \setminus j} r_{i,j}(x_i, x_j),$$

with c an arbitrary $(d-1)$ -variate copula density. Any $(d-1)$ -tuple of bivariate tail copula densities can thus be combined along any $(d-1)$ -variate copula density to form a d -variate tail copula density. The X-vine construction in Section 5 is in the same spirit but without requiring a common pivot variable j and with a further break-down of c into pair copulas.

D Exponent measure densities as X-vines

Recall the exponent measure Λ in (2.9) and its density λ in (2.12). To facilitate linking up our results with the literature on multivariate extreme value theory, we present the main formulas in the paper in terms of λ too.

For a proper, non-empty subset $J \subset [d]$, the marginal exponent measure density is

$$\lambda_J(\mathbf{y}_J) = \int_{\mathbf{y}_{\setminus J} \in (0,\infty)^{d-|J|}} \lambda(\mathbf{y}) d\mathbf{y}_{\setminus J}, \quad \mathbf{y}_J \in (0,\infty)^{|J|}.$$

For non-empty, disjoint subsets $I, J \subset [d]$, we write

$$\lambda_{I|J}(\mathbf{y}_I | \mathbf{y}_J) = \frac{\lambda_{I \cup J}(\mathbf{y}_{I \cup J})}{\lambda_J(\mathbf{y}_J)},$$

for $\mathbf{y}_I \in (0,\infty)^I$ and $\mathbf{y}_J \in (0,\infty)^J$ such that $\lambda_J(\mathbf{y}_J) > 0$. If $I = \{i\}$, we consider the conditional tail function

$$\bar{\Lambda}_{i|J}(y_i | \mathbf{y}_J) = \int_{y_i}^{\infty} \lambda_{i|J}(t | \mathbf{y}_J) dt = R_{i|J}(1/y_i | 1/\mathbf{y}_J)$$

as well as its inverse $\bar{\Lambda}_{i|J}^{-1}$ defined by

$$\bar{\Lambda}_{i|J}^{-1}(u_i | \mathbf{y}_J) = y_i \iff \bar{\Lambda}_{i|J}(y_i | \mathbf{y}_J) = u_i$$

for $u_i \in (0,1)$. The copula density in (3.1) becomes

$$c_{I;J}(\mathbf{u}_I; 1/\mathbf{y}_J) = \frac{\lambda(\mathbf{y}_I | \mathbf{y}_J)}{\prod_{i \in I} \lambda_{i|J}(y_i | \mathbf{y}_J)} \quad \text{with} \quad y_i = \bar{\Lambda}_{i|J}^{-1}(u_i | \mathbf{y}_J). \quad (\text{D.1})$$

Under the simplifying assumption, the exponent measure density λ corresponding to the X-vine tail copula density r in (5.1) is

$$\lambda(\mathbf{y}) = \prod_{j \in [d]} y_j^{2 \deg_1(j) - 2} \cdot \prod_{e \in E_1} \lambda_{a_e, b_e}(y_{a_e}, y_{b_e}) \cdot \prod_{j=2}^{d-1} \prod_{e \in E_j} c_{a_e, b_e; D_e}(\bar{\Lambda}_{a_e | D_e}(y_{a_e} | \mathbf{y}_{D_e}), \bar{\Lambda}_{b_e | D_e}(y_{b_e} | \mathbf{y}_{D_e})).$$

Here, $\deg_1(j)$ is the degree (number of neighbors) of node j in tree T_1 and the pair copula densities can be found from λ by

$$c_{a_e, b_e; D_e}(u_{a_e}, u_{b_e}) = \frac{\lambda_{a_e, b_e | D_e}(y_{a_e}, y_{b_e} | \mathbf{y}_{D_e})}{\lambda_{a_e | D_e}(y_{a_e} | \mathbf{y}_{D_e}) \cdot \lambda_{b_e | D_e}(y_{b_e} | \mathbf{y}_{D_e})}$$

for $(u_{a_e}, u_{b_e}) \in (0,1)^2$, with $y_{a_e} = \bar{\Lambda}_{a_e | D_e}^{-1}(u_{a_e} | \mathbf{y}_{D_e})$ and $y_{b_e} = \bar{\Lambda}_{b_e | D_e}^{-1}(u_{b_e} | \mathbf{y}_{D_e})$.

E X-vines as limits of regular vine copula densities

The aim of this section is to link up regular vine copula densities as introduced in Bedford and Cooke (2002) with X-vine tail copula densities, as claimed in Remark 5.8. To this end, we start from a d -variate regular vine copula density c and we show that it is attracted to a d -variate X-vine tail copula density r along exactly the same regular vine sequence as c . Moreover, c and r share the same pair copula densities on all edges except for those in the first tree. For edges in the first tree, the pair copula densities of c are to be replaced by the bivariate tail copula densities to which they are attracted.

