Estimating probabilities of multivariate failure sets based on pairwise tail dependence coefficients

Anna Kiriliouk and Chen Zhou

1 Namur Institute for Complex Systems, Université de Namur, Rempart de la vierge 8, B-5000, Namur, Belgium and 2 Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, Burg. Oudlaan 50, 3062 PA, Rotterdam, the Netherlands

Abstract

Estimating the probability of extreme events involving multiple risk factors is a critical challenge in fields such as finance and climate science. This paper proposes a semi-parametric approach to estimate the probability that a multivariate random vector falls into an extreme failure set, based on the information in the tail pairwise dependence matrix (TPDM) only. The TPDM provides a partial summary of tail dependence for all pairs of components of the random vector. We propose an efficient algorithm to obtain approximate completely positive decompositions of the TPDM, enabling the construction of a max-linear model whose TPDM approximates that of the original random vector. We also provide conditions under which the approximation turns out to be exact. Based on the decompositions, we can construct max-linear random vectors to estimate failure probabilities, exploiting its computational simplicity. The algorithm allows to obtain multiple decompositions efficiently. Finally, we apply our framework to estimate probabilities of extreme events for real-world datasets, including industry portfolio returns and maximal wind speeds, demonstrating its practical utility for risk assessment.

Key words: Failure probability, Max-linear model, Multivariate regular variation, Tail dependence, Tail pairwise dependence matrix.

1. Introduction

Extreme value statistics deal with the characterization of extreme events, such as stock market crashes, that occur with low frequency but with a potentially severe impact. A typical applied question is to estimate the (low) probability of such an extreme event. In the context of multivariate extremes, we are interested in estimating the probability that a $d$-dimensional random vector $X = (X_1, \ldots, X_d)$ falls into some “extreme region” $C$ (also called failure region). For example, in the context of portfolio risk, the components of $X$ represent negative log-returns of $d$ stocks in a portfolio. Another example is in catastrophic climate events, where $X$ may represent maximum daily rainfall amounts at $d$ monitoring stations. To estimate such a probability $\mathbb{P}[X \in C]$, it is important to take the tail dependence across the components of $X$ into account, i.e., the tendency of extremes to occur in several components simultaneously. Note that, even though our method is motivated by and formulated in an extreme-value framework, it could be easily adapted to tackle the estimation of failure probabilities for different regions $C$, as long as $\mathbb{P}[X \in C]$ is small.
In practice, the extreme region of interest might contain few or even no observed data points, especially when the dimension \( d \) is large. One potential solution is to employ appropriate parametric models for multivariate extremes and to estimate the model parameters using intermediate level observations. However, as the dimension of the vector \( \mathbf{X} \) increases, such a parametric approach becomes increasingly complex, which leads to the following dilemma. On the one hand, models with a small number of parameters tend to oversimplify the tail dependence structure. On the other hand, models with many parameters are difficult to estimate, especially without additional information (e.g. geographical location of weather stations).

This paper provides a parametric model for the case where the dimension of the underlying random vector is moderate to high. The model can capture the so-called tail dependence coefficients for all pairs of components of the \( d \)-dimensional random vector. Despite the large number of parameters (up to \( d(d+1)/2 \)), estimation is based on a simple and fast algorithm, allowing one to approximate the failure probability using the estimated model. The parametric model we use is asymptotically justified: parallels can be drawn with the (centered) multivariate normal distribution, also summarized by bivariate dependence coefficients, and motivated by the multivariate central limit theorem. Other estimators of multivariate failure probabilities are proposed in Drees and de Haan (2015) and Valk (2016). While these estimators are based on fewer assumptions than ours, their computational complexity does not allow estimation of failure probabilities for, say, \( d \geq 5 \). Other related literature (Cai et al., 2011; Einmahl et al., 2013; He and Einmahl, 2017) focuses on the semi- or non-parametric estimation of multivariate extreme quantile regions (also limited to low dimensions), thus employing a narrower definition of a failure region. To the best of our knowledge, a general approach for estimating \( \mathbb{P}[\mathbf{X} \in C] \) where the dimension of \( \mathbf{X} \) is high and \( C \) is user-defined does not exist up to date.

In the extreme-value literature, tail dependence of \( \mathbf{X} \) is characterized in various ways: by the exponent measure, by the angular measure, and by the stable tail dependence function, among others (Beirlant et al., 2004; de Haan and Ferreira, 2006). While such objects fully characterize the extremal behavior of a random vector, their estimation in high dimensions suffers from the "curse of dimensionality". Instead, one may consider a partial summary of tail dependence by focusing on the tail dependence coefficients \( \alpha \) or the extremal coefficients \( \theta \) for all pairs, see e.g. Coles et al. (1999). This is similar to summarizing the dependence structure of a random vector by the covariance or correlation matrix.

In this paper, we focus on a generalization of the so-called \textit{tail pairwise dependence matrix} (TPDM), which collects all pairwise tail dependence coefficients. This dependence matrix was originally proposed in the bivariate case by Larsson and Resnick (2012) and has gained in popularity since the work of Cooley and Thibaud (2019); for example, Klüppelberg and Kräli (2021) use it to estimate the parameters of an extreme Bayesian network, Fomichov and Ivanovs (2023) propose a clustering algorithm inspired by spherical \( k \)-means, where the means are replaced by the Perron–Frobenius eigenvectors of the TPDM, and Kim and Kokoszka (2022) study the pairwise tail dependence between vectors of scores of functional data.

In previous studies, the TPDM has only been defined for vectors \( \mathbf{X} \) that are multivariate regularly varying with tail index \( \alpha = 2 \). By contrast, our approach allows for multivariate regular variation with a general tail index \( \alpha \), without the need of standardizing the data. Since marginal standardization does not affect dependence modelling, it is common practice to separate marginal and dependence modelling by first estimating the marginal distributions (parametrically or non-parametrically), standardizing each marginal to a common distribution, and then modelling the dependence structure only. However, such an approach can be suboptimal in applications. For example, in the portfolio risk example, if one intends to estimate the Value-at-Risk or the Expected Shortfall of a portfolio loss \( L = \sum_{j=1}^{d} v_j X_j \), where \( v_1, \ldots, v_d \) represent weights of the stocks (see, e.g., Mainik and Embrechts, 2013), the failure region is of the form \( \{ \sum_{j} v_j X_j > x \} \) with \( x \) large. In this case, we need to directly model the original vector \( \mathbf{X} \): a nonparametric transformation for the marginals will lead to untractable failure regions. Our approach without standardizing a priori is therefore necessary. Such failure regions are also of interest in climate applications, where \( \sum_{j=1}^{d} v_j X_j \) represents the spatial aggregation of an environmental variable (see, e.g. Engelke et al., 2019).

The tail pairwise dependence matrix is closely related to the max-linear model, a simple but rather flexible multivariate extreme-value model. In such a model, each component of a \( d \)-dimensional vector can be interpreted as the maximum shock among a set of \( q \) independent heavy-tailed factors. The max-linear model has \( d \times q \) parameters (or \( d \times (q-1) \) when marginals are standardized), usually stocked in a parameter
matrix $A \in \mathbb{R}^{d \times q}$. Max-linear models are, among other applications, used to model extremes on directed acyclic graphs (Gissibl and Klüppelberg, 2018) or latent factor structures for financial returns (Cui and Zhang, 2018). If the number of parameters is limited, one can use minimum distance estimation (Einmahl et al., 2018) or spherical k-means (Janßen and Wan, 2020) to estimate a max-linear model. Kiriliouk (2020) presents hypothesis tests aiding in deciding the size of $q$.

Theoretically, Fougères et al. (2013) proved that the max-linear model is dense in the class of $d$-dimensional multivariate extreme-value distributions. Moreover, Cooley and Thibaud (2019) showed that there exists a finite $q \in \mathbb{N}$ such that the tail pairwise dependence matrix $\Sigma_X$ of any multivariate regularly varying random vector $X$ with $\alpha = 2$ is equal to that of a max-linear model with $q$ factors via the relation $\Sigma_X = AA^T$. In practice, the parameter matrix $A$ of the max-linear model can be estimated through a completely positive decomposition of $\Sigma_X$. A matrix $S$ is called completely positive if there exists a non-negative matrix $A$ such that $S = AA^T$; finding this matrix $A$ amounts to finding a completely positive decomposition of $S$. Note that the completely positive decomposition is not unique, i.e., multiple max-linear models may share the same TPDM. Positive decompositions are well studied in the linear algebra literature (Groetzner and Dürr, 2020; Zhang, 2018). If the number of parameters is limited, one can use minimum distance estimation (Einmahl et al., 2018) or spherical $k$-means (Janßen and Wan, 2020) to estimate a max-linear model. Kiriliouk (2020) presents hypothesis tests aiding in deciding the size of $q$.

In this paper, we propose an algorithm for an approximate completely positive decomposition $\Sigma_X \approx AA^T$ with $A \in \mathbb{R}^{d \times d}$. In other words, we aim at choosing $q = d$ in the approximate decomposition. We show that the decomposition is exact when $X$ follows a max-linear model with a triangular coefficient matrix $A$. Achieving $q = d$ with a triangular coefficient matrix $A$ in a max-linear model is related to the causal extremal dependence between components of $X$, supported on directed acyclic graphs. In addition, having $q = d$ is in line with classical principal component analysis and is more efficient in simulating the constructed max-linear model. When the decomposition is not exact, we can still use it to construct a max-linear model, for which the TPDM matches $\Sigma_X$ for all off-diagonal entries. By contrast, the TPDM of the constructed max-linear model may have higher values on its diagonal than the original $\Sigma_X$. In the latter case, for a specific type of failure regions, the corresponding probability of a failure set based on the new model can serve as a conservative estimate for the one based on the original random vector $X$. Our proposed decomposition algorithm is easy to compute and applicable for large $d$.

The outline of the paper is as follows. Section 2 presents some necessary background and introduces a generalization of the TPDM, illustrating some of its convenient properties. Next, Section 3 introduces the decomposition algorithm and demonstrates its characteristics. Estimation of the TPDM and of associated failure probabilities is discussed in Section 4. We apply the proposed decomposition algorithm to a given max-linear model and to real data of industry portfolio returns and maximal wind speeds in Section 5. Finally, Section 6 concludes. All proofs are deferred to Appendix A. An additional application, for heavy rainfall in Switzerland, can be found in Appendix B. All results can be reproduced by downloading the code from https://github.com/akiriliouk/FailureProbTPDM.

2. Background

2.1. Multivariate regular variation and the tail pairwise dependence matrix

Let $X \in [0, \infty)^d$ be a random vector with joint cumulative distribution function $F$ and continuous marginal distributions $F_1, \ldots, F_d$. Here we consider non-negative random vectors only because in applications of tail risk analysis, one often focuses on downside risks or losses. Write $\mathbb{E}_0 = [0, \infty]^d \setminus \{0\}$. We assume that $X$ is multivariate regularly varying, i.e., there exists a sequence $b_n \to \infty$ and a non-degenerate limit measure $\nu_X$ such that

$$n \mathbb{P} \left[ b_n^{-1} X \in \cdot \right] \overset{\text{w}}{\to} \nu_X(\cdot), \quad \text{as } n \to \infty,$$

(1)

where $\overset{\text{w}}{\to}$ denotes vague convergence in the space of non-negative Radon measures on $\mathbb{E}_0$ and $\nu_X$ is called the exponent measure (Resnick, 2007, Chapter 6). Then there exists an $\alpha > 0$, called the tail index of $X$, such that the limit measure $\nu_X$ satisfies the homogeneity property of order $\alpha$, i.e., $\nu_X(tC) = t^{-\alpha} \nu_X(C)$ for
any constant $t > 0$ and any ($\nu_X$-measurable) Borel set $C \subset \mathbb{E}_0$. The sequence $b_n$ can be shown to satisfy $b_n \sim L(n)^{1/\alpha}$, where $L$ denotes a slowly varying function.

