Extremes manuscript No.
(will be inserted by the editor)

Multivariate extreme value copulas with factor and tree dependence structures

David Lee · **Harry Joe**

Received: date / Accepted: date

Abstract Parsimonious extreme value copula models with $O(d)$ parameters for d observed variables of extrema are presented. These models utilize the dependence characteristics, including factor and tree structures, assumed on the underlying variables that give rise to the data of extremes. For factor structures, a class of parametric models is obtained by taking the extreme value limit of factor copulas with non-zero tail dependence. An alternative model suitable for both factor and tree structures imposes constraints on the parametric Hüsler-Reiss copula to get representations in terms of $O(d)$ other parameters. Dependence properties are discussed. As the full density is often intractable, the method of composite (pairwise) likelihood is used for model inference. Procedures to improve the stability of bivariate density evaluation are also developed. The proposed models are applied to two data examples — one for annual extreme river flows and one for monthly extremes of daily stock returns.

Keywords Extreme value limit · Gaussian quadrature · Hüsler-Reiss distribution · Parsimonious dependence · Vine graphical model

Mathematics Subject Classification (2000) 62H12 · 62H20 · 62P12

1 Introduction

In multivariate statistics based on the Gaussian distribution, the factor model is a parsimonious model when the dependence among the observed variables can be explained from latent variables. In the more general multivariate modelling based on copulas, factor copula models have been developed in recent years as well as parsimonious copula models based on Markov tree structures and their vine extensions. The goal of this paper is to use theory from copula modelling to construct multivariate extreme value copulas with factor and tree dependence structures. They offer better interpretation than saturated dependence models and are less prone to over-fitting, especially when the dimension is high.

To model the relationship among multivariate observations, the two simplest structures are factor and tree dependence (Markov trees). A 1-factor model assumes that variables are linked to a common, single latent factor, through which dependence is generated. Meanwhile, a Markov tree assumes the d observed variables are connected through $d-1$ bivariate acyclic linkages, and non-neighbouring variables are conditionally independent given the variables along the tree path. An example for time-ordered observations is the autoregressive model with order 1. In spatial applications where data are recorded from stations at different geographic locations, the tree can be drawn according to their locations so that nearest neighbours are linked.

The factor structure can be extended to general p-factor models ($p \geq 2$) and the bi-factor model and its generalizations where variables are linked to both common and group-specific factors. The Markov tree structure can be extended by adding layers of trees which introduce conditional dependencies. Using copulas, these dependence structures have been studied beyond the Gaussian context, see, e.g., Krupskii and Joe (2013) for factor copulas and Brechmann et al (2012) for truncated vine copulas, where the multiple tree dependence known as vines comes from Bedford and Cooke (2001, 2002). The cases with discrete or mixed continuous/discrete response types have also been dealt with (see Sections 3.9 and 3.10 of Joe (2014) and the references contained therein).

David Lee
Department of Statistics, University of British Columbia
Vancouver, BC, Canada V6T 1Z4
E-mail: david.lee@stat.ubc.ca, dav001@gmail.com

Harry Joe
Department of Statistics, University of British Columbia
Vancouver, BC, Canada V6T 1Z4

Copula models with factor or Markov tree dependence structures have conditional independence relations. For the 1-factor structure with d observed variables $Y_1, \dots, Y_d$ and a latent variable W, the joint density has the form

$$f_{Y_1,\dots,Y_d}(y_1,\dots,y_d) = \int \prod_{i=1}^{d} f_{Y_i|W}(y_i|w)\,\mathrm{d}F_W(w). \tag{1}$$

For the Markov tree structure when variables have been indexed so that the edges of the tree are (Y_{k_i}, Y_i), $k_i \in \{1,\dots,i-1\}$, $i = 2,\dots,d$, the joint density has the form

$$f_{Y_1,\dots,Y_d}(y_1,\dots,y_d) = f_{Y_1} \prod_{i=2}^{d} f_{Y_i|Y_{k_i}}(y_i|y_{k_i}) = \frac{\prod_{i=2}^{d} f_{Y_{k_i},Y_i}(y_{k_i}, y_i)}{\prod_{i=1}^{d} [f_{Y_i}(y_i)]^{\deg(i)-1}}, \tag{2}$$

where $\deg(i)$ is the number of edges of the tree with Y_i at one edge, based on the Hammersley-Clifford theorem (Besag (1974)). However, direct construction of families of multivariate extreme value distributions that satisfy (1) and (2) appears difficult. Instead, we construct classes of multivariate extreme value (EV) models through (a) extreme value limits of multivariate parametric copula families with a factor structure, and (b) parsimonious representations of existing extreme value copulas with $\binom{d}{2}$ dependence parameters, by expressing each as functions of other parameters of order $O(d)$; it is applicable to both factor and tree dependence structures. For (b), we consider, in particular, the Hüsler-Reiss distribution (Hüsler and Reiss (1989)) which was obtained as a non-standard EV limit of the multivariate Gaussian distribution. This distribution is chosen for its flexibility in imposing various dependence structures. Structured Hüsler-Reiss models can be considered as EV limits of Gaussian factor or vine models. We consider the multivariate EV models as theoretically valid for extreme observations with underlying processes that have a factor or tree dependence interpretation.

Inference for models being the EV limits of factor copulas is complicated by the fact that the evaluation of copula densities could be numerically unstable even in bivariate margins. One contribution of this paper is on the development of appropriate techniques, such as integral transformations and quadrature methods, to stabilize computation to the accuracy required for the estimation of standard errors.

The paper is organized as follows. Section 2 provides an overview of factor and vine copulas as well as tail dependence functions that characterize the tail behaviour of a copula. We introduce the EV factor copula model in Section 3 and give examples with specific linking copulas. The class of structured Hüsler-Reiss distributions is discussed in Section 4. Sections 3 and 4 contain the main new contributions for parsimonious EV copula models and their dependence properties. A brief outline on parameter estimation using composite likelihood methods (Lindsay (1988); Cox and Reid (2004)) is given in Section 5. A simulation study is provided in Section 6, and two data examples are presented in Section 7. Concluding remarks are in Section 8. The technical details on integral stabilization for computing the bivariate densities of EV factor copulas are given in Appendix B.

2 Overview of parsimonious copula structures and tail dependence function

Copulas are multivariate distributions on the unit hypercube with Uniform(0,1) (U(0,1) hereafter) margins. For a random vector $(X_1,\dots,X_d) \sim F$ with marginal distributions $F_1,\dots,F_d$, Sklar (1959) shows that there exists a corresponding copula C such that $F(x_1,\dots,x_d) = C(F_1(x_1),\dots,F_d(x_d))$, and that C is unique if F is continuous. Any multivariate distribution can thus be decomposed into its marginal and dependence components. In this section, multivariate extreme value distributions are specified with copulas for the dependence structure, with generalized extreme value distributions as univariate margins not mentioned. Extreme value copulas have a stability property that is presented below in (4). We will focus on the dependence properties of EV copulas, assuming that univariate margins have been transformed to U(0,1). It is noted that the unit Fréchet distribution is also widely used in the EV literature.

2.1 Factor and tree structures

In traditional multivariate analysis, one uses the Gaussian factor model when it is plausible to assume that the dependence behaviour of the observed variables is driven by one or more latent factors. Examples include students' scores on various subjects being driven by intrinsic qualities like analytical reasoning ability and creativity, or returns on stocks in the same sector being driven by the global economic environment and sector-specific idiosyncrasies. The factor copula model (Krupskii and Joe (2013)) is an extension that relaxes the assumption of Gaussianity between latent and observed variables. Figure 1 presents a schematic diagram of the dependence layers for 1- and 2-factor copulas. Each edge represents one bivariate linking copula with the distribution indicated, for example, the edge $V_2U_2; V_1$ indicates that the copula links the conditional distributions

$F_{V_2|V_1}(v_2|v_1)$ (which equals $F_{V_2}(v_2)$ due to the assumed independence between the latent variables V_1 and V_2) and $F_{U_2|V_1}(u_2|v_1)$. The 2-factor model can be interpreted as nesting a 1-factor model on the observed variables $U_1, \ldots, U_d$ given V_1 within the 1-factor model for $U_1, \ldots, U_d$ unconditionally. The observed variables are assumed to be independent of each other upon conditioning on the latent variable(s). Specification of a p-factor copula thus requires $p \times d$ bivariate copulas in total, or $O(d)$ dependence parameters.

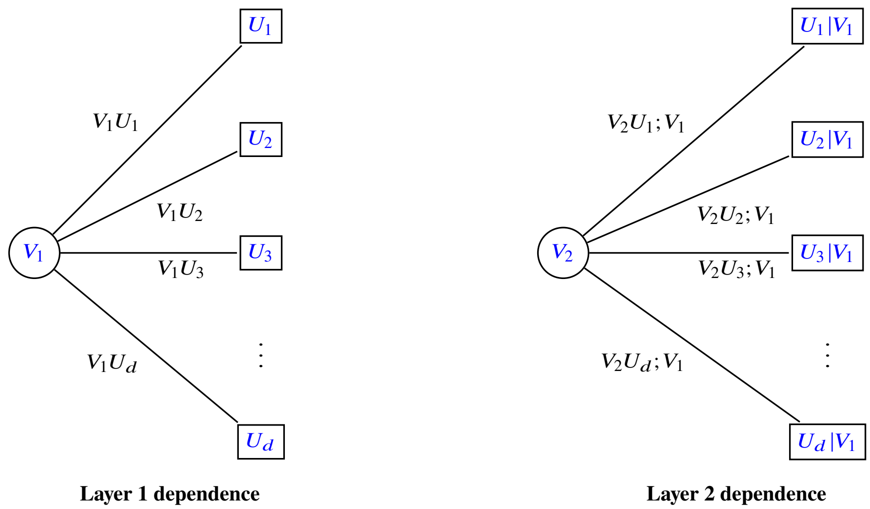

Fig. 1 First layer dependence between observed and latent variables (left) for both 1- and 2-factor copula models, and the additional second layer dependence (right) for the 2-factor model. The observed variables are denoted by $U_1, \ldots, U_d$ while the latent variables are denoted by V_1 and V_2. There are edges for linking copulas between each observed variable and the latent variable V_1 on the left plot. For the 2-factor copula, there are additional edges for linking copulas between the conditional variable $U_i|V_1$ and the latent variable V_2 to summarize the dependence between U_i and V_2 conditional on V_1. See (5) and (8) for the copula of the observed variables.

On the other hand, a tree structure connects variables directly and is plausible when there is a natural complete or partial ordering among the variables. Instances include spatially connected monitoring stations and temporally related observations. Figure 2 shows two examples of Markov trees. In vine copulas or the pair-copula construction (Bedford and Cooke (2001, 2002); Aas et al (2009)), each connection is represented by a bivariate copula. Markov trees have the conditional independence property given variables along the path, and so for example variables 1 and 3 are independent given 2 for both trees in Figure 2. When conditional independence is an unrealistic assumption, layers of trees can be built on the connecting edges of the previous tree, resulting in a vine. For example, adding a copula linking the edges (1,2) and (2,3) introduces dependence between the conditional distributions 1|2 and 3|2.

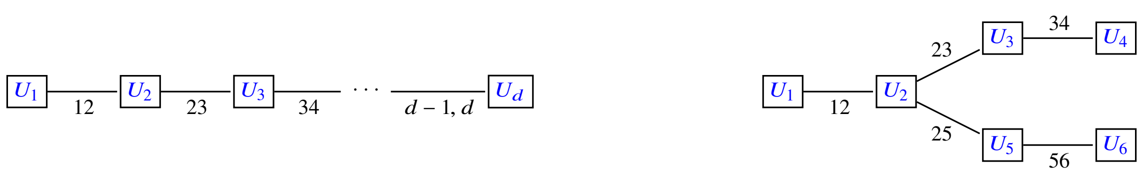

Fig. 2 Illustration of Markov tree structures. The U's in the connecting edges are dropped for brevity.

