X-Vine Models for Multivariate Extremes

Anna Kiriliouk* Jeongjin Lee* Johan Segers†

October 13, 2024

Abstract

Regular vine sequences permit the organisation of variables in a random vector along a sequence of trees. Vine-based dependence models have become greatly popular as a way to combine arbitrary bivariate copulas into higher-dimensional ones, offering flexibility, parsimony, and tractability. In this project, we use regular vine sequences 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

*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. Corresponding author. E-mail: johan.segers@uclouvain.be

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 these 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 sequence 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

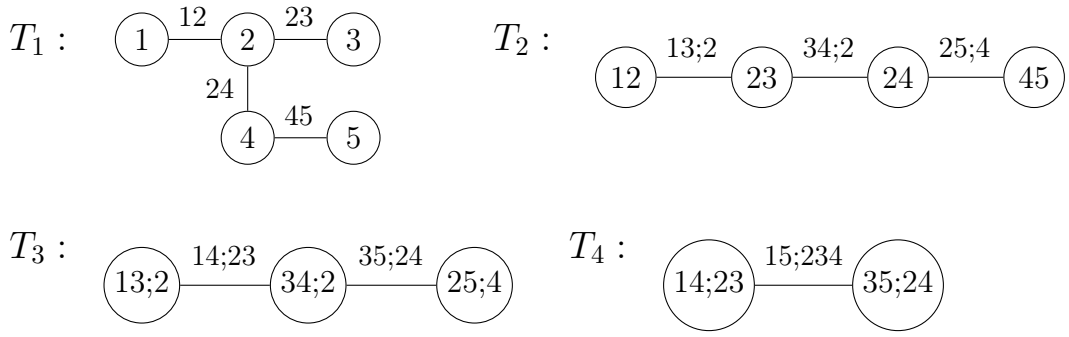


Figure 1: A regular vine sequence $\mathcal{V}$ in dimension $d = 5$ consisting of trees $T_1, \ldots, T_4$, where T_j has $d - j + 1$ nodes and $d - j$ edges. The nodes of T_1 are $\{1, \ldots, 5\}$, while the edges of T_j become the nodes of T_{j+1}.

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.

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. Li and Wu (2013) define tail densities and compute those of D-vine copulas in dimensions three and four. 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, 2024). Markov trees (Segers, 2020; Hu et al., 2024) 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 parametric examples are worked out in Section 4, with a focus on the simplifying assumption that the conditional copula densities only depend on the indices of the conditioning variables, 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. (2024). Section 10 concludes. The supplementary material contains a detailed example to illustrate our theory, the proofs of the paper's results, expressions based on exponent measure densities, and additional numerical results. The methods are implemented in the R (R Core Team, 2023) package Xvine[1], relying in particular on packages graphicalExtremes (Engelke et al., 2022) and VineCopula (Nagler et al., 2023).

2 Background

We write $[d] = \{1, \dots, d\}$ for the index set of the variables. Bold symbols refer to multivariate quantities. For a point $\boldsymbol{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ and a subset $J \subseteq [d]$, write subvectors as $\boldsymbol{x}_J = (x_j)_{j \in J}$ and $\boldsymbol{x}_{\setminus J} = (x_j)_{j \in [d] \setminus J}$. 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 $\boldsymbol{X} = (X_1, \dots, X_d)$ is in the max-domain of attraction of a multivariate extreme value distribution (Beirlant et al., 2004; de Haan and Ferreira, 2007). This assumption concerns the tails of the univariate marginal distribution functions $F_1, \dots, F_d$ and the tail dependence structure of $\boldsymbol{X}$. Our focus is on the latter aspect, that is, on probabilities of high values occurring jointly among the variables X_j. Let C be the survival copula of $\boldsymbol{X}$: under the standing assumption that $F_1, \dots, F_d$ are continuous, C is the distribution function of the random vector $\boldsymbol{U} = (U_1, \dots, U_d)$ with uniformly distributed components $U_j = 1 - F_j(X_j)$

[1]Available from `https://github.com/JeongjinLee88/Xvine`

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 $\mathbb{E} = (0,\infty]^d \setminus \{\boldsymbol{\infty}\}$ defined by

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

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 $R((\mathbf{0}, \boldsymbol{x}]) = R(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathbb{E}$.

The term 'tail copula' stems from Schmidt and Stadtmüller (2006) in case $d = 2$. The tail copula measure already appears in Einmahl et al. (2001) and is closely linked to the exponent measure Λ in Eq. (2.9) below, introduced in de Haan and Resnick (1977). The limit (2.1) appears in Jaworski (2006) and in Joe et al. (2010), who call it 'tail dependence function'.

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

$$\forall j \in [d],\ \forall x_j \in (0,\infty), \qquad R\left(\{\boldsymbol{y} \in \mathbb{E} : y_j \leqslant x_j\}\right) = x_j. \tag{2.2}$$

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

$$R(s\boldsymbol{x}) = s\, R(\boldsymbol{x}) \qquad \text{and} \qquad R(sB) = s\, R(B). \tag{2.3}$$

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; see Lemma B.1 in the supplement.

Tail copula densities. Throughout, we assume that the tail copula measure R has no mass on the hyperplanes through infinity, $R\left(\{\boldsymbol{x} \in \mathbb{E} : x_j = \infty\}\right) = 0$ for all $j \in [d]$, so R is supported on $(0,\infty)^d$. Further, we assume that R is absolutely continuous with respect to the d-dimensional

Lebesgue measure with continuous density $r : (0,\infty)^d \to [0,\infty)$, that is, $R(B) = \int_B r(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$ for all Borel sets $B \subseteq (0,\infty)^d$. Choosing $B = (\boldsymbol{0}, \boldsymbol{x}]$ for $\boldsymbol{x} \in (0,\infty)^d$, we recover r from R by $r(\boldsymbol{x}) = \frac{\partial^d}{\partial x_1 \cdots \partial x_d} R(\boldsymbol{x})$ for all $\boldsymbol{x} \in (0,\infty)^d$. The marginal constraint (2.2) implies

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

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

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

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

By Scheffé's lemma, if the copula C has density c and the tail copula measure R has density r, then (2.1) is implied by $\lim_{t \searrow 0} t^{1-d} c(t\boldsymbol{x}) = r(\boldsymbol{x})$ for all $\boldsymbol{x} \in (0,\infty)^d$. Li and Wu (2013) call r the (lower) tail density function of C. A word of caution: for a given copula density c, the limit $\lim_{t \searrow 0} t^{1-d} c(t\boldsymbol{x})$ may exist for all $\boldsymbol{x} \in (0,\infty)^d$ 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 non-empty $J \subseteq [d]$, let $\pi_J : (0,\infty)^d \to (0,\infty)^J$ denote the coordinate projection $\boldsymbol{x} \mapsto \pi_J(\boldsymbol{x}) = \boldsymbol{x}_J$ and let $R_J = R \circ \pi_J^{-1}$ denote the J-th marginal measure

$$R_J(B) = R\left(\pi_J^{-1}(B)\right) = R\left(\left\{\boldsymbol{x} \in (0,\infty)^d : \boldsymbol{x}_J \in B\right\}\right), \tag{2.6}$$

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 $\boldsymbol{U}_J = \pi_J(\boldsymbol{U})$, as in Definition 2.1. For $\boldsymbol{x}_J \in (0,\infty]^J \setminus \{\boldsymbol{\infty}\}$, we have $R_J(\boldsymbol{x}_J) = R(\tilde{\boldsymbol{x}}_J)$ with $\tilde{x}_j = x_j$ if $j \in J$ and $\tilde{x}_j = \infty$ if $j \in [d] \setminus J$. For $\boldsymbol{x}_J \in (0,\infty)^J$,

the density r_J of the measure R_J is

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

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. 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 to 5 do not depend on it.

