Partial Tail Correlation for Extremes

Lee, Jeongjin; Cooley, Daniel

Publication date:
2022

Partial Tail Correlation for Extremes

A Preprint

Jeongjin Lee*
Namur Institue for Complex Systems
University of Namur
Namur, 5000
jeongjin.lee@unamur.be

Daniel Cooley
Department of Statistics
Colorado State University
Fort Collins, CO 80526
cooleyd@rams.colostate.edu

October 6, 2022

Abstract

We develop a method for investigating conditional extremal relationships between variables at their extreme levels. We consider an inner product space constructed from transformed-linear combinations of independent regularly varying random variables. By developing the projection theorem for the inner product space, we derive the concept of partial tail correlation via projection theorem. We show that the partial tail correlation can be understood as the inner product of the prediction errors associated with the best transformed-linear prediction. Similar to Gaussian cases, we connect partial tail correlation to the inverse of the inner product matrix and show that a zero in this inverse implies a partial tail correlation of zero. We develop a hypothesis test for the partial tail correlation of zero and demonstrate the performance in a simulation study as well as in two applications: high nitrogen dioxide levels in Washington DC and extreme river discharges in the upper Danube basin.

Keywords First keyword · Second keyword · More

1 Motivation

For Gaussian random vectors, the covariance matrix provides complete information about dependence between variables. Even so, conditional relationships, which are a key concept for understanding causal structures between variables, are not directly apparent from the covariance matrix. In Gaussian cases, conditional relationships can be completely specified since conditional distributions are obtainable and remain Gaussian. Conditional relationships are more readily apparent from the the precision matrix (the inverse of the covariance matrix). The conditional relationship between X_i and X_j given all other elements of a Gaussian random vector (denoted by $\boldsymbol{X}_{\setminus(i,j)}$) is related to the (i,j)th element of the precision matrix. Specifically, if the (i,j)th element of the precision matrix is zero, that is $\Sigma^{-1}_{i,j} = 0$, then X_i and X_j are conditionally independent given $\boldsymbol{X}_{\setminus(i,j)}$.

When a distributional assumption is not made, one cannot fully characterize conditional relationships. However, the notion of partial correlation provides a measure of the strength of the conditional relationships between two variables. Consider a centered p-dimensional random vector $\boldsymbol{X}_p$ with covariance matrix Σ. Partitioning into two subvectors, let $\boldsymbol{X}_p = ({\boldsymbol{X}^{(1)}}^T, {\boldsymbol{X}^{(2)}}^T)^T$, where $\boldsymbol{X}^{(1)} = (X_i, X_j)^T$ and $\boldsymbol{X}^{(2)} = \boldsymbol{X}^T_{\setminus(i,j)}$, and partition the covariance matrix accordingly

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

The partial correlation can be connected to the idea of residuals. Consider the matrix

$$\Sigma_{1|2} = E[(\boldsymbol{X}^{(1)} - \hat{\boldsymbol{X}})(\boldsymbol{X}^{(1)} - \hat{\boldsymbol{X}})^T],$$

*Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledging funding agencies.

where $\hat{\boldsymbol{X}} = (\hat{X}_i, \hat{X}_j)^T$ is the vector of best linear predictors in terms of mean squared prediction errors. The partial correlation between X_i and X_j given $\boldsymbol{X}_{\setminus(i,j)}$ is given by

$$\rho_{ij} = \frac{[\Sigma_{1|2}]_{12}}{\sqrt{[\Sigma_{1|2}]_{11}[\Sigma_{1|2}]_{22}}}.$$

Note that $\rho_{ij} = 0$ if and only if $[\Sigma_{1|2}]_{12} = 0$. By matrix inversion, we can show that if the (i, j)th element in the precision matrix is zero, then the partial correlation between X_i and X_j is also zero.

To illustrate conditional relationships between variables, consider the simple 4-dimensional linear model

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \phi & 1 & 0 & 0 \\ \phi^2 & \phi & 1 & 0 \\ \phi^3 & \phi^2 & \phi & 1 \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{bmatrix},$$

where $|\phi| < 1$ and Z_i are uncorrelated noise terms with mean 0 and variance 1. Another way of thinking of this model is through the equation

$$X_i = \phi X_{i-1} + Z_i,$$

which if $X_0 = 0$ $a.s.$, can generate X_i sequentially for $i = 1, \ldots, 4$. The precision matrix of $\boldsymbol{X}_4 = (X_1, ..., X_4)^T$ is

$$Q := \Sigma_{\boldsymbol{X}_4}^{-1} = \begin{bmatrix} 1 & -\phi & 0 & 0 \\ -\phi & 1+\phi^2 & -\phi & 0 \\ 0 & -\phi & 1+\phi^2 & -\phi \\ 0 & 0 & -\phi & 1 \end{bmatrix}.$$

The sparsity seen in the precision matrix can lead to model simplification. In the Gaussian setting, precision matrices have been linked to Gaussian Markov random fields, which in turn can be linked to graphical representations for models [Rue and Held, 2005]. Following the convention in [Rue and Held, 2005] of connecting graph nodes for non-zero entries of the precision matrix yields the graph in Figure 1 for the illustrative model. Since we have not specified the X_i's to be Gaussian, the graph does not imply truly Markov behavior. However, in terms of linear prediction, the predicted value of X_i given only its neighbors X_{i-1} and X_{i+1} is the same as if predicted on $\boldsymbol{X}_4 \setminus X_i$.

Figure 1: The graph given by the precision matrix of the illustrative model.

As it is based on covariance, partial correlation describes conditional relationships at the center of the distribution and is not well-suited for describing relationships in the tail. In the past few years, there has been a concerted effort to develop simplified models for high dimensional extremes based on graphical models and conditional relationships at extreme levels. Gissibl and Klüppelberg [2018] develop causal structure for max linear models via directed acyclic graphs. Directed graphs differ from Figure 1 in that the graph edges have direction. Via max linear operations, Gissibl and Klüppelberg [2018] connect directed acyclic graphs to max stable models, and model sparsity is achieved from the graph structure simplifying high-dimensional models. In other work, Engelke and Hitz [2020] develop the notion of conditional independence for a multivariate Pareto distribution. In particular, Engelke and Hitz [2020] focus on the Hüsler and Reiss [1989] model which is characterized by a variogram. The graphical structure of the Hüsler and Reiss [1989] can be described by a sparse pattern from precision matrices. Engelke and Hitz [2020] use AIC to perform likelihood-based model selection, and use a greedy algorithm to stepwise search of graphical models.

In this work, we develop a novel method for characterizing and investigating extremal conditional relationships between pairs of variables. We rely on multivariate regular variation on the positive orthant to describe extremal dependence in the upper tail, which is assumed to be the direction of interest. We develop the projection theorem for the inner product space defined in Lee and Cooley [2021], and we consider subspaces spanned by a collection of p variables.

Via the projection theorem, we develop the idea of *partial tail correlation*. We show that partial tail correlation can be understood as the inner product of the prediction errors associated with the transformed linear prediction. Similar to the Gaussian case, we connect partial tail correlation to the inverse of the inner product matrix, and show that a zero in this inverse implies a partial tail correlation of zero. Our approach differs from Gissibl and Klüppelberg [2018] in that our approach is more closely linked to ideas from linear models in non-extreme statistics. Our approach is less model-based than that of Engelke and Hitz [2020] in that we do not specify the full model, but instead only work from summaries of pairwise dependence.

In terms of inference, we connect the matrix of inner products to the tail pairwise dependence matrix (TPDM) in Cooley and Thibaud [2019]. We define the observed residuals, which when considered in pairs are regularly varying in $\mathbb{R}^2$ rather than on the positive orthant. Finally, we develop a test for the hypothesis that the partial tail correlation is zero. We demonstrate the performance of this test via a simulation study as well as in two applications: high nitrogen dioxide levels in Washington D.C. to explore conditional extremal relationships between stations ,and assessing flood risk in application to extreme river discharges in the upper Danube basin which was studied in Engelke and Hitz [2020].

2 Background

2.1 Multivariate Regular Variation

Our framework assumes multivariate regular variation, which is closely related to classical multivariate extreme value analysis [De Haan and Ferreira, 2007, Appendix B]. Let $\boldsymbol{X}$ be a p-dimensional regularly varying random vector in $\mathbb{R}_+^p = [0,\infty]^p$ (denoted by $RV_+^p(\alpha)$). A formal definition is that $\boldsymbol{X} \in RV_+^p(\alpha)$ is regularly varying if there exists a normalizing function $b(s) \to \infty$ as $s \to \infty$ and a non-degenerate limit measure $\nu_{\boldsymbol{X}}$ for sets in $E := [0,\infty)^p \setminus \{\boldsymbol{0}\}$ such that

$$sPr(b(s)^{-1}\boldsymbol{X} \in \cdot) \xrightarrow{v} \nu_{\boldsymbol{X}}(\cdot) \tag{1}$$

as $s \to \infty$, where $\xrightarrow{v}$ indicates vague convergence in the space of non-negative Radon measures on $[0,\infty]^p \setminus \{\boldsymbol{0}\}$ [Resnick, 2007]. The normalizing function is of the form $b(s) = U(s)s^{1/\alpha}$ where $U(s)$ is slowly varying, and the tail index α determines the power law of the tail. Applying the same normalization $b(s)$ for all components of $\boldsymbol{X}$, we assume $\boldsymbol{X}$ has common marginal distributions throughout.

Following notations in Resnick [2007], given any norm $\|\cdot\|$, let $T : \boldsymbol{X} \mapsto (\|\boldsymbol{X}\|, \boldsymbol{X}/\|\boldsymbol{X}\|) = (R, \boldsymbol{W})$ be the polar coordinate transformation. We can equivalently formulate the regular variation by Resnick [2007],

$$sPr((b(s)^{-1}R, \boldsymbol{W}) \in \cdot) \xrightarrow{\nu} c\nu_\alpha \times H_{\boldsymbol{X}}, \tag{2}$$

where ν_α is the measure on $(0,\infty]$ and $H_{\boldsymbol{X}}$ is the angular (or spectral measure) on the unit ball $\Theta_{p-1}^+ = \{\boldsymbol{w} \in E : ||\boldsymbol{w}|| = 1\}$. The angular measure $H_{\boldsymbol{X}}$ fully characterizes tail dependence in the limit; however, modeling $H_{\boldsymbol{X}}$ is challenging in high dimensions.