The homogeneity property suggests that it is possible to express $X$ and correspondingly the limit relation in (1) in polar coordinates. Let $\| \cdot \|$ denote any norm on $\mathbb{R}^d$ and consider the transformation $T : \mathbb{E}_0 \mapsto (0, \infty) \times S_{\| \cdot \|}$, where $S_{\| \cdot \|} = \{ w \in \mathbb{E}_0 : \|w\| = 1 \}$, and

$$T(X) = \left( \|X\|, \frac{X}{\|X\|} \right) =: (R, W).$$

Then (1) is equivalent to the existence of a finite measure $H_X, \| \cdot \|$ (called the angular measure) on $S_{\| \cdot \|}$ and a sequence $b_n \to \infty$ such that

$$n \mathbb{P}\left[ \left( \frac{b_n^{-1} R, W}{\mu_\alpha} \right) \in \cdot \right] \to \mu_\alpha \times H_X, \| \cdot \| (\cdot), \quad \text{as } n \to \infty,$$

in the space of non-negative Radon measures on $(0, \infty) \times S_{\| \cdot \|}$, where $\mu_\alpha$ is a measure on $[0, \infty)$ given by $\mu_\alpha((x, \infty]) = x^{-\alpha}$ for $x > 0$ (Resnick, 2007, Theorem 6.1). We will sometimes use the $L_\alpha$ norm, denoted $\| \cdot \|_\alpha$, i.e., for $x \in [0, \infty)^d$, $\|x\|_\alpha = \left( \sum_{j=1}^d x_j^\alpha \right)^{1/\alpha}$ for $\alpha \geq 1$. Whenever the choice of norm is obvious or does not matter, we will write $S$ and $H_X$ without the subscript $\| \cdot \|$.

Let $B \subset S$ be $H_X$-measurable and consider the set

$$C_{s,B} := \{ x \in \mathbb{E}_0 : \|x\| > s, x/\|x\| \in B \}.$$  

Equation (3) implies $\nu_X(C_{s,B}) = s^{-\alpha} H_X(B)$ and $\nu_X(dr \times dw) = \alpha r^{-\alpha} \, dr \, dH_X(w)$. Denote $m := \nu(C_{1,S}) = H_X(S) > 0$. Then immediately we have that

$$\lim_{n \to \infty} n \mathbb{P}[R > b_n x] = mx^{-\alpha}. \quad (4)$$

This relation implies that the choice of $b_n$ is not unique. Nevertheless, it must be a regularly varying function of $n$ with index $1/\alpha$. For simplicity, we choose $b_n = n^{1/\alpha}$ in the rest of the paper and assume that (4) holds, which describes the tail of $R$. Such a choice slightly limits the potential distributions of $R$ but is nevertheless in line with the max-linear models presented below.

We remark that the angular measure $H_X$ is related to the probability measure $S_X$ in Resnick (2007, Chapter 6) through $H_X = m S_X$ and that the marginals of $X$ satisfy

$$\lim_{n \to \infty} n \mathbb{P}\left[ X_j > n^{1/\alpha} \right] = x^{-\alpha} \int_S w_j^\alpha \, dH_X(w). \quad (5)$$

To summarize the dependence of a multivariate regularly-varying vector $X$, we define the tail pairwise dependence matrix (TPDM) $\Sigma_X$ as

$$\Sigma_X = (\sigma_X)_{j,k=1, \ldots, d}, \quad \text{with } \sigma_X = \int_S w_j^\alpha w_k^\gamma \, dH_X(w). \quad (6)$$

We will write $X \sim \Sigma_X$ if $\Sigma_X$ is the TPDM of $X$. When no confusion can arise, we will write the elements of $\Sigma_X$ as $\sigma_{jk}$. Two variables $X_j, X_k$ are tail dependent if and only if $\sigma_{jk} > 0$. Recall that tail independence corresponds to the scenario where the components of $X$ are dependent but the exponent measure $\nu_X$ assigns no mass off the axes. From (5), we observe that

$$\lim_{n \to \infty} n \mathbb{P}\left[ X_j > n^{1/\alpha} \right] = x^{-\alpha} \sigma_{jj}. \quad (7)$$

Hence, the (limiting) scale of $X_j$ is $(\sigma_{jj})^{1/\alpha}$, and we say that the margins of $X$ are standardized with index $\alpha$ whenever $\sigma_{jj} = 1$ for all $j = 1, \ldots, d$. 


Note that our definition of a TPDM is different from the one in Larsson and Resnick (2012), whose integrand does not depend on $\alpha$ (i.e. the two definitions coincide only when the margins of $X$ have been standardized to $\alpha = 2$); see also Samorodnitsky and Taqqu (1994). The following proposition illustrates some convenient properties of the TPDM.

**Proposition 1** Let $\Sigma_X$ be as defined in (6) and let $R$ and $W$ be as defined in (2).

1. The elements of the TPDM can be written as
   \[
   \sigma_{X,jk} = m \lim_{x \to \infty} E \left[ \frac{W_j^{\alpha/2} W_k^{\alpha/2}}{\alpha^2} \right] E \left[ R > x \right].
   \]

2. The TPDM does not depend on the choice of norm; i.e., for any two norms $\| \cdot \|$ and $\| \cdot \|'$, we have
   \[
   \sigma_{X,jk} = \int_{S_{\| \cdot \|}} \frac{w_j^{\alpha/2} w_k^{\alpha/2}}{\alpha^2} H_X(\| \cdot \|)(dw) = \int_{S_{\| \cdot \|'}} \frac{(w_j')^{\alpha/2} (w_k')^{\alpha/2}}{\alpha^2} H_X(\| \cdot \|')(dw').
   \]

3. Let $X$ be multivariate regularly varying with tail index $\alpha$ and angular measure $H_X$. If
   \[
   X^* = (X_1^*, \ldots, X_d^*) = \left( X_1^{\alpha/\alpha^*}, \ldots, X_d^{\alpha/\alpha^*} \right),
   \]
   then $X^*$ is multivariate regularly varying with tail index $\alpha^*$. In addition, if $H_{X^*}$ is the angular measure of $X^*$ obtained using $b_n^* = n^{1/\alpha}$, we have $\Sigma_X = \Sigma_{X^*}$.

4. Let $X$ and $Y$ be two independent multivariate regularly varying vectors with the same tail index $\alpha$. The corresponding angular measures are $H_X$, and $H_Y$ respectively, when choosing $b_n = n^{1/\alpha}$. Then the component-wise maximum $\max(X, Y)$ is multivariate regularly varying with the same tail index $\alpha$ and with angular measure $H_X + H_Y$. Moreover, the TPDM of $\max(X, Y)$ is
   \[
   \Sigma_{\max(X, Y)} = \Sigma_X + \Sigma_Y.
   \]

Proposition 1 shows that the TPDM is invariant both to the choice of norm and to the choice of the (marginal) tail index. Using the $L_\alpha$ norm, we see that the total mass of the angular measure equals

\[
m = H_X(S) = \int_{\mathbb{R}^d} \|w\|_\alpha^\alpha dH_X(w) = \sum_{j=1}^d \sigma_{jj}.
\]

2.2. Max-linear model

We are interested in calculating the probability that $X$ falls into some extreme failure region $C \subset \mathbb{R}_0$. In practice, to estimate such probabilities $P[X \in C]$, we need to assume a parametric model for the exponent measure $\nu$, or, equivalently, for the angular measure $H$. A simple yet effective parametric model for the goal of estimating failure probabilities is the max-linear model. Let $A$ denote a $d \times q$ matrix with non-negative entries, with columns $a_1, \ldots, a_q$, and suppose that $\max(a_1) = \max(a_{11}, \ldots, a_{dl}) > 0$ for all $l \in \{1, \ldots, q\}$. Let $Z_1, \ldots, Z_q$ be independent random variables whose distribution is Fréchet with scale 1 and shape $\alpha$, $P[Z_l \leq z] = \exp \left\{-z^{-\alpha} \right\}$, and define

\[
A \times_{\max} Z := \left( \max_{l=1, \ldots, q} a_{1l} Z_l, \ldots, \max_{l=1, \ldots, q} a_{dl} Z_l \right)^T,
\]

where $X = A \times_{\max} Z$.
for $Z = (Z_1, \ldots, Z_q)$. Denoting $Y = A \times_{\text{max}} Z$, the angular measure of $Y$ is

$$H_Y(\cdot) = \sum_{l=1}^{q} \|a_l\|_{\alpha} \delta_{a_l/\|a_l\|_{\alpha}}(\cdot). \quad (8)$$

Note that the order of the columns of $A$ does not matter. The marginal distribution of $Y_j$ is Fréchet with scale $(\sum_{l=1}^{q} a_{jl})^{1/\alpha}$ and shape $\alpha$ for $j = 1, \ldots, d$. The TPDM of $Y$ has elements

$$\Sigma_{Y_{jk}} = \sum_{l=1}^{q} \|a_l\|_{\alpha} \left( \frac{a_{jl}}{\|a_l\|_{\alpha}} \right)^{\alpha/2} \left( \frac{a_{kl}}{\|a_l\|_{\alpha}} \right)^{\alpha/2} = \sum_{l=1}^{q} a_{jl}^{\alpha/2} a_{kl}^{\alpha/2},$$

and hence

$$\Sigma_Y = A_* A_{*}^{T}, \quad \text{where } A_* := \left( a_{jl}^{\alpha/2} \right)_{j=1,\ldots,d,l=1,\ldots,q}.$$

Cooley and Thibaud (2019) show that the class of max-linear angular measures, as defined in (8), is dense in the class of possible angular measures (for any norm), i.e., any multivariate regularly varying vector $X$ can be approximated by a max-linear model with a $(d \times q)$ parameter matrix as $q \rightarrow \infty$. Moreover, if attention is restricted to the TPDM, a max-linear model with finite $q$ is sufficient to exactly match $\Sigma_X$; this result has been proven in Cooley and Thibaud (2019) for $\alpha = 2$, but can be easily generalized to any $\alpha > 0$.

**Proposition 2** (Cooley and Thibaud (2019), Propositions 4 & 5)

1. Let $Z_1, \ldots, Z_q$ be Fréchet(1, $\alpha$) random variables. Given any angular measure $H$, there exists a sequence of non-negative $(d \times q)$ matrices $\{A_q\}$, for $q = 1, 2, \ldots$ such that $H_{A_q \times_{\text{max}} Z} \rightarrow H$ weakly as $q \rightarrow \infty$.

2. If $X \in \mathbb{R}^d$ has TPDM $\Sigma_X$, there exists a non-negative $(d \times \bar{q})$ matrix $A_*$ with $\bar{q} < \infty$, such that $\Sigma_X = A_* A_{*}^{T}$.

Whenever a matrix $\Sigma$ can be decomposed as $\Sigma = AA^{T}$, and the matrix $A$ has non-negative entries, we say that $\Sigma$ is completely positive. The max-linear model construction implies that any completely positive matrix can be a TPDM. Conversely, the second part of Proposition 2 shows that any TPDM is completely positive. In fact, for any vector $X$, there exists a max-linear vector $Y$ whose TPDM is equal to that of $X$. This result is key for the estimation of failure probabilities.

**Remark 1** The class of completely positive matrices coincides with the class of TPDMs. In particular, this implies that a TPDM is always positive semi-definite and element-wise non-negative (Shaked-Monderer and Berman, 2021, Chapter 3). Note that Embrechts et al. (2016) characterized the class of matrices of tail dependence coefficients $(\chi_{jk})_{j,k=1,\ldots,d}$, defined by

$$\chi_{jk} = \lim_{t \downarrow 0} t^{-1} \mathbb{P}[F_j(X_j) > 1 - t, F_k(X_j) > 1 - t],$$

and found that it is a strict subset of the class of completely positive matrices. A related problem, focusing on the sub-class of max-stable processes that is characterized through the so-called extremal coefficients, is studied in Strokorb et al. (2015) and Fiebig et al. (2017).