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

$$\lim_{n\to\infty} \mathsf{P}(\boldsymbol{V} \leqslant n\boldsymbol{z})^n = G(\boldsymbol{z}) = \exp\left\{-\ell\left(1/\boldsymbol{z}\right)\right\}, \qquad \boldsymbol{z} \in (0,\infty)^d, \tag{2.8}$$

that is, $\boldsymbol{V}$ 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* $\ell(\boldsymbol{x}) = R\left(\mathbb{E} \setminus [\boldsymbol{x}, \boldsymbol{\infty}]\right)$ for $\boldsymbol{x} \in [0,\infty)^d$. The margins of $\boldsymbol{V}$ are unit-Pareto, $\mathsf{P}(V_j > x) = 1/x$ for $x \geqslant 1$, while those of G are unit-Fréchet, $G_j(x) = \exp(-1/x)$ for $x > 0$. Equivalently, the distribution of $\boldsymbol{V}$ is *multivariate regularly varying* (de Haan and Resnick, 1977; Resnick, 2007) with limit measure

$$\Lambda(\,\cdot\,) = R\left(\{\boldsymbol{x} \in \mathbb{E} : (1/x_1, \ldots, 1/x_d) \in \cdot\,\}\right) \qquad \text{on } [0,\infty)^d \setminus \{\boldsymbol{0}\}, \tag{2.9}$$

that is, $t\,\mathsf{P}(\boldsymbol{V} \in tB) \to \Lambda(B)$ as $t \to \infty$ 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 $G(\boldsymbol{z}) = \exp\left\{-\Lambda\left([\boldsymbol{0}, \boldsymbol{\infty}) \setminus [\boldsymbol{0}, \boldsymbol{z}]\right)\right\}$ for $\boldsymbol{z} \in (0,\infty)^d$.

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

$$(\boldsymbol{V}/t \mid \max \boldsymbol{V} > t) \rightsquigarrow \boldsymbol{Y}, \qquad t \to \infty, \tag{2.10}$$

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

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

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

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 practice, it may be convenient to reduce the information contained in the tail copula measure to a dependence coefficient. In this paper, we will encounter the *tail dependence coefficients* $\chi_J = R(\{\boldsymbol{x} \in (0,\infty]^d : \max_{j\in J} x_j < 1\}) = R_J(1,\ldots,1)$ for index sets $J \subseteq [d]$ with two or more elements.

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 Λ is concentrated on $(0,\infty)^d$ and has density

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

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

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 $j \in \{2,\ldots,d-1\}$ (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, Czado (2019) or Czado and Nagler (2022). Our objective is to do the same for tail copula densities. All concepts introduced below are illustrated via a five-dimensional example in Section A of the supplementary material.

A *tree* $T = (N, E)$ is a connected, acyclic graph comprising a finite 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 $\mathcal{V} = (T_1, \ldots, T_{d-1})$ is a *regular vine tree sequence* on $d \geqslant 3$ elements if $T_1 = (N_1, E_1)$ is a tree with node set $N_1 = [d]$, while for $j \in \{2, \ldots, d-1\}$, 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.

Definition 2.3 (Complete union, conditioning set, conditioned set)**.** Let $\mathcal{V} = (T_1, \ldots, T_{d-1})$ be a regular vine sequence, with $T_j = (N_j, E_j)$. 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 $j \in \{2, \ldots, d-1\}$. 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 $j \geqslant 2$ 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 $e = \{a, b\} \in E_1$, then $\mathcal{C}_{e,a} = \{a\}$, $\mathcal{C}_{e,b} = \{b\}$, $\mathcal{C}_e = e$ and $D_e = \varnothing$.

For an edge $e = \{a, b\} \in E_j$, the sets $\mathcal{C}_{e,a}$, $\mathcal{C}_{e,b}$ and D_e are disjoint and their union equals A_e, a node set with $j + 1$ elements. The sets $\mathcal{C}_{e,a}$ and $\mathcal{C}_{e,b}$ are singletons while D_e has $j - 1$ elements. Clearly, $D_e = A_a \setminus \mathcal{C}_e = A_b \setminus \mathcal{C}_e$. 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 $a_e, b_e \in [d]$ 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 $D_e = \{2, 3\}$. See Section B.1 in the supplement for some additional properties.

For positive scalars $\gamma_1, \ldots, \gamma_d$, the identity $\prod_{j=2}^{d} (\gamma_j / \gamma_{j-1}) = \gamma_d / \gamma_1$ is 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 $d \geqslant 3$ 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*

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

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

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

Remark B.4 in Appendix B.1 in the supplement generalises Lemma 2.4 to factorisations of γ_J for certain non-empty subsets $J \subseteq [d]$, while Remark B.5 provides a generalisation of (2.13) without the restriction that $\gamma_j = 1$ for all $j \in [d]$.

3 Sklar's theorem for tail copula densities

Let r be a d-variate tail copula density, so properties (2.4) and (2.5) hold. Recall its multivariate margins in (2.7). For non-empty, disjoint $I, J \subset [d]$ and for $\boldsymbol{x}_I \in (0,\infty)^I$ and $\boldsymbol{x}_J \in (0,\infty)^J$ such that $r_J(\boldsymbol{x}_J) > 0$, define $r_{I|J}(\boldsymbol{x}_I|\boldsymbol{x}_J) := r_{I\cup J}(\boldsymbol{x}_{I\cup J})/r_J(\boldsymbol{x}_J)$ for all $\boldsymbol{x}_{I\cup J} \in (0,\infty)^{I\cup J}$. 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 $\boldsymbol{x}_J \in (0,\infty)^J$ such that $r_J(\boldsymbol{x}_J) > 0$, the function $r_{I|J}(\,\cdot\,|\boldsymbol{x}_J) : (0,\infty)^I \to [0,\infty)$, defined via $\boldsymbol{x}_I \mapsto r_{I|J}(\boldsymbol{x}_I|\boldsymbol{x}_J)$, is a probability density on $(0,\infty)^I$. For $i \in I$, it has marginal probability density $x_i \mapsto r_{i|J}(x_i|\boldsymbol{x}_J)$ and cumulative distribution function $x_i \mapsto R_{i|J}(x_i|\boldsymbol{x}_J) := \int_0^{x_i} r_{i|J}(t|\boldsymbol{x}_J)\,\mathrm{d}t$, with quantile function $u_i \mapsto R_{i|J}^{-1}(u_i|\boldsymbol{x}_J)$. If*

$|I| \geqslant 2$, *then the copula density of* $r_{I|J}(\,\cdot\,|\boldsymbol{x}_J)$ *is*

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

The homogeneity property (2.5) induces certain scaling properties.

Lemma 3.2 (Scale equi- and invariance)**.** *In Proposition 3.1, for* $\boldsymbol{x}_J \in (0,\infty)^J$ *such that* $r_J(\boldsymbol{x}_J) > 0$*, the family* $\{r_{I|J}(\,\cdot\,|t\boldsymbol{x}_J) : t \in (0,\infty)\}$ *of probability densities is a scale family: if the random vector* $\boldsymbol{\xi}_I$ *has probability density* $r_{I|J}(\,\cdot\,|\boldsymbol{x}_J)$*, then the random vector* $t\boldsymbol{\xi}_I$ *has probability density* $r_{I|J}(\,\cdot\,|t\boldsymbol{x}_J)$*. As a consequence, for all* $t \in (0,\infty)$*,* $\boldsymbol{x}_I \in (0,\infty)^I$*,* $\boldsymbol{u}_I \in (0,1)^I$ *and* $i \in I$*, we have* $r_{i|J}(tx_i|t\boldsymbol{x}_J) = r_{i|J}(x_i|\boldsymbol{x}_J)$*,* $R_{i|J}^{-1}(u_i|t\boldsymbol{x}_J) = t \cdot R_{i|J}^{-1}(u_i|\boldsymbol{x}_J)$*, and*

$$c_{I;J}(\boldsymbol{u}_I;t\boldsymbol{x}_J) = c_{I;J}(\boldsymbol{u}_I;\boldsymbol{x}_J). \tag{3.2}$$

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 $\boldsymbol{Y}$ be the multivariate Pareto vector in (2.11), suppose $d \geqslant 3$, 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 $\boldsymbol{Y}_J$ in the sense of Definition 1 in Engelke and Hitz (2020) if and only if $c_{i,j;J}(u_i,u_j;\boldsymbol{x}_J) = 1$ for all $(u_i,u_j) \in (0,1)^2$ and $\boldsymbol{x}_J \in (0,\infty)^J$, 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 (C.1) in Appendix C in the supplement for an expression of $c_{I;J}$ in terms of the exponent measure density λ.