The right hand side in (2) is a product measure, implying that the radial and angular measure are independent of each other in the limit. For a set $C(r,B) = \{\boldsymbol{x} \in \mathbb{R}_+^p : ||\boldsymbol{x}|| > r, \boldsymbol{x}/||\boldsymbol{x}|| \in B\}$ defined with some high threshold r and a Borel set $B \subset \Theta_{p-1}^+$, the scaling property of $H_{\boldsymbol{X}}$ implies $\nu_{\boldsymbol{X}}(C(r,B)) = cr^{-\alpha}H_{\boldsymbol{X}}(B)$.

2.2 Tail Pairwise Dependence Matrix

To fully characterize the angular measure is challenging even in moderately large dimension. Instead, we use the summary information of tail dependence in a matrix of pairwise dependence measures, which is obtainable in high dimensions. We choose a bivariate dependence measure which has similar properties to covariance. Let $\boldsymbol{X} \in RV_+^p(2)$ have its angular measure $H_{\boldsymbol{X}}$. We consider the tail pairwise dependence matrix $\Sigma_{\boldsymbol{X}} = \{\sigma_{\boldsymbol{X}_{ij}}\}_{i,j=1,\cdots,p} \in \mathbb{R}_{p\times p}^+$ defined in Cooley and Thibaud [2019]. The (i,j)th element of $\Sigma_{\boldsymbol{X}}$ is $\sigma_{\boldsymbol{X}_{ij}} = \int_{\Theta_{p-1}^+} w_i w_j dH_{\boldsymbol{X}}(w)$ on $\Theta_{p-1}^+ = \{\boldsymbol{w} \in \mathbb{R}_+^{p-1} : ||\boldsymbol{w}||_2 = 1\}$ and is essentially the extremal dependence measure of Larsson and Resnick [2012].

However, unlike Larsson and Resnick [2012], we set two main different features in our framework. First of all, we require $\alpha = 2$ and the L_2 norm to make $\Sigma_{\boldsymbol{X}}$ have similar properties to a covariance matrix; $\Sigma_{\boldsymbol{X}}$ is positive semi-definite [Cooley and Thibaud, 2019]. In addition, we do not require $H_{\boldsymbol{X}}$ to be a probability measure so that the diagonal elements $\sigma_{\boldsymbol{X}_{ii}}$ imply the relative magnitudes of the respective elements X_i like a covariance matrix. The relation between the scale and the magnitude of each element of $\boldsymbol{X}$ can be readily observed by regular variation $\lim_{s\to\infty} sPr(b(s)^{-1}X_i > c) = c^{-2}\sigma_{\boldsymbol{X}ii}$. By letting $x = cU(s)s^{1/2}$, the relation can be equivalently expressed as

$$\lim_{x\to\infty} \frac{Pr(X_i > x)}{x^{-2}U(x)} = \sigma_{\boldsymbol{X}ii}. \tag{3}$$

Therefore, the magnitude of X_i tied to the diagonal element $\sigma_{\boldsymbol{X}_{ii}}$ is based on tail probabilities. Furthermore, the sum of diagonal elements is identical to the total mass of the angular measure since $\sum_{i=1}^{p} \sigma_{\boldsymbol{X}_{ii}} = \int_{\Theta_{p-1}^{+}} dH_{\boldsymbol{X}}(\boldsymbol{w})$. If the (i,j)th element of $\sigma_{ij} = 0$ is zero, then it implies the asymptotic independence of (X_i, X_j) since $H_{\boldsymbol{X}_p}(\{\boldsymbol{w} \in \Theta_{p-1}^{+} : w_i > 0, w_j > 0\}) = 0$. Unlike covariance matrices, there is an additional property of the TPDM in that it is completely positive, meaning there exists some $q_* < \infty$ and a nonnegative $p \times q_*$ matrix A_* such that $\Sigma_{\boldsymbol{X}} = A_* A_*^T$. The value of q^* is unknown, and A_* is not unique.

2.3 Transformed Linear Operations

To establish a vector space in the positive orthant, Cooley and Thibaud [2019] defined transformed linear operations. For $\boldsymbol{x} \in \mathbb{R}_+^p = [0,\infty)^p$, a key idea is to consider a monotone bijection function t mapping from $\mathbb{R}$ to $\mathbb{R}_+$, with t^{-1} its inverse. The transform t is applied to $\boldsymbol{x}$ or $\boldsymbol{y} := t^{-1}(\boldsymbol{x})$ componentwise. For $\boldsymbol{x}_1$ and $\boldsymbol{x}_2 \in \mathbb{R}_+^p = [0,\infty)^p$, consider the transformed linear operations: vector addition $\boldsymbol{x}_1 \oplus \boldsymbol{x}_2 = t\{t^{-1}(\boldsymbol{x}_1) + t^{-1}(\boldsymbol{x}_2)\}$, and scalar multiplication $a \circ \boldsymbol{x}_1 = t\{at^{-1}(\boldsymbol{x}_1)\}$ for $a \in \mathbb{R}$. We can easily show that $\mathbb{R}_+^p$ with these transformed-linear operations is a vector space [Cooley and Thibaud, 2019]. Furthermore, for $x_j \in \mathbb{R}_+$ and $a_j \in \mathbb{R}$, $j = 1, ..., q$, a transformed-linear combination is defined as $a_1 \circ x_1 \oplus \cdots \oplus a_q \circ x_q = t\Big\{ \sum_{j=1}^{q} a_j t^{-1}(x_j) \Big\}$. Let $A = (\boldsymbol{a}_1, ..., \boldsymbol{a}_q)$ be a $p \times q$ matrix of real numbers where $\boldsymbol{a}_j \in \mathbb{R}^p$ is a p-dimensional column vector for $j = 1, \ldots, q$. Matrix multiplication is defined as $A \circ \boldsymbol{x} = \boldsymbol{a}_1 \circ x_1 \oplus \cdots \oplus \boldsymbol{a}_q \circ x_q = t\{At^{-1}(x)\}$ and note that $A \circ \boldsymbol{x} \in \mathbb{R}_+^p$. For a matrix $B \in \mathbb{R}^{r \times p}$, $B \circ A \circ x = B \circ t\{At^{-1}(x)\} = t\{BAt^{-1}(x)\} = BA \circ x$. It coincides with the standard matrix multiplication.

More importantly, to preserve regular variation on the positive orthant, Cooley and Thibaud [2019] consider the specific transform $t : \mathbb{R} \to (0,\infty)$, $t(y) = \log\{\exp(y) + 1\}$ and its inverse $t^{-1} : (0,\infty) \to \mathbb{R}$, $t^{-1} = \log\{\exp(x) - 1\}$ under some conditions. This transform t is called the softplus function widely used in neural networks. The important property that the transform must meet is such that $\lim_{y\to\infty} t(y)/y = \lim_{x\to\infty} t^{-1}(x)/x = 1$. The condition implies that the transform negligibly affects large values. Consider $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{i,p})^T \in RV_+^p(\alpha)$, another condition we require is a lower tail condition $P(X_{i,j} < x) \to 0$ as $x \to 0$. This condition ensures that $a \circ \boldsymbol{X}_i$ does not affect the upper tail for $a < 0$. For the softplus t, the lower tail condition is met since $sPr\{X_{i,j} \le \exp(-kb(s))\} \to 0$ as $s \to \infty$ for all $k > 0, j = 1, \cdots, p$. This lower tail condition ensures that none of marginals have enough non-zero mass at 0. The lower tail condition is met by standard regularly varying distributions such as the Pareto and the Fréchet. Other transforms $t : \mathbb{R} \to \mathbb{R}_+$ meeting both the aforementioned limiting properties and the lower tail condition can be used to preserve regular variation on the positive orthant.

More precisely, Cooley and Thibaud [2019] show the following.

Proposition 2.1. *Let* $sPr(b(s)^{-1}\boldsymbol{X}_i \in \cdot) \xrightarrow{\nu} \nu_{\boldsymbol{X}_i}(\cdot)$, $i = 1, 2$ *and* $\boldsymbol{X}_1, \boldsymbol{X}_2$ *be independent, then*

$$sPr(b(s)^{-1}(\boldsymbol{X}_1 \oplus \boldsymbol{X}_2) \in \cdot) \xrightarrow{\nu} \nu_{\boldsymbol{X}_1}(\cdot) + \nu_{\boldsymbol{X}_2}(\cdot)$$

Proposition 2.2. *Let* $sPr(b(s)^{-1}\boldsymbol{X} \in \cdot) \xrightarrow{\nu} \nu_{\boldsymbol{X}}(\cdot)$, *then for* $a \in \mathbb{R}$,

$$sPr[b(s)^{-1}(a \circ \boldsymbol{X}) \in \cdot] \xrightarrow{\nu} \begin{cases} a^{\alpha}\nu_{\boldsymbol{X}(\cdot)} & if\, a > 0 \\ 0 & if\, a \le 0 \end{cases}$$

Furthermore, Cooley and Thibaud [2019] consider a simple and useful model framework for $\boldsymbol{X} \in RV_+^p(\alpha)$ via transformed linear combinations of independent regularly varying random variables. Under the aforementioned propositions, we can construct a regularly varying random vector $\boldsymbol{X}$ by applying an arbitrary matrix A to a vector of independent regularly varying random variables $\boldsymbol{Z}$. Let $A = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_q)$ with $\max_{i=1,\ldots p} a_{ij} > 0$ for all $j = 1, ..., q$, where $\boldsymbol{a}_j \in \mathbb{R}^p$ and hence $A \in \mathbb{R}^{p \times q}$. Let

$$\boldsymbol{X} = A \circ \boldsymbol{Z} = t(At^{-1}(\boldsymbol{Z})), \tag{4}$$

where $\boldsymbol{Z} = (Z_1, \ldots Z_q)^T$ is a vector of independent and identically distributed regularly varying random variables meeting $sPr(b(s)^{-1}Z_j > z) \to z^{-\alpha}$ as $s \to \infty$ for $j = 1, \ldots, q$ and the lower tail condition. Then, $\boldsymbol{X} \in RV_+^p(\alpha)$, and when normalized by $b(s)$, its angular measure is $H_{\boldsymbol{X}} = \sum_{j=1}^{q} \|\boldsymbol{a}_j^{(0)}\|^{\alpha} \delta_{\boldsymbol{a}_j^{(0)}/\|\boldsymbol{a}_j^{(0)}\|}(\cdot)$, where δ is the Dirac mass function. The angular measure $H_{\boldsymbol{X}}$ is consistent with point masses corresponding to the normalized columns of A. The zero operation $a^{(0)} := \max(a, 0)$ is applied to vectors or matrices componentwise throughout.