The dependence structure is sometimes well approximated by a few trees. If high order residual dependence can be ignored, a parsimonious vine model can be obtained by truncation. A p-truncated vine with $p < d - 1$ is one with all linking copulas beyond the pth tree being the independence copula. With p fixed, the number of linking copulas is of the order $O(d)$. Note that a p-factor copula is structurally the same as a p-truncated C-vine (i.e., one variable serves as the connection hub for all other variables in each tree) if the latent variables are included.

2.2 Tail dependence function and EV limit

Tail dependence functions (Nikoloulopoulos et al (2009); Joe et al (2010)) describe the joint lower or upper tail behaviour of a copula. For a d-dimensional copula C, the lower and upper tail dependence functions, denoted respectively as $b(\cdot)$ and $b^*(\cdot)$, are defined as

$$b(w_1,\dots,w_d)=\lim_{u\to 0^+}u^{-1}C(uw_1,\dots,uw_d);\quad b^*(w_1,\dots,w_d)=\lim_{u\to 0^+}u^{-1}\overline{C}(1-uw_1,\dots,1-uw_d),$$

where $\overline{C}$ is the survival function of C. The tail dependence function is also known as the tail copula (see, e.g., Schmidt and Stadtmüller (2006)), although it is not a distribution function. When C is bivariate, $b(1,1)$ and $b^*(1,1)$ reduce to the lower and upper tail dependence coefficient, respectively. These quantities are extensively used in extreme value theory as measures of the strength of asymptotic dependence (see, e.g., Coles (2001)). The b^* function of C is equivalent to the b function of its reflected or survival copula $\widehat{C}(u_1,\dots,u_d)=\overline{C}(1-u_1,\dots,1-u_d)$; if $(U_1,\dots,U_d)\sim C$, then the reflection $(1-U_1,\dots,1-U_d)\sim\widehat{C}$. Without loss of generality, we focus on the lower tail in the rest of this section.

Other quantities we will use in the later sections include the marginal tail dependence function

$$b_S(w_{k_1},\dots,w_{k_m})=\lim_{u\to 0^+}u^{-1}C_S(uw_{k_1},\dots,uw_{k_m}),$$

where $\varnothing\neq S=\{k_1,\dots,k_m\}\subset I_d\triangleq\{1,\dots,d\}$, $m=|S|$, and $C_S(u_{k_1},\dots,u_{k_m})=\mathbb{P}\{U_j\le u_j:j=k_1,\dots,k_m\}$ is a marginal copula of $(U_1,\dots,U_d)\sim C$, as well as the conditional tail dependence function

$$b_{k_1|k_2,\dots,k_m}(w_{k_1}|w_{k_2},\dots,w_{k_m})=\lim_{u\to 0^+}C_{U_{k_1}|U_{k_2},\dots,U_{k_m}}(uw_{k_1}|uw_{k_2},\dots,uw_{k_m}),$$

where $k_i\in I_d$ and $k_i\neq k_j$ if $i\neq j$. Another useful quantity involves the probability of the union of events rather than the intersection:

$$a(w_1,\dots,w_d)=\lim_{u\to 0^+}u^{-1}\mathbb{P}\left\{\bigcup_{i=1}^{d}\{U_i\le uw_i\}\right\}=\sum_{\varnothing\neq S\subset I_d}(-1)^{|S|-1}b_S(w_i,i\in S)\tag{3}$$

from the inclusion-exclusion principle. For EV copulas, $a(w_1,\dots,w_d)$ is known as the stable tail dependence function (Huang (1992); Drees and Huang (1998)). It follows from the results in Galambos (1987) and Nikoloulopoulos et al (2009) that the lower EV limit of a copula C is given by $C_{EV}(u_1,\dots,u_d)=\lim_{n\to\infty}\widehat{C}^n(u_1^{1/n},\dots,u_d^{1/n})=\exp\{-a(w_1,\dots,w_d)\}$, where $w_i=-\log u_i$, $i=1,\dots,d$. If C has lower tail dependence, then $a(w_1,\dots,w_d)\neq w_1+\cdots+w_d$ and C_{EV} has positive dependence. The upper EV limit of C can be obtained as the lower EV limit of $\widehat{C}$. An EV copula satisfies the max-stability condition (see Galambos (1987))

$$C^n(u_1^{1/n},\dots,u_d^{1/n})=C(u_1,\dots,u_d),\tag{4}$$

and is suitable for modelling joint extrema if the extremes of different variables can be attributed to common events.

The EV limit of p-factor copulas and p-truncated C-vines (a boundary class of vines) has p-dimensional integrals. For general truncated vines, the EV limit is intractable in the sense of having $O(d)$-dimensional integrals (see Section 3.15 of Joe (2014) for an example on 1-truncated D-vine). Numerically, 1- and 2-dimensional integrals can be handled within likelihood optimization, so EV limits of 1- and 2-factor copulas with tail dependence are numerically tractable but EV limits of truncated vine copulas with tail dependence cannot be considered for arbitrary dimensions. For EV copulas that are limits of the truncated vine structure, we instead adapt, in Section 4, the derivation of the Hüsler-Reiss model.

3 Construction and parametric examples of EV factor copula models

In this section, we obtain the EV limit of factor copulas, suitable for modelling multivariate extremes with an underlying factor structure. We name the resulting models as EV factor copulas. Section 3.1 has details for one factor, Section 3.2 has the extension to more factors and Section 3.3 has the bi-factor extension for the case where variables can be divided into non-overlapping groups. Bivariate dependence properties are discussed in Section 3.4, and some parametric examples of the construction are given in Section 3.5. For notational simplicity, we replace the subscript U_i for C and b by i in the following discussion, for instance $C_{i|V}=C_{U_i|V}$ and $b_{i|V}=b_{U_i|V}$ representing the conditional cdf and conditional tail dependence function of the linking copulas, respectively. All copulas are assumed to be twice continuously differentiable in the following.

3.1 1-factor EV copula

The 1-factor copula of Krupskii and Joe (2013) is

$$C(u_1,\dots,u_d)=\int_0^1\prod_{i=1}^{d}C_{i|V}(u_i|v)\,\mathrm{d}v,\tag{5}$$

where $C_{iV}(u_i, v)$ is the bivariate copula linking U_i to the latent variable V and $C_{i|V}(u_i|v) = \partial C_{iV}(u_i, v)/\partial v$. Assuming that $b_{i|V}(w|z) = \lim_{u\to 0^+} C_{i|V}(uw|uz)$ exists for $i = 1, \ldots, d$, its lower tail dependence function is

$$b(w_1, \ldots, w_d) = \lim_{u\to 0^+} \int_0^{1/u} \prod_{i=1}^{d} C_{i|V}(uw_i|uz)\,\mathrm{d}z = \int_0^{\infty} \lim_{u\to 0^+} \left[1\{z \le u^{-1}\} \cdot \prod_{i=1}^{d} C_{i|V}(uw_i|uz)\right] \mathrm{d}z = \int_0^{\infty} \prod_{i=1}^{d} b_{i|V}(w_i|z)\,\mathrm{d}z, \tag{6}$$

where $1\{A\}$ is the indicator function for the event A. For a given set of conditional distributions $C_{1|V}, \ldots, C_{d|V}$, it may be possible to apply Lebesgue's dominated convergence theorem to justify the exchange of the limit and integral in (6); an example with C_{iV} being Mardia-Takahasi-Clayton-Cook-Johnson (MTCJ) (Mardia (1962); Takahasi (1965); Clayton (1978); Cook and Johnson (1981)) copulas (see Section 3.5) is given in Appendix A. Alternatively, using results from Joe et al (2010) (see Section 3.16 and Theorem 8.76 of Joe (2014)), a sufficient condition for the validity of this result is that each $b_{i|V}(\cdot|z)$ is a proper distribution on $[0, \infty)$ for all $z > 0$. This condition is usually much easier to check and is satisfied by many bivariate copulas with lower tail dependence, such as the reflected Gumbel (Gumbel (1960)) and MTCJ copulas.

Similarly, $b_S(w_i, i \in S) = \int_0^\infty \prod_{i\in S} b_{i|V}(w_i|z)\,\mathrm{d}z$ for any non-empty subset S of I_d. By letting $m_i = b_{i|V}(w_i|z)$ and using the multinomial formula, we obtain

$$a(w_1, \ldots, w_d) = \int_0^{\infty} \left(\sum_i m_i - \sum_{i<j} m_i m_j + \cdots + (-1)^{d-1} m_1 \cdots m_d\right) \mathrm{d}z = \int_0^{\infty} \left[1 - \prod_{i=1}^{d} (1 - m_i)\right] \mathrm{d}z \tag{7}$$

for (3). The copula of the lower EV limit of C is given by

$$C_{EV}(u_1, \ldots, u_d) = \exp\left\{-\int_0^{\infty} \left[1 - \prod_{i=1}^{d} \left(1 - b_{i|V}(w_i|z)\right)\right] \mathrm{d}z\right\},$$

where $w_i = -\log u_i$ (this relationship emphasizes the min-stable representation of the cdf, and will be omitted thereafter when we mention EV copulas). Note that C_{EV} depends only on the conditional tail dependence functions of the linking copulas. Similarly, the upper EV limit of C can be obtained as the lower EV limit of its reflected copula $\widehat{C}$.

3.2 2-factor EV copula and higher order generalization

A similar technique applies to 2-factor copulas, with linking copulas C_{i,V_1} between the ith observed and first latent variable, and $C_{i,V_2;V_1}$ between the ith observed and second latent variable conditional on the first latent variable. Let $C_{i|V_1}$ and $C_{i|V_2;V_1}$ $(= \partial C_{i,V_2;V_1}(x, y)/\partial y)$ be the respective conditional cdf's. The 2-factor copula of Krupskii and Joe (2013) is

$$C(u_1, \ldots, u_d) = \int_0^1 \int_0^1 \prod_{i=1}^{d} C_{i|V_2;V_1}(C_{i|V_1}(u_i|v_1)|v_2)\,\mathrm{d}v_2\mathrm{d}v_1. \tag{8}$$

The tail dependence function of the 2-factor copula can be derived as in the preceding subsection. Assuming that $b_{i|V_1} = b_{i|V_1}(w_i|z_1) = \lim_{u\to 0^+} C_{i|V_1}(uw_i|uz_1)$, $i = 1, \ldots, d$, are proper distributions on $[0, \infty)$ for all $z_1 > 0$,

$$b(w_1, \ldots, w_d) = \lim_{u\to 0^+} \int_0^1 \int_0^{1/u} \prod_{i=1}^{d} C_{i|V_2;V_1}\left(C_{i|V_1}(uw_i|uz_1)|v_2\right) \mathrm{d}z_1\mathrm{d}v_2 = \int_0^1 \int_0^{\infty} \prod_{i=1}^{d} C_{i|V_2;V_1}\left(b_{i|V_1}(w_i|z_1)|v_2\right) \mathrm{d}z_1\mathrm{d}v_2.$$

The marginal tail dependence function is $b_S(w_i, i \in S) = \int_0^1 \int_0^\infty \prod_{i\in S} C_{i|V_2;V_1}\left(b_{i|V_1}(w_i|z_1)|v_2\right) \mathrm{d}z_1\mathrm{d}v_2$. With a similar approach as in (7), the stable tail dependence function of the lower EV limit is

$$a(w_1, \ldots, w_d) = \int_0^1 \int_0^{\infty} \left(1 - \prod_{i=1}^{d} \left[1 - C_{i|V_2;V_1}\left(b_{i|V_1}(w_i|z_1)|v_2\right)\right]\right) \mathrm{d}z_1\mathrm{d}v_2, \tag{9}$$

and the copula is $C_{EV}(u_1, \ldots, u_d) = \exp\{-a(w_1, \ldots, w_d)\}$. It involves (a) the conditional tail dependence functions for the copulas linked to the first latent factor, and (b) the (conditional cdf's of) copulas that link the conditional distributions of observed variables given the first latent factor and that of the second latent factor given the first.