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

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.

As a converse to 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 $(\boldsymbol{u}, \theta) \mapsto c(\boldsymbol{u}; \theta)$ is Borel measurable on $(0,1)^k \times \Theta$.

Proposition 3.5 (Sklar's theorem: converse part)**.** *Let $I \cup J$ be a partition of $[d]$ into two disjoint, non-empty sets, with $d \geqslant 3$ and $|I| \geqslant 2$. 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 $\left\{c_{I;J}(\,\cdot\,;\boldsymbol{x}_J) : \boldsymbol{x}_J \in (0,\infty)^J\right\}$ be a jointly measurable family of $|I|$-variate copula densities such that* (3.2) *holds. Then the function $r : (0,\infty)^d \to [0,\infty)$ defined by*

$$r(\boldsymbol{x}) := r_J(\boldsymbol{x}_J) \cdot \prod_{i\in I} r_{i|J}(x_i|\boldsymbol{x}_J) \cdot c_{I;J}\left(R_{i|J}(x_i|\boldsymbol{x}_J), i \in I; \boldsymbol{x}_J\right), \tag{3.3}$$

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

4 Parametric families of tail copula densities

We will compute the copula densities $c_{I;J}(\,\cdot\,;\boldsymbol{x}_j)$ 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 $d \geqslant 2$, consider parameters $\alpha_1, \ldots, \alpha_d > 0$ and $\rho > -\min(\alpha_1, \ldots, \alpha_d)$ 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.2 in the supplement, the corresponding tail copula density is

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

for $\boldsymbol{x} \in (0,\infty)^d$, where $\bar{\alpha} = \alpha_1 + \cdots + \alpha_d$ and $c(\alpha,\rho) = \Gamma(\alpha+\rho)/\Gamma(\alpha)$. The model is closed under marginalisation: for $J \subseteq [d]$ with $|J| \geqslant 2$, 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 $\alpha_1 = \ldots = \alpha_d = 1$ and $\rho > 0$, we get the negative logistic model (Joe, 1990);
- if $\alpha_1 = \ldots = \alpha_d = 1$ 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 $\boldsymbol{Y} = S\boldsymbol{W}$, where $\boldsymbol{W}$ has a Dirichlet distribution on the unit simplex and is independent of the positive random variable S. The special case where $\boldsymbol{W}$ 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 $d \geqslant 3$. For every pair of disjoint, non-empty subsets $I, J \subset [d]$ with $|I| \geqslant 2$, the copula densities $c_{I;J}(\,\cdot\,;\boldsymbol{x}_J)$ in* (3.1) *satisfy the simplifying assumption (Definition 3.4). If $\rho > 0$, then $c_{I;J}$ is equal to the density of 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 in the supplement):

- if $\alpha_1 = \ldots = \alpha_d = 1$ and $\rho > 0$, then $c_{I;J}$ is the density of the $|I|$-variate Clayton copula with parameter $\theta/\left(1 + |J|\theta\right)$ with $\theta = 1/\rho > 0$;
- if $\alpha_1 = \ldots = \alpha_d = 1$ and $-1 < \rho < 0$, then $c_{I;J}$ is the density of the $|I|$-variate Clayton survival copula with parameter $\theta/\left(|J|\theta - 1\right)$ 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 $\Gamma_{ij} = \mathsf{Var}(A_i - A_j)$ for a d-variate random vector $\boldsymbol{A} = (A_1, \ldots, A_d)$ 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

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

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

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

where ϕ_{d-1} is the centered $(d-1)$-dimensional Gaussian density with the stated covariance matrix and $\bar{\boldsymbol{x}}_{\setminus k} = \left(\log(x_i/x_k) - \frac{1}{2}\Gamma_{ik}\right)_{i:i\neq k}$. 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 $\boldsymbol{R}$ evaluated in $\boldsymbol{u} \in (0,1)^m$ is

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

Proposition 4.2. *Let Γ be a $d \times d$ variogram with $d \geqslant 3$ 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 $|I| \geqslant 2$, the copula densities $c_{I;J}(\,\cdot\,; \boldsymbol{x}_J)$ in* (3.1) *satisfy the simplifying assumption (Definition 3.4). The copula density $c_{I;J}$ is equal to the $|I|$-variate Gaussian copula density* (4.4) *with correlation matrix $\boldsymbol{R}_{I|J} = (\Delta_{I|J}^{(k)})^{-1/2}\, \Sigma_{I|J}^{(k)}\, (\Delta_{I|J}^{(k)})^{-1/2}$ 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 matrix $\boldsymbol{R}_{I|J}$ in Proposition 4.2 is the correlation matrix of the conditional Gaussian distribution with covariance matrix $\Sigma_{I|J}^{(k)}$, the corresponding Schur complement. Its expression depends on the choice of $k \in J$, although the actual matrix does not. It remains to be investigated how to express it in terms of the Hüsler–Reiss precision matrix Θ introduced in Hentschel et al. (2024).

5 Vine decompositions of tail copula densities

5.1 X-vines: density decomposition and construction

In this section, we show that any tail copula density r in dimension $d \geqslant 3$ 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, copula densities with only a single conditioning variable satisfy the simplifying assumption (Definition 3.4). Conversely, starting from any regular vine sequence on $d \geqslant 3$ 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*. Appendix C in the supplement states the main results in terms of exponent measure densities.

Theorem 5.1 (Tail copula density decomposition along a regular vine)**.** *Let r be a d-variate ($d \geqslant 3$) 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* $\boldsymbol{x} \in (0,\infty)^d$, *we have*

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

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

$$c_{a_e,b_e;D_e}\left(u_{a_e}, u_{b_e}; \boldsymbol{x}_{D_e}\right) = \frac{r_{a_e,b_e|D_e}(x_{a_e}, x_{b_e}|\boldsymbol{x}_{D_e})}{r_{a_e|D_e}(x_{a_e}|\boldsymbol{x}_{D_e}) \cdot r_{b_e|D_e}(x_{b_e}|\boldsymbol{x}_{D_e})}, \tag{5.2}$$

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

The decomposition (5.1) also applies to marginal tail copula densities r_J for index sets $J = A_f$ for some edge f in the vine; see Remark B.7 in the supplement. For the pair copula $C_{a_e,b_e;D_e}(\,\cdot\,,\,\cdot\,;\boldsymbol{x}_{D_e})$ associated to the edge $e = (a_e, b_e; D_e)$ in the regular vine sequence, consider the first-order partial derivatives

$$\begin{aligned} C_{a_e|b_e;D_e}(u_{a_e} \mid u_{b_e}; \boldsymbol{x}_{D_e}) &:= \frac{\partial}{\partial u_{b_e}} C_{a_e,b_e;D_e}(u_{a_e}, u_{b_e}; \boldsymbol{x}_{D_e}) = \int_{v=0}^{u_{a_e}} c_{a_e,b_e;D_e}(v, u_{b_e}; \boldsymbol{x}_{D_e})\,\mathrm{d}v, \\ C_{b_e|a_e;D_e}(u_{b_e} \mid u_{a_e}; \boldsymbol{x}_{D_e}) &:= \frac{\partial}{\partial u_{a_e}} C_{a_e,b_e;D_e}(u_{a_e}, u_{b_e}; \boldsymbol{x}_{D_e}) = \int_{v=0}^{u_{b_e}} c_{a_e,b_e;D_e}(u_{a_e}, v; \boldsymbol{x}_{D_e})\,\mathrm{d}v. \end{aligned} \tag{5.3}$$