Cooley and Thibaud [2019] show that as $q \to \infty$, the class of angular measures constructed by this matrix multiplication is dense in the class of possible angular measures, implying we are only required to consider a nonnegative matrix A to construct the dense class.

If $\boldsymbol{X} = A \circ \boldsymbol{Z}$ as in (4), the TPDM of the resulting vector is $\Sigma_{\boldsymbol{A}\circ\boldsymbol{Z}} = A^{(0)} {A^{(0)}}^T$. Further, if $\boldsymbol{X} \in RV_+^p(2)$ has TPDM $\Sigma_{\boldsymbol{X}}$, the completely positive decomposition implies that there exists a $0 < q_* < \infty$ and a nonnegative $p \times q_*$ matrix A_* such that $\boldsymbol{X}_* := A_* \circ \boldsymbol{Z}$ has TPDM $\Sigma_{\boldsymbol{X}_*} = \Sigma_{\boldsymbol{X}}$.

3 Projection Theorem in Inner Product Space $\mathcal{V}^q$

3.1 Inner Product Space $\mathcal{V}^q$

We briefly review the vector space of regularly varying random variables constructed from a transformed-linear combinations that was introduced by Lee and Cooley [2021]. For $\boldsymbol{a} \in \mathbb{R}^q$, the vector space is,

$$\mathcal{V}^q = \{X; X = \boldsymbol{a}^T \circ \boldsymbol{Z} = a_1 \circ Z_1 \oplus \cdots \oplus a_q \circ Z_q\}, \tag{5}$$

where $Z_j \in RV_+^1(2)$, $j = 1, \ldots, q$, are independent regularly varying random variables meeting lower tail condition $sP(Z_j \leq \exp(-kb(s))) \to 0$ as $s \to \infty$ with a common normalization $\lim_{z\to\infty} \frac{P(Z_j > z)}{z^{-2}L(z)} = 1$. We note that $\mathcal{V}^q$ is a stochastic vector space and a Hilbert space [Lee, 2022]. For any X_1, X_2 in $\mathcal{V}^q$, the inner product of $X_1 = \boldsymbol{a}_1^T \circ \boldsymbol{Z}$ and $X_2 = \boldsymbol{a}_2^T \circ \boldsymbol{Z}$ is defined as

$$\langle X_1, X_2 \rangle := \boldsymbol{a}_1^T \boldsymbol{a}_2 = \sum_{i=1}^{q} a_{1i} a_{2i}.$$

We define the angle between X_1 and X_2 to be

$$\theta = \cos^{-1}[\langle X_1, X_2 \rangle / (||X_1|| \, ||X_2||)],$$

where $\theta \in [0, \pi]$. We say $X_1, X_2 \in \mathcal{V}^q$ are orthogonal if $\langle X_1, X_2 \rangle = 0$. The norm of X is defined as $\|X\|_{\mathcal{V}^q} = \sqrt{\langle X, X \rangle}$. We use the subscript $\mathcal{V}^q$ to remind that the norm is based on the coefficients which determine the random variable and to distinguish from the usual Euclidean norm based on a location in space. The norm induces the metric of X_1 and X_2 as $d(X_1, X_2) = \|X_1 \ominus X_2\|_{\mathcal{V}^q} = \sqrt{\sum_{i=1}^q (a_{1i} - a_{2i})^2}$.

We recall the tail ratio of $X = \boldsymbol{a}^T \circ \boldsymbol{Z} \in \mathcal{V}^q$,

$$TR(X) := \lim_{z\to\infty} \frac{P(X > z)}{P(Z_1 > z)} = \sum_{j=1}^{q} \left(a_j^{(0)}\right)^2,$$

in Lee and Cooley [2021] where only the positive elements of $\boldsymbol{a}$ contribute. The square in the exponent arises because we assume $\alpha = 2$. Unlike the norm which is not estimable since the random variable's coefficients are not observable from data, the tail ratio is estimable. However, the metric can be connected to the tail ratio

$$TR\left(\max((X_1 \ominus X_2), (X_2 \ominus X_1))\right) = \sum_{j=1}^{q} (a_{1j} - a_{2j})^2 = d^2(X_1, X_2),$$

because $P(\max(Z_1, Z_2) > z) \sim P(Z_1 > z) + P(Z_2 > z)$ as $z \to \infty$ for independent regularly varying random variables Z_1 and Z_2 [cf. Jessen and Mikosch, 2006, Lemma 3.1]. This equality still holds for non-independent regularly varying random variables $X_1 \ominus X_2$ and $X_2 \ominus X_1$ because of the max operation. See proofs in Lee [2022]. In general, it is required for $\alpha = 2$ to connect the inner products of $\mathcal{V}^q$ to quantities which are observable from the tail behavior of the data. We will return to this discussion in Section 5.

For a random vector whose elements are in $\mathcal{V}^q$: $\boldsymbol{X}_p = (X_1, \ldots, X_p)^T$ where $X_i = \boldsymbol{a}_i^T \circ \boldsymbol{Z} \in \mathcal{V}^q$ for $i = 1, \ldots, p$, it was shown that $\boldsymbol{X}_p \in RV_+^p(2)$ and $\boldsymbol{X}_p$ is of the form $A \circ \boldsymbol{Z}$. We denote the matrix of inner products by

$$\Gamma_{\boldsymbol{X}_p} = \langle X_i, X_j \rangle_{i,j=1,\ldots p} = AA^T. \tag{6}$$

The inner product matrix $\Gamma_{\boldsymbol{X}}$ for $X_i \in \mathcal{V}^q$ will be connected to the TPDM $\Sigma_{\boldsymbol{X}}$ for $\boldsymbol{X} \in RV_+^p(2)$ in section 5

3.2 Projection Theorem in $\mathcal{V}^q$

As any $X \in \mathcal{V}^q$ is uniquely identifiable by its vector of coefficients $\boldsymbol{a}$, $\mathcal{V}^q$ is isomorphic to $\mathbb{R}^q$ with the same inner product. Therefore, $\mathcal{V}^q$ is complete and is Hilbert space [Lee, 2022]. Let $X_i = \boldsymbol{a}_i^T \circ \boldsymbol{Z} \in \mathcal{V}^q$, $i = 1, \ldots, p$, where p is assumed to be less than q. We consider the subspace $\mathcal{V}_A$ spanned by a finite set $\{X_1, \ldots, X_p\}$, where A refers to the matrix which generates $\boldsymbol{X}_p = (X_1, \ldots, X_p)^T$. Thus,

$$\mathcal{V}_A = \{\boldsymbol{b}^T \circ \boldsymbol{X}_p; \boldsymbol{b} \in \mathbb{R}^p\}.$$

We develop the projection theorem in the vector space $\mathcal{V}^q$ constructed from transformed-linear combinations. For any $X \in \mathcal{V}^q$, we define a transformed-projection mapping $P_{\mathcal{V}_A}$ by

$$P_{\mathcal{V}_A} X = \{\boldsymbol{b}^T \circ \boldsymbol{X}_p \text{ such that } ||X \ominus (\boldsymbol{b}^T \circ \boldsymbol{X}_p)||_{\mathcal{V}^q} = \inf_{Y \in \mathcal{V}_A} ||X \ominus Y||_{\mathcal{V}^q}\}.$$

We say $P_{\mathcal{V}_A}$ is a transformed-linear projection mapping of $\mathcal{V}^q$ onto $\mathcal{V}_A$. We define the orthogonal complement of a subset $\mathcal{V}_A^\perp \subset \mathcal{V}^q$ as

$$\mathcal{V}_A^\perp = \{X \in \mathcal{V}^q; \langle X, Y \rangle = 0 \quad \forall Y \in \mathcal{V}_A\};$$

that is, $\mathcal{V}_A^\perp$ is the set of all elements of $\mathcal{V}^q$ which are orthogonal to all elements of $\mathcal{V}_A$.

Lee and Cooley [2021] briefly mentioned the projection theorem as an alternative method to find the optimal transformed linear predictor of an unobserved X_{p+1} given $\boldsymbol{X}_p$. Here, we present a more thorough development of the projection theorem. The following development of the projection theorem and its properties are similar to the the presentation in Brockwell et al. [1991] and Cline [1983]. Instead of considering transformed linear operations of nonnegative regularly varying random variables as we do, Cline [1983] considered standard linear combinations of symmetric regularly varying random variables with any tail index. We defer all proofs to the Appendix A and B.

Theorem 3.1. *(Projection theorem) Let $\mathcal{V}_A$ be the previously defined subspace of the Hilbert space $\mathcal{V}^q$ and $X \in \mathcal{V}^q$. Let $X_i = \sum_{j=1}^q a_{ij} \circ Z_j \in \mathcal{V}^q$, $i = 1, \dots, p$, and let $X = \sum_{j=1}^q a_j^* \circ Z_j \in \mathcal{V}^q$. Then*
i) $\hat{X} := P_{\mathcal{V}_A} X$ ($\hat{X}$ is the projection of X onto $\mathcal{V}_A$) has a unique element in $\mathcal{V}_A$ such that

$$||X \ominus \hat{X}||_{\mathcal{V}^q} = \inf_{Y \in \mathcal{V}_A} ||X \ominus Y||_{\mathcal{V}^q}, \textit{ and}$$

ii) $\hat{X} \in \mathcal{V}_A$ such that $||X \ominus \hat{X}||_{\mathcal{V}^q} = \inf_{Y \in \mathcal{V}_A} ||X \ominus Y||_{\mathcal{V}^q}$ if and only if $\hat{X} \in \mathcal{V}_A$ and $(X \ominus \hat{X}) \in \mathcal{V}_A^\perp$.