2.3. The probability of a failure region under the max-linear model

We intend to estimate the failure probability $\mathbb{P}[Y \in C]$ for a max-linear vector $Y$. Indeed, whenever the set $C \subset \mathbb{E}_0$ is “extreme enough”, we have

$$\mathbb{P}[Y \in C] = \frac{1}{n} \left\{ n \mathbb{P}\left[n^{-1/\alpha} Y \in n^{-1/\alpha} C \right] \right\} \approx \frac{1}{n} \nu_Y \left( n^{-1/\alpha} C \right). \quad (9)$$

The choice of the region $C$ in (9) depends on the application of interest. Nevertheless, for some practically relevant failure regions, the probability can be calculated directly from the parameters of the max-linear
model. For environmental data, e.g., if $Y$ represents a vector of maximum daily wind speeds, rainfall amounts, or temperatures in $d$ locations, the region

$$C_{(\text{max})}(x) = \{ y \in \mathbb{E}_0 : y_1 > x_1 \text{ or } \ldots \text{ or } y_d > x_d \},$$

is a common choice if an extreme event at at least one location can lead to a climatological catastrophe. The exponent measure for this region equals

$$\nu(C_{(\text{max})}(x)) = \sum_{l=1}^{q} \max_{j=1,\ldots,d} \left( \frac{a_{jl}}{x_j} \right)^{\alpha}.$$  

On the other hand, if $Y$ represents for example the negative log-returns of $d$ stocks in a portfolio, one might be interested in estimating the probability of all stocks crashing together,

$$C_{(\text{min})}(x) = \{ y \in \mathbb{E}_0 : y_1 > x_1 \text{ and } \ldots \text{ and } y_d > x_d \}.$$  

Similarly to $C_{(\text{max})}$, we can calculate

$$\nu(C_{(\text{min})}(x)) = \sum_{l=1}^{q} \min_{j=1,\ldots,d} \left( \frac{a_{jl}}{x_j} \right)^{\alpha}.$$  

Many simple regions (including the ones above when $x_1 = \ldots = x_d$) can be written as follows. Let $C_f(x) = \{ y \in \mathbb{E}_0 : f(y) > x \}$ denote a failure region where $f : \mathbb{R}^d \to \mathbb{R}$ is a measurable function and is non-decreasing in all dimensions. For fixed $w$, let $r^*_w = r^*_w(w)$ be such that $f(r^*_w w) = x$. Then

$$\nu(C_f(x)) = \sum_{l=1}^{q} \|a_l\|^{\alpha} \{ r^*_w (a_l/\|a_l\|) \}^{-\alpha}.$$  

Hence, we can obtain an explicit expression for the exponent measure of any failure region that can be written in the form of $C_f$. For example, if $f(y) = v^T y$ for $v \in \mathbb{E}_0$ such that $v_1 + \ldots + v_d = 1$, we get

$$C_{(\text{sum})}(v, x) = \{ y \in \mathbb{E}_0 : v_1 y_1 + \ldots + v_d y_d > x \}.$$  

Such a region is both of interest for climate events (e.g. in attribution studies such as Kiriliouk and Naveau (2020), among others) and in the financial context (e.g. to calculate the Value-at-Risk for a portfolio), and we calculate

$$\nu(C_{(\text{sum})}(v, x)) = x^{-\alpha} \sum_{l=1}^{q} \left( v^T a_l \right)^{\alpha}.$$  

Moreover, we find

$$\nu(C_{(\text{sum})}(v, x)) = \begin{cases} x^{-1} \sum_{j=1}^{d} v_j \sigma_{jj} & \text{if } \alpha = 1, \\ x^{-2} \left( v^T \Sigma_Y v \right) & \text{if } \alpha = 2. \end{cases}$$  

In these two special cases, $\nu(C_{(\text{sum})}(v, x))$ can be calculated based on the TPDM only. In other words, the random variable $v^T Y$ approximately follows a Pareto distribution whose scale is a function of the TPDM. By contrast, when $\alpha \notin \{1, 2\}$, the parameter matrix $A$ needs to be recovered first. For more complex regions we can still calculate the exponent measure by simulating observations from the max-linear model.
3. Decomposing the pairwise tail dependence matrix

Proposition 2 shows that, for any random vector \( \mathbf{X} \), a max-linear vector \( \mathbf{Y} \) exists such that \( \Sigma_{\mathbf{X}} = \Sigma_{\mathbf{Y}} \). Obtaining such a model requires a completely positive decomposition of \( \Sigma_{\mathbf{X}} \), i.e., finding a \((d \times q)\) matrix \( A_* \) with non-negative entries such that \( \Sigma_{\mathbf{X}} = A_* A_*^T \) (see also Section 2.2). Recall that the matrix \( A_* \) is not unique, i.e., there may exist multiple max-linear models with the same TPDM. Barioli and Berman (2003) approximate completely positive decomposition in the following sense. Firstly, we define the notation “\( \preceq \)”: for two TPDMs, we write \( \sigma_{\mathbf{X},jk} = \sigma_{\mathbf{Y},jk} \) for all \( j, k \in \{1, \ldots, d\} \), \( j \neq k \), and \( \sigma_{\mathbf{X},jj} \preceq \sigma_{\mathbf{Y},jj} \) for all \( j \in \{1, \ldots, d\} \).

Then, for any random vector \( \mathbf{X} \) with TPDM \( \Sigma_{\mathbf{X}} \), we aim at constructing a max-linear vector \( \mathbf{Y} \) with a \((d \times d)\) coefficient matrix \( A_* \), such that \( \Sigma_{\mathbf{X}} \preceq \Sigma_{\mathbf{Y}} \). The max-linear model can be viewed as an approximate model for the original random vector \( \mathbf{X} \), “matching” the pairwise dependence structure of \( \mathbf{X} \).

In this section, we propose a simple algorithm to construct a \((d \times d)\) matrix \( A_* \), which leads to an approximate completely positive decomposition in the following sense. Firstly, we define the notation “\( \preceq \)”: for two TPDMs, we write \( \Sigma_{\mathbf{X}} \preceq \Sigma_{\mathbf{Y}} \) if

\[
\sigma_{\mathbf{X},jk} = \sigma_{\mathbf{Y},jk} \quad \text{for all } j, k \in \{1, \ldots, d\}, \ j \neq k, \quad \text{and} \quad \sigma_{\mathbf{X},jj} \preceq \sigma_{\mathbf{Y},jj} \quad \text{for all } j \in \{1, \ldots, d\}.
\]

Then, for any random vector \( \mathbf{X} \) with TPDM \( \Sigma_{\mathbf{X}} \), we aim at constructing a max-linear vector \( \mathbf{Y} \) with a \((d \times d)\) coefficient matrix \( A_* \), such that \( \Sigma_{\mathbf{X}} \preceq \Sigma_{\mathbf{Y}} \). The max-linear model can be viewed as an approximate model for the original random vector \( \mathbf{X} \), “matching” the pairwise dependence structure of \( \mathbf{X} \).

The details of the proposed algorithm are as follows. Define, for any TPDM \( \Sigma_{\mathbf{X}} = (\sigma_{jk})_{j,k=1,\ldots,d} \) and any \( i \in \{1, \ldots, d\} \),

\[
D_i(\Sigma_{\mathbf{X}}) := \max \left\{ j, k \in \{1, \ldots, d\} \setminus \{i\} : \frac{\sigma_{ji}\sigma_{ki}}{\sigma_{jk}} \right\}.
\]

We write \( D_i = D_i(\Sigma_{\mathbf{X}}) \) when no confusion can arise. Let \( \tau_i = (\tau_{1,i}, \ldots, \tau_{d,i})^T \) with

\[
\tau_{j,i} = \begin{cases} 
\sigma_{ji} (\sigma_{ii} \max(D_i,1))^{-1/2} & \text{if } j \neq i, \\
(\sigma_{ii} \max(D_i,1))^{1/2} & \text{if } j = i.
\end{cases}
\]

Let \( \tau_i = (\tau_{1,i}, \ldots, \tau_{i-1,i}, \tau_{i+1,i}, \ldots, \tau_{d,i})^T \in \mathbb{R}^{d-1} \) and define

\[
\Sigma_{-i}^{(i)} := \Sigma_{\mathbf{X}}^{(-i,-i)} - \tau_i \tau_i^T \in \mathbb{R}^{(d-1) \times (d-1)},
\]

where \( \Sigma_{\mathbf{X}}^{(-i,-i)} \in \mathbb{R}^{(d-1) \times (d-1)} \) is the matrix obtained by removing the \( i \)-th row and column from \( \Sigma_{\mathbf{X}} \).

**Lemma 1** For all \( i \in \{1, \ldots, d\} \) with \( D_i \neq 1 \), the matrix \( \Sigma_{-i}^{(i)} \) in (11) satisfies that

1. \( \Sigma_{-i}^{(i)} \geq 0 \) componentwise;
2. \( \Sigma_{-i}^{(i)} \) is positive semi-definite.

The next proposition shows that if we can construct a \((d-1)\)-dimensional random vector with TPDM \( \Sigma_{-i}^{(i)} \geq \Sigma_{-i} \), we can then use this vector to construct \( \mathbf{Y} \) such that \( \Sigma_{\mathbf{X}} \preceq \Sigma_{\mathbf{Y}} \). In addition, the proposition illustrates the circumstances under which \( \Sigma_{\mathbf{X}} = \Sigma_{\mathbf{Y}} \) is achieved.

**Proposition 3** Let \( \Sigma_{\mathbf{X}} = (\sigma_{jk})_{j,k=1,\ldots,d} \in \mathbb{R}^{d \times d} \) be a TPDM. Let \( i \in \{1, \ldots, d\} \). Consider a random vector \( \mathbf{Y}_{-i} = (\mathbf{Y}_{1}, \ldots, \mathbf{Y}_{i-1}, \mathbf{Y}_{i+1}, \ldots, \mathbf{Y}_{d})^T \) with TPDM \( \Sigma_{\mathbf{Y}_{-i}} \geq \Sigma_{\mathbf{X}} \), where \( \Sigma_{\mathbf{X}} \) is defined as in (11). Let \( Z_i \) be a Fréchet random variable with scale \( 1 \) and shape \( \alpha \), independent of \( \mathbf{Y}_{-i} \). Defining a new vector \( \mathbf{Y} \) via

\[
Y_j := \begin{cases} 
\max \left( \frac{\tau_{j,i}^2}{\alpha} Z_i, Y_j \right) & \text{if } j \in \{1, \ldots, d\} \setminus \{i\}, \\
\tau_{i,i}^2 \alpha Z_i & \text{if } j = i,
\end{cases}
\]

with \( \tau_{j,i} \) as in (10), we have that \( \Sigma_{\mathbf{X}} \preceq \Sigma_{\mathbf{Y}} \). Moreover, \( \Sigma_{\mathbf{X}} = \Sigma_{\mathbf{Y}} \) if and only if \( D_i(\Sigma_{\mathbf{X}}) \leq 1 \) and \( \Sigma_{\mathbf{Y}} = \Sigma_{\mathbf{X}} \).
Based on Lemma 1 and Proposition 3, we can repeatedly apply the procedure in expression (11) to $\Sigma_{X}^{(i)}$ to further decrease its dimension. This leads to an iterative algorithm, throughout which we obtain the matrix $A_\ast$. First, let $i_1 \mapsto i_2 \mapsto \ldots \mapsto i_d$ denote a path, where $(i_1, \ldots, i_d)$ is a permutation of $(1, \ldots, d)$. The order determines which column of $\Sigma_{X}$ will be treated first. For a given path, the iterative algorithm is as follows.