Theorem 5.2 (Recursion and uniqueness)**.** *In the setting of Theorem 5.1, for any* $e = (a_e, b_e; D_e) \in E_2 \cup \cdots \cup E_{d-1}$, *we have*

$$\begin{aligned} R_{a_e|D_e \cup b_e}(x_{a_e}|\boldsymbol{x}_{D_e \cup b_e}) &= C_{a_e|b_e;D_e}\left(R_{a_e|D_e}(x_{a_e}|\boldsymbol{x}_{D_e}) \mid R_{b_e|D_e}(x_{b_e}|\boldsymbol{x}_{D_e}); \boldsymbol{x}_{D_e}\right), \\ R_{b_e|D_e \cup a_e}(x_{b_e}|\boldsymbol{x}_{D_e \cup a_e}) &= C_{b_e|a_e;D_e}\left(R_{b_e|D_e}(x_{b_e}|\boldsymbol{x}_{D_e}) \mid R_{a_e|D_e}(x_{a_e}|\boldsymbol{x}_{D_e}); \boldsymbol{x}_{D_e}\right). \end{aligned} \tag{5.4}$$

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* $c_{a_e,b_e;D_e}(\,\cdot\,;\boldsymbol{x}_{D_e})$ *for* $e \in E_2 \cup \cdots \cup E_{d-1}$ *and* $\boldsymbol{x}_{D_e} \in (0,\infty)^{D_e}$.

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 5.3 (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 $e \in E_2 \cup \cdots \cup E_{d-1}$, the pair copula densities $c_{a_e,b_e;D_e}(\,\cdot\,,\,\cdot\,;\boldsymbol{x}_{D_e})$ do not depend on the value of $\boldsymbol{x}_{D_e} \in (0,\infty)^{D_e}$.

Example 5.4 (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 $\{1,2,3\}$. For the vine determined by $E_1 = \{\{1,2\},\{2,3\}\}$, for instance, we have

$$r(x_1,x_2,x_3) = r_{12}(x_1,x_2)\, r_{23}(x_2,x_3) \cdot c_{13;2}\left(R_{1|2}(x_1|x_2),\, R_{3|2}(x_3|x_2)\right), \tag{5.5}$$

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

The function r is thus completely specified by the two bivariate tail copula densities r_{12} and r_{23} and one bivariate copula density $c_{13;2}$. The form (5.5) was already discovered for tail copula densities of D-vine copulas in Li and Wu (2013).

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 $c_{I;J}(\,\cdot\,;\boldsymbol{x}_J)$ 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 5.5 (X-vine specification)**.** The triplet $(\mathcal{V},\mathcal{R},\mathcal{C})$ is an *X-vine specification* on d elements ($d \geqslant 3$) 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. $\mathcal{C} = \left\{c_{a_e,b_e;D_e} :\ e = (a_e,b_e;D_e) \in \bigcup_{j\geqslant 2} E_j\right\}$ is a family of bivariate copula densities.

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

Theorem 5.6 (X-vine tail copula density construction)**.** *Let $(\mathcal{V}, \mathcal{R}, \mathcal{C})$ be an X-vine specification on $d \geqslant 3$ elements. Then the function r defined by*

$$r(\boldsymbol{x}) = \prod_{e \in E_1} r_{a_e,b_e}(x_{a_e}, x_{b_e}) \cdot \prod_{j=2}^{d-1} \prod_{e \in E_j} c_{a_e,b_e;D_e}\left(R_{a_e|D_e}(x_{a_e}|\boldsymbol{x}_{D_e}), R_{b_e|D_e}(x_{b_e}|\boldsymbol{x}_{D_e})\right) \tag{5.6}$$

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

$$\left.\begin{aligned} R_{a_e|D_e \cup b_e}(x_{a_e}|\boldsymbol{x}_{D_e \cup b_e}) &= C_{a_e|b_e;D_e}\left(R_{a_e|D_e}(x_{a_e}|\boldsymbol{x}_{D_e}) \mid R_{b_e|D_e}(x_{b_e}|\boldsymbol{x}_{D_e})\right) \\ R_{b_e|D_e \cup a_e}(x_{b_e}|\boldsymbol{x}_{D_e \cup a_e}) &= C_{b_e|a_e;D_e}\left(R_{b_e|D_e}(x_{b_e}|\boldsymbol{x}_{D_e}) \mid R_{a_e|D_e}(x_{a_e}|\boldsymbol{x}_{D_e})\right) \end{aligned}\right\} \tag{5.7}$$

is a d-variate tail copula density. For $e \in E_1$, the bivariate margin of r is equal to $r_e \in \mathcal{R}$, while for $e = (a_e, b_e; D_e) \in E_2 \cup \cdots \cup E_{d-1}$, the pair copula density $c_{a_e,b_e;D_e}(\,\cdot\,,\,\cdot\,;\boldsymbol{x}_{D_e})$ in (5.2) is equal to $c_{a_e,b_e;D_e} \in \mathcal{C}$. In particular, r is an X-vine.

In Example 5.4 with $d = 3$, the vine $\mathcal{V} = (T_1, T_2)$ is determined by $E_1 = \{\{1,2\},\{2,3\}\}$, while $\mathcal{R} = \{r_{12}, r_{23}\}$ and $\mathcal{C} = \{c_{13;2}\}$, and (5.6) reduces to (5.5).

5.2 X-vines as limits of regular vine copula densities

A natural question is whether X-vine tail copula densities arise as the lower tail dependence limits of regular vine copula densities, as introduced in Bedford and Cooke (2002). Below, we show that this is indeed the case, 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.

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 $d \geqslant 3$. 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

$$c(\boldsymbol{u}) = \prod_{e \in E_1} c_{a_e,b_e}(u_{a_e}, u_{b_e}) \cdot \prod_{j=2}^{d-1} \prod_{e \in E_j} c_{a_e,b_e;D_e}\left(C_{a_e|D_e}\left(u_{a_e}|\boldsymbol{u}_{D_e}\right), C_{b_e|D_e}\left(u_{b_e}|\boldsymbol{u}_{D_e}\right)\right), \tag{5.8}$$

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

$$C_{a_e,b_e|D_e}\left(u_{a_e}, u_{b_e}|\boldsymbol{u}_{D_e}\right) = C_{a_e,b_e;D_e}\left(C_{a_e|D_e}\left(u_{a_e}|\boldsymbol{u}_{D_e}\right), C_{b_e|D_e}\left(u_{b_e}|\boldsymbol{u}_{D_e}\right)\right). \tag{5.9}$$

Assumption 5.7. *Let c be a d-variate regular vine copula density as in* (5.8)*, with $d \geqslant 3$.*

(i) For every edge $e \in E_1$, there exists a bivariate tail copula density r_{a_e,b_e} such that $\lim_{t \searrow 0} t \cdot c_{a_e,b_e}(tx, ty) = r_{a_e,b_e}(x, y)$, for $(x, y) \in (0, \infty)^2$.

(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 5.8. *Under Assumption 5.7, we have $\lim_{t \searrow 0} t^{d-1} \cdot c(t\boldsymbol{x}) = r(\boldsymbol{x})$ for $\boldsymbol{x} \in (0, \infty)^d$, where the X-vine tail copula density r is generated by the triple $(\mathcal{V}, \mathcal{R}, \mathcal{C})$ as in Theorem 5.6 with*

- *$\mathcal{V}$ the same regular vine sequence as in* (5.8)*,*
- *$\mathcal{R} = \{r_{a_e,b_e} : e \in E_1\}$ for r_{a_e,b_e} in Assumption 5.7(i), and*
- *$\mathcal{C}$ the same bivariate copula densities for edges $e \in E_j$ with $j \geqslant 2$ as in* (5.8)*.*

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

Proposition 5.8 is foreshadowed by Theorem 3.4 and Example 3.5 in Li and Wu (2013), who consider D-vine copulas and who require stronger convergence properties.