Let $\mathcal{V} = (T_j)_{j=1}^{d-1}$ with $T_j = (N_j, E_j)$ be a regular vine sequence on d elements, for $d \geq 3$. For every edge $e = (a_e, b_e; D_e) \in \bigcup_{j=1}^{d-1} E_j$, let $c_{a_e, b_e; D_e}$ be a bivariate copula density, with copula $C_{a_e, b_e; D_e}$ and conditional distribution functions $C_{a_e|b_e; D_e}$ and $C_{b_e|a_e; D_e}$. For $e \in E_1$, the conditioning set D_e is empty and we simply write c_{a_e, b_e} and so on. Then there is a d -variate regular vine copula density c given by

$$c(\mathbf{u}) = \prod_{e \in E_1} c_{a_e, b_e}(u_{a_e}, u_{b_e}) \cdot \prod_{j=2}^{d-1} \prod_{e \in E_j} c_{a_e, b_e; D_e}(C_{a_e|D_e}(u_{a_e}|\mathbf{u}_{D_e}), C_{b_e|D_e}(u_{b_e}|\mathbf{u}_{D_e})), \quad (\text{E.1})$$

where, for a random vector $\mathbf{U}$ with density c , the conditional distribution function of (U_{a_e}, U_{b_e}) given $\mathbf{U}_{D_e} = \mathbf{u}_{D_e}$ is

$$C_{a_e, b_e|D_e}(u_{a_e}, u_{b_e}|\mathbf{u}_{D_e}) = C_{a_e, b_e; D_e}(C_{a_e|D_e}(u_{a_e}|\mathbf{u}_{D_e}), C_{b_e|D_e}(u_{b_e}|\mathbf{u}_{D_e})). \quad (\text{E.2})$$

Assumption E.1. Let c be a d -variate regular vine copula density as in (E.1), with $d \geq 3$.

(i) For every edge $e \in E_1$, there exists a bivariate tail copula density r_{a_e, b_e} such that

$$\lim_{t \searrow 0} t \cdot c_{a_e, b_e}(tx, ty) = r_{a_e, b_e}(x, y), \quad (x, y) \in (0, \infty)^2.$$

(ii) For every edge $e \in \bigcup_{j=2}^{d-1} E_j$, the pair copula density $c_{a_e, b_e; D_e}$ is continuous.

Proposition E.2. Under Assumption E.1, we have

$$\lim_{t \searrow 0} t^{d-1} \cdot c(t\mathbf{x}) = r(\mathbf{x}), \quad \mathbf{x} \in (0, \infty)^d,$$

where the X-vine tail copula density r is generated by the triple $(\mathcal{V}, \mathcal{R}, \mathcal{C})$ as in Theorem 5.7 with

- $\mathcal{V}$ the same regular vine sequence as in (E.1),
- $\mathcal{R} = \{r_{a_e, b_e} : e \in E_1\}$ for r_{a_e, b_e} in Assumption E.1(i), and
- $\mathcal{C}$ the same bivariate copula densities for edges $e \in E_j$ with $j \geq 2$ as in (E.1).

In particular, the copula C of c has tail copula R with tail copula density r .

Proof of Proposition E.2. Let $\mathbf{x} \in (0, \infty)^d$. As there $d - 1$ edges in the first tree, we have

$$t^{d-1} \cdot c(t\mathbf{x}) = \prod_{e \in E_1} (t \cdot c_{a_e, b_e}(tx_{a_e}, tx_{b_e})) \cdot \prod_{j=2}^{d-1} \prod_{e \in E_j} c_{a_e, b_e; D_e}(C_{a_e|D_e}(tx_{a_e}|t\mathbf{x}_{D_e}), C_{b_e|D_e}(tx_{b_e}|t\mathbf{x}_{D_e})).$$

The first product on the right-hand side converges to $\prod_{e \in E_1} r_{a_e, b_e}(x_{a_e}, x_{b_e})$ as $t \rightarrow 0$ by Assumption E.1(i). It remains to show that for every $j \in \{2, \dots, d-1\}$, every $e \in E_j$ and every $i \in \mathcal{C}_e = \{a_e, b_e\}$, we have

$$\lim_{t \searrow 0} C_{i|D_e}(tx_i|t\mathbf{x}_{D_e}) = R_{i|D_e}(x_i|\mathbf{x}_{D_e}). \quad (\text{E.3})$$

By properties of regular vine sequences as explained in Section 2.2, there exists an edge $f = (a_f, b_f; D_f) \in E_{j-1}$ such that $f \in e$ and $i \in \mathcal{C}_f = \{a_f, b_f\}$ and $A_f = D_e \cup \{i\}$. The proof of (E.3) is by recursion on j .

If $j = 2$, then $f \in E_1$ so that $D_f = \emptyset$ while D_e is a singleton, say $D_e = \{d_e\}$, and thus $f = \{i, d_e\}$. But then (E.3) follows from Assumption E.1(i) and Lemma E.3 below.