- In the first step, we obtain the vector $\tau_{i_1}$ from (10) by taking $i = i_1$. We fill in the first column of the matrix $A_\ast$ with the obtained vector $\tau_{i_1}$ (as the order of the columns of $A_\ast$ does not matter, we do not necessarily need to fill the $i_1$-th column). The targeted TPDM is then reduced to a $(d-1) \times (d-1)$ matrix $\Sigma_{X}^{(i_1)}$ as defined in (11).

- In the second step, we apply expressions (10) and (11) to $\Sigma_{X}^{(i_1)}$ with the focal dimension now indexed by $i_2$. This step results in a $(d-1)$-dimensional vector. For the second column of the matrix $A_\ast$, we set the $i_1$-th row to zero and fill in the rest with the obtained $(d-1)$-dimensional vector. After the second step, the targeted TPDM is reduced to a $(d-2) \times (d-2)$ matrix.

- The previous step is iterated as follows. In each step $j$, we aim at filling in the $j$-th column of the matrix $A_\ast$. First, we set the elements in the $i_1, i_2, \ldots, i_{j-1}$-th row to zero. Then, we fill in the other $(d-j+1)$ elements by the obtained $(d-j+1)$-dimensional vector of this iteration. After each step, the dimension of the targeted TPDM is reduced by one.

- At step $d$, the targeted TPDM is reduced to $1 \times 1$, i.e. a single positive number. We simply fill in the square root of that value to the $d$-th column and $i_d$-th row of $A_\ast$, while setting all other elements in the $d$-th column to zero.

Proposition 3 guarantees that the matrix $A_\ast$ obtained with this algorithm has only non-negative elements and satisfies $\Sigma_{X} \preceq A_\ast A_\ast^T$. In addition, define $A = A_\ast^{2/\alpha}$ (i.e. each element of the matrix $A_\ast$ is raised to the power $2/\alpha$) and let $Z = (Z_1, \ldots, Z_d)^T$ where $Z_1, \ldots, Z_d$ are i.i.d. standard Fréchet distributed random variables with scale 1 and shape $\alpha$. The max-linear vector $Y = A \times \max Z$ has a TPDM that satisfies $\Sigma_{X} \preceq \Sigma_{Y}$. In other words, we can construct a max-linear vector $Y$ based on only $d$ independent factors to approximate the TPDM of $X$, with at least all off-diagonal elements exactly matched.

Below is an example for this procedure when $d = 3$ and the chosen path is $1 \mapsto 2 \mapsto 3$.

Example 1 $(d = 3, \alpha = 2, \text{path } 1 \mapsto 2 \mapsto 3)$ Let $\Sigma_{X} = (\sigma_{jk})_{j,k = 1,2,3}$ be a known (or estimated) TPDM. We want to find a coefficient matrix $A$ (since $\alpha = 2$, we have $A = A_\ast$) such that the corresponding max-linear model $Y = A \times \max (Z_1, Z_2, Z_3)^T$ has TPDM $\Sigma_{Y}$ satisfying $\Sigma_{X} \preceq \Sigma_{Y}$. Starting with $i = 1$, we obtain from (10) that

$$
\begin{pmatrix}
\frac{\tau_{1,1}}{
\tau_{2,1}\tau_{3,1}
}
\end{pmatrix}
= \frac{1}{\sqrt{\sigma_{11} \max(D_{1,1})}}
\begin{pmatrix}
\sigma_{11} \max(D_{1,1})
\sigma_{12}
\sigma_{13}
\end{pmatrix}
$$

and

$$
\Sigma_{X}^{(-1)} :=
\begin{pmatrix}
\sigma_{22} & \sigma_{23} \\
\sigma_{23} & \sigma_{33}
\end{pmatrix}
:=
\begin{pmatrix}
\frac{\sigma_{22}}{\sigma_{11} \max(D_{1,1})} & \sigma_{23} - \frac{\sigma_{22}\sigma_{13}}{\sigma_{11} \max(D_{1,1})} \\
\sigma_{23} - \frac{\sigma_{22}\sigma_{13}}{\sigma_{11} \max(D_{1,1})} & \frac{\sigma_{33}}{\sigma_{11} \max(D_{1,1})}
\end{pmatrix}.
$$

Hence, the first column of $A$ is given by $(\tau_{1,1}, \tau_{2,1}, \tau_{3,1})^T$.

To find the second column of $A$, we iterate the same algorithm to the newly obtained TPDM $\Sigma_{X}^{(-1)}$ to get that

$$
\begin{pmatrix}
\frac{\tau_{2,2}}{\tau_{3,2}}
\end{pmatrix}
= \frac{1}{\sqrt{\sigma_{22}}}
\begin{pmatrix}
\sigma_{22}
\end{pmatrix}.
$$

Note that since $\Sigma_{X}^{(-1)}$ is positive semi-definite, in this case we have $D_2(\Sigma_{X}^{(-1)}) = \frac{s_{22}^2}{s_{33}s_{22}} \leq 1$. After this iteration, the targeted TPDM is reduced to a single value $s_{33} - \frac{s_{23}^2}{s_{33}s_{22}} \geq 0$. Hence, we can define $\tau_{3,3} := \sqrt{s_{33} - \frac{s_{23}^2}{s_{33}s_{22}}}$.
Eventually, we can construct the max-linear model as $Y = A \times_{\max} (Z_1, Z_2, Z_3)$ where

$$A = \begin{pmatrix} \tau_{1,1} & 0 & 0 \\ \tau_{2,1} & \tau_{2,2} & 0 \\ \tau_{3,1} & \tau_{3,2} & \tau_{3,3} \end{pmatrix}, \quad \text{and} \quad \Sigma_Y = AA^T = \begin{pmatrix} \sigma_{11} \max(D_1, 1) & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix}.$$

We have that $\Sigma_X \preceq \Sigma_Y$, and $\Sigma_Y = \Sigma_X$ if $D_1 \leq 1$.

In Example 1, we applied the iterative algorithm implied by Proposition 3 with the path $1 \mapsto 2 \mapsto 3$. The resulting matrix $A_1$ (and eventually $A$) is then lower triangular. By considering other paths we will obtain different matrices $A_\ell$. Hence, in general, for different choices of paths, a completely positive decomposition (approximate or not) is not unique. It is also possible that one path leads to an exact decomposition while another path leads to an approximation in the sense of $\preceq$. Proposition 4 shows that if the vector $X$ indeed follows a max-linear model constructed from a lower-triangular parameter matrix $A$, by choosing the path appropriately, it is possible to recover the exact max-linear model. Given that the order of columns of $A$ does not matter, the proposition can of course be generalized to parameter matrices $A$ in which each column $j$ has at least one more zero than column $(j - 1)$, for $j = 2, \ldots, d$.

**Proposition 4** Suppose that $X = A \times_{\max} Z$ follows a max-linear model with lower-triangular parameter matrix $A \in \mathbb{R}^{d \times d}$ and $Z = (Z_1, \ldots, Z_q)$, where $Z_1, \ldots, Z_q$ are iid Fréchet with scale 1 and shape $\alpha$. If we choose the canonical path, i.e., $1 \mapsto \ldots \mapsto d$ in the decomposition of $\Sigma_X$ then we obtain the matrix $A_* = A^{\alpha/2}$, with $A_*A_*^T = \Sigma_X$.

Remark 2 When the matrix $A$ is a lower-triangular matrix, the max-linear model $X = A \times_{\max} Z$ corresponds to a fully connected directed acyclic graph (DAG), where each node $i$ points to all nodes $j > i$; see e.g. Kräi et al. (2023). In other words, with a known DAG structure among the components, our algorithm provides the edge weights leading to a given TPDM. When all edge weights are non-zero, the DAG is a complete graph. Each zero element in the lower-triangular part of $A$ corresponds to removing an edge in the DAG.

When the graphical structure is unknown, if the proposed algorithm leads to an exact decomposition together with an appropriate path, it reveals one possible DAG structure among the components of $X$ that can have the same TPDM. The path reveals a potential causal structure among the components.

Remark 3 If the matrix $\Sigma_X$ has a zero eigenvalue, it means that the components of $X$ are linearly related when the radius $R$ is high. In practice, an eigenvalue of $\Sigma_X$ can indeed be close to zero when the variables are strongly tail dependent. This is comparable with the concept of multicolinearity in multivariate statistics. In that case, we will encounter a zero element in $\Sigma_X^{(i)}$ after a certain number of steps (say, $d'$) of the iterative algorithm. As a consequence, the matrix $A_*$ will have $d'$ columns, where $d'$ equals the rank of $\Sigma_X$.

When handling a TPDM $\Sigma_X$ with a large $d$, it is not feasible to exhaust all $d!$ possible paths in search for an exact decomposition. In practice, we need an efficient searching algorithm to find a path that leads to a max-linear model $Y$ such that $\Sigma_Y = \Sigma_X$, if possible. We consider three potential algorithms to achieve the goal.

1. A simple approach: in step $j$ we pick the next dimension $i_j$ corresponding to the lowest value of $D_i$ among all remaining dimensions.
2. An exhaustive approach: in step $j$, we let $T_j$ denote the set of remaining dimensions corresponding to a value of $D_i$ below 1, i.e. $T_j := \{i : D_i \leq 1, i \neq i_1, i_2, \ldots, i_{j-1}\}$. For each $i \in T_j$, we proceed with the iterative process for $\Sigma_X^{(i)}$. This will result in a tree of possibilities after each step. For some “branches” of the tree, the algorithm may end at a “leaf” where no $D_i$ lies below 1, i.e. $T_j = \emptyset$. If some “branch” can iterate $d$ steps, it leads to a path that corresponds to an exact decomposition. We retain all paths (and hence all possible matrices $A_\ell$) that lead to an exact decomposition.
3. A pragmatic approach: in step $j$, we denote $T_j$ as the set of remaining dimensions corresponding to a value of $D_i$ below 1, i.e. $T_j := \{i : D_i < 1, i \neq i_1, i_2, \ldots, i_{j-1}\}$. We randomly choose an element from
this set and use it to proceed with the iteration. If the procedure stops after less than \( d \) iterations and ends with a scenario where no \( D_i \) lies below 1, we restart the entire procedure from the first iteration.

We repeat the algorithm until finding an exact decomposition.

Note that the simple approach may not lead to an exact decomposition because there is no guarantee that the minimum of the \( D_i \)'s is below 1 in all steps. The exhaustive approach will find all exact decompositions if there exists at least one. However, it is limited to moderate dimensions \( d \) because of its computational intensity. The pragmatic approach does not guarantee finding an exact decomposition either. Nevertheless, for most examples in Section 5, an exact decomposition is quickly found after a few “restarts”, and hence it is the approach that we will most use in practice.

We conclude this section by discussing the usefulness of a non-exact decompositions in downstream tasks such as estimating the probability of a failure set. Consider two TPDMs, \( \Sigma_Y \succeq \Sigma_X \), with \( \Sigma_Y \neq \Sigma_X \). Denote \( w = (\text{diag}(\Sigma_Y) - \text{diag}(\Sigma_X))^{1/\alpha} \), where the diag operator is defined by taking the diagonal elements of a square matrix. Then \( w \) is a non-negative \( d \)-dimensional vector with at least one element being positive. Suppose a random vector \( X \) is multivariate regularly varying with index \( \alpha \) and a TPDM \( \Sigma_X \). Consider \( d \) i.i.d. independent Fréchet distributed random variables \( Z = (Z_1, \ldots, Z_d)^T \) with scale 1 and shape \( \alpha \). Then define \( Y = \max(X, w \cdot Z) \), where both the product and the maximum are taken component-wise. From Proposition 1, Statement 4, we can obtain the TPDM of \( Y \) as \( \Sigma_X + \text{Diag}(w^\alpha) = \Sigma_Y \), where the Diag operator constructs a diagonal matrix using elements in \( w^\alpha \). In other words, there exists a random vector \( Y \) with the TPDM \( \Sigma_Y \) such that \( Y \succeq X \) element-wise.