5.3 Truncated X-vines

If the bivariate copula density $c_{a_e,b_e;D_e}$ associated with an edge $e \in E_2 \cup \cdots \cup E_{d-1}$ 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 (5.7) simplify to $R_{a_e|D_e \cup b_e}\left(x_{a_e}|\boldsymbol{x}_{D_e \cup b_e}\right) = R_{a_e|D_e}\left(x_{a_e}|\boldsymbol{x}_{D_e}\right)$ and $R_{b_e|D_e \cup a_e}\left(x_{b_e}|\boldsymbol{x}_{D_e \cup a_e}\right) = R_{b_e|D_e}\left(x_{b_e}|\boldsymbol{x}_{D_e}\right)$. *Sparse* X-vine specifications arise when $c_{a_e,b_e;D_e} \equiv 1$ for many edges $e \in E_2 \cup \cdots \cup E_{d-1}$. 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 $d \geqslant 3$ 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 ($d \geqslant 3$) 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 $\mathcal{C} = \varnothing$.

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 $e \in E'_{q+1} \cup \cdots \cup E'_{d-1}$. 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

$$r(\boldsymbol{x}) = \prod_{e \in E_1} r_{a_e,b_e}(x_{a_e}, x_{b_e}) \cdot \prod_{j=2}^{q} \prod_{e \in E_j} c_{a_e,b_e;D_e}\left(R_{a_e|D_e}(x_{a_e}|\boldsymbol{x}_{D_e}), R_{b_e|D_e}(x_{b_e}|\boldsymbol{x}_{D_e})\right), \tag{5.10}$$

The right-hand side of (5.10) involves $\sum_{j=2}^q (d-j)$ bivariate copula densities. The truncation level q allows to tune the trade-off between sparsity and flexibility. If $q = 1$, the second product is empty and the model is a Markov tree as in Engelke and Hitz (2020), Segers (2020) and Engelke and Volgushev (2022). Increasing q and adding trees ('layers') yields more complex dependence models. Truncated X-vine specifications will be shown at work 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 $\boldsymbol{Z}$ has an *inverted multivariate Pareto distribution* if there exists a d-variate tail copula measure R such that $\mathsf{P}(\boldsymbol{Z} \in B) = R(B \cap \mathbb{L})/R(\mathbb{L})$, for Borel sets B, with $\mathbb{L} = \big\{\boldsymbol{x} \in (0,\infty]^d : \min \boldsymbol{x} < 1\big\}$; equivalently, if there exists a multivariate Pareto random vector $\boldsymbol{Y}$ such that $\boldsymbol{Z} = 1/\boldsymbol{Y}$.

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

$$\mathsf{P}(\boldsymbol{Z} \in B \mid Z_j < 1) = R\left(B \cap \mathbb{L}^{(j)}\right), \tag{6.1}$$

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

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

$$(t\boldsymbol{U} \mid \min \boldsymbol{U} < 1/t) \rightsquigarrow \boldsymbol{Z}, \qquad t \to \infty. \tag{6.2}$$

For non-empty $J \subseteq [d]$, we have $(t\boldsymbol{U}_J \mid \min \boldsymbol{U}_J < 1/t) \rightsquigarrow \boldsymbol{Z}_{|J} \stackrel{d}{=} (\boldsymbol{Z}_J \mid \min \boldsymbol{Z}_J < 1)$, as $t \to \infty$. The conditional marginal $\boldsymbol{Z}_{|J}$ has a $|J|$-variate inverted multivariate Pareto distribution associated with the marginal tail copula measure R_J in (2.6). In contrast, $\boldsymbol{Z}_J$ does *not* necessarily have an inverted multivariate Pareto distribution, as $\min \boldsymbol{Z}_J < 1$ is not guaranteed.

6.2 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 sampling a multivariate Pareto random vector $\boldsymbol{Y}$ to sampling the conditional distributions $(\boldsymbol{Y} \mid Y_j > 1)$ for every $j \in [d]$. Below, we study the equivalent problem of simulating from the conditional distribution of $(\boldsymbol{Z} \mid Z_j < 1)$ for X-vine tail copula densities.

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 *sampling order* (Cooke et al., 2015) is encoded by the permutation constructed in the next lemma.

Lemma 6.2. *Let $\mathcal{V}$ be a regular vine sequence on d elements. For all $j \in [d]$, 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 $\sigma_j(k) \in \mathcal{C}_{e_{k-1}}$ and $\{\sigma_j(i) : i = 1, \dots, k\} = A_{e_{j,k-1}}$ for all $k \in \{2, \dots, d\}$.*

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

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

Let $W_1, \ldots, W_d$ 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 $\boldsymbol{Z}^{(j)} = (Z_1^{(j)}, \ldots, Z_d^{(j)})$ recursively as follows:

$$Z_j^{(j)} = W_j \text{ and } Z_{\sigma_j(k)}^{(j)} = R^{-1}_{\sigma_j(k)|\sigma_j(\{1,\ldots,k-1\})}\left(W_{\sigma_j(k)} \mid \boldsymbol{Z}^{(j)}_{\sigma_j(\{1,\ldots,k-1\})}\right), \quad k \in \{2, \ldots, d\}. \tag{6.4}$$

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

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

7 Estimation and model selection

Let $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,d})$ for $i \in \{1, \ldots, n\}$ be an independent random sample from a distribution function F with continuous but unspecified margins $F_1, \ldots, F_d$ 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 5.3 and 5.5 and Theorem 5.6). 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 $j \geqslant 2$ 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 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 $\boldsymbol{Z}$ 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 $\boldsymbol{Z}$ be an inverted multivariate Pareto random vector associated to r. Let $I, J \subset [d]$ be non-empty and disjoint.*

(i) For $\boldsymbol{z}_J \in (0,\infty)^J$ such that $\min \boldsymbol{z}_J < 1$ and $r_J(\boldsymbol{z}_J) > 0$, the conditional density of $\boldsymbol{Z}_I$ given $\boldsymbol{Z}_J = \boldsymbol{z}_J$ is $r_{I|J}(\,\cdot\,|\boldsymbol{z}_J)$.

(ii) Suppose $|I| \geqslant 2$. If $c_{I;J}(\,\cdot\,;\,\cdot\,)$ satisfies the simplifying assumption (Definition 3.4), then, conditionally on the event $\min \boldsymbol{Z}_J < 1$, the random vector $\big(R_{i|J}(Z_i|\boldsymbol{Z}_J)\big)_{i\in I}$ is independent of $\boldsymbol{Z}_J$ and its density is $c_{I;J}$.

By statement (ii), the density of $\big(R_{i|J}(Z_i|\boldsymbol{Z}_J)\big)_{i\in I}$ given $\boldsymbol{Z}_J \in A$ is equal to $c_{I;J}$ for any non-empty set $A \subseteq \big\{\boldsymbol{z} \in (0,\infty)^J : \min \boldsymbol{z} < 1\big\}$. 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).

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 5.6. Assume that the bivariate tail copula densities ($\mathcal{R}$) and bivariate copula densities ($\mathcal{C}$) belong to specified parametric families. Let the parameter vector be denoted by $\boldsymbol{\theta} = (\theta_{\mathcal{R}}, \theta_{\mathcal{C}})$: here, $\theta_{\mathcal{R}} = (\theta_e)_{e \in E_1}$ 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 $\theta_{\mathcal{C}} = (\theta_{\mathcal{C},j})_{j=2}^{d-1}$ for $\theta_{\mathcal{C},j} = (\theta_e)_{e \in E_j}$ denotes the 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 is challenging. Instead, using the X-vine decomposition into *bivariate* components and recursively defined quantities (Theorem 5.6), we outline a *sequential* procedure for parameter estimation, 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 that $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ is an independent random sample from F, with survival copula C, tail copula R and tail copula density r. For $i = 1, \dots, n$ and $j \in [d]$, let $\widehat{U}_{i,j} = 1 - \widehat{F}_j(X_{i,j})$, where $\widehat{F}_j$ denotes any estimator of the marginal distribution function $F_j(x) = \mathsf{P}(X_{i,j} \leqslant x)$. One possibility is the empirical distribution function, and to avoid boundary effects, we set

$$\widehat{U}_{i,j} = 1 - (\mathrm{rnk}_{i,j} - 0.5)/n, \tag{7.1}$$

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

By Eq. (6.2), for large $t > 0$, the rescaled points $t\widehat{\boldsymbol{U}}_i$ for $i \in \{1, \dots, n\}$ such that $\min \widehat{\boldsymbol{U}}_i < 1/t$ constitute pseudo-observations from a distribution that approximates the inverted multivariate Pareto distribution associated with r. We set $t = n/k$ where, in an asymptotic setting,