Let $j \geq 3$ and assume (E.3) has been shown for edges in $E_1 \cup \dots \cup E_{j-1}$. Let k be the other node in $\mathcal{C}_f$ but i , that is, $\{a_f, b_f\} = \mathcal{C}_f = \{i, k\}$. Then by (E.2) applied to f , we have

$$C_{i, k|D_f}(tx_i, tx_k|t\mathbf{x}_{D_f}) = C_{i, k; D_f}(C_{i|D_f}(tx_i|t\mathbf{x}_{D_f}), C_{k|D_f}(tx_k|t\mathbf{x}_{D_f}))$$

and thus

$$\begin{aligned} C_{i|D_e}(tx_i|t\mathbf{x}_{D_e}) &= C_{i|D_f \cup k}(tx_i|t\mathbf{x}_{D_f \cup k}) \\ &= C_{i|k; D_f}(C_{i|D_f}(tx_i|t\mathbf{x}_{D_f}), C_{k|D_f}(tx_k|t\mathbf{x}_{D_f})). \end{aligned}$$

By the induction hypothesis, the two arguments of $C_{i|k;D_f}$ converge to $R_{i|D_f}(x_i|\mathbf{x}_{D_f})$ and $R_{k|D_f}(x_k|\mathbf{x}_{D_f})$, respectively. As $f \in E_{j-1}$ and $j \geq 3$, the pair copula density $c_{i,k;D_f}$ is continuous by Assumption E.1(ii). In view of Corollary E.5 below, $C_{i|k;D_f}$ is continuous in its two arguments too, and thus

$$\lim_{t \searrow 0} C_{i|D_e}(tx_i|t\mathbf{x}_{D_e}) = C_{i|k;D_f}(R_{i|D_f}(x_i|\mathbf{x}_{D_f}), R_{k|D_f}(x_k|\mathbf{x}_{D_f})).$$

By the recursive property (5.6), the last expression is equal to $R_{i|D_f \cup k}(x_i|\mathbf{x}_{D_f \cup k})$. As $D_f \cup k = D_e$, the convergence (E.3) follows. $\square$

Lemma E.3. *Let c be a bivariate copula density and r a bivariate tail copula density. Assume that*

$$\lim_{t \searrow 0} t \cdot c(tx, ty) = r(x, y), \quad (x, y) \in (0, \infty)^2.$$

The copula C of c and the tail copula R of r then satisfy

$$\begin{aligned} \lim_{t \rightarrow 0} C_{1|2}(tx|ty) &= R_{1|2}(x|y), \\ \lim_{t \rightarrow 0} C_{2|1}(ty|tx) &= R_{2|1}(y|x). \end{aligned}$$

Proof of Lemma E.3. By symmetry, it is sufficient to show the first relation. Fix $y \in (0, \infty)$ and define $f_t(x) = t \cdot c(tx, ty)$ for $x \in (0, \infty)$. By assumption, $\lim_{t \rightarrow 0} f_t(x) = r(x, y)$ for all $x \in (0, \infty)$. Moreover, for all $t \in (0, 1/y)$, we have

$$\int_0^\infty f_t(x) dx = \int_0^\infty c(u, ty) du = 1 = \int_0^\infty r(x, y) dx.$$

By Scheffé's lemma applied to the functions f_t and their point-wise limit $x \mapsto f(x) = r(x, y)$, it follows that

$$\lim_{t \rightarrow \infty} \int_0^\infty |t \cdot c(tx, ty) - r(x, y)| dx = 0.$$

But then

$$\begin{aligned} |C_{1|2}(tx|ty) - R_{1|2}(x|y)| &= \left| \int_0^{tx} c(s, ty) ds - \int_0^x r(s, y) ds \right| \\ &\leq \int_0^x |t \cdot c(ts, ty) - r(s, y)| ds \rightarrow 0, \quad t \rightarrow 0. \end{aligned} \quad \square$$

Lemma E.4. *Let I and J be non-empty open intervals and let $f : I \times J \rightarrow [0, \infty)$ be a continuous function such that $f_2 : J \rightarrow [0, \infty) : y \mapsto \int_I f(x, y) dx$ is continuous. Then the function*

$$(x, y) \mapsto \int_I f(s, y) \mathbb{1}\{s \leq x\} ds$$

is continuous on $\bar{I} \times J$ too, where $\bar{I}$ is the closure of I in $[-\infty, +\infty]$.

Proof of Lemma E.4. For $y \in J$, define $f_y : I \rightarrow [0, \infty) : x \mapsto f_y(x) := f(x, y)$. Then $\lim_{y' \rightarrow y} f_{y'}(x) = f_y(x)$ for all $(x, y) \in I \times J$ and $\int_I f_{y'}(x) dx = f_2(y') \rightarrow f_2(y) = \int_I f_y(x) dx$ as $y' \rightarrow y$ in J . By Scheffé's lemma, it follows that

$$\lim_{y' \rightarrow y} \int_I |f_{y'}(x) - f_y(x)| dx = 0. \quad (\text{E.4})$$

But then, as $(x', y') \rightarrow (x, y)$ in $\bar{I} \times J$, we have

$$\begin{aligned} &\left| \int_I f(s, y') \mathbb{1}\{s \leq x'\} ds - \int_I f(s, y) \mathbb{1}\{s \leq x\} ds \right| \\ &\leq \int_I |f(s, y') - f(s, y)| ds + \int_I f(s, y) |\mathbb{1}\{s \leq x\} - \mathbb{1}\{s \leq x'\}| ds \rightarrow 0. \end{aligned}$$

The first term on the right-hand side converges to zero by (E.4) and the second term vanishes asymptotically by the dominated convergence theorem. $\square$

Corollary E.5. *Let c be a bivariate copula density and let $C_{1|2}(u|v) = \int_0^u c(w, v) dw$ for $(u, v) \in [0, 1] \times (0, 1)$. If c is continuous, then so is $C_{1|2}(\cdot|\cdot)$.*

Proof of Corollary E.5. Apply Lemma E.4 to $f = c$ and $I = J = (0, 1)$. $\square$

F Additional numerical results

F.1 Parametric relationships between tail dependence coefficients and bivariate tail copula parameters

Dependence measures facilitate the comparison of parameter estimates across different models on a consistent scale. In particular, closed-form expressions for the bivariate tail dependence coefficient $\chi = R(1, 1)$ are well-established for widely used parametric families of tail copula densities R , albeit it in the form of exponent measures.