Now, let I be the identity mapping on $\mathcal{V}^q$. The proposition below shows there is a unique mapping $P_{\mathcal{V}_A}$ of $\mathcal{V}^q$ onto $\mathcal{V}_A$ such that $I - P_{\mathcal{V}_A}$ maps $\mathcal{V}^q$ onto $\mathcal{V}_A^\perp$ by Theorem 3.1.

Proposition 3.1. *(Property of Projection Mappings) Let $P_{\mathcal{V}_A}$ be the projection mapping of $\mathcal{V}^q$ onto a subspace $\mathcal{V}_A$. Then,*
i) $P_{\mathcal{V}_A}(\alpha \circ X \oplus \beta \circ Y) = \alpha \circ P_{\mathcal{V}_A} X \oplus \beta \circ P_{\mathcal{V}_A} Y, \quad X, Y \in \mathcal{V}^q, \quad \alpha, \beta \in \mathbb{R}.$
[That is, the projection mapping $P_{\mathcal{V}_A}$ is a linear mapping.]
ii) For every $X \in \mathcal{V}^q$, there exist an element of $\mathcal{V}_A$ and an element of $\mathcal{V}_A^\perp$ such that

$$X = P_{\mathcal{V}_A} X \oplus (I - P_{\mathcal{V}_A}) X$$

and this decomposition is unique.

Theorem 3.1 shows that $\hat{X} \in \mathcal{V}_A$ is the unique element closest to X such that

$$\langle X \ominus \hat{X}, Y \rangle = 0 \tag{7}$$

for all $Y \in \mathcal{V}_A$. The equation (7) is called the prediction equation and makes $\hat{X}$ as the best predictor of $X \in \mathcal{V}^q$. When we consider the problem of predicting an unobserved X_{p+1} by using the transformed-linear predictor of $(X_1, \dots, X_p)$, the goal is to find $\hat{X}_{p+1} \in \mathcal{V}_A$ that minimizes $||\hat{X}_{p+1} \ominus X_{p+1}||_{\mathcal{V}^q}$. The prediction equation is written as $\langle X_{p+1} \ominus \hat{X}_{p+1}, X_i \rangle = 0$, for $i = 1, \cdots, p$. This condition can equivalently be expressed with the matrix notation by the linearity of the inner product.

$$[< X_{p+1}, X_i >]_{i=1}^p = [< X_i, X_j >]_{i,j=1}^p [b_i]_{i=1}^p = \left[\sum_{k=1}^q a_{ik} a_{jk}\right]_{i,j=1}^p [b_i]_{i=1}^p \tag{8}$$

Solving this equation, this in turn yields the form of the best transformed linear predictor of $\hat{X} = \boldsymbol{b}^T \circ \boldsymbol{X}_p$ in terms of minimizing the tail metric.

3.3 Inner Product Matrix of Prediction Errors

Changing focus from the setting where $\boldsymbol{X}_p$ is observed and X_{p+1} is unobserved, we continue to assume $\boldsymbol{X}_p = (X_1, \dots, X_p)^T$ where $X_i \in \mathcal{V}^q$, for $i = 1, \dots, p$,, but assume we partition the vector so that $\boldsymbol{X}_p = (\boldsymbol{X}^{(1)^T}, \boldsymbol{X}^{(2)^T})^T$, where $\boldsymbol{X}^{(1)}$ has dimension $p_1 < p$ and $\boldsymbol{X}^{(2)}$ has dimension $p - p_1$. Without loss of generality, $\boldsymbol{X}_p$ can be reordered so that $\boldsymbol{X}^{(1)}$ is any subvector of elements of $\boldsymbol{X}_p$. Partitioning A yields

$$\begin{bmatrix} \boldsymbol{X}^{(1)} \\ \boldsymbol{X}^{(2)} \end{bmatrix} = \begin{bmatrix} A^{(1)} \\ A^{(2)} \end{bmatrix} \circ \boldsymbol{Z}_q.$$

The matrix of inner products of $(\boldsymbol{X}^{(1)^T}, \boldsymbol{X}^{(2)^T})^T$ is

$$\Gamma_{(\boldsymbol{X}^{(1)^T}, \boldsymbol{X}^{(2)^T})^T} = \begin{bmatrix} A^{(1)}A^{(1)^T} & A^{(1)}A^{(2)^T} \\ A^{(2)}A^{(1)^T} & A^{(2)}A^{(2)^T} \end{bmatrix} := \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}. \tag{9}$$

We now consider the problem of finding $P_{\mathcal{V}_A}\boldsymbol{X}^{(1)}$ via projection theorem. Minimizing $d(\boldsymbol{b}^T \circ \boldsymbol{X}^{(2)}, \boldsymbol{X}^{(1)})$ is identical to minimization of $\|\boldsymbol{b}^T A^{(2)} - A^{(1)}\|^2_{\mathcal{V}^q}$. Taking derivatives with respect to $\boldsymbol{b}$ and setting equal to zero, the minimizer $\hat{\boldsymbol{b}}$ solves $(A^{(2)}A^{(2)^T})\hat{\boldsymbol{b}} = A^{(2)}A^{(1)^T}$. If $(A^{(2)}A^{(2)^T})$ is invertible, then the solution $\hat{\boldsymbol{b}}$ is,

$$\hat{\boldsymbol{b}} = (A^{(2)}A^{(2)^T})^{-1}A^{(2)}A^{(1)^T} = \Gamma_{22}^{-1}\Gamma_{21}. \tag{10}$$

With the best linear predictor, we can then consider the vector of *prediction errors* $\boldsymbol{X}^{(1)} \ominus \hat{\boldsymbol{X}} = (A^{(1)} - \boldsymbol{b}^T A^{(2)}) \circ \boldsymbol{Z} \in RV^+_{p_1}(2)$, and whose elements are in $\mathcal{V}^q$. The inner product of these prediction errors has a similar form to the conditional covariance matrix under Gaussian assumptions.

$$\begin{aligned} \|\boldsymbol{X}^{(1)} \ominus \hat{\boldsymbol{X}}\|^2_{\mathcal{V}^q} &:= \langle \boldsymbol{X}^{(1)} \ominus \hat{\boldsymbol{X}}, \boldsymbol{X}^{(1)} \ominus \hat{\boldsymbol{X}} \rangle \\ &= (A^{(1)} - \hat{\boldsymbol{b}}^T A^{(2)})(A^{(1)} - \hat{\boldsymbol{b}}^T A^{(2)})^T \\ &= \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}. \end{aligned} \tag{11}$$

4 Partial Tail Correlation

4.1 Partial Tail Correlation via the Projection Theorem

We now turn attention to developing the notion of partial tail correlation between pairs of elements of a vector $\boldsymbol{X}_p = (X_1, \ldots, X_p)^T$ where $X_i \in \mathcal{V}^q$ for $i = 1, \ldots, p$. Let $\boldsymbol{X}^{(1)} = (X_i, X_j)^T$ and $\boldsymbol{X}^{(2)} = (\boldsymbol{X}_p \setminus (X_i, X_j))^T$. From a geometric perspective the projection theorem provides a way of defining the partial tail correlation between X_i and X_j given $\boldsymbol{X}^{(2)}$ as the cosine of the angle between the prediction errors.

Because we aim to project $\boldsymbol{X}^{(1)}$ onto the space spanned by $\boldsymbol{X}^{(2)}$, we consider the subspace $\mathcal{V}_{A_2}$ spanned by a finite set $\{X_1, \ldots, X_p\} \setminus \{X_i, X_j\}$. Note that $\mathcal{V}_{A_2} \subset \mathcal{V}_A$. We define $P_{\mathcal{V}_{A_2}}$ as the projection mapping of $\mathcal{V}^q$ onto $\mathcal{V}_{A_2}$. We denote by $P_{\mathcal{V}_{A_2}}\boldsymbol{X}^{(1)}$ the projection of $\boldsymbol{X}^{(1)}$ onto the space $\mathcal{V}_{A_2}$. We call $(\boldsymbol{X}^{(1)} \ominus P_{\mathcal{V}_{A_2}}\boldsymbol{X}^{(1)})$ prediction errors obtained by projecting $\boldsymbol{X}^{(1)}$ onto the space $\mathcal{V}_{A_2}$. The orthogonality condition says $P_{\mathcal{V}_{A_2}}\boldsymbol{X}^{(1)} = \boldsymbol{b}^T \circ \boldsymbol{X}^{(2)}$ is such that $\boldsymbol{X}^{(1)} \ominus P_{\mathcal{V}_{A_2}}\boldsymbol{X}^{(1)}$ is orthogonal to the space $\mathcal{V}_{A_2}$. Proposition 3.1 says that $\boldsymbol{X}^{(1)}$ can be uniquely expressed as the sum of $P_{\mathcal{V}_{A_2}}\boldsymbol{X}^{(1)}$ and $(I - P_{\mathcal{V}_{A_2}})\boldsymbol{X}^{(1)}$.

Definition 4.1. Let $X_i \in \mathcal{V}^q$ for $i = 1, ..., p$. Denote by $\mathcal{V}_{A_2}$ the space spanned by the set of variables $\boldsymbol{X}^{(2)} = (\boldsymbol{X}_p \setminus \{X_i, X_j\})^T$. Let $X_i \ominus P_{\mathcal{V}_{A_2}}X_i$ and $X_j \ominus P_{\mathcal{V}_{A_2}}X_j$ be prediction errors obtained after projecting X_i and X_j onto the space $\mathcal{V}_{A_2}$, respectively. Then, the partial tail correlation between X_i and X_j is defined as

$$\rho^E_{ij} = \frac{< X_i \ominus P_{\mathcal{V}_{A_2}}X_i, X_j \ominus P_{\mathcal{V}_{A_2}}X_j >}{\|X_i \ominus P_{\mathcal{V}_{A_2}}X_i\|_{\mathcal{V}^q}\|X_j \ominus P_{\mathcal{V}_{A_2}}X_j\|_{\mathcal{V}^q}}, \tag{12}$$

where the superscript E in ρ^E_{ij} stands for "extreme". $\langle X_i \ominus P_{\mathcal{V}_{A_2}}X_i, X_j \ominus P_{\mathcal{V}_{A_2}}X_j \rangle = 0$ iff $\rho^E_{ij} = 0$, which we denote by $X_i \perp X_j | \boldsymbol{X}^{(2)}$.