$k = k_n \in \{1, \ldots, n\}$ satisfies $k \to \infty$ and $k/n \to 0$. For $j \in [d]$ and $\widehat{U}_{i,j}$ as in (7.1), let

$$K_j = \Big\{ i = 1, \ldots, n : \widehat{U}_{i,j} < k/n \Big\} \tag{7.2}$$

be the set of indices i corresponding to the k largest observations for the jth component. Further, put $K_J = \bigcap_{j \in J} K_j$ and $K = \bigcup_{j \in [d]} K_j$, 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 $\widehat{\boldsymbol{Z}}_i = (\widehat{Z}_{i,j})_{j \in [d]}$ where $\widehat{Z}_{i,j} = (n/k)\, \widehat{U}_{i,j}$. In view of Eq. (6.2), we treat $\{\widehat{\boldsymbol{Z}}_i\}_{i \in K}$ as a sample of $|K|$ pseudo-observations of the inverted multivariate Pareto distribution associated with r. The sample size $|K|$ is random, and from (7.1), we have $k \leqslant |K| \leqslant dk$. For non-empty $J \subseteq [d]$ and $i \in K$, we write $\widehat{\boldsymbol{Z}}_{i,J} = (\widehat{Z}_{i,j})_{j \in J}$.

(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 $r_{a_e,b_e}(z_{a_e}, z_{b_e}; \theta_e)\, \mathbb{1}(z_j < 1)$. We use maximum pseudo-likelihood estimation to fit this density to the sub-samples $(\widehat{Z}_{i,a_e}, \widehat{Z}_{i,b_e})$ for both $i \in K_{a_e}$ and $i \in K_{b_e}$, with K_j as in Eq. (7.2). More precisely, we maximise each of the two pseudo-likelihoods

$$\mathcal{L}_{\mathcal{R},e}\Big(\theta_e^{(j)};\, (\widehat{Z}_{i,a_e}, \widehat{Z}_{i,b_e}), i \in K_j\Big) = \prod_{i \in K_j} r_{a_e,b_e}\Big(\widehat{Z}_{i,a_e}, \widehat{Z}_{i,b_e}; \theta_e^{(j)}\Big), \qquad j \in \{a_e, b_e\}, \tag{7.3}$$

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 proposed for the Hüsler–Reiss model (Engelke et al., 2015, 2022). Other estimation approaches include censored likelihoods (Ledford and Tawn, 1996; de Haan et al., 2008) or empirical stable tail dependence functions (Einmahl et al., 2008, 2018).

(3) Estimating the copula parameters $\theta_{\mathcal{C}}$ in $T_2, \ldots, T_{d-1}$. Estimation of the parameters associated with edges $e = (a_e, b_e; D_e)$ in E_j for $j \in \{2, \ldots, d-1\}$ is based on a similar procedure

as the one employed in regular vine copulas (Czado, 2019, Section 7.2). The main difference concerns the definition of the pseudo-observations: if $\hat{\theta}(E_{1:(j-1)})$ denotes the parameter estimates associated with the edges in $E_{1:(j-1)} = E_1 \cup \cdots \cup E_{j-1}$, by Proposition 7.1(i), for $i \in K_{D_e}$, pseudo-observations $(\widehat{U}_{i,a_e;D_e}, \widehat{U}_{i,b_e;D_e})$ from $c_{a_e,b_e;D_e}$ can be defined by

$$\widehat{U}_{i,a_e;D_e} = R_{a_e|D_e}\left(\widehat{Z}_{i,a_e} \mid \widehat{\boldsymbol{Z}}_{i,D_e};\ \hat{\theta}(E_{1:j-1})\right), \quad \widehat{U}_{i,b_e;D_e} = R_{b_e|D_e}\left(\widehat{Z}_{i,b_e} \mid \widehat{\boldsymbol{Z}}_{i,D_e};\ \hat{\theta}(E_{1:j-1})\right). \quad (7.4)$$

The full procedure is detailed in Section D of the supplement.

7.3 Model selection for X-vines

Model selection for X-vines based on a random sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ from F involves:

(1) selecting a regular vine sequence $\mathcal{V} = (T_1, \ldots, T_{d-1})$;

(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 5.5 in tree T_1. Let $\mathcal{B}_{\mathcal{R},1:T} = (\mathcal{B}_{\mathcal{R},1}, \ldots, \mathcal{B}_{\mathcal{R},T})$ 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 $\mathcal{B}_{\mathcal{R},t}$, $t = 1, \ldots, T$, 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 (7.3). The averaged-AIC value is $\mathrm{AIC}_e\left(\mathcal{B}_{\mathcal{R},t}\right) = 2\nu^{(t)} - \frac{1}{2}\left\{\log \mathcal{L}_{\mathcal{R},e}\left(\hat{\theta}_e^{(t,a_e)}\right) + \log \mathcal{L}_{\mathcal{R},e}\left(\hat{\theta}_e^{(t,b_e)}\right)\right\}$, 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 $\mathrm{AIC}_e\left(\mathcal{B}_{\mathcal{R},1}\right), \ldots, \mathrm{AIC}_e\left(\mathcal{B}_{\mathcal{R},T}\right)$.

Similarly, to specify $\mathcal{C}$ and select the bivariate parametric copula family among a list of

candidates for each edge in $E_2 \cup \cdots \cup E_{d-1}$, we follow common practice in vine copula modelling and choose the family with the lowest AIC; see Brechmann (2010) and Czado (2019, Section 8.1).

It is worthwhile to note that when implementing the model selection step for an edge e in E_j, only the trees $T_1, \ldots, T_j$ need to have been selected, but not the trees $T_{j+1}, \ldots, T_{d-1}$. 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 regular vine sequences on d elements is equal to $d!\, 2^{\binom{d-2}{2}-1}$, making it impossible to go through all possible vine sequences. We adopt a 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, on the node set $N_1 = [d]$, we follow a procedure as in Engelke and Volgushev (2022) and Hu et al. (2024). For every pair $\{a, b\}$ of distinct elements in $[d]$, let $w_{a,b}$ be a nonnegative weight derived from the data. We use an empirical version of the tail dependence coefficient $\chi_{a,b} = R_{a,b}(1,1)$, setting $w_{a,b}$ equal to $\widehat{\chi}_{a,b} = \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}\left\{\widehat{U}_{i,a} \leqslant k/n, \widehat{U}_{i,b} \leqslant k/n\right\}$, where $\widehat{U}_{i,a}$ and $\widehat{U}_{i,b}$ are defined as in (7.1). Another possible edge weight could be the empirical extremal variogram as in Engelke and Volgushev (2022). For subsequent trees $T_2, \ldots, T_{d-1}$, the edge weight is chosen to be the absolute value of the empirical Kendall's tau. All trees are selected as maximum spanning trees; see, for example, Czado (2019, Section 8.3).

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 $q \in \{1, \ldots, d-2\}$. 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, or 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 small with respect to the dimension d.