Consider the type of failure set \( C \) defined by a measurable risk function \( f \) as \( C_f(c) = \{ x \in \mathbb{R}^d : f(x) > c \} \) where \( f \) is non-decreasing in all dimensions. We immediately get that \( f(Y) > c \geq f(X) \), which implies that \( \mathbb{P}(f(Y) > c) \geq \mathbb{P}(f(X) > c) \), i.e., the probability of failure set using the constructed vector \( Y \) can only overestimate the intended probability of failure set based on \( X \). For failure sets as \( C_f(c) \), using a non-exact decomposition to estimate its probability can only lead to a conservative estimate, which is desired from the perspective of risk management.

4. Estimation

4.1. Estimation of the TPDM

Let \( X_i = (X_{i1}, \ldots, X_{id}) \) for \( i = 1, \ldots, n \) be iid copies of \( X \). First, we estimate the tail indices \( \alpha_j \), or, equivalently, the extreme value indices \( \gamma_j = 1/\alpha_j \), for each \( j = 1, \ldots, d \) using the sample \( \{X_{1j}, \ldots, X_{nj}\} \).

Note that the assumption of regular variation (1) implies that all margins have equal tail indices, \( \alpha_1 = \ldots = \alpha_d := \alpha \), which is a common assumption, since different tail indices will lead to the smallest one dominating the others and masking tail dependence.

Let \( X_{1(j)} \leq \ldots \leq X_{(n),j} \) denote the order statistics of \( X_{1j}, \ldots, X_{nj} \) and let \( k = k_n \) be an intermediate sequence such that \( k \to \infty \) and \( k/n \to 0 \) as \( n \to \infty \). The Hill estimator,

\[
\hat{\gamma}_j = \frac{1}{k} \sum_{i=1}^{k} \log \left( \frac{X_{(n-i+1),j}}{X_{(n-k),j}} \right),
\]

is the best-known estimator of the extreme-value index (Hill, 1975), and will be used to obtain the estimated tail indices \( \hat{\alpha}_j = 1/\hat{\gamma}_j \). If the confidence intervals of \( \hat{\alpha}_1, \ldots, \hat{\alpha}_d \) are overlapping, a single tail index estimate \( \hat{\alpha} \) might be justified. In that case, the entire sample \( X_{11}, X_{n1}, X_{12}, \ldots, X_{n2}, \ldots, X_{1d}, \ldots, X_{nd} \) of length \( nd \) can be used to obtain the estimate \( \hat{\alpha} \) (Einmahl et al., 2022).

Next, for \( i = 1, \ldots, n \), set \( R_i = \|X_i\|_\alpha \) and \( W_i = X_i / R_i \). By Statement 1 in Proposition 1, we estimate the TPDM by focusing on observations corresponding to high values of the radius component, \( R \). More specifically, we take a high quantile of the empirical distribution of \( R_1, \ldots, R_n \), denoted as \( R_0 \). Note that \( m = \lim_{n \to \infty} n x^{\alpha} \mathbb{P}(R > n^{1/\alpha} x) \). By replacing \( n^{1/\alpha} x \) with \( r_0 \), \( m \) can be estimated by \( \hat{m} = r_0 \left( n_{\text{exc}} / n \right) \) for \( n_{\text{exc}} = \sum_{i=1}^{n} I\{R_i > r_0\} \). Since \( \sigma_{jk} = m \lim_{x \to \infty} \mathbb{E} \left[ W_j^{\alpha/2} W_k^{\alpha/2} \mid R > x \right] \), with replacing \( x \) by \( r_0 \), the
elements of the TPDM can be estimated by
\[
\hat{\sigma}_{jk} = \frac{\hat{m}}{n_{\text{exc}}} \sum_{i=1}^{n} W_{ij}^{\hat{\alpha}/2} W_{ik}^{\hat{\alpha}/2} 1 \{R_i > r_0\} = \frac{r_0}{n} \sum_{i: R_i > r_0} (W_{ij} W_{ik})^{\hat{\alpha}/2},
\]
giving \( \hat{\Sigma}_X = (\hat{\sigma}_{jk})_{j,k=1,\ldots,d} \). Larsson and Resnick (2012) show the asymptotic normality of \( \hat{\sigma}_{jk} \) when \( \alpha = 2 \) is known. In practice, the value for \( r_0 \) could be chosen by plotting different quantiles against the estimates of \( m \) and searching for a region where estimates are stable.

4.2. Estimation of failure probabilities and standardization

Once the TPDM is estimated, we calculate the \( d \times d \) matrix \( \hat{A}_* \) obtained by applying the decomposition algorithm outlined in Section 3 to \( \hat{\Sigma}_X \), and compare it with the empirical \((d \times n_{\text{exc}}) \) estimate
\[
\hat{A}_* = \left( \frac{\hat{m}}{n_{\text{exc}}} \right)^{1/2} \hat{W}^{\hat{\alpha}/2},
\]
where \( \hat{W}^{\hat{\alpha}/2} \) is the matrix whose columns are the vectors \( W^{\hat{\alpha}/2}_i \) for which \( R_i > r_0 \). The matrix \( \hat{A}_* \) is “inefficient” in the sense that it has much more columns than \( \hat{A}_* \), especially when \( n_{\text{exc}} \) is large. However, it is easy to calculate and satisfies \( \hat{\Sigma}_X = \hat{A}_* \hat{A}_*^T \) and we will use it as a benchmark. Finally, we obtain two estimators of \( A \)
\[
\hat{A} = \left( \hat{A}_* \right)^{2/\hat{\alpha}} \quad \text{and} \quad \hat{A} = \left( \hat{A}_* \right)^{2/\hat{\alpha}}.
\]

Now, \( \hat{A} \) and \( \hat{\Sigma}_X \) can be used to estimate the probability of falling into one of the failure regions presented in Section 2.3. We will denote the estimators of these probabilities by \( \hat{p} \) for \( \hat{A} \) and \( \hat{p} \) for \( \hat{A}_* \). Let \( \| \cdot \|_F \) denote the Frobenius norm. In the applications, we will compare exact decompositions of \( \hat{\Sigma}_X \), satisfying \( \| \hat{\Sigma}_X - \hat{A}_* \hat{A}_*^T \|_F \leq 10^{-12} \) with approximate decompositions of \( \hat{\Sigma}_X \), satisfying \( \| \hat{\Sigma}_X - \hat{A}_* \hat{A}_*^T \|_F \leq 5 \).

Finally, note that for certain failure regions we could standardize the margins of \( X_1, \ldots, X_n \) without losing tractability of the shape of the failure set. Standardizing the margins is necessary when the estimated tail indices \( \hat{\alpha}_1, \ldots, \hat{\alpha}_d \) differ too much or whenever (some) tail indices are negative, since our framework requires an equal positive marginal tail index. Note that, as shown in Proposition 1, the TPDM will be the same regardless of the choice of \( \alpha \). We set
\[
X_{ij}^\# = \left( -\log \hat{F}_j^P (X_{ij}) \right)^{-1/\alpha}, \quad j = 1, \ldots, d, \quad i = 1, \ldots, n,
\]
where \( \hat{F}_j^P \) denotes the semi-parametric estimate of \( F_j \),
\[
\hat{F}_j^P (y) = \begin{cases} 
\hat{F}_j(y) & \text{if } y \leq u_j; \\
1 - \left( 1 - \hat{F}_j(u_j) \right) \hat{H}_{\hat{\sigma}_j, \hat{\gamma}_j} (y - u_j) & \text{if } y > u_j;
\end{cases}
\]
where \( \hat{F}_j \) denotes the empirical distribution function of \( X_{1j}, \ldots, X_{nj} \) and \( \hat{H}_{\hat{\sigma}_j, \hat{\gamma}_j} \) the survival function of a generalized Pareto distribution (GPD) with estimated scale \( \hat{\sigma}_j \) and shape \( \hat{\gamma}_j \), obtained by fitting \( (X_{ij} - u_j \mid X_{ij} > u_j)_{i=1,\ldots,n} \) to the GPD for some high threshold \( u_j \). The distribution of \( X_{1j}^\#, \ldots, X_{nj}^\# \) is approximately Fréchet with scale 1 and shape \( \alpha \). Hence, we have \( \sigma_{jj} = 1 \) for \( j \in \{1, \ldots, d\} \) and \( m = H_X(S) = d \) does not need to be estimated. Subsequently, we can decompose the TPDM \( \hat{\Sigma}_X^\# \) and construct the associated max-linear model to proceed with estimating the probability of the transformed failure set.

Note that part of the marginal transformation relies on the empirical distribution function which is not continuous. Therefore, even if the original failure set is measurable, the transformed failure set might not be measurable. Recalling the failure sets introduced in Section 2.3 for example, if the original set is \( C_{(\text{max})}(x) \) or \( C_{(\text{min})}(x) \), the transformed set is still measurable. However, for \( C_{(\text{sum})}(x) \), the transformed set might not be measurable. For that reason, marginal transformations should only be considered for the former type of failure region.
5. Applications

5.1. Failure probabilities computed from repeated decompositions

The TPDM forms an incomplete characterization of tail dependence; hence, we would like to assess by how much the exponent measure of a given failure region varies if it is computed from repeated decompositions. Such an investigation will give us insight into how much information is lost when using pairwise tail dependence coefficients instead of a full $d$-dimensional dependence model. We proceed as follows, inspired by Section 2 of the supplementary material of Cooley and Thibaud (2019). Define

\[
A_1 = \begin{pmatrix}
1.00 & 0.50 & 0.00 & 0.25 & 1.75 & 0.50 & 0.75 & 1.00 & 1.00 & 1.00 & 1.75 & 0.25 & 1.50 & 0.50 & 0.25 & 1.50 \\
2.00 & 0.00 & 1.50 & 1.00 & 1.00 & 0.25 & 1.00 & 0.00 & 1.75 & 0.25 & 1.00 & 0.25 & 0.50 & 0.25 & 1.75 \\
1.75 & 1.25 & 0.75 & 0.25 & 0.50 & 2.00 & 1.75 & 0.25 & 1.25 & 0.25 & 1.00 & 0.25 & 0.50 & 0.25 & 1.75 \\
1.25 & 0.25 & 2.00 & 0.25 & 1.25 & 2.00 & 0.50 & 0.25 & 0.50 & 0.50 & 0.00 & 0.75 & 0.50 & 0.25 & 1.75 \\
1.75 & 0.50 & 0.75 & 1.25 & 0.25 & 0.50 & 1.75 & 0.00 & 2.00 & 1.00 & 1.50 & 0.50 & 0.00 & 0.50 & 1.25
\end{pmatrix},
\]

and let $A_2$ consist of the first eight columns of $A_1$, and $A_3$ of the first four columns of $A_1$. Note that matrix $A_2$ is the same as the matrix considered in Cooley and Thibaud (2019). Let $\nu_i$ denote the exponent measure of a max-linear random vector with coefficient matrix $A_i$, $i = 1, 2, 3$ and $\alpha = 4$. Let $\Sigma_i = A_i A_i^T$; we obtain the $d! = 5! = 120$ decompositions of $\Sigma_i$ and consider the 120 max-linear vectors with corresponding coefficient matrices for each $i = 1, 2, 3$. From the 120 possible paths, we use 98 paths for $A_1$, 894 for $A_2$, and 80 for $A_3$, since the others paths lead to (near) zero eigenvalues in $\Sigma_i$; see Remark 3.