As before we denote the inner product matrix of prediction errors by

$$\begin{aligned} \Gamma_{1|2} &:= < \boldsymbol{X}^{(1)} \ominus P_{\mathcal{V}_{A_2}}\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(1)} \ominus P_{\mathcal{V}_{A_2}}\boldsymbol{X}^{(1)} > \\ &= \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}. \end{aligned}$$

We define $\Gamma_{1|2}$ as the *conditional* inner product matrix (IPM). The partial tail correlation can be represented by elements of the conditional IPM,

$$\rho^E_{i,j} = \frac{a_{ij}}{\sqrt{a_{ii}a_{jj}}}, \quad i, j = 1, 2, \tag{13}$$

where $\Gamma_{1|2} = [a_{ij}]_{i,j=1,2}$. Note that $\Gamma_{1|2}$ is positive semi-definite but not completely positive.

4.2 Partial Tail Correlation and Transformed Linear Prediction

We return temporarily to the problem of predicting one variable $X_{p+1} \in \mathcal{V}^q$ given $\boldsymbol{X}_p \in RV_+^p(2)$. In this setting, the partial tail correlation corresponds to the coefficients of the vector $\boldsymbol{b}$ for the best transformed-linear predictor. Importantly, if b_i denotes the ith element of $\boldsymbol{b}$, $b_i = 0$ iff $X_{p+1} \perp X_i | \boldsymbol{X}_p \setminus X_1$. This implies that if X_{p+1} and X_i given $\boldsymbol{X}_p \setminus X_i$ have partial tail correlation of zero, then X_i adds no additional information to the transformed-linear prediction of X_{p+1}. Without loss of generality, below we consider the specific case where $i = 1$.

Proposition 4.1. *Let $\mathcal{V}_A$ be the previously defined subspace of the Hilbert space $\mathcal{V}^q$. Assume $X_i \in \mathcal{V}^q$, $i = 1, ..., p+1$. Then the partial tail correlation between X_{p+1} and X_1 is zero if and only if the ith coefficient of $\boldsymbol{b}$ in the best transformed-linear predictor $\hat{X}_{p+1} = \boldsymbol{b}^T \circ \boldsymbol{X}_p = b_1 \circ X_1 \oplus \cdots \oplus b_p \circ X_p$ is zero.*

Proof. By projection theorem, the space $\mathcal{V}_A$ can be decomposed into two orthogonal subspaces $\mathcal{V}_{A_1}$ spanned by $(X_2, \cdots, X_p)$ and $\mathcal{V}_{A_1^\perp}$ spanned by $(X_1 \ominus P_{\mathcal{V}_{A_1}} X_1)$, respectively. Thus, the projection of X_{p+1} onto the space $\mathcal{V}_A$ can also be split into two parts,

$$\hat{X}_{p+1} = P_{\mathcal{V}_A} X_{p+1} = P_{\mathcal{V}_{A_1}} X_{p+1} \oplus P_{\mathcal{V}_{A_1^\perp}} X_{p+1} = P_{\mathcal{V}_{A_1}} X_{p+1} \oplus c \circ (X_1 \ominus P_{\mathcal{V}_{A_1^\perp}} X_1), \tag{14}$$

where $c = \frac{\langle X_{p+1}, X_1 \ominus P_{\mathcal{V}_{A_1}} X_1 \rangle}{||X_1 \ominus P_{\mathcal{V}_{A_1}} X_1||^2} = \frac{\langle X_{p+1} \ominus P_{\mathcal{V}_{A_1}} X_{p+1}, X_1 \ominus P_{\mathcal{V}_{A_1}} X_1 \rangle}{||X_1 \ominus P_{\mathcal{V}_{A_1}} X_1||^2}$ since $P_{\mathcal{V}_{A_1}} X_{p+1} \perp X_1 \ominus P_{\mathcal{V}_{A_1}} X_1$. We show that c is related to the partial tail correlation between X_1 and X_{p+1}. To find the form of c, we note that the projection of any variable in $\mathcal{V}^q$ onto the space $\mathcal{V}_{A_1}$ is represented by the transformed-linear combination of the remaining variables $\{X_2, \cdots, X_p\}$. The projection of X_1 onto $\mathcal{V}_{A_1}$ is $P_{\mathcal{V}_{A_1}} X_1 = \bigoplus_{i=1}^{p-1} d_i \circ X_{i+1}$ and the projection of X_{p+1} onto $\mathcal{V}_{A_1}$ is $P_{\mathcal{V}_{A_1}} X_{p+1} = \bigoplus_{i=1}^{p-1} e_i \circ X_{i+1}$. Substituting these projections into (14), $\hat{X}_{p+1} = c \circ X_1 \oplus \big(\sum_{i=1}^{p-1} (d_i - c e_i) \circ X_{i+1} \big)$. By matching the coefficient c of X_1 in (14) with the b_1 of the best transformed-linear predictor $\hat{X}_{p+1} = b_1 \circ X_1 \oplus \cdots b_p \circ X_p$, the coefficient c can be expressed in terms of the inner product of residuals,

$$c = b_1 = \frac{\langle X_{p+1}, X_1 \ominus P_{\mathcal{V}_{A_1}} X_1 \rangle}{\|X_1 \ominus P_{\mathcal{V}_{A_1}} X_1\|^2_{\mathcal{V}^q}} = \frac{\langle X_{p+1} \ominus P_{\mathcal{V}_{A_1}} X_{p+1}, X_1 \ominus P_{\mathcal{V}_{A_1}} X_1 \rangle}{\|X_1 \ominus P_{\mathcal{V}_{A_1}} X_1\|^2_{\mathcal{V}^q}}.$$

Thus, if b_1 is zero, then the partial tail correlation between X_{p+1} and X_1 is zero. $\square$

Another way of understanding $P_{\mathcal{V}_{A_1^\perp}} X_{p+1}$ is through the regression setting. We consider a simple linear regression with no intercept $Y = X\beta + \epsilon$. The projection of Y onto the space spanned by X is given by $\hat{Y} = P_X Y = X\hat{\beta}$ where $\beta = (X^T X)^{-1} X^T y$. Note that $\hat{Y}$ can be expressed in terms of the inner products, we have $\hat{Y} = \frac{\langle X, y \rangle}{\langle X, X \rangle} X$. By replacing Y and X with X_{p+1} and $X_1 \ominus P_{\mathcal{V}_{A_1}} X_1$ respectively, $\hat{X}_{p+1} = P_{\mathcal{V}_{A_1}} X_{p+1} = \frac{\langle X_{p+1}, X_1 \ominus P_{\mathcal{V}_{A_1}} X_1 \rangle}{||X_1 \ominus P_{\mathcal{V}_{A_1}} X_1||^2} (X_1 \ominus P_{\mathcal{V}_{A_1}} X_1)$.

4.3 Relation between Partial Tail Correlation and the Inverse Inner Product Matrix

In non-extreme analysis of dependence, the precision matrix (the inverse of the covariance matrix) provides information about conditional relationships between variables. In the non-extreme setting, the partial correlation between X_i and X_j given all other elements of $\boldsymbol{X}_{\setminus(i,j)}$ is related to the (i,j)th element of the precision matrix. Specifically, $\Sigma_{ij}^{-1} = 0 \Leftrightarrow X_i \perp X_j | \boldsymbol{X}_{\setminus\{X_i, X_j\}}$. Analogously, we connect the idea of partial tail correlation in (12) to the inverse of the inner product matrix. The relation between the partial tail correlation and the inverse inner product matrix can be shown by matrix inversion.

Let $\boldsymbol{X}_p \in RV_p^+(2)$ be a p-dimensional regularly varying random vector where $X_i \in \mathcal{V}^q$, $i = 1, \ldots, p$. As in Section 4.1, partition $\boldsymbol{X}_p$ into two subvectors $\boldsymbol{X}^{(1)} := (X_i, X_j)^T$ and $\boldsymbol{X}^{(2)} := \boldsymbol{X}^T_{\setminus(i,j)}$. Recall the block form of the inner product matrix of $\boldsymbol{X}_p = (\boldsymbol{X}^{(1)^T}, \boldsymbol{X}^{(2)^T})^T$

$$\Gamma_{\boldsymbol{X}_p} := \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}, \tag{15}$$

By the matrix inversion in block form,

$$\Gamma^{-1}_{\boldsymbol{X}_p} = \begin{bmatrix} \Gamma^{-1}_{1|2} & -\Gamma^{-1}_{1|2}\Gamma_{12}\Gamma^{-1}_{22} \\ -\Gamma^{-1}_{22}\Gamma_{21}\Gamma^{-1}_{1|2} & \Gamma^{-1}_{22} + \Gamma^{-1}_{22}\Gamma_{21}\Gamma^{-1}_{1|2}\Gamma_{12}\Gamma^{-1}_{22}, \end{bmatrix} \tag{16}$$

where $\Gamma_{1|2}^{-1} \in \mathbb{R}^{2\times 2}$ is the inverse of the *conditional* IPM. Note that the matrix $\Gamma_{1|2}^{-1}$ could have negative off-diagonal elements. Since both of $\Gamma_{1|2}$ and $\Gamma_{1|2}^{-1}$ are a 2 by 2 matrix, the relation between $\Gamma_{1|2}$ and $\Gamma_{1|2}^{-1}$ can be readily shown by the inversion formula,

$$\Gamma_{1|2}^{-1} = \frac{1}{|\Gamma_{1|2}|}\begin{bmatrix}[\Gamma_{1|2}]_{22} & -[\Gamma_{1|2}]_{12} \\ -[\Gamma_{1|2}]_{21} & [\Gamma_{1|2}]_{11}\end{bmatrix}, \tag{17}$$

where $[\Gamma_{1|2}]_{i,j=1,2}$ is the element of $\Gamma_{1|2}$ and a determinant $|\Gamma_{1|2}| = [\Gamma_{1|2}]_{11}[\Gamma_{1|2}]_{22} - [\Gamma_{1|2}]_{12}[\Gamma_{1|2}]_{21}$. Thus

$$\rho_{i,j}^{E} = \frac{[\Gamma_{1|2}]_{12}}{\sqrt{[\Gamma_{1|2}]_{11}[\Gamma_{1|2}]_{22}}} = \frac{-[\Gamma_{1|2}^{-1}]_{12}}{\sqrt{[\Gamma_{1|2}^{-1}]_{11}[\Gamma_{1|2}^{-1}]_{22}}}. \tag{18}$$

Hence, the partial tail correlation between X_i and X_j given $\boldsymbol{X}^{(2)}$ can be represented by the first block matrix of the inverse IPM. Note that the direction of the partial tail correlation is of the opposite sign of $[\Gamma_{1|2}^{-1}]_{12}$. If $[\Gamma_{1|2}^{-1}]_{12} = 0$ it implies that X_i and X_j given $\boldsymbol{X}_{\setminus(i,j)}$ are partially uncorrelated in terms of tail behavior.