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 $j \geqslant 2$, 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 $\mathrm{mBIC}(1) = 0$ and, for $q \in \{2, \dots, d-1\}$, put

$$\mathrm{mBIC}(q) = \sum_{j=2}^{q} \sum_{e \in E_j} \left[\mathbb{1}(\theta_e \neq 0) \left\{ \log n_{D_e} - 2 \log \left(\frac{\psi_e}{1 - \psi_e} \right) \right\} - 2 \log \left\{ \frac{\mathcal{L}_{\mathcal{C},e}\left(\hat{\theta}_e\right)}{(1 - \psi_e)^{-1}} \right\} \right] \tag{7.5}$$

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

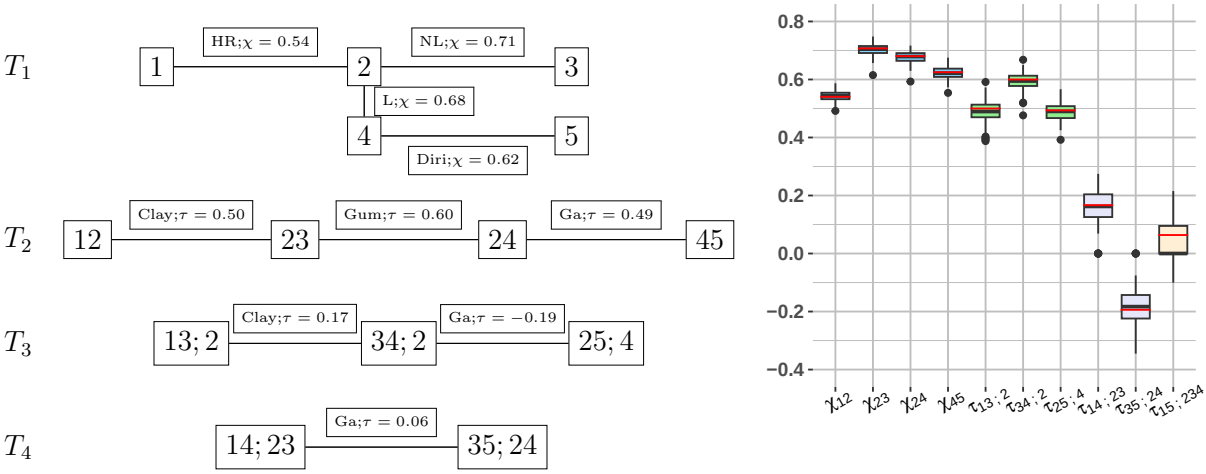
8 Simulation study

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. 2a. 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. Fig. 2a shows the values specified for the tail dependence coefficient $\chi = R(1, 1)$ or for Kendall's τ for each edge. The formulas connecting the parameters of the families of tail copula densities with χ are given in Appendix E.1 in the supplement.

Implementing simulation algorithms as detailed in Section 6 and Appendix A in the sup-



(a) (b)

Figure 2: (a) 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. (b) Box-plots (■ for T_1 ■ for T_2 ■ for T_3 ■ for T_4) of dependence measure estimates from bivariate parametric families selected sequentially from the data for the X-vine specification in (a). The red lines indicate the specified parameter values.

plement, we generate multivariate inverted Pareto random samples $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ associated with r. As in Eq. (7.1), we transform to $\widehat{\boldsymbol{U}}_1, \ldots, \widehat{\boldsymbol{U}}_n$, with $\hat{U}_{i,j}$ based on the rank of $Z_{i,j}$ among $Z_{1,j}, \ldots, Z_{n,j}$. We then take the sub-sample $\{\widehat{\boldsymbol{Z}}_i\}_{i \in K}$ with $K = \bigcup_{j \in [d]} K_j$ and K_j as in (7.2).

We set $(n, k) = (4\,000, 200)$ and perform sequential parameter estimation as in Section 7.2 over 200 repetitions. We obtain maximum pseudo-likelihood estimates and corresponding dependence measures, $\widehat{\chi}$ and $\widehat{\tau}$, using the relations in Appendix E.1 of the supplement, for each of the ten edges in the vine. In Fig. 2b, box-plots present dependence measure estimates of parametric families based on the X-vine specification in Fig. 2a. The four left-most box-plots (■) show the tail dependence coefficient estimates $\widehat{\chi}$ for the four edges $e \in E_1$, and the remaining plots display Kendall's tau estimates $\widehat{\tau}$ for the six edges $e \in E_2 \cup E_3 \cup E_4$. The plot supports the validity of the sequential method for estimating tail dependence measures. We see that estimation uncertainty becomes larger at higher tree levels. Sections F.1.1–F.1.3 in the supplement include additional box-plots: first, of dependence measures as in Fig. 2b, but

supposing that the bivariate parametric families are known, second, of dependence measures with varying sample sizes and threshold exceedances, and third, 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 X-vine specification in Fig. 2a. 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, logistic, negative logistic, and Dirichlet models, 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 R package VineCopula (Nagler et al., 2023).

Using again 200 repetitions with $(n, k) = (4\,000, 200)$, 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 trees is 60%. For individual trees, the proportions are 62% for T_1, 78% for T_2, 55% for T_3, and 10% for T_4. The specific proportions for each edge in each tree are

T_1	99%	40%	60%	49%
T_2		58%	93%	82%
T_3			42%	68%
T_4				11%

The lower proportions observed for e_{23} and e_{45} in E_1 result from the challenge in distinguishing between the logistic and negative 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. For this edge, the algorithm selects the independence copula 110 times out of 200.

In Section F.1.4 of the supplement, we investigate the effective sample sizes for each tree.

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 $j \geqslant 10$. This X-vine specification allows us to explore truncated X-vine models by setting pair-copulas with weak dependence to independence copulas.

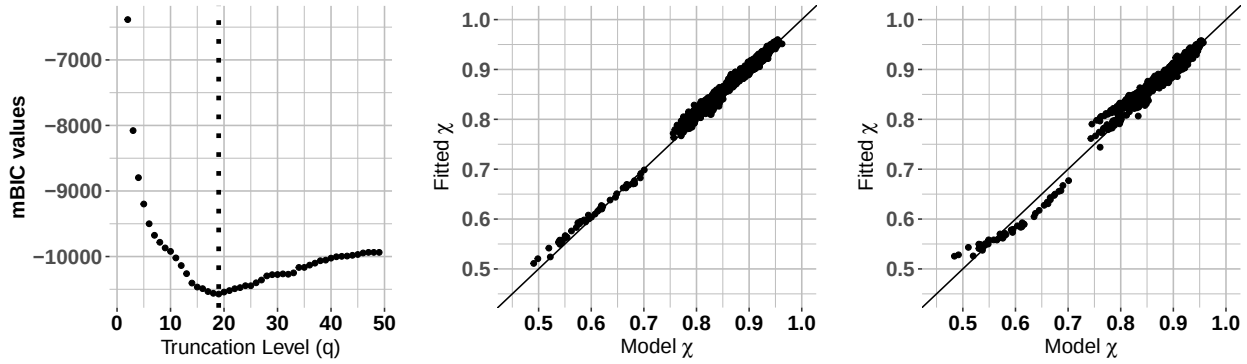
We use a single inverted multivariate Pareto sample of size $n = 1\,000$ from the X-vine model directly, i.e., without rank transformation and thresholding. As in Sections 7.2 and 7.3, we sequentially select trees $T_1, \dots, T_{d-1}$ using $\widehat{\chi}_e$ for $e \in E_1$ and $\widehat{\tau}_e$ for $e \in E_j$, $j \geqslant 2$, as edge weights, with $\widehat{\chi}_e$ as in Eq. (E.1) in the supplement. For each selected tree, we choose the bivariate parametric families with the lowest (averaged) AIC for each edge and estimate the associated parameters. Additionally, independence copulas are chosen in subsequent trees if either $|\widehat{\tau}_e| < 0.05$ or $n_{D_e} < 10$ for each edge.

To explore truncated X-vine models, we use the mBIC in (7.5) with $\psi_0 = 0.9$ as in Nagler et al. (2019). The estimated mBIC-optimal truncation level is $q^* = 19$ (dotted line in Fig. 3a). We assess the goodness of fit by comparing pairwise χ-values from the true X-vine model with those from the fitted 50-dimensional X-vine model via Monte Carlo simulations (as explained in Section E.2 in the supplement) in Fig. 3b. For the truncated X-vine model, the χ-plot in Fig. 3c 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 Statistics[2].

[2] `https://www.bts.dot.gov`



(a) (b) (c)

Figure 3: Simulation study: (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 true X-vine model to those obtained from the fitted 50-dimensional X-vine model via Monte Carlo simulations. (c) Similar to (b) but for the truncated fitted X-vine model.