- Hüsler–Reiss model, $r(x_1, x_2; \theta) = x_1^{-1} \frac{1}{\sqrt{2\pi\theta}} \exp \left[-\frac{1}{2} \left\{ \frac{\log(x_1/x_2) - \theta/2}{\sqrt{\theta}} \right\}^2 \right]$ for $\theta > 0$: we have $\chi = 2 - 2\Phi(\sqrt{\theta}/2)$ where Φ is the standard normal cumulative distribution function.
- Negative logistic model, $r(x_1, x_2; \theta) = (1 + \theta)(x_1 x_2)^{-(\theta+1)}(x_1^{-\theta} + x_2^{-\theta})^{-1/\theta-2}$ for $\theta > 0$: we have $\chi = 2^{-1/\theta}$.
- Logistic model, $r(x_1, x_2; \theta) = (\theta - 1)(x_1 x_2)^{\theta-1}(x_1^\theta + x_2^\theta)^{1/\theta-2}$ for $\theta > 1$: we have $\chi = 2 - 2^{1/\theta}$.
- Dirichlet model, $r(x_1, x_2; \theta) = \frac{2\Gamma(2\theta)}{\Gamma(\theta)^2} (x_1 + x_2)^{-2\theta-1} (x_1 x_2)^\theta$ for $\theta > 0$: we have $\chi = \int_0^1 I_{\frac{1}{1+x}}(\theta + 1, \theta) dx$ where $I_{\frac{1}{1+x}}$ is the regularised (incomplete) Beta function.

For the parametric relationships between Kendall’s tau τ_e and pair-copula parameters, see for instance Czado (2019, Section 3.5).

F.2 Additional simulation results to those in Section 8

F.2.1 Box-plots of dependence measure estimates with varying (n, k)

Given the X-vine specification in Fig. 2, we create box-plots of dependence measure estimates from the specified bivariate parametric families for each edge across the four trees, using various (n, k) , following the same set-up as in Section 8.1. We consider three different sample sizes, $n \in \{1\,000, 2\,000, 4\,000\}$, and quantiles, $k/n \in \{0.02, 0.05, 0.1\}$.

F.2.2 Box-plots of pseudo-maximum likelihood estimates of parameters of bivariate tail copulas

Using the same X-vine specification in Fig. 2, we evaluate bias and variance of maximum pseudo-likelihood parameter estimates of tail copula densities. Specifically, we create box-plots of estimates for the Hüsler–Reiss tail copula density associated with the edge $e_{12} \in E_1$, using the same combinations of (n, k) defined in Appendix F.2.1. An increase in the effective sample size n_{D_e} reduces variability, while a decreasing threshold k/n leads to smaller bias, as expected. Similar results are observed with other tail copula densities (figures are omitted).

F.3 Additional results for US flight delay data in Section 9

F.3.1 Estimated regular vine sequence of the X-vine model from T_1 to T_7

Assigning $\hat{\chi}_e$ as edge weight for $e \in E_1$ and $\hat{\tau}_e$ for $e \in E_j$ with $j \geq 2$, we obtain the sequence of maximum spanning trees that satisfy the proximity condition. With the mBIC-optimal truncation level of $q^* = 7$, Figure 9 displays the trees of the X-vine model from T_1 to T_7 sequentially. For trees T_2 to T_7 , the edges for which the independence copula is selected are not shown.

F.3.2 Goodness of fit for the fitted (truncated) X-vine model

To assess the goodness of fit for the full X-vine model, we compare empirical pairwise $\hat{\chi}$ values with those derived from the full X-vine model in Fig. 10a. As an indicator of capturing a higher-dimensional extremal dependence, we use a trivariate (lower) tail dependence coefficient $\chi_{abc} = R_{abc}(1, 1, 1)$ for distinct $a, b, c \in [d]$. Its empirical estimate is

$$\hat{\chi}_{abc} = \frac{1}{k} \sum_{i=1}^n \mathbb{1} \left\{ \hat{U}_{i,a} \leq k/n, \hat{U}_{i,b} \leq k/n, \hat{U}_{i,c} \leq k/n \right\}.$$

The empirical trivariate tail dependence coefficients are compared with those derived from the full X-vine model in Fig. 10b and from the truncated X-vine model in Fig. 10c, respectively.