We can also consider the case where we predict one variable $X_{p+1} \in \mathcal{V}^q$ conditioned on $\boldsymbol{X}_p \in RV_+^p(2)$. Similarly, let $\Gamma_{(X_{p+1},\boldsymbol{X}_p^T)^T}$ be partitioned in block form,

$$\Gamma_{(X_{p+1},\boldsymbol{X}_p^T)^T} = \begin{bmatrix}\Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22}\end{bmatrix}, \tag{19}$$

where $\Gamma_{11} \in \mathbb{R}$ is a scale of X_{p+1} and $\Gamma_{22} \in \mathbb{R}^{p\times p}$ is the IPM of $\boldsymbol{X}_p$. Suppose Γ_{22} is invertible, then by the inverse formula,

$$\Gamma_{(X_{p+1},\boldsymbol{X}_p^T)^T}^{-1} = \begin{bmatrix}\frac{1}{k} & -\frac{1}{k}\Gamma_{12}\Gamma_{22}^{-1} \\ -\frac{1}{k}\Gamma_{22}^{-1}\Gamma_{21} & (\Gamma_{22} - \Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12})^{-1}\end{bmatrix} = \begin{bmatrix}\frac{1}{k} & -\frac{1}{k}\boldsymbol{b}^T \\ -\frac{1}{k}\boldsymbol{b} & (\Gamma_{22} - \Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12})^{-1}\end{bmatrix} \tag{20}$$

where $k = \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21} \in \mathbb{R}$. Thus, the inverse IPM can be expressed in terms of the vector $\boldsymbol{b} = \Gamma_{22}^{-1}\Gamma_{21}$, and we see that if the element of Γ^{-1} relating X_{p+1} to X_i, $\Gamma_{1,i+1}^{-1}$, equals zero, then $b_i = 0$.

For illustration, we now consider the transformed-linear model

$$\begin{bmatrix}X_1 \\ X_2 \\ X_3 \\ X_4\end{bmatrix} = \begin{bmatrix}1 & 0 & 0 & 0 \\ \phi & 1 & 0 & 0 \\ \phi^2 & \phi & 1 & 0 \\ \phi^3 & \phi^2 & \phi & 1\end{bmatrix} \circ \begin{bmatrix}Z_1 \\ Z_2 \\ Z_3 \\ Z_4\end{bmatrix} \tag{21}$$

where $\{Z_i\}$ is a sequence of independent regularly varying $\alpha = 2$ with unit scale. We set $\phi \in (0,1)$ to induce a positive dependence in the $\{X_i\}$. The sequential generating equation is

$$X_i = \phi \circ X_{i-1} \oplus Z_i, \text{ for } i = 1,2,3,4,$$

where $X_0 = 0$ $a.s.$

By the matrix inversion in (20), the inverse IPM shows a sparse pattern.

$$\Sigma_{(X_{p+1},\boldsymbol{X}_p^T)^T}^{-1} = \begin{bmatrix}1 & -\phi & 0 & 0 \\ -\phi & 1+\phi^2 & -\phi & 0 \\ 0 & -\phi & 1+\phi^2 & -\phi \\ 0 & 0 & -\phi & 1\end{bmatrix}, \tag{22}$$

The partial tail correlation between X_i and X_{i-k} is zero for $|i| > 1$. In terms of transformed linear prediction, consider $\hat{X}_4 = \boldsymbol{b}^T \circ \boldsymbol{X}_3$ where $\boldsymbol{b} = \Gamma_{22}^{-1}\Gamma_{21} = (0,0,\phi)^T$. These optimized weights imply that given X_3, knowledge of X_1 or X_2 does not provide additional information about X_4 in terms of tail behaviors.

5 Positive Subset $\mathcal{V}_+^q$ as a Modeling Framework

In the vector space $\mathcal{V}^q$ in (5), we essentially require that the elements of the coefficient vectors $\boldsymbol{a}$ take negative values for $\mathcal{V}^q$ to be a vector space. However, we note that $X = \boldsymbol{a} \circ \boldsymbol{Z} \in \mathcal{V}^q$ and $X^+ = \boldsymbol{a}^{(0)} \circ \boldsymbol{Z}$ have the same tail ratio because negative values in $\boldsymbol{a}$ do not contribute to tail behaviors. Thus, $\boldsymbol{X}_p = A \circ \boldsymbol{Z}_q$ and $\boldsymbol{X}_p^+ = A^{(0)} \circ \boldsymbol{Z}$ are indistinguishable in terms of their tail behavior because they both have the same angular measure: $H_{\boldsymbol{X}} = H_{\boldsymbol{X}_p^+} = \sum_{j=1}^q \|a_j^{(0)}\|^\alpha \delta_{a_j^{(0)}/\|a_j^{(0)}\|}(\cdot)$.

In terms of modeling and inference, it is reasonable to restrict attention to the positive subset $\mathcal{V}_+^q = \{X; X = \boldsymbol{a}^T \circ \boldsymbol{Z} = a_1 \circ Z_1 \oplus \cdots \oplus a_q \circ Z_q\}$, where $a_j \in [0, \infty)$, $j = 1, \ldots, q$, and $\boldsymbol{Z} = (Z_1, \ldots Z_q)^T$ meeting a lower tail condition and having a common unit scale. Critically, if $\boldsymbol{X}_p = A \circ \boldsymbol{Z}$ where $A \geq 0$, then $\Sigma_{\boldsymbol{X}} = \Gamma_{\boldsymbol{X}} = AA^T$. The assumption $X_i \in \mathcal{V}_+^q$ for $i = 1, \ldots, p$ is really essential for modeling as the inner product matrix which forms the basis is not estimable as the coefficients which determine X_1 are not observable; however, the TPDM is estimable.

Perhaps most importantly, q is not identifiable, nor does it need to be in order to use the framework for modeling. In fact, we do not need to believe that our data arise from a linear combination of q regularly varying random variables. Even if we do not believe that our data are constructed by this linear combinations, our summary matrix or its inverse is still obtainable from the data and provides a useful information about extreme dependence between variables. The definition of the TPDM is not tied to $\mathcal{V}^q$, these pairwise dependence summaries can be estimated for any regularly varying random vector in $RV_+^p(2)$. If we are willing to make the modeling assumption that $X_i = \boldsymbol{a}_i^T \circ \boldsymbol{Z}_q \in \mathcal{V}_+^q$, for $i = 1, \ldots, p$, we then have all the tools that arise from this inner product space. This is not such a strong assumption since angular measures arising from $p \times q$ matrices A are dense in the class of angular measures for p-dimensional regularly varying random vectors as $q \to \infty$ [Cooley and Thibaud, 2019].

6 Hypothesis Testing for Zero Elements in the Inverse TPDM

6.1 Asymptotic Normality of TPDM Estimates

We aim to develop a hypothesis test for $H_0 : \rho_{ij}^E = 0$ versus $H_1 : \rho_{ij}^E \neq 0$. Towards that aim, we first review the asymptotic normality of the sample TPDM $\hat{\Sigma}$ using results for the extremal dependence measure by (Resnick [2004];Larsson and Resnick [2012]).

Let $\boldsymbol{X}_p \in RV_+^p(2)$ be a p-dimensional regularly varying random vector such that $nP(n^{-1/2}\boldsymbol{X}_p \in \cdot) \xrightarrow{\nu} \nu_{\boldsymbol{X}_p}(\cdot)$ and have the angular measure $H_{\boldsymbol{X}}$. Unlike Larsson and Resnick [2012], we do not require $H_{\boldsymbol{X}}$ to be a probability measure so that the scale of the components of $\boldsymbol{X}_p$ is retained in the angular measure. We can find the equivalent form of the extremal dependence measure given as $\sigma_{ij} = \lim_{x\to\infty} mE[W_i W_j | R > x]$, where $R = ||(X_i, X_j)||$, x is a high threshold, and $m = H_{\boldsymbol{X}_p}(\Theta_{p-1}^+)$ is the total mass of the angular measure by [Larsson and Resnick, 2012, Proposition4]. This equivalent form provides a natural estimator for σ_{ij}.