These flight delay data were analysed by Hentschel et al. (2024) who first selected airports with a minimum of 1 000 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. (2024) 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. 4a.

Let $\widehat{\boldsymbol{x}}_i$, for $i = 1, \ldots, n$, denote the measurements. Following the standardisation process with the rank transformation (7.1) as in Section 7.2, we take the sub-sample $\widehat{\boldsymbol{z}}_i = (n/k)\widehat{\boldsymbol{u}}_i$ for $i \in K = \bigcup_{j\in[d]} K_j$ as in (7.2). We choose $k = 0.13n$ in order to have a large enough effective sample size with respect to the number of variables.

Assigning $\widehat{\chi}_e$ for $e \in E_1$ and $\widehat{\tau}_e$ for $e \in E_j$, $j \geqslant 2$, as edge weights, we select the trees of the

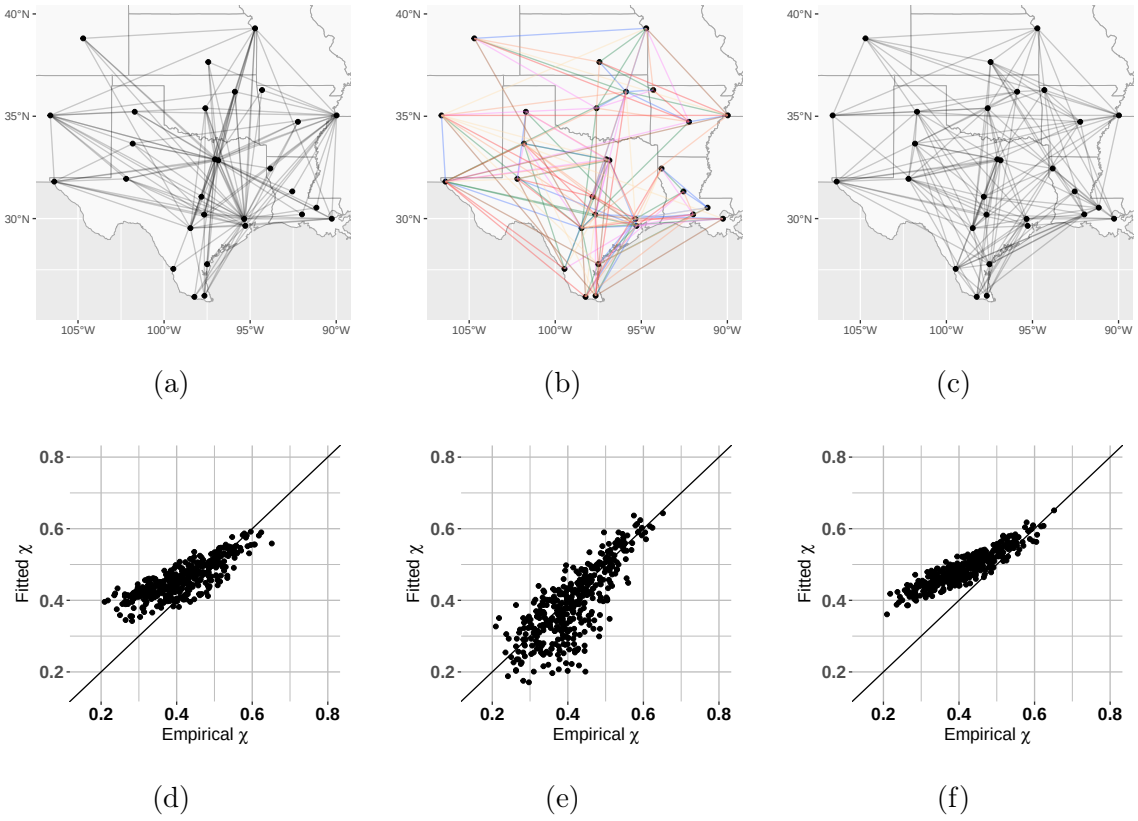


Figure 4: (a) Flight graph illustrating connection flights between airports joined by edges in the Texas cluster. (b) The X-vine graph with the first seven trees superimposed, showing for trees T_2 to T_7 only the 113 out of 147 edges for which a copula other than the independence copula is selected, 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.

regular vine sequentially as in Section 7.3. Fig. F.5a in the supplement shows the first maximum spanning tree. The corresponding estimated tail dependence coefficients vary between 0.42 and 0.65, with a mean of 0.55. 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 chosen with the following pair-copula families (number of edges in 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 T_{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, we compare bivariate and trivariate empirical tail dependence coefficients with those from the fitted X-vine models in Section F.2.2 in the supplement. Fig. F.6 indicates a satisfactory fit, both for the full 29-dimensional model and for the truncated model.

To further explore truncated X-vine models with a lower truncation level, we use the mBIC in Eq. (7.5). Fig. F.7a in the supplement 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 in Fig. **??**. 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. 4b, 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 vine tree sequence from T_1 to T_7 is presented in Fig. F.5.

Returning to model comparison, we consider extremal Hüsler–Reiss graphical models (Engelke et al., 2021; Hentschel et al., 2024). Hentschel et al. (2024) obtained a sparse Hüsler–Reiss graphical model using the EGlearn algorithm (Engelke et al., 2021). The tuning parameter $\rho \geqslant 0$ controls sparsity. The empirical extremal variogram matrix $\widehat{\Gamma}$ is used as edge weight for the *minimum* spanning tree with smaller elements in $\widehat{\Gamma}$ indicating stronger extremal dependence. Following 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 across tuning parameter values ρ from the test set. The resulting sparse extremal graph with 148 edges and $\rho^* = 0.1$ is shown in Fig. 4c.

We assess the goodness of fit using the entire sub-sample and compare empirical tail dependence coefficients to those obtained from the fitted graphical models as explained in Appendix E.2 in the supplement: the Hüsler–Reiss extremal graphical model for the flight graph in Fig. 4d, the truncated X-vine model with $q^* = 7$ in Fig. 4e, and the extremal Hüsler–Reiss

graphical model with $\rho^* = 0.1$ in Fig. 4f, respectively.

The flexibility of a vine-based dependence model 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 $j \geqslant 2$, a total of 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 biased towards higher tail dependence. Fig. F.8 in the supplement presents a model comparison similar to Figs. 4d–4f, but focusing exclusively on the Hüsler–Reiss model and using empirical tail dependence coefficients derived from the empirical extremal variogram matrix.

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, together with a telescoping product formula for regular vine sequences.

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 sequence 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., 2024) 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

The authors thank Manuel Hentschel for providing the US flight delay data and gratefully acknowledge two reviewers and an associate editor for valuable comments. This work is supported financially by the *Fonds de la Recherche Scientifique – FNRS*, Belgium (grant T.0203.21).

Data availability

The R package Xvine, available from `https://github.com/JeongjinLee88/Xvine`, permits to replicate all results, including tables and figures, in this article. The flight data used in Section 9 is also included in the package.

Conflict of interest

The authors declare no competing interests.

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 Muünchen.

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.

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

Czado, C. (2019). *Analyzing dependent data with vine copulas*, 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.

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., M. Hentschel, M. Lalancette, and F. Röttger (2024). Graphical models for multivariate extremes. Available at `https://arxiv.org/abs/2402.02187`.

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. Available at `https://arxiv.org/abs/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 (2024). Statistical inference for Hüsler–Reiss graphical models through matrix completions. *Journal of the American Statistical Association*, `https://doi.org/10.1080/01621459.2024.2371978`.

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

Hu, S., Z. Peng, and J. Segers (2024). Modelling multivariate extreme value distributions via Markov trees. *Scandinavian Journal of Statistics 51*(2), 760–800.

Hüsler, J. and R.-D. Reiss (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statistics & Probability Letters 7*(4), 283–286.

Jaworski, P. (2006). On uniform tail expansions of multivariate copulas and wide convergence of measures. *Applicationes Mathematicae 33*, 159–184.

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.

Li, H. and P. Wu (2013). Extremal dependence of copulas: A tail density approach. *Journal of Multivariate Analysis 114*, 99–111.

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.

List of Figures