This can be generalized to p-factor copulas with $p \ge 3$. With latent factors being $V_1, \ldots, V_p$, we have

$$a(w_1, \ldots, w_d) = \underbrace{\int_0^1 \cdots \int_0^{\infty}}_{p \text{ integrals}} \left(1 - \prod_{i=1}^{d} \left[1 - C_{i|V_p;V_1,\ldots,V_{p-1}}\left(\cdots C_{i|V_2;V_1}\left(b_{i|V_1}(w_i|z_1)|v_2\right) \cdots |v_p\right)\right]\right) \mathrm{d}z_1\mathrm{d}v_2 \cdots \mathrm{d}v_p.$$

3.3 Bi-factor EV copula

The bi-factor model with G groups of non-overlapping variables is a special case of the general p-factor model with $p = G + 1$. Each observed variable is related to the *common* latent variable, denoted now as V_0. Each variable in group g (for $g = 1, \ldots, G$) is additionally linked to latent variable V_g. An example would be stocks in different sectors: A latent factor common to all stocks can be assumed to describe the overall state of the economy, while other latent factors are sector-specific whose effects are more local to the sectors concerned. Table 1 shows a possible dependence structure between the observed and latent variables in a bi-factor model. The presence of a checkmark indicates dependence between the elements of that pair, conditional on the previous latent variables. For example, the checkmark in the cell linking U_4 and V_2 means that $U_4|V_0, V_1$ and $V_2|V_0, V_1$ are dependent. Note that each observed variable depends on exactly two latent variables, one of them being V_0.

Observed Variable	Latent Variable V_0	V_1	V_2	$\cdots$	V_G
U_1	✓	✓			
U_2	✓	✓			
U_3	✓		✓		
U_4	✓		✓		
$\vdots$				$\ddots$	
U_d	✓				✓

Table 1 An example of the dependence structure between observed and latent variables in a bi-factor model

The latent variables $V_0, V_1, \ldots, V_G$ are assumed to be independent. Two variables in group g are conditionally independent given V_0, V_g and two variables in different groups are conditionally independent given V_0. Suppose the indices for group g variables are $k_{g-1} + 1, \ldots, k_g$, for $g = 1, \ldots, G$ where $k_0 = 0$ and $k_G = d$. Let $C_{i|V_g;V_0}$ be the conditional cdf of the bivariate copula $C_{i,V_g;V_0}$, where the V_0 after the semicolon signifies that the arguments of the cdf involve conditional distributions given V_0. The bi-factor copula is

$$C_{1,\ldots,d}(u_1, \ldots, u_d) = \underbrace{\int_0^1 \cdots \int_0^1}_{G+1 \text{ integrals}} \prod_{g=1}^{G} \prod_{i \in \text{group } g} C_{i|V_g;V_0}\left(C_{i|V_0}(u_i|v_0)\,|v_g\right) \mathrm{d}v_0 \mathrm{d}v_1 \cdots \mathrm{d}v_G$$

$$= \int_0^1 \prod_{g=1}^{G} \left[\int_0^1 \prod_{i=k_{g-1}+1}^{k_g} C_{i|V_g;V_0}\left(C_{i|V_0}(u_i|v_0)\,|v_g\right) \mathrm{d}v_g\right] \mathrm{d}v_0,$$

with numerical complexity of a 2-dimensional nested integral. Since the bi-factor copula is a special case of the general $(G+1)$-factor copula model, so is the tail dependence function. With the same assumption on $b_{i|V_0}$ as in the preceding case, we have

$$a(w_1, \ldots, w_d) = \int_0^\infty \left(1 - \prod_{g=1}^{G}\left[\int_0^1 \prod_{i=k_{g-1}+1}^{k_g} \left(1 - C_{i|V_g;V_0}\left(b_{i|V_0}(w_i|z_0)|v_g\right)\right) \mathrm{d}v_g\right]\right) \mathrm{d}z_0,$$

and the corresponding lower EV limit is again $C_{EV}(u_1, \ldots, u_d) = \exp\{-a(w_1, \ldots, w_d)\}$.

3.4 Dependence properties for bivariate margins

The stable tail dependence function for the (i, j) bivariate marginal distribution of a p-factor EV copula is

$$a_{ij}(w_i, w_j) = \int_0^1 \cdots \int_0^\infty \left(m_i + m_j - m_i m_j\right) \mathrm{d}z_1 \mathrm{d}v_2 \cdots \mathrm{d}v_p = w_i + w_j - b_{ij}(w_i, w_j),$$

where $m_i = C_{i|V_p;V_1,\ldots,V_{p-1}}\left(\cdots C_{i|V_2;V_1}\left(b_{i|V_1}(w_i|z_1)|v_2\right)\cdots|v_p\right)$, and b_{ij} is the bivariate marginal tail dependence function of the parent factor copula. With this relationship, we list some dependence properties applicable to the EV factor copula.

- **Concordance ordering**. A bivariate distribution F is more concordant than G, written as $F \succ_c G$, if $F \geq G$ or equivalently $\overline{F} \geq \overline{G}$ pointwise. Increasing in the concordance ordering means that increasing the dependence parameter of a copula leads to stronger dependence, i.e., the copula becomes more concordant. For the 1-factor case, with $V = V_1$, it is given in Krupskii and Joe (2013) that, assuming (a) the linking copula C_{jV} is fixed and $C_{j|V}$ is stochastically increasing (that is, $1 - C_{j|V}(u|v) = \Pr(U_j > u|V = v)$ is increasing in v for all $0 < u < 1$), and (b) C_{iV} increases in the concordance ordering, then the factor copula C_{ij} increases in the concordance ordering. When this is the case, its corresponding tail dependence function b_{ij} is increasing in the parameter(s) of the linking copula C_{iV} by definition. This in turn implies that the EV limit $C_{EV}(u_i, u_j) = u_i u_j \exp\left\{b_{ij}(-\log u_i, -\log u_j)\right\}$ increases in the concordance ordering.
- **1-Factor dependence structure.** Consider variables i, j, k linked to a latent variable V in the 1-factor copula in (5), and assume that the 1-factor copula has lower tail dependence. Suppose that each linking copula with V is stochastically increasing. If $C_{iV} \succ_c C_{jV} \succ_c C_{kV}$ are ordered in the concordance ordering $\succ_c$, then the preceding item implies that $C_{ij} \succ_c C_{ik} \succ_c C_{jk}$ for the bivariate margins of the 1-factor copula, and the same concordance ordering holds for the corresponding bivariate margins of the EV limit. If the variables are indexed so that the strength of dependence (concordance) is decreasing as i increases from 1 to d, then the bivariate dependence of the (i, j) bivariate margin for the 1-factor copula and its EV limit decreases as i, j increase. This is the typical pattern of dependence for the 1-factor structure. By a matrix of empirical bivariate dependence measures, it is possible to assess whether the 1-factor structure is a good approximation. If not, one could consider 2-factor or bi-factor structures assuming there are plausible latent variables.
- **Dependence measures**. The tail dependence function of a factor copula is equal to that of its EV limit. Other measures commonly used in the EV literature include the Pickands dependence function $B_{ij}(w) \triangleq a_{ij}(w, 1-w), 0 \leq w \leq 1$ (Pickands (1981)) and extremal coefficient $\vartheta_{ij} = a_{ij}(1, 1)$ (Smith (1990)). The extremal coefficient takes the range $[1, 2]$ and can be interpreted as the effective number of independent variables. These measures are readily obtained as functions of b_{ij}.
- **Dependence boundaries**. The full d-dimensional EV factor copula becomes the independence (resp. comonotonicity) copula when all linking copulas are tail independent (resp. comonotonicity). The bivariate marginal copula attains these limits when all linking copulas related to the variables concerned do. The EV limit is not joint independence if there is tail dependence for every linking copula connected to the first latent factor.

Therefore, to construct parametric EV copulas with factor structure and interpretable parameters, we start with factor copulas where the bivariate linking copulas are in parametric families that increase in concordance, cover the range of independence to comonotonicity, and satisfy the property of stochastically increasing. These properties are satisfied by 1-parameter families of bivariate copulas that are mentioned in the next subsection.

3.5 Examples of 1-factor and 2-factor EV copulas

In this subsection, we present some examples of 1- and 2-factor EV copula families as the EV limits of some bivariate parametric copula families where properties, including tail dependence functions, are given in Chapter 4 of Joe (2014).

1-factor with Dagum (inverse Burr) conditional tail dependence functions

Let $b_{i|V}(w_i|z) = \left[1 + (w_i/z)^{-\delta_i}\right]^{-1/\delta_i - 1}$ with $\delta_i > 0$, $i = 1, \ldots, d$. This is a special case of the Dagum (Dagum (1975)) or inverse Burr distribution and can be derived as the lower conditional tail dependence function of the MTCJ copula, or the upper one for the Galambos copula (Galambos (1975)) with dependence parameters δ_i that increase in the concordance ordering. The copula is $C_{EV} = \exp\{-a(w_1, \ldots, w_d)\}$, where

$$a(w_1, \ldots, w_d) = \int_0^\infty \left[1 - \prod_{i=1}^{d}\left(1 - \left[1 + \left(\frac{w_i}{z}\right)^{-\delta_i}\right]^{-1/\delta_i - 1}\right)\right] \mathrm{d}z.$$

1-factor with Burr (Singh-Maddala) conditional tail dependence functions

Alternatively, let $b_{i|V}(w_i|z) = 1 - \left[(w_i/z)^{\theta_i} + 1\right]^{1/\theta_i - 1}$ with $\theta_i > 1$, $i = 1, \ldots, d$. This is a special case of the Burr Type XII (Burr (1942)) or Singh-Maddala (Singh and Maddala (1976)) distribution and can be derived as the upper conditional tail dependence function of the Gumbel or Joe/B5 (Joe (1993)) copulas with dependence parameters θ_i, again increasing in the concordance ordering. The resulting stable tail dependence function is

$$a(w_1, \ldots, w_d) = \int_0^\infty \left(1 - \prod_{i=1}^{d}\left[\left(\frac{w_i}{z}\right)^{\theta_i} + 1\right]^{1/\theta_i - 1}\right) \mathrm{d}z.$$

2-factor with Dagum conditional tail dependence functions for factor 1

For the 2-factor EV copula with Dagum conditional tail dependence functions and dependence parameters δ_i, $i = 1, \ldots, d$, for the first factor, we obtain from (9) that

$$a(w_1, \ldots, w_d) = \int_0^1 \int_0^\infty \left(1 - \prod_{i=1}^{d} \left[1 - C_{i|V_2;V_1}\left(\left[1 + (w_i/z_1)^{-\delta_i}\right]^{-1/\delta_i - 1} \middle| v_2; \theta_i\right)\right]\right) \mathrm{d}z_1 \mathrm{d}v_2,$$

where $C_{i|V_2;V_1}$ is the conditional cdf of the linking copula family between the ith observed variable and the second latent factor with dependence parameter θ_i.

4 Structured Hüsler-Reiss copula

In this section, we propose another class of parsimonious EV dependence models based on the multivariate Hüsler-Reiss copula (Hüsler and Reiss (1989)). This is an alternative to EV limits of vine copulas, which involve high-dimensional integrals even in bivariate margins and are computationally intractable. For parsimonious submodels, we impose a structure on the parameter matrix of the Hüsler-Reiss copula according to the desired dependence structure. This approach also applies to the factor structure. Note that parsimonious forms of the Hüsler-Reiss copula have been developed in the context of multivariate spatial extremes (see, e.g., Smith (1990) and Davison et al (2012)). The model formulation and bivariate dependence properties of the proposed parsimonious structuring are given in Sections 4.1 and 4.2, respectively, while comparisons with the EV factor copula are in Section 4.3.