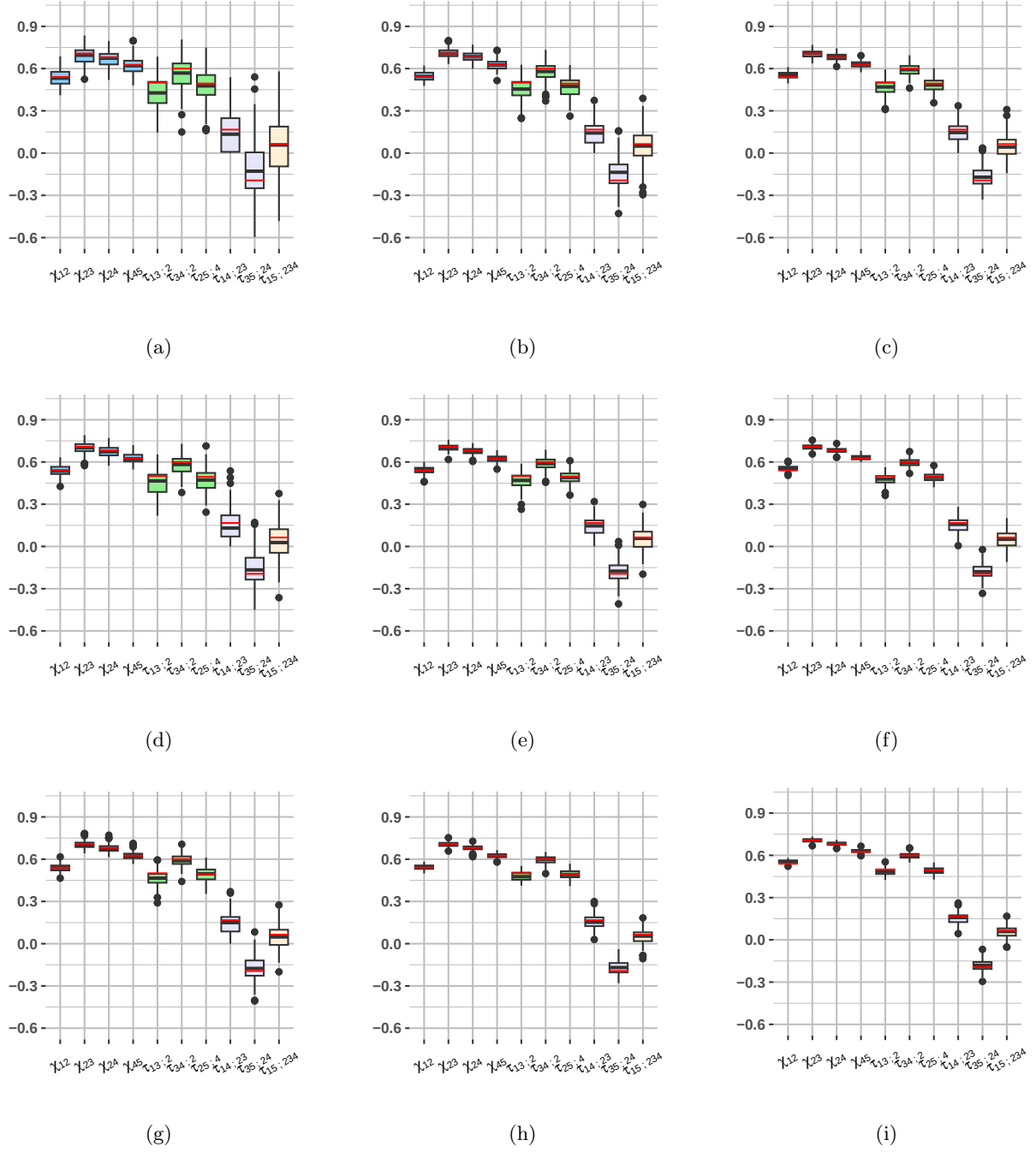


Figure 7: Box-plots of dependence measure estimates from specified parametric models in the X-vine specification shown in Fig. 2, varying with (n, k) : (a) $(n, k) = (1000, 20)$ (b) $(n, k) = (1000, 50)$ (c) $(n, k) = (1000, 100)$ (d) $(n, k) = (2000, 40)$ (e) $(n, k) = (2000, 100)$ (f) $(n, k) = (2000, 200)$ (g) $(n, k) = (4000, 80)$ (h) $(n, k) = (4000, 200)$ (i) $(n, k) = (4000, 400)$

F.3.3 Sensitivity analysis for mBIC truncation levels over quantiles

We conduct a sensitivity analysis to observe how the mBIC-optimal truncation level changes over quantiles. We observe a decline in the mBIC-optimal truncation level as the quantile k/n decreases. The range of mBIC-optimal truncation levels is from 4 to 8 with a median of 6.

F.3.4 Model comparison: X-vine Hüsler–Reiss model versus Hüsler–Reiss extremal graphical model from EGlearn

Focusing on the Hüsler–Reiss distribution, we compare model assessments between the X-vine Hüsler–Reiss model (all tail copulas in T_1 are Hüsler–Reiss and all bivariate copulas in the subsequent trees are Gaussian) and the Hüsler–Reiss extremal graphical model obtained with EGlearn algorithm (Engelke et al., 2021). In