We assess the limiting exponent measure of $C_f(x) = \{(y \in E_0 : f(y) > x)\}$ where $f$ is some function; see Section 2.3. We use the functions $f_1(y) = \frac{1}{d} \sum_{j=1}^d y_j$, $f_2(y) = \prod_{j=1}^d y_j$, $f_3(y) = \min(y_1, \ldots, y_d)$ and $f_4(y) = \max(y_1, \ldots, y_d)$, and we write the associated limiting exponent measures as $C_{(\text{sum})} = C_{f_1}$, $C_{(\text{prod})} = C_{f_2}$, $C_{(\text{min})} = C_{f_3}$, and $C_{(\text{max})} = C_{f_4}$. We consider $\alpha = 4$ to be known and we fix $x$ such that $\nu_i(C_f) = 0.1$. Figure 1 shows boxplots for the measures $\nu_i$, $i = 1, 2, 3$, of the sets $C_{(\text{sum})}$, $C_{(\text{prod})}$, $C_{(\text{min})}$ and $C_{(\text{max})}$, obtained through the different paths. We show both the results when taking only exact decompositions into account (38 for $A_1$, 12 for $A_2$, and 16 for $A_3$) and when using approximate decompositions, i.e., with Frobenius norm $\leq 5$ (68 for $A_1$, 58 for $A_2$, and 72 for $A_3$). Note that for the product-region ($f_2$) and the min-region ($f_3$), only the first column of the resulting coefficient matrices is used in the calculation of $\nu_i$ (since all subsequent columns contain at least one zero; see also Section 2.3).

This can be explained by the fact that the regions $C_{(\text{sum})}$ and $C_{(\text{max})}$ “cut” the axes and hence concern more scenarios driven by a single extreme event, compared to the regions $C_{(\text{prod})}$ and $C_{(\text{min})}$ that only concern scenarios involving simultaneous extremes. In general, variability increases when using approximate decompositions rather than exact decompositions.

5.2. Financial industry portfolios

We focus on the daily negative value-averaged returns of $d = 30$ industry portfolios, downloaded from the Kenneth French Data Library. Data is available for the period January 1970 until December 2019, leading to a sample size $n = 12613$.

We start by estimating the marginal tail indices $\alpha_j = 1/\gamma_j$ using the standard Hill estimator with automated threshold selection (Danielsson et al., 2016). Figure 2 (left) show the estimates, with upper and lower bootstrap confidence limits for a 95% confidence level. Overlapping confidence intervals suggest that the hypothesis of a single index of regular variation $\alpha$ is not rejected. The marginal estimates are rather similar, and the straight line corresponding to the estimate $\hat{\alpha} = 3.479$, based on all data combined, is within all confidence intervals. The next step concerns the choice of $r_0$. Figure 2 (right) show estimates of $m$, the total mass of the angular measure, as a function of $r_0$. We continue with $r_0$ corresponding to the 97.5% quantile and hence $\hat{m} = 29.99$, corresponding to using the 316 largest values.

Finally, we obtain the estimates $\hat{\alpha}$ and $\hat{\tilde{\alpha}}$ based on a (single) decomposition of $\hat{\Sigma}_X$, as described in Section 4.2. In order to characterize the uncertainty stemming from the estimation of $\alpha$ and $\Sigma$, we bootstrap the data 100 times and calculate $\hat{\alpha}$ and $\hat{\tilde{\alpha}}$ for each bootstrapped sample (such that $\alpha$ is re-estimated for each sample using the Hill estimator with automated threshold selection, and $m$ is re-estimated for each sample based on a 97.5% threshold). As in Section 5.1, we present results based on both exact and approximate decompositions. Note that the restriction to approximate decompositions allows for a major time gain (i.e. two minutes for
the 100 approximate decompositions versus 24 hours for the exact ones). Then, using \( A_1, \ldots, A_{100} \) and \( \tilde{A}_1, \ldots, \tilde{A}_{100} \), we estimate the following quantities:

1. the probability \( p_{\text{sum}} = \mathbb{P}[X \in C_{(\text{sum})}(v, x)] = \mathbb{P}[v^T X > x] \), where \( x \) corresponds to the empirical 0.995 quantile of \( (v^T X_i)_{i=1,...,n} \); in other words, we validate the empirical 99.5% value-at-risk (VaR) associated with a weighted sum of industry portfolio returns.

Fig. 1: Boxplots of limiting exponent measures of exact and approximate decompositions of \( \Sigma_1, \Sigma_2, \Sigma_3 \) and \( \Sigma_4 \) for the four functions \( f_1, f_2, f_3, \) and \( f_4 \), and \( \alpha = 4 \). The numbers in parentheses represent the number of decompositions on which the boxplot is based. The true value based on the full 5-dimensional dependence model is equal to 0.1 for all cases.
Estimating probabilities of multivariate failure sets

Fig. 2: Financial industry portfolios. Left: estimates $\hat{\alpha}_j$ with upper and lower 95% confidence limits; the horizontal line represents the estimate $\hat{\alpha}$. Right: estimated mass $\hat{m}$ of the angular measure as a function of $r_0$, a high quantile of $R_1, \ldots, R_n$; the horizontal line represents the estimate for $r_0 = 0.975$.

2. the probability 

$$p_{\text{minsum}} = \mathbb{P}[X \in C_{f_1}(x)]$$

associated with $C_{f_1}(x) = \{y \in \mathbb{E}_0 : f_1(y) > x\}$ and

$$f_1(y) = \min \left( \sum_{j=1}^{10} v_j y_j, \sum_{j=11}^{20} v_j y_j, \sum_{j=21}^{30} v_j y_j \right),$$

where $x$ is the empirical 0.995 quantile of $f_1(X_1), \ldots, f_1(X_n)$; i.e., we validate the empirical 99.5% VaR associated with the minimum of three weighted sums of industry portfolio returns.

3. the probability 

$$p_{\text{maxsum}} = \mathbb{P}[X \in C_{f_2}(x)]$$

associated with $C_{f_2}(x) = \{y \in \mathbb{E}_0 : f_2(y) > x\}$ and

$$f_2(y) = \max \left( \sum_{j=1}^{10} v_j y_j, \sum_{j=11}^{20} v_j y_j, \sum_{j=21}^{30} v_j y_j \right),$$

where $x$ is the empirical 0.995 quantile of $f_2(X_1), \ldots, f_2(X_n)$.

For all three failure regions, we compare equal weights, $v = (1/d, \ldots, 1/d)$, and unequal weights, $v = (0.6, 1.5, 0.9, 0.6, 1.5, 0.9, \ldots, 0.6, 1.5, 0.9)$. All probabilities can be calculated analytically from the estimates of $A$; see Section 2.3.

Figure 3 shows boxplots of the estimates based on $\hat{A}_1, \ldots, \hat{A}_{100}$ (exact), $\hat{A}_1, \ldots, \hat{A}_{100}$ (approximate), and $\hat{A}_1, \ldots, \hat{A}_{100}$. We note that variability is lowest for the sum-region, and highest for the minsum-region. We see that the approximate decompositions lead to almost identical estimates as the ones based on exact decompositions.

In general, $N$ point estimates of failure probabilities (based on $N$ exact decompositions of $\hat{\Sigma}_X$) can be used to validate a VaR as follows. Let $\hat{\text{VaR}}$ denotes a $(1-\alpha)100\%$ VaR estimate from an external model (for $\alpha$ small) and let $\hat{p}_1, \ldots, \hat{p}_N$ be the estimated failure probabilities of $\mathbb{P}[v^T X > \hat{\text{VaR}}]$ (or a region defined via $f_1, f_2$ or similar as above). Then the initial VaR estimate is validated if $\alpha$ falls into the empirical confidence interval computed from $\hat{p}_1, \ldots, \hat{p}_N$. 
5.3. Wind gusts in the Netherlands

Extreme windstorms are one of the important natural hazards affecting Europe. Impacts are not primarily caused by wind speed itself but by the gusts, i.e., the maxima of the wind speed during a few seconds. In February 2022, storm Eunice caused massive damages in Western, Central, and Northern Europe. In the Netherlands, the maximum wind gust measured during storm Eunice was 144 km/h.

We consider daily maximal speeds of wind gusts, measured in km/h, observed at $d = 35$ weather stations in the Netherlands during extended winter (October–March), from October 2001 up to and including March 2022. The data set is freely available from the Royal Netherlands Meteorological Institute (KNMI), see https://climexp.knmi.nl/. Please note that up to 50 stations are available in the Netherlands, but 15 were removed because of a large amount of missing data. The remaining 35 stations have between 0 and 15 missing data points out of $n = 3827$; these do not occur around clusters of extreme wind gusts and hence they do not impact the following analysis. During the extended winter season, the KNMI issues a code red alarm (for any Dutch province) when there is at least 90 % certitude that wind gusts will exceed 120 km/h. The 35 weather stations are divided into coastal stations (17) and inland stations (18) based on the same criteria used in Lenderink et al. (2009), because most environmental variables are known to behave differently over land or close to the sea; see Figure 4 (left) for a map of the stations used in this study.

Inspection of the data suggest that the assumption of positive tail indices is not justified. Hence, we standardize the data as described in Section 4.2. Recall that the choice of $\alpha$ in the standardization procedure does not matter.

Next, we estimate pairwise tail dependence coefficients non-parametrically to examine the difference in pairwise dependence structure between the coastal and the inland stations. Figure 4 shows both the pairwise tail dependence coefficients $\chi_{jk}$ (middle) and the tail pairwise dependence matrix entries $\sigma_{jk}$ (right), for
Estimating probabilities of multivariate failure sets

$j, k = 1, \ldots, d$, both calculated based on a 95% threshold. The plot suggests that there is no significant difference between pairs of inland or pairs of coastal stations, and that the wind gusts are asymptotically dependent, even for large distances.

![Fig. 4: Left: map of the Netherlands with 17 coastal stations in brown and 18 inland stations in green. Middle: empirical coefficients of tail dependence $\chi_{jk}$ for all pairs (black), pairs of coastal stations (brown), and pairs of inland stations (green), with loess curves. Right: same but for the TPDM entries $\sigma_{jk}$.](image)

Finally, we obtain the estimates $\tilde{A}$ and $\hat{A}$ based on a (single) decomposition of $\tilde{X}_X$, as described in Section 4.2. In order to characterize the uncertainty stemming from the standardization and the estimation of $\Sigma$, we bootstrap the data (on the original scale) 100 times and calculate $\tilde{A}$ and $\hat{A}$ for each bootstrapped sample (i.e. each bootstrapped sample is first standardized to $\alpha = 2$). We present results for exact decompositions only because the dimensions are sufficiently low to ensure quick calculations. Then, using $\tilde{A}_1, \ldots, \tilde{A}_{100}$ and $\hat{A}_1, \ldots, \hat{A}_{100}$, we estimate the probability $p_{(\text{max})} = \mathbb{P}[X \in C_{\text{max}}(x)]$ for $x = x_1d$ and $x \in \{1, 1.2, 1.3\}$ (on the original scale); the first two values are the thresholds for a code orange and a code red alarm respectively (issued by the KNMI during winter season).

Figure 5 shows boxplots of the estimates based on $\tilde{A}_1, \ldots, \tilde{A}_{100}$ and $\hat{A}_1, \ldots, \hat{A}_{100}$ for the three thresholds (from left to right). Upper panels represent the analysis using coastal stations only, while lower panels are for inland stations. Even though no apparent difference between coastal and inland stations was visible based on pairwise coefficients, when estimating failure probabilities based on the 17- and 18-dimensional datasets, we observe that values for coastal stations are up to eight times as high as for inland stations. In general, the estimates based on $\hat{A}$ show a slightly larger discrepancy with the empirical estimates than the estimates based on the completely positive decomposition.

6. Discussion

We proposed a general method for the estimation of failure probabilities based on pairwise measures of tail dependence. Our approach is fast and convenient for relatively large $d$, e.g., $d \leq 30$. When $d > 30$, it becomes increasingly hard to find exact decompositions; hence, the (asymptotic) scales of the variables are overestimated, and the approach is appropriate for certain classes of failure regions only.

When summarizing the tail dependence structure through bivariate summary measures, it is natural that some information is lost. However, the ability to obtain repeated decompositions helps in quantifying such uncertainty. Another source of error stems from the estimation of the tail index $\alpha$ and the tail pairwise dependence matrix $\Sigma$. Although a multitude of estimators exist for $\alpha$, estimation of $\Sigma$ has yet to be investigated further and a better estimator could potentially improve our failure probability estimates.
A. Proofs

Proof of Proposition 1 We prove the four statements as follows.