Let $\boldsymbol{x}_\ell = (x_{\ell,1}, \ldots, x_{\ell,p})^T$, $\ell = 1, \ldots, n$ be realizations of iid copies of $\mathbb{R}_+^p$-valued regularly varying vectors with the tail index $\alpha = 2$. Letting $r_\ell = \|\boldsymbol{x}_\ell\|$ and $\boldsymbol{w}_\ell = r_\ell^{-1}\boldsymbol{x}_\ell$, the natural estimator for σ_{ij} is

$$\hat{\sigma}_{ij}(n) = \frac{\hat{m}}{k} \sum_{\ell=1}^n w_{\ell,i} w_{\ell,j} \mathbb{I}[r_\ell > r_{(k)}], \tag{23}$$

where $\hat{m}$ is an estimate of $H_{\boldsymbol{X}_p}(\Theta_{p-1}^+)$, $k := k(n)$ is such that $\lim_{n\to\infty} k/n = 0$ as $k \to \infty$, and $r_{(k)}$ is the k^{th} upper order statistic in the sample of size n. If we preprocess the data to have a common unit scale, then m is identical to p. When the data are not preprocessed to have a common unit scale, an estimator for m is given as $\hat{m} = (r_{(k)}^2/n)k$ by Cooley and Thibaud [2019].

Asymptotic normality is shown for the estimator of the extremal dependence measure in the case of iid observations by Resnick [2004] and Larsson and Resnick [2012]. The asymptotic normality of $\hat{\sigma}_{ij}(n)$ is proven under the following condition. Let F be the distribution function of R and $\bar{F}$ be its tail probability.

$$\lim_{n\to\infty} \sqrt{k}\Big(\frac{n}{k} mE[W_i W_j \mathbb{I}[R/b(n/k) \geq t^{-\gamma}]] - E[mW_i W_j \frac{n}{k}\bar{F}(b(n/k)t^{-\gamma}]\Big) = 0, \tag{24}$$

holds locally uniformly for $t \in [0, \infty)$, and assume that $\tau^2 = Var(W_i W_j) > 0$. Larsson and Resnick [2012] notes that $\tau^2 = 0$ implies asymptotic independence, meaning that the rate factor $\sqrt{k}$ increases too slowly to obtain a non-degenerate limit. This condition implies that the dependence between $(W_{i,l}W_{j,l})$ and R_l must decay fast enough with n as R_l is conditioned to lie above $b(n/k)$. The condition (24) does not require the second-order regular variation condition and the use of the order statistic does not require to know the normalization $b(\cdot)$. Under the condition (24), the estimator $\hat{\sigma}_{ij}(n)$ is asymptotically normal by Larsson and Resnick [2012].

$$\sqrt{k}(\hat{\sigma}_{ij}(n) - mE[W_i W_j]) \sim N(0, \tau_{ij}^2),$$

where m is the total mass of $H_{\boldsymbol{X}}(\cdot)$ and $\tau_{ij}^2 = Var(\hat{\sigma}_{ij})$.

Following a construction method for $\boldsymbol{X}_p$ as in (4), that is, $\boldsymbol{X}_p = A \circ \boldsymbol{Z} \in RV_+^p(2)$, we can specify an explicit form of the variance τ_{ij}^2 in terms of the angular measure $H_{\boldsymbol{X}}$ being consistent with point masses. For $X_i = \boldsymbol{a}_i^T \circ \boldsymbol{Z}$ and $X_j = \boldsymbol{a}_j^T \circ \boldsymbol{Z}$ in $\mathcal{V}^q$, the (i,j)th element of $\Sigma_{A\circ \boldsymbol{Z}}$ is given by $\sigma_{ij} = \int_{\Theta_{p-1}^+} w_i w_j dH_{A\circ\boldsymbol{Z}}(\boldsymbol{w}) = \sum_{l=1}^q a_{il}^{(0)} a_{jl}^{(0)}$, where $H_{\boldsymbol{X}}(\cdot) = \sum_{l=1}^q ||\boldsymbol{a}_l^{(0)}||^2 \delta_{\boldsymbol{a}_l/||\boldsymbol{a}_l||}(\cdot)$. To find $\tau_{ij}^2 = Var(\hat{\sigma}_{ij})$, we first consider

$$Var(W_iW_j) = \int_{\Theta_{p-1}^+} (w_iw_j - E[W_iW_j])^2 dN_{\boldsymbol{X}}(\boldsymbol{w}) = E[W_i^2W_j^2] - E[W_iW_j]^2,$$

where $N_{\boldsymbol{X}}(\cdot) = m^{-1}H_{\boldsymbol{X}}(\cdot)$ indicates a probability measure, $E[W_iW_j] = \frac{1}{m}\sum_{l=1}^q a_{il}^{(0)}a_{jl}^{(0)}$, and $E[W_i^2W_j^2] = \frac{1}{m}\sum_{l=1}^q \frac{a_{il}^{(0)^2}}{||\boldsymbol{a}_l^{(0)}||}\frac{a_{jl}^{(0)^2}}{||\boldsymbol{a}_l^{(0)}||}$. We thus have $Var(\hat{\sigma}_{ij}) = Var\Big(\frac{m}{k}\sum_{l=1}^n W_{il}W_{jl}\mathbb{I}[R_l > R_{(k)}]\Big) = \frac{m^2}{k}Var[W_{i,1}W_{j,1}]$ since $(W_{i,l}, W_{j,l})$'s are i.i.d. We can obtain an estimate of $\widehat{Var}(\hat{\sigma}_{ij})$ by estimating $Var(W_{i1}W_{j1})$. To obtain $\widehat{Var}(W_{i1}W_{j1})$, a natural estimate for $E[W_iW_j]$ and $E[W_i^2W_j^2]$ are $\hat{E}[W_iW_j] = \frac{1}{k-1}\sum_{l=1}^k w_{i,l}w_{j,l}\mathbb{I}[R_l > R_{(k)}]$ and $\hat{E}[W_i^2W_j^2] = \frac{1}{k-1}\sum_{l=1}^k w_{il}^2 w_{jl}^2 \mathbb{I}[R_l > R_{(k)}]$, respectively.

6.2 Residuals and Asymptotic Normality of the Conditional Inner Product Matrix

The ultimate goal is to derive the asymptotic normality of the sample *conditional* inner product matrix. We assume that we observe iid copies of $\boldsymbol{X}_p$ whose elements are in $\mathcal{V}^q$, from which we obtain $\hat{\Sigma}$, an estimate of the TPDM. A straightforward estimator of the conditional inner product matrix is

$$\hat{\Gamma}_{1|2} = [\hat{\Sigma}_{11} - \hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}], \tag{25}$$

where $\hat{\Sigma}_{ij}$ for $i,j = 1,2$ are sample block matrices in (15). However, the distribution of $\hat{\Gamma}_{1|2}$ is not straightforward to obtain from (25).

As the partial tail correlation is tied to the inner product of prediction errors, it is natural to consider using the observed 'residuals' to understand the properties of the sample conditional inner product matrix. The prediction errors are in $\mathcal{V}^q$ and $\boldsymbol{X}^{(1)} \ominus \hat{\boldsymbol{X}} = (A^{(1)} - \boldsymbol{b}^T A^{(2)}) \circ \boldsymbol{Z}_q$. Thus

$$\Gamma_{1|2} = (A^{(1)} - \boldsymbol{b}^T A^{(2)})(A^{(1)} - \boldsymbol{b}^T A^{(2)})^T,$$

and note that $\Gamma_{1|2}$ is not necessarily completely positive. Unlike the original data where we can assume away the importance of any negative coefficients as described in Section 5, here negative coefficients are consequential. If we consider the TPDM of the prediction errors,

$$\Sigma_{\boldsymbol{X}^{(1)}\ominus\hat{\boldsymbol{X}}} = (A^{(1)} - \boldsymbol{b}^T A^{(2)})^{(0)}(A^{(1)} - \boldsymbol{b}^T A^{(2)})^{(0)^T} \neq \Gamma_{1|2}.$$

Furthermore, the order of the definition of the prediction errors matters as the scale of $\boldsymbol{X}^{(1)} \ominus \hat{\boldsymbol{X}}$ is $(A^{(1)} - \boldsymbol{b}^T A^{(2)})^{(0)}$, and this differs from the scale of $\hat{\boldsymbol{X}} \ominus \boldsymbol{X}^{(1)}$ which is $(\boldsymbol{b}^T A^{(2)} - A^{(1)})^{(0)}$.

As the conditional inner product matrix is not completely positive, the direct use of transformed-residuals is not suitable for estimation. Instead, we consider the preimages of the prediction errors in (4.1) to account for negative coefficients. Assuming Z_j, $j = 1, \ldots, q$, is independent and identically distributed regularly varying random variable with unit scale meeting lower tail condition $nP(Z_j \leq \exp(-kn^{1/2})) \to 0$ for any $k > 0$, we define the preimages of the transformed-residuals by

$$\boldsymbol{U} := t^{-1}(\boldsymbol{X}^{(1)} \ominus \hat{\boldsymbol{X}}) = (A^{(1)} - \boldsymbol{b}^T A^{(2)})t^{-1}(\boldsymbol{Z}_q)$$

which are not restricted to the positive orthant. Lemma A4 in the appendix in Cooley and Thibaud [2019] implies that $\boldsymbol{U} \in RV_2(2)$. Let $\boldsymbol{U} = (U_1, U_2)^T$ and continue to let T denote the polar coordinate transformation, $T(U_1, U_2) = (R, \boldsymbol{W})$, where $R = ||(U_1, U_2)||_2$ and $\boldsymbol{W} = (U_1/R, U_2/R)$. We can summarize the second-order behaviors of $\boldsymbol{U}$ with respect to the angular measure $H_{\boldsymbol{U}}$. We define

$$\sigma_{\boldsymbol{U}_{ij}} = \int_{\Theta_1} w_i w_j dH_{\boldsymbol{U}}(\boldsymbol{w}), \quad i,j = 1,2,$$

where $\Theta_1 = \{w \in \mathbb{R} : ||\boldsymbol{w}||_2 = 1\}$, and $H_{\boldsymbol{U}}$ is the angular measure of $\boldsymbol{U}$. Thus, the pairwise tail dependence matrix of $\boldsymbol{U}$ is $\Sigma_{\boldsymbol{U}} := (A^{(1)} - \boldsymbol{b}^T A^{(2)})(A^{(1)} - \boldsymbol{b}^T A^{(2)})^T$ identical to $\Gamma_{1|2}$. In contrast to the fact that $\sigma_{\boldsymbol{X}_{12}} = 0$

implies asymptotic independence of X_1 and X_2, $\sigma_{\boldsymbol{U}_{12}} = 0$ does not necessarily mean that elements U_1 and U_2 are asymptotically independent. Instead this implies

$$\int_{\Theta_1: w_1 w_2 > 0} w_1 w_2 dH_{\boldsymbol{U}}(\boldsymbol{w}) = \int_{\Theta_1: w_1 w_2 < 0} w_1 w_2 dH_{\boldsymbol{U}}(\boldsymbol{w}),$$

meaning that angular components of (w_1, w_2) are uncorrelated on the L_2 unit ball and quadrants of (w_1, w_2) plane are balanced.