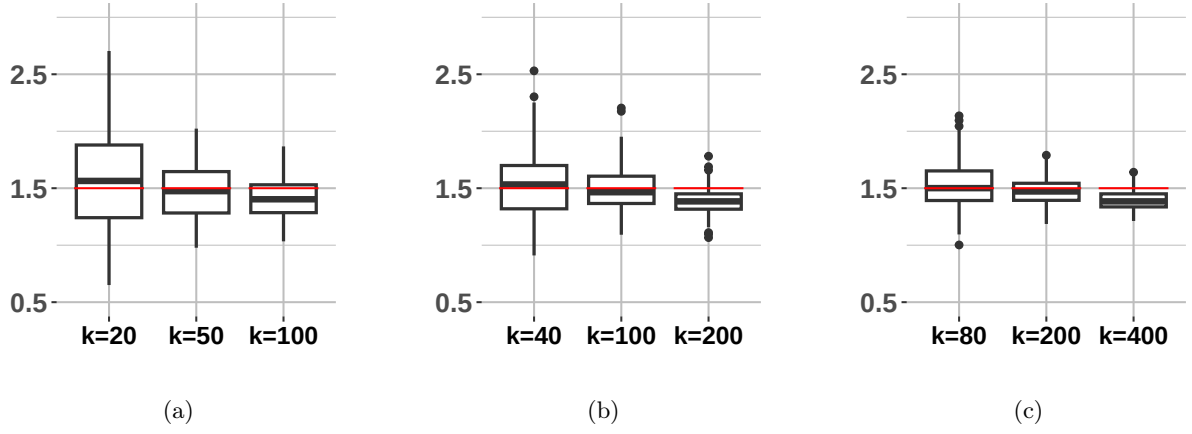


Figure 8: Box-plots of maximum pseudo-likelihood parameter estimates for the bivariate Hüsler-Reiss tail copula density with varying (n, k) : (a) $n = 1000$ (b) $n = 2000$, and (c) $n = 4000$. The red line indicates the specified parameter value of $\theta = 1.5$.

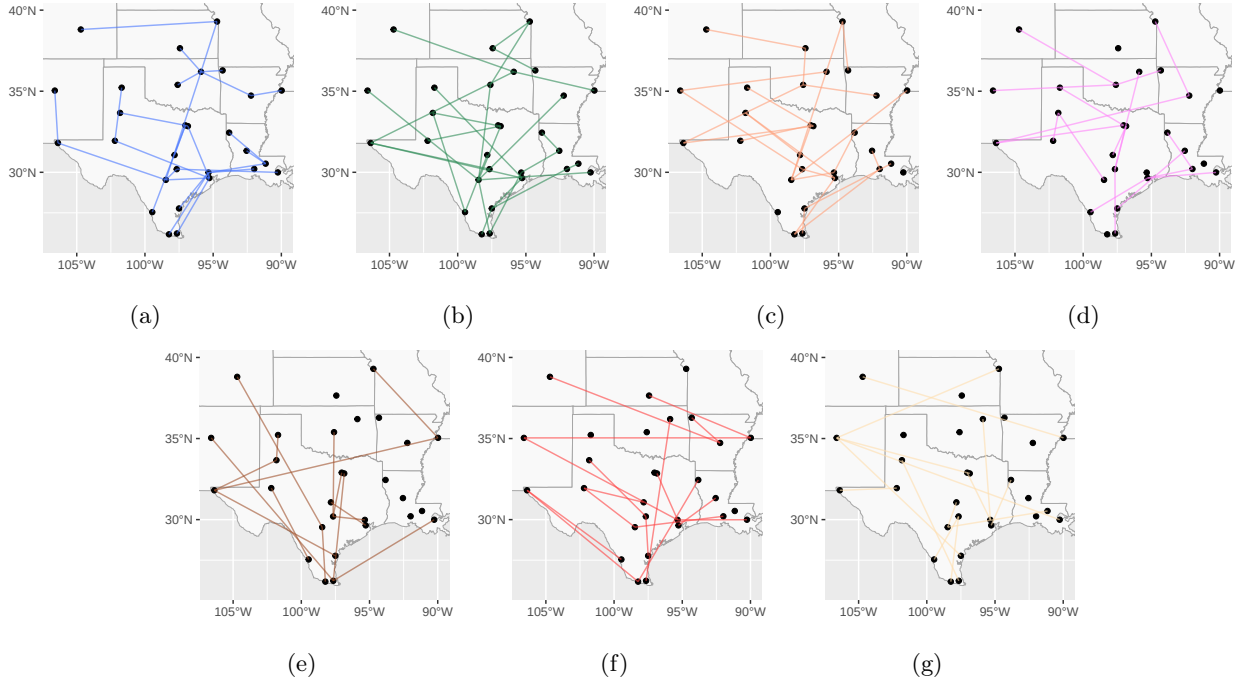


Figure 9: Sequential display of subsequent trees of the truncated X-vine model. For trees T_2 to T_7 , the edges for which the independence copula is selected are not shown. (a) T_1 (b) T_2 (c) T_3 (d) T_4 (e) T_5 (f) T_6 (g) T_7 .

both models, within T_1 , we use the empirical extremal variogram matrix $\hat{\Gamma}$ as edge weight for the minimum spanning tree. The resulting tree must be identical for both models, as confirmed in Fig. 13. As described in Section 9, with the tuning parameter value of $\rho^* = 0.1$ selected by the algorithm, the resulting sparse extremal graph has 148 edges, shown in Fig. 5f. for the Hüsler-Reiss X-vine, the algorithm sets all bivariate copulas to the independence copula from tree level 19 on (i.e., the truncation level is $q = 18$), selecting 217 independence copulas out of 406 edges in trees T_1 to T_{28} (recall $d = 29$). Using the mBIC, we further consider the truncated Hüsler-Reiss X-vine with a lower truncated level. The optimal mBIC truncation level of $q^* = 7$ results in a total of 175 edges in edges T_1 to T_7 , shown in Fig. 11a. In trees T_2 to T_7 , there are 46 edges out of 147 for which the independence copula is selected.