1. The proof follows from that of Larsson and Resnick (2012, Proposition 4): define \( b(t) = t^{1/\alpha} \) and rewrite equation (3) to get that, as \( x \to \infty \),

\[
b^\rightarrow(x) \mathbb{P} \left[ \left( \frac{R}{x}, W \right) \in \cdot \right] \to \mu_{\alpha} \times H_X(\cdot),
\]

where the normalizing sequence \( b^\rightarrow(x) = x^\alpha \sim m/\mathbb{P}(R > x) \). Hence, we have that, as \( x \to \infty \),

\[
H_{X,x}(\cdot) := m \mathbb{P}[R/x > 1, W \in \cdot] = m \mathbb{P}[W \in \cdot | R/x > 1] \to H_X(\cdot).
\]

For \( j, k \in \{1, \ldots, d\} \), define the function \( f_{jk} : S \to \mathbb{R}, w \mapsto w_j^{\alpha/2} w_k^{\alpha/2} \). We find \( H_X(f_{jk}) = \sigma_{X,ik} \) and

\[
H_{X,x}(f_{jk}) = m \int_S w_j^{\alpha/2} w_k^{\alpha/2} \mathbb{P}[W \in \cdot | R > x] = m \mathbb{E}[W_j^{\alpha/2} W_k^{\alpha/2} | R > x],
\]

which completes the proof.

2. Larsson and Resnick (2012, Proposition 2) prove the following change-of-norm conversion formula: for any \( f : S_{\| \cdot \|} \to (0, \infty) \), we have

\[
\int_{S_{\| \cdot \|}} f(w) H_{X,\| \cdot \|}(dw) = \int_{S_{\| \cdot \|}} \|w\|^{\alpha} f(w/\|w\|) H_{X,\| \cdot \|}(dw').
\]
The result follows directly by defining $f$ as $w \mapsto w_j^{\alpha/2}w_k^{\alpha/2}$ for some $j, k \in \{1, \ldots, d\}$.

3. Following statement 2 above, the choice of norm does not affect the calculation of TPDM. Thus we choose $L_\alpha$ and $L_{\alpha^*}$ norms for $X$ and $X^*$ respectively. Correspondingly, we define $R = \|X\|_\alpha$ and $R^* = \|X^*\|_{\alpha^*}$. Then immediately, we get that $(R^*)^{\alpha^*} = R^\alpha$, which implies that

$$W^* = \frac{X^*}{R^*} = \left(\frac{X}{R}\right)^{\alpha/\alpha^*} = W^{\alpha/\alpha^*}.$$  

The multivariate regular variation of $X^*$ is shown as follows. Note that

$$nP \left[ \left( n^{-1/\alpha^*}R^*, W^* \right) \in \cdot \right] = nP \left[ \left( \left(n^{-1/\alpha}R\right)^{\alpha/\alpha^*}, W^{\alpha/\alpha^*} \right) \in \cdot \right].$$

By taking $n \to \infty$, due to the multivariate regular variation definition in (3), we get that $X^*$ is multivariate regularly varying with index $\alpha^*$.

In addition, the two radius factor $R$ and $R^*$ share the same scale $m$: since

$$nx^{\alpha^*} P[R^* > n^{1/\alpha^*}x] = nx^{\alpha^*} P[R^* > nx^{\alpha^*}] = n \left( x^{\alpha^*/\alpha} \right)^{\alpha} P \left[ R > n^{1/\alpha}x^{\alpha*/\alpha} \right],$$

we can replace $x^{\alpha^*/\alpha}$ with $y$ and take $n \to \infty$ on both sides, obtaining the same limit $m$.

Finally, using the expression for the TPDM found in statement 1, we get

$$\sigma_{X^*_{jk}} = m \lim_{x \to \infty} \mathbb{E} \left[ \left( \frac{X^*_jX^*_k}{(R^*)^2} \right)^{\alpha/2} \mid R^* > x \right] = m \lim_{y \to \infty} \mathbb{E} \left[ W^*_{jk}^{\alpha/2} \mid R > y \right] = \sigma_{X^*_{jk}},$$

where we again set $y = x^{\alpha^*/\alpha}$ in the last line.

4. For a given $x \in [0, \infty)^d \setminus \{0\}$, consider all sets in the form of

$$L(x) := \{v \in [0, \infty)^d : v_1 \geq x_1 \text{ or } \ldots \text{ or } v_d \geq x_d\}.$$  

We have that, as $n \to \infty$,

$$nP \left[ n^{-1/\alpha}X \in L \right] \overset{u}{\to} \nu_X(L), \quad nP \left[ n^{-1/\alpha}Y \in L \right] \overset{u}{\to} \nu_Y(L).$$

Clearly $\{n^{-1/\alpha}\max(X, Y) \in L\} = \{n^{-1/\alpha}X \in L\} \cup \{n^{-1/\alpha}Y \in L\}$, which implies that

$$P \left[ n^{-1/\alpha}\max(X, Y) \in L \right] = P \left[ n^{-1/\alpha}X \in L \right] + P \left[ n^{-1/\alpha}Y \in L \right] - P \left[ n^{-1/\alpha}X \in L \right] \cdot P \left[ n^{-1/\alpha}Y \in L \right].$$

By taking $n \to \infty$ and noting that the last term is of a smaller order, we obtain that $\nu_{\max(X, Y)}(L) = \nu_X(L) + \nu_Y(L)$. This relation can then be generalized to the $\sigma-$algebra generated by all sets $L$, i.e. all Borel sets in $[0, \infty)^d \setminus \{0\}$. In other words, we have shown that $\nu_{\max(X, Y)}(\cdot) = \nu_X(\cdot) + \nu_Y(\cdot)$.

Due to the decomposition of the exponent measure $\nu$ to the angular measure $H$ and the relation between the TPDM and the angular measure, we immediately get the remaining parts of the statement.

$\square$
Proof of Lemma 1 We have \( \Sigma^{(i)}_X \geq 0 \) if and only if
\[
\sigma_{jk} - \frac{\sigma_{jj}\sigma_{kk}}{\sigma_{jj} \max(D_i,1)} \geq 0 \iff \max(D_i,1) \leq \frac{\sigma_{jj}\sigma_{kk}}{\sigma_{jj} \sigma_{jk}}, \quad \text{for all } j,k \in \{1, \ldots, d\} \setminus \{i\},
\]
which is true by definition of \( D_i \). Next, for \( x_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d)^T \in \mathbb{R}^{d-1} \) we intend to show that \( x_{-i}^T \Sigma^{(i)}_X x_{-i} \geq 0 \). Let \( y = (y_1, \ldots, y_d)^T \) with
\[
y_j = \begin{cases} x_j & \text{if } j \in \{1, \ldots, d\} \setminus \{i\}; \\ -x_{-i} x_{-i}^{-1} \sigma_{-i} & \text{if } j = i; \end{cases}
\]
where \( \sigma_{-i} = (\sigma_{1,i}, \ldots, \sigma_{i-1,i}, \sigma_{i+1,i}, \ldots, \sigma_{d,i})^T \), the \( i \)-th column of \( \Sigma_X \) with the \( i \)-th element removed. Because \( \Sigma_X \) is positive semi-definite, we have \( y^T \Sigma y \geq 0 \). Denote \( \Sigma_{(-i,-i)} \) as the matrix \( \Sigma_X \) with removing the \( i \)-th row and column. Then, observe that
\[
0 \leq y^T \Sigma y = x_{-i}^T \Sigma_{(-i,-i)} x_{-i} + 2y_i \left( x_{-i}^T \sigma_{-i} \right) + y_i^2 \sigma_{ii} \\
= x_{-i}^T \Sigma_{(-i,-i)} x_{-i} - \frac{2 \left( x_{-i}^T \sigma_{-i} \right)^2}{\sigma_{ii} \max(D_i,1)} + \frac{\left( x_{-i}^T \sigma_{-i} \right)^2}{\sigma_{ii} \max(D_i,1)^2} \\
= x_{-i}^T \Sigma_{(-i,-i)} x_{-i} + \left( x_{-i}^T \tau_{-i} \right)^2 \left( \frac{1}{\max(D_i,1)} - 2 \right) \\
\leq x_{-i}^T \Sigma_{(-i,-i)} x_{-i} - \left( x_{-i}^T \tau_{-i} \right)^2 = x_{-i}^T \Sigma^{(i)}_X x_{-i},
\]
where the third equality is obtained by observing that \( \frac{(x_{-i}^T \sigma_{-i})^2}{\sigma_{ii} \max(D_i,1)} = \left( x_{-i}^T \tau_{-i} \right)^2 \). Hence, we have shown that \( \Sigma^{(i)}_X \) is positive semi-definite. \( \square \)

Proof of Proposition 3 Suppose without loss of generality that \( i = 1 \). Recall that \( \hat{Y}_{-1} \) has a TPDM \( \hat{\Sigma}^{(1)}_Y \). We can view the vector \( (0, \hat{Y}_{-1}) \) as a degenerated \( d \)-dimensional random vector which is also multivariate regularly varying with index \( \alpha \). Its corresponding TPDM is a \( (d \times d) \) matrix, where the first row and column are zero and the lower-right \( (d-1) \times (d-1) \) submatrix is identical to \( \hat{\Sigma}^{(1)}_Y \). Next, the vector \( (\tau_{j,1}^2 Z_1^T)_{j=1}^d \) is also a \( d \)-dimensional multivariate regularly varying random vector with an index \( \alpha \) and a TPDM \( \tau_1^T \tau_1^T \). In addition, these two vectors are independent by construction.

By Statement 4 in Proposition 1, we get that the constructed \( Y \) is multivariate regularly varying with index \( \alpha \). In addition, its TPDM, \( \Sigma_Y \), reads as follows.
\[
\Sigma_Y = \begin{pmatrix} \tau_{1,1}^2 & \tau_{1,1} \tau_{2,1} & \cdots & \tau_{1,1} \tau_{d,1} \\ \tau_{1,1} \tau_{2,1} & \tau_{2,1}^2 & \cdots & \tau_{2,1} \tau_{d,1} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{1,1} \tau_{d,1} & \vdots & \vdots & \tau_{d,1}^2 \end{pmatrix} = \begin{pmatrix} \sigma_{11} \max(D_1,1) & \sigma_{12} & \cdots & \sigma_{1d} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1d} & \sigma_{2d} & \cdots & \sigma_{dd} \end{pmatrix},
\]
where on the diagonal, we have \( \tilde{\sigma}_{ii} = \tau_{i,1}^2 + \tilde{\Sigma}^{(1)}_{Y,ii} \geq \tau_{i,1}^2 + \tilde{\Sigma}^{(1)}_{X,ii} = \sigma_{X,ii} \). All off-diagonal elements are equal to \( \sigma_{X,ij} \) by the fact that \( \tilde{\Sigma}^{(1)}_Y \geq \tilde{\Sigma}^{(1)}_X \) and the definition of \( \tilde{\Sigma}^{(1)}_Y, \tau_1 \) and \( \tau_{-1} \). Hence, \( \Sigma_Y \geq \Sigma_X \). To obtain the exact equation \( \Sigma_Y = \Sigma_X \), the necessary and sufficient condition is that all diagonal elements match. That is, \( D_1 \leq 1 \) and \( \tilde{\Sigma}^{(1)}_Y = \tilde{\Sigma}^{(1)}_X \). \( \square \)
We prove the Proposition by mathematical induction with respect to $d$. First, we prove for $d=2$. Since $A$ is lower triangular, so is the corresponding matrix $A^{\alpha/2}$. For $d=2$, assume that

$$A^{\alpha/2} = \begin{pmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{pmatrix}.$$

Then

$$\Sigma_X = \begin{pmatrix} a_{11}^2 & a_{11}a_{21} \\ a_{11}a_{21} & a_{21}^2 + a_{22} \end{pmatrix}.$$

By taking the path $1 \rightarrow 2$, we calculate

$$D_1(\Sigma_X) = \frac{(a_{11}a_{21})^2}{a_{11}(a_{21} + a_{22})} = \frac{a_{21}^2}{a_{21} + a_{22}}.$$

Since $D_1(\Sigma_X) < 1$, we find $\tau_1 = (a_{11}, a_{21})^T$. Correspondingly, we calculate that

$$\Sigma_X^{(1)} = a_{21}^2 + a_{22}^2 - a_{21}^2 = a_{22}^2.$$

Therefore, the iterative algorithm leads to the matrix

$$A_* = \begin{pmatrix} \tau_1 \\ a_{22} \end{pmatrix} = A^{\alpha/2}.$$

The proposition is thus proven for $d=2$.