4.1 Model formulation

The d-dimensional Hüsler-Reiss copula is

$$C(u_1, \ldots, u_d) = \exp\left\{-\sum_{j=1}^{d} w_j \Phi_{d-1,\Gamma_j}\left(\frac{1}{\delta_{ij}} + \frac{\delta_{ij}}{2}\log\left(\frac{w_j}{w_i}\right), i \neq j\right)\right\}, \quad u_1, \ldots, u_d \in [0, 1], \tag{10}$$

where $w_j = -\log u_j$ and Φ_{d-1,Γ_j} is the $(d-1)$-dimensional Gaussian cdf with zero mean and correlation matrix

$$\Gamma_j = (\gamma_{ik})_{i,k\neq j}, \quad \gamma_{ik} = (\delta_{ij}^{-2} + \delta_{kj}^{-2} - \delta_{ik}^{-2})/(2\delta_{ij}^{-1}\delta_{kj}^{-1}), \tag{11}$$

with $\delta_{ij} = \delta_{ji}$ if $i \neq j$, and $\delta_{ii}^{-1} = 0$ for all i. The multivariate copula in the compact form (10) is attributed to Nikoloulopoulos et al (2009). The Hüsler-Reiss copula is derived as a non-standard EV limit of the Gaussian copula, so that $\left[1 - \rho_{ij}(n)\right] \log n \to \delta_{ij}^{-2} \in (0, \infty)$ as the sample size $n \to \infty$, where $\Sigma(n) = \left(\rho_{ij}(n)\right)_{1\le i,j\le d}$ is the correlation matrix, dependent on n, of a d-variate Gaussian distribution with zero mean and unit variance.

Nikoloulopoulos et al (2009) show that the Hüsler-Reiss copula can be obtained as the limit of the t-EV copula when the dispersion matrix $\Sigma(\nu) = (\rho_{ij}(\nu))$ is such that $\rho_{ij}(\nu) = 1 - 2\delta_{ij}^{-2}/\nu$, and the limit is taken as $\nu \to \infty$. This and the limit argument of the Hüsler-Reiss copula provide a link between the correlations ρ_{ij} of the underlying Gaussian variates and the parameters δ_{ij}. We propose an EV model whereby the ρ_{ij}'s are structured according to the dependence pattern assumed for the data, so as to facilitate model interpretation. This structure is translated to the parameters δ_{ij}. Specifically, we let $\delta_{ij} = \gamma(1 - \rho_{ij})^{-1/2}$, where $\gamma > 0$ is a proportionality constant. Different structures can be applied to ρ_{ij}, for example:

- **Factor structure**. For a given number of factors p, define the $d \times p$ matrix of parameters $\mathbf{L} = (\alpha_{ij})$, $1 \le i \le d; 1 \le j \le p$, where $\alpha_{ij} \in [-1, 1]$ is the parameter for the ith variable and the jth factor. Its role is analogous to the loadings in classical factor analysis. The correlations are given by $\rho_{ij} = \sum_{k=1}^{p} \alpha_{ik}\alpha_{jk}$ for $i \neq j$, and $\rho_{ii} = 1$. For computation, the following alternative parametrization can be more convenient:

$$\mathbf{L} = \begin{bmatrix} \alpha_{11} & \alpha_{12}\sqrt{1-\alpha_{11}^2} & \alpha_{13}\sqrt{\left(1-\alpha_{11}^2\right)\left(1-\alpha_{12}^2\right)} & \cdots & \alpha_{1p}\sqrt{\prod_{j=1}^{p-1}\left(1-\alpha_{1j}^2\right)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{d1} & \alpha_{d2}\sqrt{1-\alpha_{21}^2} & \alpha_{d3}\sqrt{\left(1-\alpha_{d1}^2\right)\left(1-\alpha_{d2}^2\right)} & \cdots & \alpha_{dp}\sqrt{\prod_{j=1}^{p-1}\left(1-\alpha_{dj}^2\right)} \end{bmatrix}.$$

  This parametrization is better suited for numerical optimization as each α_{ij} can algebraically take the range $(-1, 1)$ independently. Note that this model is different from the EV limit of factor copulas introduced in Section 3, with stable tail dependence functions being integrals that do not generally correspond to the sum of Gaussian cdf's for the Hüsler-Reiss copula.

- **Vine/tree structure**. The number of parameters (or correlations) needed to specify a Markov tree is $d-1$, each corresponding to an edge of the associated graphical model (e.g., those in Figure 2). A 2-truncated vine uses $d-2$ additional partial correlation parameters to connect the $d-1$ edges in the Markov tree, and similarly for higher-order truncated vines (see Section 3.9 of Joe (2014)). For a fixed level of truncation $p < (d-1)$, the total number of parameters is $(d-1)+(d-2)+\cdots+(d-p) = p\,[d-(p+1)/2]$ or $O(d)$. The parameters can be labelled according to the Gaussian pair-copula it represents. For example, for a d-dimensional D-vine (linear in tree 1), the parameters are $\big\{\{\alpha_{12},\alpha_{23},\ldots,\alpha_{(d-1),d}\},\{\alpha_{13;2},\alpha_{24;3},\ldots,\alpha_{(d-2),d;(d-1)}\},\ldots,\{\alpha_{1d;23\cdots(d-1)}\}\big\}$. By construction of vine copulas, the α's are within $(-1,1)$ and are algebraically independent.

The Hüsler-Reiss copula has been generalized to the t-EV copula (Nikoloulopoulos et al (2009); the bivariate version is attributed to Demarta and McNeil (2005)) and their skewed counterparts (Padoan (2011)). The t-EV copula has a parameter Ω that is a correlation matrix. A multivariate t copula with factor structure has been defined with Ω having a factor structure. It is possible to impose a factor structure directly on Ω of a t-EV copula. However, the bivariate strength of dependence corresponding to a correlation parameter of zero depends on the degree of freedom ν specified (Davison et al (2012); Ribatet (2013)); it approaches the independence limit as $\nu\to\infty$. This may complicate interpretation of the model. Nevertheless, it may be useful when the overall dependence among the variables is rather strong, so that the aforementioned boundary is not a major concern. The stock returns data example in Section 7 contains an illustration of the 1-factor t-EV model.

4.2 Dependence properties for bivariate margins

The tail dependence function of the Hüsler-Reiss copula is

$$b_{ij}(w_i,w_j) = w_i\Phi\left(\frac{\delta_{ij}}{2}\log\frac{w_j}{w_i}-\frac{1}{\delta_{ij}}\right)+w_j\Phi\left(\frac{\delta_{ij}}{2}\log\frac{w_i}{w_j}-\frac{1}{\delta_{ij}}\right).$$

The tail dependence index is given by $b_{ij}(1,1) = 2\Phi(-1/\delta_{ij})$, and the extremal coefficient is $2-b_{ij}(1,1) = 2\Phi(1/\delta_{ij})$. Increasing α and/or γ will lead to an increase in some or all of the ρ's, and consequently the parameters δ's for the Hüsler-Reiss copula, leading to stronger dependence. Therefore, the structured Hüsler-Reiss model is increasing in concordance ordering with respect to both α and γ.

For the d-dimensional copula, independence is obtained when all δ parameters are zero. With our parametrization, this will only happen when $\gamma = 0$. Care must be exercised when interpreting the fitted correlation parameters. Meanwhile, the comonotonicity copula is the limit as $\delta_{ij}\to\infty$ for every (i,j), or equivalently when $\rho_{ij}=1$ for every pair. For the factor model, this happens when all the parameters for the first factor are 1; for the vine model, all α parameters have to be 1. Regardless of the dependence structure, we obtain the same limit by letting $\gamma\to\infty$.

4.3 Some comparisons between the EV factor copula and the structured Hüsler-Reiss model

Both classes of EV copulas are potentially useful parsimonious dependence models for multivariate extremes. In this section, we contrast some differences between them.

- The EV factor copula is suitable for data exhibiting a latent or observed factor structure, whereas the structured Hüsler-Reiss model is applicable to any parsimonious dependence structure as long as it can be parametrized and represented in terms of the correlation matrix of a Gaussian distribution.
- The parameters between different linking copulas in the EV factor copula model are algebraically independent. However, for the Hüsler-Reiss model, the collection of parameters have to be such that all d constituent $(d-1)$-dimensional correlation matrices Γ_j in (11) are positive definite. Also, some parameters may approach the boundary, i.e., 1 or -1, during model fitting of the structured Hüsler-Reiss model, especially when the sample size is small. If that occurs, one can fit again with such parameters fixed at the boundary.
- The Hüsler-Reiss copula with factor structure does not coincide with the class of EV factor copulas, but with careful choice of parameters, the pairwise extremal coefficients can be well approximated. In general, the approximation becomes better when the ρ_{ij}'s are close to 1; this is not surprising given the limit argument in deriving the Hüsler-Reiss copula.
- Model estimation, via pairwise likelihood as in Section 5, is faster and generally more stable for the structured Hüsler-Reiss model as it only involves Gaussian densities and cdfs. For EV factor copulas, one must evaluate the stable tail dependence function and its derivatives through numerical integration of the conditional tail dependence functions. This procedure is not trivial and will be addressed in the next section.

We give an example to compare these models through their pairwise dependence characteristics. Table 2 displays the Spearman's ρ for each bivariate margin of different parsimonious models with 5 observed variables. The top left panel corresponds to the 1-factor copula with MTCJ linkages and dependence parameters $(\delta_1, \dots, \delta_5) = (1, 2, 3, 4, 5)$ (dependence is stronger as the index increases); the factor structure is apparent in the sense that the strength of dependence is similarly ordered across columns and rows. This property is carried over to its EV limit shown on the top right panel; this EV 1-factor copula can be closely approximated by the Hüsler-Reiss 1-factor structure with parameters $(\tanh^{-1} \alpha_{11}, \dots, \tanh^{-1} \alpha_{15}, \gamma) = (4.25, 4.83, 5.19, 5.45, 5.67, 0.0297)$.

The bottom panel of Table 2 suggests that the dependence pattern of the structured Hüsler-Reiss model follows that of the tree imposed on the underlying correlation matrix. This can be observed in the bottom left panel, where a linear tree (1-truncated D-vine) with correlation parameters α all equal to 0.5 and $\gamma = 1$ is imposed. The adjacent variables are the most correlated with dependence strength tapering off between variables that are further apart. The bottom middle panel results from the same vine but with all $\alpha = 0.875$ and $\gamma = 0.5$ instead, so that dependence between adjacent variables remains the same. We can see that a smaller γ results in a more flexible coverage of the strength of dependence. Finally, the bottom right panel represents the dependence structure of a tree with variables 1 to 4 linearly connected, but variable 5 is connected to 3 instead. The parameters are all $\alpha = 0.875$ and $\gamma = 0.5$. Since variables 4 and 5 are symmetric, they have the same bivariate dependence with the other three variables.

Because relative pairwise dependence strengths are preserved, one may thus apply the assumed structure directly to the proposed models. The magnitudes of the parameters, however, should not be interpreted directly. For example, even if $\alpha < 0$ implying possible negative values in the underlying correlation matrix, the actual correlation of the observed EV variables is still positive, albeit with a smaller magnitude.