We assess the goodness of fit between the different graphical models using the entire sub-samples, that is, no longer splitting the data in a training and test set. Fig. 12 compares empirical tail dependence coefficients to those obtained from the fitted models, specifically for the extremal Hüsler-Reiss graphical model based on the flight graph, the truncated Hüsler-Reiss X-vine, and the extremal Hüsler-Reiss graph derived from the EGlearn algorithm, respectively. Similarly to the truncated X-vine model, the truncated Hüsler-Reiss X-vine exhibits improved fit but more variability compared to the extremal Hüsler-Reiss graphical model.

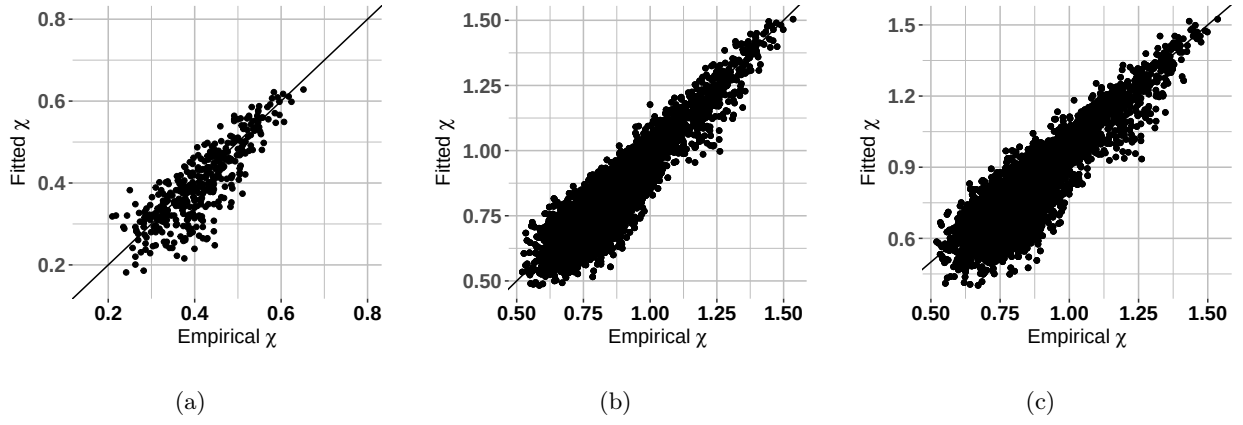


Figure 10: A χ -plot comparing empirical pairwise tail dependence coefficients $\hat{\chi}_{a,b}$ for distinct $a, b \in [d]$ with those derived from the full X-vine model (a), and a plot comparing empirical trivariate tail dependence coefficients $\hat{\chi}_{a,b,c}$ for distinct $a, b, c \in [d]$ with those obtained from the fitted full X-vine model (b); from the fitted truncated X-vine model with a truncation level of $q^* = 7$ (c), respectively.

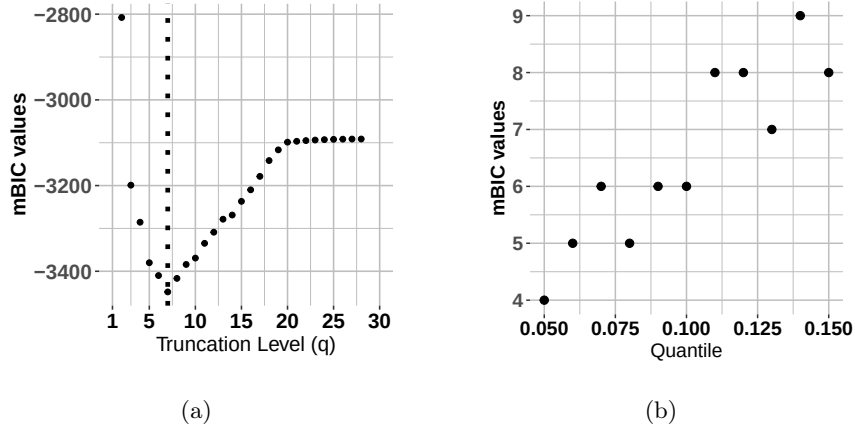


Figure 11: (a) mBIC-values plotted across tree levels with a dotted line indicating a selected mBIC-optimal truncation level of $q^* = 7$. (b) Plot of the mBIC-optimal truncation level across a sequence of quantiles $k/n \in (0.05, 0.15)$.