Next, assume that the Proposition is valid for $d-1$ with $d \geq 3$, we show that it is also valid for $d$. Denote the lower triangular matrix $A^{\alpha/2}$ as

$$A^{\alpha/2} = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} a_{11} \\ a_{21}^T \\ a_{31}^T \\ \vdots \\ a_{n1}^T \end{pmatrix}.$$

where $0$ is a $(d-1)-$ dimensional vector with all elements zero, $a_{-1} = (a_{21}, \cdots, a_{n1})^T$ and $\bar{A}^{(-1,-1)}$ is a $(d-1) \times (d-1)$ lower-triangular matrix. Then

$$\Sigma_X = \begin{pmatrix} a_{11}^2 & a_{11}a_{-1}^T \\ a_{11}a_{-1}^T & a_{-1}^T + \bar{A}^{(-1,-1)} \bar{A}^{(-1,-1)^T} \end{pmatrix}.$$

Because $\bar{A}^{(-1,-1)}$ contains only non-negative elements, we get $\sigma_{jk} > a_{j1}a_{k1}$ for any $j, k \neq 1$. In addition, $\sigma_{j1} = a_{11}a_{j1}$. Since we consider the path $1 \rightarrow 2 \rightarrow \cdots \rightarrow d$, we first calculate $D_1$; for any $j, k \neq 1$,

$$\frac{\sigma_{jk}\sigma_{k1}}{\sigma_{j1}\sigma_{k1}} = \frac{a_{j1}a_{k1}a_{k1}}{a_{j1}a_{k1}a_{k1}} = 1,$$

which implies that $D_1 < 1$ and thus $\tau_1 = (a_{11}, a_{21}, \cdots, a_{d1})^T$ following the algorithm specified in (10). Correspondingly, from (11) we get

$$\Sigma_X^{(1)} = \Sigma_X^{(-1,-1)} - \tau_{-1}^T \tau_{-1} = a_{-1}a_{-1}^T + \bar{A}^{(-1,-1)} \bar{A}^{(-1,-1)^T} - a_{-1}a_{-1}^T = \bar{A}^{(-1,-1)} \bar{A}^{(-1,-1)^T}.$$

In other words, the $(d-1) \times (d-1)$ matrix $\Sigma_X^{(1)}$ is of the form $\bar{A}^{(-1,-1)} \bar{A}^{(-1,-1)^T}$ where $\bar{A}^{(-1,-1)}$ is a lower-triangular matrix with non-negative elements only. By invoking the assumption that the proposition holds for $d-1$, we complete the proof of the induction step and conclude that the proposition hold for any positive integer $d$. □
B. Extreme precipitation in Switzerland

We focus on daily precipitation amounts in mm for June, July and August for the years 1962–2012 at \( d = 44 \) weather stations near Zürich, Switzerland, leading to a sample size \( n = 4691 \). This data was analyzed in Thibaud and Opitz (2015) and Cooley and Thibaud (2019) and found to be stationary, independent in time, and asymptotically dependent in space. Estimates of marginal tail indices tend to vary so that a common index of regular variation \( \alpha \) is not justified; we standardize the data to \( X_1^#, \ldots, X_n^# \). We take \( r_0 \) corresponding to the 95% quantile, yielding 235 “extreme” observations. Finally, we obtain the estimates \( \hat{A}^1, \ldots, \hat{A}^{100} \) and \( \hat{A}_1, \ldots, \hat{A}_{100} \), we estimate

1. the probability \( p_{(\text{min})} = P[X \in C_{\text{min}}(x)] \), where \( x \) is chosen as the 5, 10 and 30 mm threshold for precipitation on the original scale;
2. the probability \( p_{(\text{max})} = P[X \in C_{\text{max}}(x)] \), where \( x \) is chosen as the 80, 110 and 140 mm threshold for precipitation on the original scale.

In Section 4.1, we've seen that using approximate decompositions instead of exact ones has very little impact on \( p_{(\text{min})} \), but that variability increases for \( p_{(\text{max})} \); we opt for approximate decompositions in this section because of the high dimension of the data.

The probability \( p_{(\text{min})} \) has also been estimated in Cooley and Thibaud (2019), who found a value of \( 4.8 \times 10^{-4} \) and compared it to the empirical estimate of \( 4.3 \times 10^{-4} \) and the estimate based on \( \hat{A} \) which gives \( 1.7 \times 10^{-3} \). Figure 6 shows boxplots of the estimates based on \( \hat{A}^1, \ldots, \hat{A}^{100} \) and \( \hat{A}_1, \ldots, \hat{A}_{100} \), together with empirical estimates for \( p_{(\text{max})} \). We see indeed that the variability of the estimates based on the completely positive decomposition is rather high, even though they show a rather good correspondence with the empirical estimates.

7. Data availability

Code and datasets can be obtained from https://github.com/akiriliouk/FailureProbTPDM and allow the user to replicate all results in this article.

8. Acknowledgments

The authors would like to thank the associate editor and the two reviewers for their insightful remarks that greatly improved the paper. We would also like to thank Emeric Thibaud for providing the Swiss rainfall data used in Appendix B. This work is supported in part by funds from the Fonds de la Recherche Scientifique - FNRS (grant number J.0124.20).

References

F. Barioli and A. Berman. The maximal cp-rank of rank k completely positive matrices. Linear Algebra and its Applications, 363:17–33, 2003.

J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. Statistics of Extremes: Theory and Applications. Wiley, 2004.

R. I. Bot and D.-K. Nguyen. Factorization of completely positive matrices using iterative projected gradient steps. Numerical Linear Algebra with Applications, 28(6):e2391, 2021.

J.-J. Cai, J. H. Einmahl, and L. De Haan. Estimation of extreme risk regions under multivariate regular variation. Annals of Statistics, 39(3):1803–1826, 2011.

S. Coles, J. Heffernan, and J. Tawn. Dependence measures for extreme value analyses. Extremes, 2(4):339–365, 1999.

D. Cooley and E. Thibaud. Decompositions of dependence for high-dimensional extremes. Biometrika, 106(3):587–604, 2019.

Q. Cui and Z. Zhang. Max-linear competing factor models. Journal of Business & Economic Statistics, 36(1):62–74, 2018.

J. Danielsson, L. M. Ergun, L. de Haan, and C. G. de Vries. Tail index estimation: Quantile driven threshold selection. Available at SSRN 2717478, 2016.

L. de Haan and A. Ferreira. Extreme Value Theory: an Introduction. Springer-Verlag Inc, 2006.
Estimating probabilities of multivariate failure sets

Fig. 6: Boxplots of estimates of failure probabilities $p_{(\min)}$ (top) and $p_{(\max)}$ (bottom) for the Swiss precipitation data.

H. Drees and L. de Haan. Estimating failure probabilities. Bernoulli, 21(2):957–1001, 2015.
J. H. Einmahl, L. de Haan, and A. Krajina. Estimating extreme bivariate quantile regions. Extremes, 16(2):121–145, 2013.
J. H. Einmahl, A. Kiriliouk, and J. Segers. A continuous updating weighted least squares estimator of tail dependence in high dimensions. Extremes, 21:205–233, 2018.
J. H. Einmahl, A. Ferreira, L. de Haan, C. Neves, and C. Zhou. Spatial dependence and space–time trend in extreme events. The Annals of Statistics, 50(1):30–52, 2022.
P. Embrechts, M. Höfert, and R. Wang. Bernoulli and tail-dependence compatibility. The Annals of Applied Probability, 26(3):1636–1658, 2016.
S. Engelke, R. De Fondeville, and M. Oesting. Extremal behaviour of aggregated data with an application to downscaling. Biometrika, 106(1):127–144, 2019.
U.-R. Fiebig, K. Strokorb, and M. Schlather. The realization problem for tail correlation functions. Extremes, 20:121–168, 2017.
V. Fomichov and J. Ivanovs. Spherical clustering in detection of groups of concomitant extremes. Biometrika, 110(1):135–153, 2023.
A.-L. Fougeres, C. Mercadier, and J. P. Nolan. Dense classes of multivariate extreme value distributions. Journal of Multivariate Analysis, 116:109–129, 2013.
N. Gissibl and C. Klüppelberg. Max-linear models on directed acyclic graphs. Bernoulli, 24(4A):2693–2720, 2018.
P. Groetzner and M. Dür. A factorization method for completely positive matrices. Linear Algebra and its Applications, 591:1–24, 2020.
Y. He and J. H. Einmahl. Estimation of extreme depth-based quantile regions. Journal of the Royal Statistical Society, Series B (Statistical Methodology), pages 449–461, 2017.
B. M. Hill. A simple general approach to inference about the tail of a distribution. The Annals of Statistics, pages 1163–1174, 1975.
A. Janßen and P. Wan. k-means clustering of extremes. Electronic Journal of Statistics, 14(1):1211–1233, 2020.
M. Kim and P. Kokoszka. Extremal dependence measure for functional data. Journal of Multivariate Analysis, 189:104887, 2022.

A. Kiriliouk. Hypothesis testing for tail dependence parameters on the boundary of the parameter space. Econometrics and Statistics, 16:121–135, 2020.

A. Kiriliouk and P. Naveau. Climate extreme event attribution using multivariate peaks-over-thresholds modeling and counterfactual theory. The Annals of Applied Statistics, 14(3):1342–1358, 2020.

C. Klüppelberg and M. Krali. Estimating an extreme Bayesian network via scalings. Journal of Multivariate Analysis, 181:104672, 2021.

M. Krali, A. C. Davison, and C. Klüppelberg. Heavy-tailed max-linear structural equation models in networks with hidden nodes. arXiv preprint arXiv:2306.15356, 2023.

M. Larsson and S. I. Resnick. Extremal dependence measure and extremogram: the regularly varying case. Extremes, 15(2):231–256, 2012.

G. Lenderink, E. Van Meijgaard, and F. Selten. Intense coastal rainfall in the Netherlands in response to high sea surface temperatures: analysis of the event of August 2006 from the perspective of a changing climate. Climate Dynamics, 32(1):19–33, 2009.

G. Mainik and P. Embrechts. Diversification in heavy-tailed portfolios: properties and pitfalls. Annals of Actuarial Science, 7(1):26–45, 2013.

S. I. Resnick. Heavy-tail phenomena: probabilistic and statistical modeling. Springer Science & Business Media, 2007.

G. Samorodnitsky and M. S. Taqqu. Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance. Chapman & Hall, 1994.

N. Shaked-Monderer and A. Berman. Copositive and completely positive matrices. World Scientific, 2021.

K. Strokorb, M. Schlather, et al. An exceptional max-stable process fully parameterized by its extremal coefficients. Bernoulli, 21(1):276–302, 2015.

E. Thibaud and T. Opitz. Efficient inference and simulation for elliptical Pareto processes. Biometrika, 102(4):855–870, 2015.

C. d. Valk. Approximation and estimation of very small probabilities of multivariate extreme events. Extremes, 19(4):687–717, 2016.