1-factor MTCJ

$$\begin{bmatrix} 1 & 0.36 & 0.40 & 0.43 & 0.44 \\ & 1 & 0.57 & 0.60 & 0.62 \\ & & 1 & 0.69 & 0.71 \\ & & & 1 & 0.77 \\ & & & & 1 \end{bmatrix}$$

EV 1-factor Dagum

$$\begin{bmatrix} 1 & 0.51 & 0.55 & 0.56 & 0.57 \\ & 1 & 0.75 & 0.78 & 0.79 \\ & & 1 & 0.85 & 0.87 \\ & & & 1 & 0.90 \\ & & & & 1 \end{bmatrix}$$

HR (D-vine, $\gamma = 1$)

$$\begin{bmatrix} 1 & 0.57 & 0.46 & 0.41 & 0.39 \\ & 1 & 0.57 & 0.46 & 0.41 \\ & & 1 & 0.57 & 0.46 \\ & & & 1 & 0.57 \\ & & & & 1 \end{bmatrix}$$

HR (D-vine, $\gamma = 0.5$)

$$\begin{bmatrix} 1 & 0.57 & 0.39 & 0.29 & 0.23 \\ & 1 & 0.57 & 0.39 & 0.29 \\ & & 1 & 0.57 & 0.39 \\ & & & 1 & 0.57 \\ & & & & 1 \end{bmatrix}$$

HR (non-linear)

$$\begin{bmatrix} 1 & 0.57 & 0.39 & 0.29 & 0.29 \\ & 1 & 0.57 & 0.39 & 0.39 \\ & & 1 & 0.57 & 0.57 \\ & & & 1 & 0.39 \\ & & & & 1 \end{bmatrix}$$

Table 2 Matrices of Spearman's ρ for the different models mentioned in the text

5 Statistical inference via composite likelihood methods

In this section, we outline the composite likelihood method based on bivariate margins for model estimation. Maximum likelihood modelling requires the full density of the distribution, and is numerically infeasible or difficult for $d \geq 5$ for EV copulas of the form $C(u_1, \dots, u_d) = \exp\{-a(w_1, \dots, w_d)\}$, with $w_i = -\log u_i$. Composite likelihood methods (Lindsay (1988); Cox and Reid (2004); Varin et al (2011)) provide an attractive alternative as likelihood contributions involve only marginal or conditional densities of lower dimensions. Here we focus on pairwise likelihoods, which have been used in the spatial extremes context (see, e.g., Padoan et al (2010); Davison et al (2012)). Let c_{jk} be the bivariate marginal copula density for the (j, k) margin. The pairwise composite log-likelihood is $\ell_P(\theta; \mathbf{u}) = \sum_{r=1}^{n} \sum_{j<k} \log c_{j,k}(u_{rj}, u_{rk}; \theta)$, where n is the number of observations in a random sample, θ is the collection of all (dependence) parameters and $\mathbf{u} = (u_{rs})$ for $1 \leq r \leq n$ and $1 \leq s \leq d$ is the collection of all data, assumed to have U(0,1) margins. Univariate margins are first fitted by the generalized extreme value distribution, and the data are transformed to U(0,1) via probability integral transforms.

Under the null model with true parameter value θ_0 and given the usual regularity conditions, the maximizer $\hat{\theta}_P$ of $\ell_P(\theta; \mathbf{u})$ is consistent and has asymptotic distribution $\sqrt{n}\left(\hat{\theta}_P - \theta_0\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{G}^{-1}(\theta_0)\right)$, where $\mathbf{G}(\theta) = \mathbf{H}(\theta)\mathbf{J}^{-1}(\theta)\mathbf{H}(\theta)$ is the Godambe or sandwich information matrix (Godambe (1960)), with sensitivity matrix $\mathbf{H}(\theta) = \mathbb{E}\left[-\nabla^2 \ell_P(\theta; \mathbf{U}_1)\right]$ and variability matrix $\mathbf{J}(\theta) = \mathrm{Var}\left[\nabla \ell_P(\theta; \mathbf{U}_1)\right]$. Here $\mathbf{U}_1 = (U_{11}, \dots, U_{1d})^\intercal$ is the first observation as a random vector. These matrices can be estimated using sample averages; in most cases $\widehat{\mathbf{H}}(\hat{\theta})$ is conveniently obtained as the output of the optimization algorithm.

The bivariate marginal EV copula densities are given by $c_{12}(u_1, u_2) = e^{-a}(a^{(1)}a^{(2)} - a^{(12)})/u_1u_2$, where $a^{(1)}$, $a^{(2)}$ and $a^{(12)}$ are the partial derivatives of $a(w_1, w_2)$ with respect to the argument(s) indicated. The pairwise densities for the Hüsler-Reiss copula involve only univariate Gaussian densities and cdfs and are thus easy to compute. Meanwhile, the partial derivatives for the 1- and 2-factor EV copulas with respect to the set $\varnothing \neq S \subset I_2$ are:

$$a^{(S)}_{\text{1-fact}} = \int_0^\infty (-1)^{|S|-1} \prod_{j\in S} b'_{j|V}(w_j|z) \prod_{i\notin S} \left[1 - b_{i|V}(w_i|z)\right] \mathrm{d}z;$$

$$a^{(S)}_{\text{2-fact}} = \int_0^1 \int_0^\infty (-1)^{|S|-1} \left\{ \prod_{j\in S} \left[c_{j,V_2;V_1}\left(b_{j|V_1}(w_j|z_1), v_2\right) \cdot b'_{j|V_1}(w_j|z_1) \right] \prod_{i\notin S} \left[1 - C_{i|V_2;V_1}\left(b_{i|V_1}(w_i|z_1)|v_2\right)\right] \right\} \mathrm{d}z_1 \mathrm{d}v_2,$$

where $b'_{j|V}(w_j|z) = \partial b_{j|V}(w_j|z)/\partial w_j$. In most cases, a and $a^{(S)}$ have to be evaluated numerically. The pairwise likelihood method has the advantage that rounding errors are minimized. Although these functions can be evaluated by adaptive integration, the results are not precise enough for the computation of standard errors using the sandwich information matrix. This issue arises because the integrand often has a somewhat heavy tail, especially for a, and evaluation based on adaptive integration may not be sufficiently smooth within a neighbourhood of the estimated parameter value. This problem is exacerbated by the algebraic operations on these functions needed to obtain the bivariate marginal densities. We propose strategies to stabilize the numerical evaluations of a and $a^{(S)}$ through a two-step procedure that involves Gaussian quadrature and transformations of the integrands; the technical details are outlined in Appendix B.

6 Simulation study

We conduct a simulation study on the 1-factor Dagum EV factor model, based on the EV limit of a 1-factor MTCJ copula with parameters $\delta = (\delta_1, \ldots, \delta_d)$, with the following intentions: (i) to verify that the numerical evaluation of the likelihood of the model is valid, (ii) to compare the statistical efficiency between full and composite (pairwise) likelihood estimation in dimension $d = 4$, and (iii) to show that the method of pairwise likelihood yields accurate model-based standard errors. Note that maximum likelihood is impractical with $d \geq 5$. We consider 3 sets of parameters for δ with $d = 4$ and let $\zeta_i = \delta_i/(\delta_i + 2)$ be a transform of δ_i to the interval (0, 1), corresponding to the Kendall's τ value for a bivariate MTCJ copula with parameter δ_i.

- Weak dependence: $\delta = (1, 1, 1, 1)$ or $\zeta = (\zeta_1, \zeta_2, \zeta_3, \zeta_4) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. This corresponds to a Kendall's τ of 0.268 between variables of the EV factor copula, and an extremal coefficient of 1.667.
- Strong dependence: $\delta = (4, 4, 4, 4)$ or $\zeta = (\frac{2}{3}, \frac{2}{3}, \frac{2}{3}, \frac{2}{3})$. This corresponds to a Kendall's τ of 0.707 between variables of the EV factor copula, and an extremal coefficient of 1.227.
- Mixed dependence: $\delta = (1, 2, 3, 4)$ or $\zeta = (\frac{1}{3}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3})$. The pairwise Kendall's τ and extremal coefficients of the EV factor copula are in the range [0.358, 0.667] and [1.263, 1.565], respectively.

To explore the finite-sample performance, for each set of parameters we consider sample sizes $n = 100$ and 500 and number of replications $R = 1000$, where each simulated sample is fitted by both full and pairwise likelihood. In each case we compute the average value of the Kendall's τ corresponding to the estimated parameters, and the associated standard error using the relevant information matrix. The sampling standard error is also obtained for comparison.

Table 3 summarizes the results of the simulation; a histogram of the sampling distribution (not shown) suggests some skewness when $n = 100$, but is close to symmetric when $n = 500$. Estimates are reported in terms of the Kendall's τ corresponding to the parameters to reduce the right-skewness of the sampling distributions. Both the full and pairwise likelihood estimators have small bias even with $n = 100$. The pairwise likelihood estimators are in general more variable than their full likelihood counterparts, and such amplification is especially pronounced when the dependence is strong; this is likely due to a bigger loss of information by using pairwise likelihood in this case. Both estimators become less variable as the sample size increases. The departure between the model-based and sampling standard errors is quite small for both estimation methods, and is negligible when $n = 500$.

In summary, estimation using pairwise likelihood results in acceptable performance, and the model-based standard errors are an accurate representation of the sampling variability when the sample size is reasonably large. With small sample sizes, however, we note that a more stable estimate of the variance under composite likelihood may be obtained through resampling methods such as the jackknife (see, e.g., Zhao and Joe (2005)), but this will incur a significant computational cost since it requires refitting the model multiple times.

Finally, simulation based on the structured Hüsler-Reiss model is not attempted as the composite likelihood method is known to be feasible. In the context of spatial extremes with the Smith and Brown-Resnick parsimonious representation of the Hüsler-Reiss model, pairwise and triplewise likelihood estimators have been compared in Genton et al (2011) and Huser and

n	Model estimates and standard errors							
	Full likelihood				Pairwise likelihood			
	$\hat{\zeta}_1$	$\hat{\zeta}_2$	$\hat{\zeta}_3$	$\hat{\zeta}_4$	$\hat{\zeta}_1$	$\hat{\zeta}_2$	$\hat{\zeta}_3$	$\hat{\zeta}_4$
Weak dependence: $\delta = (1, 1, 1, 1)$ and $\zeta = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$								
100	.338	.342	.339	.337	.338	.343	.340	.337
	(.048)	(.048)	(.048)	(.047)	(.048)	(.049)	(.049)	(.048)
500	.333	.335	.334	.333	.334	.335	.334	.334
	(.020)	(.021)	(.021)	(.020)	(.021)	(.021)	(.021)	(.021)
Strong dependence: $\delta = (4, 4, 4, 4)$ and $\zeta = (\frac{2}{3}, \frac{2}{3}, \frac{2}{3}, \frac{2}{3})$								
100	.664	.667	.666	.665	.669	.673	.672	.669
	(.025)	(.025)	(.025)	(.025)	(.035)	(.035)	(.035)	(.035)
500	.665	.665	.665	.665	.667	.668	.667	.667
	(.011)	(.011)	(.011)	(.011)	(.014)	(.014)	(.014)	(.014)
Mixed dependence: $\delta = (1, 2, 3, 4)$ and $\zeta = (\frac{1}{3}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3})$								
100	.336	.505	.607	.657	.334	.503	.609	.672
	(.032)	(.033)	(.038)	(.042)	(.031)	(.032)	(.042)	(.053)
500	.335	.502	.602	.656	.333	.500	.600	.669
	(.014)	(.015)	(.017)	(.019)	(.014)	(.015)	(.019)	(.028)

n	Sampling standard errors											
	Full likelihood				Pairwise likelihood				Efficiency (%)			
	$\hat{\zeta}_1$	$\hat{\zeta}_2$	$\hat{\zeta}_3$	$\hat{\zeta}_4$	$\hat{\zeta}_1$	$\hat{\zeta}_2$	$\hat{\zeta}_3$	$\hat{\zeta}_4$	$\hat{\zeta}_1$	$\hat{\zeta}_2$	$\hat{\zeta}_3$	$\hat{\zeta}_4$
Weak dependence: $\delta = (1, 1, 1, 1)$ and $\zeta = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$												
100	.054	.048	.051	.049	.055	.051	.054	.049	96	90	89	100
500	.020	.021	.020	.020	.020	.021	.021	.021	96	97	94	98
Strong dependence: $\delta = (4, 4, 4, 4)$ and $\zeta = (\frac{2}{3}, \frac{2}{3}, \frac{2}{3}, \frac{2}{3})$												
100	.022	.022	.022	.021	.037	.038	.037	.036	38	33	35	35
500	.010	.010	.010	.010	.015	.015	.015	.014	46	48	47	49
Mixed dependence: $\delta = (1, 2, 3, 4)$ and $\zeta = (\frac{1}{3}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3})$												
100	.033	.031	.039	.041	.033	.031	.052	.061	100	100	57	46
500	.015	.015	.017	.018	.014	.015	.019	.028	100	96	77	41

Table 3 Mean estimated transformed ζ of the parameter values and standard errors for the 1-factor extreme value copula simulation. Mean standard errors are shown in brackets. Efficiency is calculated as the ratio of sampling variances between the estimations using full and pairwise likelihood, and is capped at 100%. The transform $\zeta = \delta/(\delta + 2)$ is used because the sampling distribution of $\hat{\delta}_i$ is strongly skewed for the sample size $n = 100$.