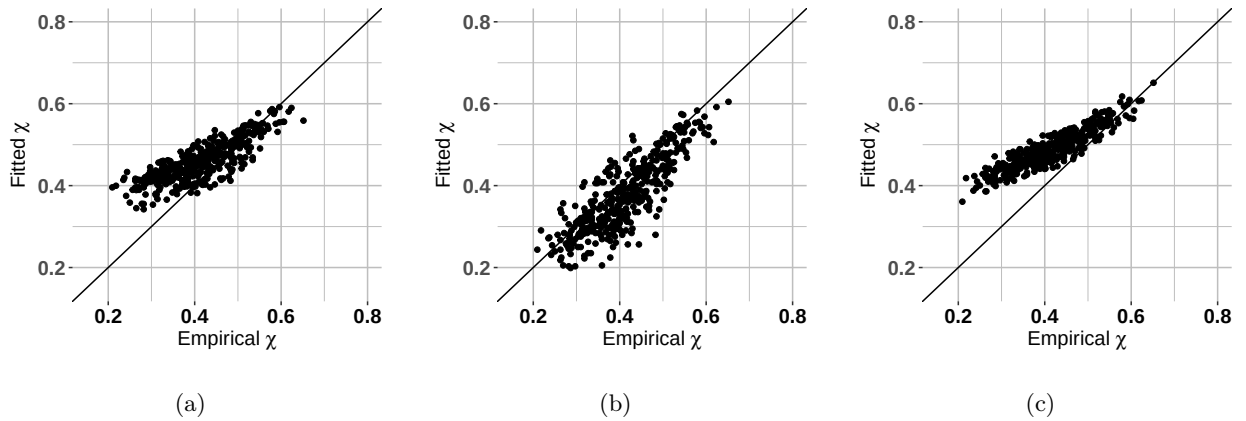


Figure 12: χ -plots comparing empirical pairwise $\hat{\chi}$ values to those from the fitted graphical models: the extremal Hüsler-Reiss graphical model for the flight graph (a), the truncated Hüsler-Reiss X-vine with $q^* = 7$ (b), and the extremal Hüsler-Reiss graphical model with $\rho^* = 0.1$ using the EGlearn algorithm (c), respectively.

F.3.5 Estimated regular vine sequence of the truncated Hüsler–Reiss X-vine

In Fig. 13, we sequentially display the selected trees T_1 to T_7 of the truncated Hüsler–Reiss X-vine in Section F.3.4. For trees T_2 to T_7 , there are 46 out of 147 edges for which the independence copula is selected, and these edges are not shown in the figure.

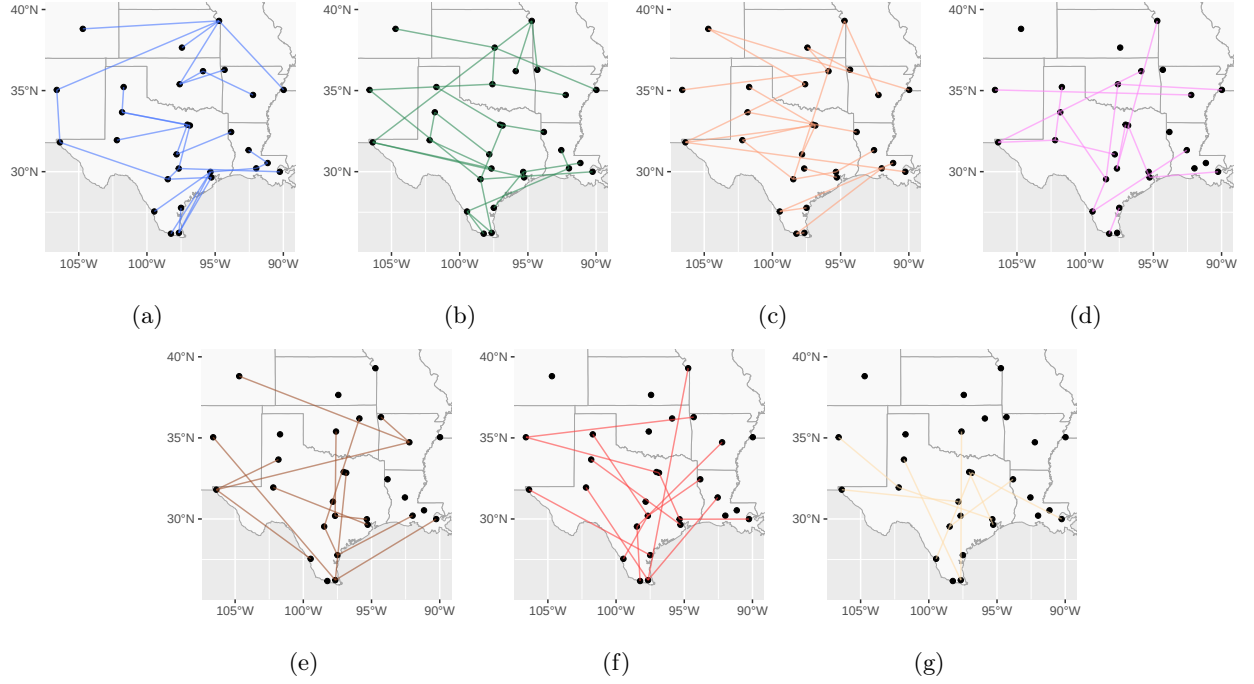


Figure 13: Sequential display of the first $q^* = 7$ trees of the truncated Hüsler–Reiss X-vine fitted to the US flight data in Section F.3.4. For trees T_2 to T_7 , the edges for which the independence copula is selected are not shown. (a) T_1 (b) T_2 (c) T_3 (d) T_4 (e) T_5 (f) T_6 (g) T_7 .