Davison (2013). The performance of pairwise likelihood estimators is satisfactory except for cases of the Smith model where spatial locations and the parametrization of the Hüsler-Reiss model leads to matrices that are singular or nearly singular.

7 Data examples

The proposed models in Sections 3 and 4 are applied to two data sets where there is a plausible latent factor: (1) The Fraser River flows data where a common source of streamflow and spatial dependence among gauge stations is likely, and (2) The United States stock returns data for selected listed companies in the same sector. For example (1), the tree dependence structure arising from the spatial arrangement of the gauging stations is also plausible.

7.1 Fraser River flows data

The Fraser River is the largest river in British Columbia (Canadian Heritage Rivers System (2016)). It originates from the Rocky Mountains near the British Columbia–Alberta border. From there, it continues downstream through Prince George and turns south, eventually reaching the Lower Mainland and discharging into the Strait of Georgia south of Richmond. The rate or volume of the flow along the river is highly dependent on the climate conditions in central and southeastern BC. Snow accumulates during the winter and melts in the spring and summer, causing an increase in river flows. This is compounded by possible rainstorms in the late spring and early summer along the Fraser River; it is the time of the year when most extreme streamflows are recorded.

To study the dependence characteristics of river flows along the Fraser River, we have downloaded such data at 8 hydrometric gauging stations (with locations indicated in the left panel of Figure 3) from WaterOffice, Environment Canada (URL: https://wateroffice.ec.gc.ca/). The observations are annual maxima of daily streamflows until 2013; most stations commenced operations in the 1950s, with the only exceptions at Hope (6) whose records date back to 1912, and Mission (8) with earliest observation in 1965 (see Table 4 for the data availability at the 8 stations)[1]. The pairwise likelihood approach thus allows us to utilize information between 1950s and 1964 when data for most but one or two stations are available. We also check that the annual maxima have insignificant autocorrelations and it is not necessary to detrend the data.

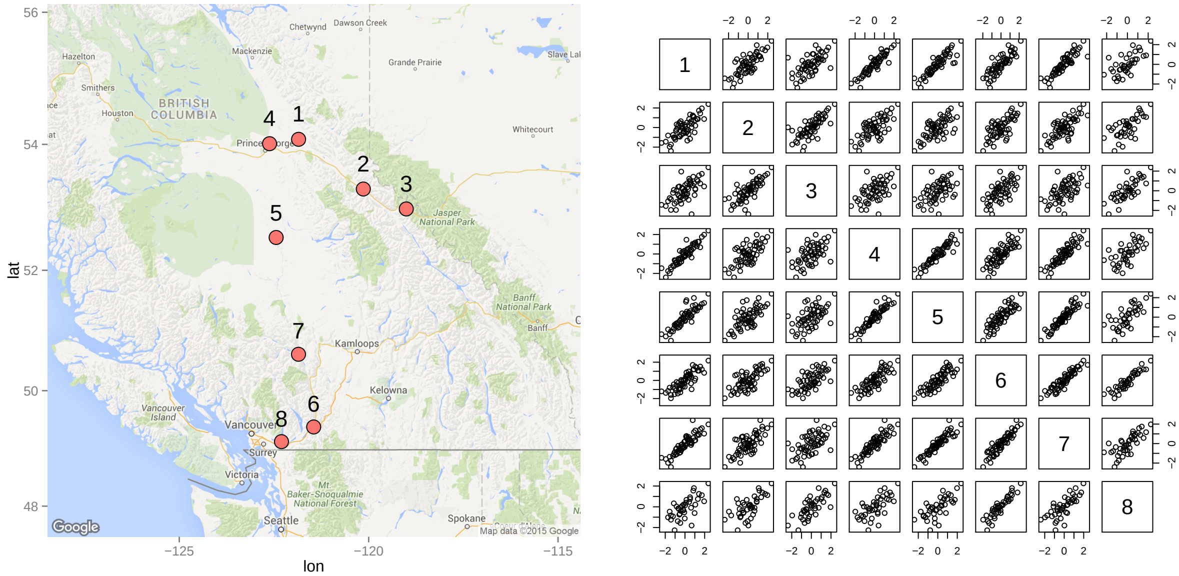

Fig. 3 Locations of the gauging stations along the Fraser River (left) and the pairwise scatterplot of the normal scores (right). The labels are: 1 – Hansard; 2 – McBride; 3 – Red Pass; 4 – Shelley; 5 – Marguerite; 6 – Hope; 7 – Texas Creek; 8 – Mission. The Google map is plotted using the `ggmap` package in `R`.

Location	Years with data	Sample size
1: Hansard	1953–2010	58
2: McBride	1953–2012	60
3: Red Pass	1955–2013	59
4: Shelley	1950–2012	63
5: Marguerite	1950–2011	62
6: Hope	1912–2012	101
7: Texas Creek	1952–2013	62
8: Mission	1965–1992; 2000–2013	42

Table 4 List of the availability of annual maximum data at the 8 gauging stations.

The right panel of Figure 3 shows a pairwise scatterplot of the data (transformed to standard normal margins); dependence among stations is obvious. Both factor and vine are plausible structural assumptions for the data. For a factor interpretation, the latent factor could be rainfall in the interior of BC — higher rainfall contributes to increased river flows at all stations. Meanwhile, there is a natural ordering of the stations along the river, and a vine structure is possible as those close to each other are likely to exhibit stronger dependence than those far apart.

We conduct an exploratory analysis using the correlation matrix of the normal scores. Results based on the discrepancy between the sample correlation matrix and fitted ones with factor and vine structures suggest similar performance between the two classes of models. We then fit EV factor copula models and the structured Hüsler-Reiss models. Each margin is first fitted by the generalized extreme value (GEV) distribution and the observations are converted to U(0,1) through the probability integral transform using the estimated parameters. Diagnostic checks (not included here) show that the GEV distribution fits well to the 8 univariate series. We consider the Dagum and Burr 1-factor EV factor copulas; for the structured Hüsler-Reiss model, we fit the 1- and 2-factor structures as well as 1- and 2-truncated linear D-vines based on the ordering of the stations along the river. The fitting results are given in Table 6, with standard errors estimated from the sandwich information matrix conditional on the univariate estimation stage. The composite likelihood information criterion (CLIC, Varin and Vidoni (2005)) and Bayesian information criterion (CLBIC, Gao and Song (2010)) are obtained; these are the composite likelihood analogues of the AIC and BIC. We observe that the two classes of models perform similarly in terms of the CLBIC. The best models

[1] At the time of our analysis, the data for the recent years had not been made available.

among those considered are the Burr 1-factor EV copula and the Hüsler-Reiss model with 1-truncated vine structure. It is therefore useful to analyze the results for both classes of models.

Both 1-factor EV factor copula models suggest that sites 4, 5 and 7 have strong dependence with the latent factor. This is in agreement with the strong pairwise correlations among these sites. They are in the midstream of the Fraser River and this suggests that the river flows there are likely to be indicative of the overall flow. Meanwhile, sites 2, 3 and 8 have the weakest dependence with the latent factor. These stations are near the ends of the river and the flows there are more likely to be influenced by other factors.

For the Hüsler-Reiss models, there are cases where the α parameters reach the boundary (1 or -1). In this situation we optimize the pairwise likelihood with those parameter(s) fixed at the boundary. We convert the parameters of the Hüsler-Reiss factor models back to the loading matrix for easier interpretation (Table 5). Both factor models point to a contrast between the midstream stations $(1, 4, 5, 7)$ and the source station 3. The 2-factor model suggests downstream stations 6 and 8 are related to each other. Based on the CLIC/CLBIC values, the 1-factor Hüsler-Reiss model may not be suitable. For vine structures, their tree diagrams are plotted in Figure 4. The correlations in the first tree are moderate to strong, with somewhat weaker dependence for the pairs (3,2) and (2,1). Note that these correlations are not directly comparable to those of the normal scores of the original data, due to the limit argument in the derivation of the Hüsler-Reiss copula and the role of the ρ_{ij}'s as the correlations of the underlying Gaussian variates instead. This can be illustrated through the fitted coefficients of the second tree, where multiple strong negative partial correlations are obtained. The partial correlations of the observed normal scores are $(r_{13;2}, r_{24;1}, r_{15;4}, r_{47;5}, r_{56;7}, r_{78;6}) = (0.181, -0.110, 0.158, 0.204, -0.031, -0.603)$. The ranks are generally preserved with the exception of $r_{13;2}$, for which the fitted coefficient is strongly negative.

1-factor		2-factor		
Station	Factor 1	Station	Factor 1	Factor 2
1	0.690	1	0.060	**0.720**
2	−0.225	2	0.140	−0.195
3	**−1.000**	3	−0.009	**−1.000**
4	**0.831**	4	0.272	**0.962**
5	**0.870**	5	−0.081	**0.832**
6	0.648	6	**−0.824**	0.549
7	**0.870**	7	−0.406	**0.841**
8	0.192	8	**−1.000**	0.000

Table 5 Loadings of the Hüsler-Reiss 1- and 2-factor models. Coefficients larger than 0.7 in absolute value are shown in boldface.

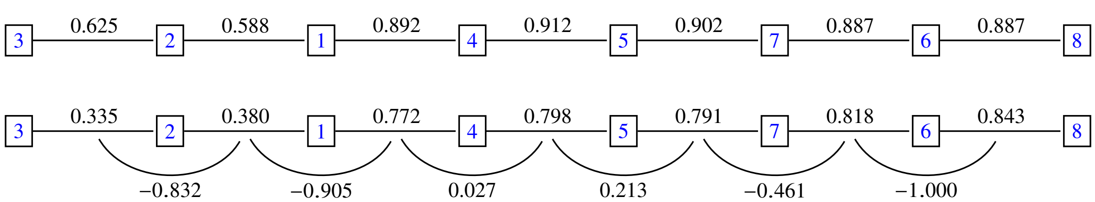

Fig. 4 Vine diagram for the fitted Hüsler-Reiss 1- (top) and 2-truncated (bottom) vine models using a D-vine following the relative positions of the stations

7.2 United States stock returns data

The second example we consider is on the returns of selected stocks traded on the US stock exchanges. Extreme value theory is widely used in the financial field, one reason being that it provides the theoretical foundation for the modelling of value-at-risk (VaR) on stock returns. The VaR is an extreme quantile of the distribution for the returns; usually the quantile corresponding to a loss is of interest. Data for the extreme tails are often sparse, making empirical quantiles unreliable. Tsay (2010) illustrates the use of univariate extreme value theory on monthly minima of daily log returns. The log return for day t is defined as $r_t = \log(P_t/P_{t-1})$ where P_t is the closing stock price on day t. When P_t/P_{t-1} is close to 1, r_t is close to the percentage change $(P_t - P_{t-1})/P_{t-1}$.

To highlight the use of the EV factor copula model, we selected 7 major stocks in the pharmaceutical sector. They are: (1) GlaxoSmithKline PLC (GSK); (2) Johnson & Johnson (JNJ); (3) Eli Lilly and Co (LLY); (4) Merck & Co., Inc. (MRK); (5) Mylan NV (MYL); (6) Novartis AG (NVS), and; (7) Pfizer Inc. (PFE). The observations are bimonthly minima of daily

Model		Estimates (SE): θ_i, δ_i or α_i								γ	Pen.	nplk	CLIC	CLBIC
EV1f (Burr)	F1	3.928	2.147	1.854	4.618	5.237	3.613	5.484	2.492		25.24	−790	−1530	−1480
		(0.489)	(0.222)	(0.200)	(0.995)	(2.076)	(0.468)	(1.690)	(0.280)					
EV1f (Dagum)	F1	3.206	1.422	1.125	3.884	4.440	2.895	4.820	1.770		27.14	−788	−1521	−1467
		(0.513)	(0.219)	(0.201)	(1.088)	(2.119)	(0.502)	(1.894)	(0.284)					
HR1f	F1	0.690	−0.225	−1.000	0.831	0.870	0.648	0.870	0.192	1.855	33.05	−785	−1503	−1437
		(0.105)	(0.179)	—	(0.108)	(0.119)	(0.177)	(0.139)	(0.290)	(0.180)				
HR2f	F1	0.060	0.140	−0.009	0.272	−0.081	−0.824	−0.406	−1.000		35.21	−808	−1546	−1476
		(0.258)	(0.287)	(0.303)	(0.244)	(0.182)	(0.058)	(0.161)	—					
	F2	0.721	−0.197	−1.000	1.000	0.835	0.968	0.920	0.000	1.838				
		(0.100)	(0.196)	—	—	(0.086)	(0.137)	(0.074)	—	(0.192)				
HR1t	T1		0.625	0.588	0.892	0.912	0.902	0.887	0.887	1.212	29.44	−798	−1538	−1479
			(0.129)	(0.229)	(0.057)	(0.056)	(0.056)	(0.057)	(0.067)	(0.282)				
HR2t	T1		0.335	0.380	0.772	0.798	0.791	0.818	0.843		35.68	−809	−1547	−1476
			(0.149)	(0.160)	(0.082)	(0.086)	(0.064)	(0.045)	(0.046)					
	T2			−0.832	−0.905	0.027	0.213	−0.461	−1.000	1.743				
				(0.541)	(0.352)	(0.219)	(0.263)	(0.331)	—	(0.149)				

Table 6 Fitting results for the Fraser River flows data using multivariate extreme value distributions. The models are EV: Extreme value factor copula and HR: Structured Hüsler-Reiss, where 1f and 2f indicate the number of factors, and 1t and 2t the level of truncation for vines. The SEs are based on the pairwise log-likelihood with the univariate parameters held fixed from the first stage GEV estimation. The last 4 columns are respectively: (1) Trace of the matrix $\mathbf{H}^{-1}\mathbf{J}$ which can be considered as the penalty on the pairwise likelihood; (2) Negative pairwise log-likelihood; (3) Composite likelihood information criterion; (4) Composite likelihood Bayesian information criterion.

log returns between January 1997 and October 2016, for a total of 119 observations. We choose bimonthly minima as the resulting series show weaker autocorrelations than monthly minima. Returns prior to 1997 are not used to reduce the effect of potential nonstationarity. For example, the advent of modern computing technology in the 1980s has seen a rapid growth in trading volume and frequency (and thus volatility), and is thought to be a potential factor underlying the 1987 stock market crash (Carlson (2007)). A joint extreme treatment of stock returns may be of use when the weights of different assets are to be selected to achieve an objective, such as maximizing expected returns, subject to restrictions on the VaR of the portfolio (Smith (2002)).

We negate the minimum returns in the subsequent analysis. They generally exhibit upper tail dependence, i.e., extreme losses tend to occur together. The pairwise sample correlations of the normal scores are mostly between 0.4 and 0.6. As a preliminary analysis, Gaussian factor and vine models are fitted to the correlation matrix of normal scores. Unlike the preceding example, in this case the factor model is better than the vine model in terms of smaller average absolute differences between observed and fitted correlations. This result is consistent with intuition — there is no obvious ordering or tree structure among the stocks, but the returns of stocks in the same sector can be thought of driven by some common factors such as the general economic environment and sector-specific drivers.

The GEV distribution is fitted to the univariate margins, and the transformed data is used to fit the EV factor copula and structured Hüsler-Reiss models. The truncated vines used for the Hüsler-Reiss model are those suggested from the partial correlation vines from the correlation matrix of normal scores. We additionally fit a 1-factor t-EV model with 3 degrees of freedom. The fitting results are reported in Table 7; the 1-factor EV copulas fit better than the Hüsler-Reiss models, but the 1-factor t-EV model appears to be superior. From the fitted parameters of the 1-factor models, the returns of Johnson & Johnson are more strongly correlated to the latent factor. Among the Hüsler-Reiss models, the 1-truncated vine (Figure 5) appears to be the best. Both Novartis and Pfizer connect to three other stocks due to their stronger dependence in extreme returns.

We run separate analyses on stocks in the banking and consumer staples sectors and find that the factor structure fits better than the vine structure in general. These results suggest that the factor structure for daily returns is a plausible modelling assumption among major stocks in the same sector, and hence diversification within one sector is less useful. With the fitted models, one can arrive at a model-based VaR for a portfolio consisting of these stocks. The dependence model also allows us to gain insight on the joint tail behaviour, for all or a subset of the stocks.

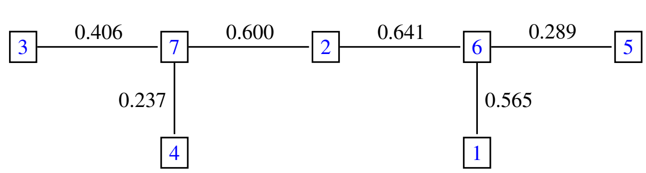

Fig. 5 Diagram for the fitted Hüsler-Reiss 1-truncated vine model using the vine suggested from Gaussian analysis.

Model		Estimates (SE): θ_i, δ_i or α_i							γ	Pen.	nplk	CLIC	CLBIC
EV1f	F1	1.827	2.744	1.673	1.578	1.589	2.179	2.017		22.34	−331	−617	−555
(Burr)		(0.185)	(0.758)	(0.188)	(0.165)	(0.134)	(0.283)	(0.247)					
EV1f	F1	1.093	1.857	0.936	0.830	0.853	1.470	1.277		24.29	−321	−594	−526
(Dagum)		(0.189)	(0.613)	(0.204)	(0.179)	(0.140)	(0.288)	(0.250)					
HR1f	F1	0.720	0.891	0.580	0.500	0.540	0.824	0.773	0.768	24.81	−306	−562	−493
		(0.224)	(0.089)	(0.447)	(0.451)	(0.475)	(0.183)	(0.245)	(0.308)				
HR2f	F1	−0.148	−0.428	−0.293	−1.000	−0.292	−0.394	−0.485		28.09	−306	−555	−477
		(0.288)	(0.261)	(0.161)	—	(0.230)	(0.259)	(0.164)					
	F2	0.768	0.856	0.508	0.000	0.459	0.785	0.692	0.775				
		(0.179)	(0.046)	(0.154)	—	(0.298)	(0.108)	(0.064)	(0.151)				
tEV1f	F1	0.790	0.915	0.749	0.720	0.692	0.859	0.841		17.31	−345	−656	−608
($\nu = 3$)		(0.058)	(0.041)	(0.069)	(0.087)	(0.077)	(0.041)	(0.047)					
HR1t	T1		0.641	0.600	0.565	0.406	0.237	0.289	0.918	21.26	−307	−572	−513
(Gauss. vine)			(0.122)	(0.130)	(0.128)	(0.164)	(0.265)	(0.283)	(0.117)				
HR2t	T1		0.593	0.452	0.742	0.382	0.680	0.477		29.79	−309	−558	−475
(Gauss. vine)			(0.269)	(0.347)	(0.172)	(0.363)	(0.181)	(0.290)					
	T2			0.470	0.534	0.460	0.309	0.261	0.754				
				(0.153)	(0.120)	(0.242)	(0.158)	(0.213)	(0.222)				

Table 7 Fitting results for the US stocks minimum returns data using multivariate extreme value distributions. The models are EV: Extreme value factor copula, HR: Structured Hüsler-Reiss, and tEV: Structured t-EV, where 1f and 2f indicate the number of factors, and 1t and 2t the level of truncation for vines. The SEs are based on the pairwise log-likelihood with the univariate parameters held fixed from the first stage GEV estimation. The last 4 columns are respectively: (1) Trace of the matrix $\mathbf{H}^{-1}\mathbf{J}$ which can be considered as the penalty on the pairwise likelihood; (2) Negative pairwise log-likelihood; (3) Composite likelihood information criterion; (4) Composite likelihood Bayesian information criterion.

8 Conclusion

In this paper, we consider parsimonious multivariate EV models for factor and tree dependence structures. The constructions are based on EV limits of factor copula models with tail dependence, or the Hüsler-Reiss copula where the $\binom{d}{2}$ dependence parameters are expressed in terms of $O(d)$ other parameters. The emphasis is on the interpretation of these structured models. Similar constructions could be considered for other parsimonious dependence structures. If a multivariate copula family is represented by a p-dimensional integral, then typically this is the case for its EV limit.

Due to the intractability of the full copula likelihood for dimensions $d \geq 5$, we use composite (pairwise) likelihood for parameter estimation. While the main considerations are computational accuracy and feasibility, using pairwise likelihood allows us to utilize observations with missing data that are frequently encountered in environmental data. For the EV factor copula model, we illustrate through a simulation study that pairwise likelihood results in reliable estimates of parameter values, albeit at a cost of reduced efficiency. Two data examples based on environmental and financial EV data are presented to demonstrate the applicability of and interpretations offered by the proposed models.

For EV factor copula models, we demonstrate that suitable quadrature methods and integral transformations are necessary for accurate evaluation of the bivariate density. Often, these techniques are specific to the conditional tail dependence functions involved. Potential future research includes exploring the possibility of more tractable EV factor copula models. Another topic is on the gain in efficiency that may be achieved by suitably modifying the weights of the pairwise likelihood contributions from different margins, based on data-driven measures such as the strength of dependence of the variables involved.

Acknowledgements

This research has been supported by UBC's Four Year Doctoral Fellowship and NSERC Discovery Grant 8698. We would like to thank the associate editor and the anonymous referees for their helpful comments, which improved the presentation of the paper.

A Tail dependence function of the 1-factor copula with MTCJ linking copulas

In this appendix, we demonstrate that the limit and integral operators in (6) can be interchanged when the linking copulas are MTCJ, where $C_{iV}(u_i, v) = (u_i^{-\delta_i} + v^{-\delta_i} - 1)^{-1/\delta_i}$ with dependence parameters $\delta_i > 0$ for $i = 1, \ldots, d$.

The conditional distribution of U_i given V is given by

$$C_{i|V}(u_i|v) = \frac{\partial}{\partial v} C_{iV}(u_i, v) = \left[1 + \left(\frac{u_i}{v}\right)^{-\delta_i} - v^{\delta_i}\right]^{-1/\delta_i - 1}.$$

Let $w_M = w_1 + \epsilon$ for some small, fixed $\epsilon > 0$ and $m = w_1^{-\delta_1} - w_M^{-\delta_1} > 0$. Note that, for all $z > 0$ and as $u \to 0^+$,

$$\left| 1\{z \le u^{-1}\} \prod_{i=1}^{d} C_{i|V}(uw_i|uz) \right| \le C_{1|V}(uw_1|uz) = \left[1 + \left(\frac{w_1}{z}\right)^{-\delta_1} - (uz)^{\delta_1} \right]^{-1/\delta_1-1} \le \left(1 + mz^{\delta_1}\right)^{-1/\delta_1-1}.$$

This upper bound is integrable with respect to z:

$$\int_0^\infty \left(1 + mz^{\delta_1}\right)^{-1/\delta_1-1} \mathrm{d}z = m^{-1/\delta_1} < \infty.$$

Therefore, (6) is valid by Lebesgue's dominated convergence theorem.

B Gaussian quadrature methods for evaluating the density of 1-factor EV copulas

In this appendix, we provide results on integral transformations for the stable tail dependence function a and its derivatives. These techniques improve the stability in evaluating the bivariate densities of the EV factor copula via a two-step procedure: (a) We standardize the arguments of the a function and its derivatives using their homogeneity property, to avoid boundary problems, and (b) transform the integral into an expected value of bounded functions of gamma or beta random variables, on which Gaussian quadrature (Stroud and Secrest (1966)) is applied. This helps improve precision as it avoids dealing with a heavy-tailed integrand directly. With these procedures, we are able to obtain sufficient smoothness for the numerical evaluation of the gradient and Hessian matrices. Here we give the details for the 1-factor parametric models in Section 3.5. The techniques are also applicable to 2-factor and bi-factor parametric models.

The first step is to transform w_1 and w_2, the arguments of a and its derivatives, to $w_1^* = w_1/(w_1+w_2)$ and $w_2^* = w_2/(w_1+w_2)$ so that $w_1^*+w_2^* = 1$. Since a, $a^{(i)}$ $(i = 1, 2)$, $a^{(12)}$ are respectively homogeneous functions of orders 1, 0 and -1, we have $a(w_1, w_2) = (w_1 + w_2)a\left(w_1^*, w_2^*\right)$, $a^{(i)}(w_1, w_2) = a^{(i)}\left(w_1^*, w_2^*\right)$ for $i = 1, 2$, and $a^{(12)}(w_1, w_2) = (w_1 + w_2)^{-1} a^{(12)}\left(w_1^*, w_2^*\right)$. This is the most useful when w_1 and w_2 are close to 0, i.e., when u_1 and u_2 are close to 1, since this will bring them away from the boundary where numerical difficulties are the most likely to happen. We henceforth assume that w_1 and w_2 have already been transformed in this manner.

Due to the difference in asymptotic behaviour of the integrands as $z \to 0$ and $z \to \infty$, we split the integral into two regions: $z \in [0, 1)$ and $z \in [1, \infty)$, to which the Gauss-Laguerre quadrature may be applied separately. For integrand $h(z)$ with the transformation $z = e^{\eta y}$, $\mathrm{d}z = \eta e^{\eta y}\,\mathrm{d}y$, where η is a non-negative constant, we have

$$\int_0^\infty h(z)\mathrm{d}z = \int_0^\infty h(e^{-\eta y}) \cdot \eta e^{-(\eta-1)y} \cdot e^{-y}\mathrm{d}y + \int_0^\infty h(e^{\eta y}) \cdot \eta e^{(\eta+1)y} \cdot e^{-y}\mathrm{d}y, \tag{12}$$

so that each integral becomes an expectation of a standard exponential random variable. Since $h(z)$ is bounded in all cases, the first integral in (12) does not cause instability in general. However, the $e^{(\eta+1)y}$ term in the second integral can grow rapidly and this may lead to numerical overflow. A different transformation is thus applied to the region $z \in [1, \infty)$ in the original integral. We illustrate using the Dagum 1-factor EV copula. Without loss of generality, assume $\delta_1 \le \delta_2$. Using the transformation $s = z^{-\delta_1}$, we have

$$\begin{aligned}
a_+(w_1, w_2) &= \int_1^\infty \left\{ 1 - \left(1 - \left[1 + \left(\frac{w_1}{z}\right)^{-\delta_1} \right]^{-1/\delta_1-1} \right)\left(1 - \left[1 + \left(\frac{w_2}{z}\right)^{-\delta_2} \right]^{-1/\delta_2-1} \right) \right\} \mathrm{d}z \\
&= w_1 \left[1 - \left(w_1^{\delta_1} + 1\right)^{-1/\delta_1} \right] + \frac{1}{1+\delta_1} w_2^{\delta_2+1} \mathbb{E}\left[S^{(\delta_2-\delta_1+1)/\delta_1} \left(w_2^{\delta_2} S^{\delta_2/\delta_1} + 1\right)^{-1/\delta_2-1} \right] \\
&\quad - \frac{1}{1+\delta_1} w_1^{\delta_1+1} w_2^{\delta_2+1} \mathbb{E}\left[S^{\delta_2/\delta_1} \left(w_1^{\delta_1} S + 1\right)^{-1/\delta_1-1} \left(w_2^{\delta_2} S^{\delta_2/\delta_1} + 1\right)^{-1/\delta_2-1} \right],
\end{aligned}$$

where a_+ denotes the integral in $[1, \infty)$, and $S \sim \mathrm{Beta}(1 + \delta_1^{-1}, 1)$. The original integral is thus transformed into one with both finite integrand and range. For $a^{(1)}$ (that of $a^{(2)}$ is similar) and $a^{(12)}$, we obtain

$$\begin{aligned}
a_+^{(1)}(w_1, w_2) &= \int_1^\infty \left\{ \frac{1+\delta_1}{z} \left(\frac{w_1}{z}\right)^{-\delta_1-1} \left[1 + \left(\frac{w_1}{z}\right)^{-\delta_1} \right]^{-1/\delta_1-2} \left(1 - \left[1 + \left(\frac{w_2}{z}\right)^{-\delta_2} \right]^{-1/\delta_2-1} \right) \right\} \mathrm{d}z \\
&= 1 - \left(w_1^{\delta_1} + 1\right)^{-1/\delta_1-1} - w_1^{\delta_1} w_2^{\delta_2+1} \mathbb{E}\left[S^{\delta_2/\delta_1} \left(w_1^{\delta_1} S + 1\right)^{-1/\delta_1-2} \left(1 + w_2^{\delta_2} S^{\delta_2/\delta_1}\right)^{-1/\delta_2-1} \right]; \\
a_+^{(12)}(w_1, w_2) &= -\int_1^\infty \prod_{i=1}^{2} \left\{ \frac{1+\delta_i}{z} \left(\frac{w_i}{z}\right)^{-\delta_i-1} \left[1 + \left(\frac{w_i}{z}\right)^{-\delta_i} \right]^{-1/\delta_i-2} \right\} \mathrm{d}z \\
&= -(1+\delta_2)\, w_1^{\delta_1} w_2^{\delta_2} \mathbb{E}\left[S^{\delta_2/\delta_1} \left(w_1^{\delta_1} S + 1\right)^{-1/\delta_1-2} \left(w_2^{\delta_2} S^{\delta_2/\delta_1} + 1\right)^{-1/\delta_2-2} \right].
\end{aligned}$$

Each expectation can be evaluated using the Gauss-Jacobi quadrature. The number of quadrature points needed depends on the strength of dependence of the EV factor copula (and hence the tail heaviness of the integrand). In general, more quadrature points are needed for weaker dependence. In the simulation study in Section 6, we use 35 quadrature points to ensure sufficient smoothness for a grid size of 10^{-3}, in the scale of the dependence parameter.

References

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

Bedford T, Cooke RM (2001) Probability density decomposition for conditionally dependent random variables modeled by vines. Annals of Mathematics and Artificial Intelligence 32:245–268

Bedford T, Cooke RM (2002) Vines — a new graphical model for dependent random variables. Annals of Statistics 30:1031–1068

Besag J (1974) Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society Series B 36:192–236

Brechmann EC, Czado C, Aas K (2012) Truncated regular vines in high dimensions with application to financial data. Canadian Journal of Statistics 40:68–85, DOI 10.1002/cjs.10141

Burr IW (1942) Cumulative frequency functions. The Annals of Mathematical Statistics 13:215–232

Canadian Heritage Rivers System (2016) *Fraser River*. URL: http://chrs.ca/the-rivers/fraser/. Accessed on August 14, 2016

Carlson M (2007) A brief history of the 1987 stock market crash with a discussion of the Federal Reserve response. Finance and Economics Discussion Series, Federal Reserve Board of Governors

Clayton DG (1978) A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika 65:141–151

Coles S (2001) An Introduction to Statistical Modeling of Extreme Values. Springer, London

Cook RD, Johnson ME (1981) A family of distributions for modelling non-elliptically symmetric multivariate data. Journal of the Royal Statistical Society Series B 43:210–218

Cox DR, Reid N (2004) A note on pseudolikelihood constructed from marginal densities. Biometrika 91:729–737

Dagum C (1975) A model of income distribution and the conditions of existence of moments of finite order. In: Proceedings of the 40th session of the International Statistical Institute, vol 46, pp 199–205

Davison AC, Padoan SA, Ribatet M (2012) Statistical modeling of spatial extremes. Statistical Science 27:161–186

Demarta S, McNeil AJ (2005) The t copula and related copulas. International Statistical Review 73:111–129

Drees H, Huang X (1998) Best attainable rates of convergence for estimators of the stable tail dependence function. Journal of Multivariate Analysis 64:25–46

Galambos J (1975) Order statistics of samples from multivariate distributions. Journal of the American Statistical Association 70:674–680

Galambos J (1987) The Asymptotic Theory of Extreme Order Statistics, 2nd edn. Krieger Publishing Co., Malabar

Gao X, Song PXK (2010) Composite likelihood Bayesian information criteria for model selection in high-dimensional data. Journal of the American Statistical Association 105:1531–1540

Genton MG, Ma Y, Sang H (2011) On the likelihood function of Gaussian max-stable processes. Biometrika 98:481–488

Godambe VP (1960) An optimum property of regular maximum likelihood estimation. The Annals of Mathematical Statistics 31:1208–1211

Gumbel EJ (1960) Distributions des valeurs extrêmes en plusieurs dimensions. Publications de l'Institut de Statistique de l'Université de Paris 9:171–173

Huang X (1992) Statistics of bivariate extremes. PhD thesis, Erasmus University Rotterdam, Tinbergen Institute Research Series 22

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

Joe H (1993) Parametric families of multivariate distributions with given margins. Journal of Multivariate Analysis 46:262–282

Joe H (2014) Dependence Modeling with Copulas. Chapman & Hall, Boca Raton

Joe H, Li H, Nikoloulopoulos AK (2010) Tail dependence functions and vine copulas. Journal of Multivariate Analysis 101:252–270

Krupskii P, Joe H (2013) Factor copula models for multivariate data. Journal of Multivariate Analysis 120:85–101

Lindsay BG (1988) Composite likelihood methods. Contemporary Mathematics 80:221–239

Mardia KV (1962) Multivariate Pareto distributions. The Annals of Mathematical Statistics 33:1008–1015

Nikoloulopoulos AK, Joe H, Li H (2009) Extreme value properties of multivariate t copulas. Extremes 12:129–148

Padoan SA (2011) Multivariate extreme models based on underlying skew-t and skew-normal distributions. Journal of Multivariate Analysis 102:977–991

Padoan SA, Ribatet M, Sisson SA (2010) Likelihood-based inference for max-stable processes. Journal of the American Statistical Association 105:263–277

Pickands J (1981) Multivariate extreme value distributions. In: Proceedings 43rd Session International Statistical Institute, vol 2, pp 859–878

Ribatet M (2013) Spatial extremes: Max-stable processes at work. Journal de la Société Française de Statistique 154:156–177

Schmidt R, Stadtmüller U (2006) Non-parametric estimation of tail dependence. Scandinavian Journal of Statistics 33:307–335

Singh SK, Maddala GS (1976) A function for size distribution of incomes. Econometrica 44:963–970

Sklar A (1959) Fonctions de répartition à n dimensions et leurs marges. Publications de l'Institut de Statistique de l'Université de Paris 8:229–231

Smith RL (1990) Max-stable processes and spatial extremes. *Unpublished manuscript*

Smith RL (2002) Measuring risk with extreme value theory. In: Dempster MAH (ed) Risk Management: Value at Risk and Beyond, Cambridge University Press, pp 224–246

Stroud A, Secrest D (1966) Gaussian Quadrature Formulas. Prentice-Hall, Englewood Cliffs, NJ

Takahasi K (1965) Note on the multivariate Burr's distribution. Annals of the Institute of Statistical Mathematics 17:257–260

Tsay RS (2010) Analysis of Financial Time Series, 3rd edn. Wiley, Hoboken

Varin C, Vidoni P (2005) A note on composite likelihood inference and model selection. Biometrika 92:519–528

Varin C, Reid N, Firth D (2011) An overview of composite likelihood methods. Statistica Sinica 21:5–42

Zhao Y, Joe H (2005) Composite likelihood estimation in multivariate data analysis. Canadian Journal of Statistics 33:335–356