The off-diagonal element $\sigma_{\boldsymbol{U}_{12}}$ in $\Gamma_{1|2}$ is of primary interest because it is tied to the idea of the partial tail correlation ρ^E_{ij}. Following similar steps above, let $\boldsymbol{U}_\ell = (U_{\ell,1}, U_{\ell,2})$, $\ell = 1, \dots, n$, be iid copies of $\boldsymbol{U} \in RV_2(2)$. We set $R_\ell = ||\boldsymbol{U}_\ell||_2$, $\boldsymbol{W}_\ell = (U_{\ell,1}/R_\ell, U_{\ell,2}/R_\ell)$, and $k(n) = \sum_{\ell=1}^n \mathbb{I}[R_\ell > R_{(k)}]$, is the number of exceedances over the k^{th} upper order statistic.

Since $\boldsymbol{U}$ is a linear combination of independent Z_j's, its angular measure is discrete and σ^u_{12} has a simple form. Let $A^{(1)} - \boldsymbol{b}^T A^{(2)} := \boldsymbol{C} = (\boldsymbol{c}_1^T, \cdots, \boldsymbol{c}_q^T)^T \in \mathbb{R}^{2\times q}$. The $(1,2)$ element of $\Sigma_{\boldsymbol{U}}$ is

$$\sigma^u_{12} = \int_{\Theta_1} w_1 w_2 dH_{\boldsymbol{U}}(\boldsymbol{w}) = \sum_{i=1}^q c_{1i} c_{2i},$$

where $H_{\boldsymbol{U}}(\cdot) = \sum_{j=1}^q ||\boldsymbol{c}_j||^2 \delta_{\boldsymbol{c}_j/||\boldsymbol{c}_j||}(\cdot)$ and $\delta(\cdot)$ is a Dirac mass function.

The natural estimator for $\sigma^u_{12} = \int_{\Theta_1} w_1 w_2 dH_{\boldsymbol{U}}(\boldsymbol{w}) = \tilde{m} \int_{\Theta_1} w_1 w_2 dN_{\boldsymbol{U}}(\boldsymbol{w})$ is given by

$$\hat{\sigma}^u_{12,n} = \tilde{m} \int_{\Theta_1} w_1 w_2 d\hat{N}_{\boldsymbol{U}}(\boldsymbol{w}) = \frac{\tilde{m}}{k} \sum_{l=1}^n w_{1l} w_{2l} \mathbb{I}[R_l > R_{(k)}], \tag{26}$$

where $\tilde{m}$ is the total mass of the angular measure $H_{\boldsymbol{U}}(\cdot)$ and $N_{\boldsymbol{U}}(\cdot) = \tilde{m}^{-1} H_{\boldsymbol{U}}(\cdot)$ is a probability measure. $k = \sum_{\ell=1}^n \mathbb{I}[R_\ell > R_{(k)}]$ is such that $\lim_{n\to\infty} k/n = 0$ as $k \to \infty$ and $R_{(k)}$ is the k^{th} upper order statistic in the sample of size n.

Under the condition (24), the estimator $\hat{\sigma}^u_{12,n}$ is asymptotically normal by Larsson and Resnick [2012],

$$\sqrt{k}(\hat{\sigma}^u_{12,n} - \tilde{m} E[W_{1,1} W_{2,1}]) \sim N(0, \tau^{u^2}),$$

where $\tilde{m}$ is the total mass of the angular measure $H_{\boldsymbol{U}}$ identical to the sum of diagonal elements of the conditional TPDM $\Sigma_{1|2}$. To obtain $\tau^{u^2} = \tilde{m}^2 Var(W_1 W_2)$, we first consider $Var(W_1 W_2)$,

$$Var(W_1 W_2) = \int_{\Theta_1} (w_1 w_2 - E[W_1 W_2])^2 dN_{\boldsymbol{U}}(\boldsymbol{w}),$$

where $N_{\boldsymbol{U}}(\cdot) = \tilde{m}^{-1} H_{\boldsymbol{U}}(\cdot)$.

$$\begin{aligned} Var(\hat{\sigma}^u_{12,n}) &= Var\Big(\frac{\tilde{m}}{k} \sum_{l=1}^n W_{1l} W_{2l} \mathbb{I}[R_l > R_{(k)}]\Big) \\ &= \frac{\tilde{m}^2}{k} Var[W_{1,1} W_{2,1}] \quad \text{since} \quad (W_{1,l}, W_{2,l})\text{'s are iid} \end{aligned}$$

Our estimate $\hat{\tau}^{u^2}$ for $\tilde{m}^2 Var(\sigma^u_{12,n})$ is obtained in the same manner as above. Under the null hypothesis $H_0 : \rho^E_{ij} = 0$, since $\sqrt{k}(\hat{\sigma}^u_{12,n} - \sigma^u_{12}) \sim N(0, \tau^{u^2})$, we have

$$\frac{\hat{\sigma}^u_{12,n}}{\sqrt{\hat{\tau}^{u^2}/k}} \sim T_{k-1}, \tag{27}$$

where T_{k-1} denotes a t-distribution with $k-1$ degrees of freedom. With the asymptotic result, we can construct confidence intervals and perform a hypothesis test for zero elements in the inverse TPDM.

6.3 Asymptotic Normality for the Transformed-linear Extreme Illustrative Model

We use a simulation study to illustrate asymptotic normality for the sample conditional inner product matrix and perform a hypothesis test for zero elements in its inverse. We again consider the four-dimensional transformed-linear extreme model in (4.3) with generating equation

$$X_i = \phi \circ X_{i-1} \oplus Z_i, \quad i = 1, 2, 3, 4,$$

where $\{Z_i\}$ is a sequence of independent regularly varying random variables meeting lower tail condition, $\phi \in (0,1)$ and $X_0 = 0 \; a.s.$

Our simulation study aims to estimate the partial correlation between X_2 and X_4 given X_1 and X_3, and to test whether this is significantly different from zero. We set $\phi = 0.7$ and generate $n = 10,000$ four dimensional vectors $\boldsymbol{X}_4$. The largest 2% of the samples is used to find the estimated TPDM $\hat{\Sigma}_{\boldsymbol{X}_4}$.

Let $\boldsymbol{X}^{(1)} = (X_2, X_4)^T$ and $\boldsymbol{X}^{(2)} = (X_1, X_3)^T$. We find $\hat{\boldsymbol{b}} = \hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}$ and subsequently find $\hat{\boldsymbol{X}}^{(1)}$. We then obtain two dimensional vectors of residuals $\boldsymbol{U} = t^{-1}(\boldsymbol{X}^{(1)}) - t^{-1}(\hat{\boldsymbol{X}}^{(1)})$. We have two methods for estimating the conditional inner product matrix. The first is to use the partitions of the estimated TPDM $\hat{\Gamma}_{1|2} = \hat{\Sigma}_{11} - \hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}$. The second is to estimate $\Gamma_{1|2}$ from the residuals. For this method, we use the largest 2% of angular components. We focus on the off diagonal element $[\Sigma_{1|2}]_{12}$. Figure 2 shows the comparison between the kernel density of estimates obtained from the residuals (solid line) and the kernel density from the partition of the TPDM (dashed line) under repeated simulations. The figure shows little difference in these methods, and we suggest using the estimate from the partition as this is immediately available from the estimated TPDM.

Importantly, Figure 2 indicates that the variance of the residuals does in fact capture the uncertainty in the estimates of the conditional inner product matrix. Following the procedure in Section 6.1, we obtain estimated variances for τ^{u^2}. From the equation in (27), we construct a 95% confidence interval for each iteration and achieve a coverage rate of 0.95.

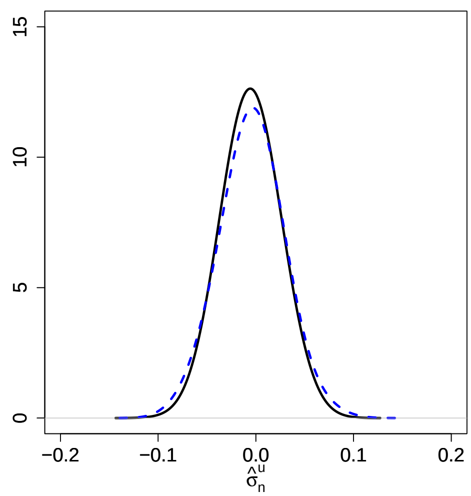

Figure 2: The kernel density based on the residuals (solid line) versus the kernel density from the partition of the TPDM (dashed line).

7 Application

7.1 Nitrogen Dioxide Air Pollution

We apply the idea of partial tail correlation to daily EPA NO_2 data from five stations in the Washington DC metropolitan area (see Figure 3). We analyze 5163 daily NO_2 data between 1995 and 2016 where all five stations have measurements. We follow the same preprocessing process by Lee and Cooley [2021] so that we can assume each variable $X_i \in RV_+^1(2)$ for $i = 1, \ldots, 5$. Let $X_i^{(orig)}$ denote the random variable for detrended NO_2 at the ith location. We apply the empirical CDF to perform marginal transformation $X_i = 1/\sqrt{1 - \hat{F}_i(X_i^{(orig)})} - \delta$ for each location so that X_i follows a 'shifted' Pareto distribution for $i = 1, \ldots 5$. Each $X_i \in RV_+^1(2)$ and the shift $\delta = 0.9352$ is such that $E[t^{-1}(X_i)] = 0$. This shift makes the preimages of the transformed data centered which helps reduce bias in the estimation of the TPDM [Lee and Cooley, 2021]. Assuming $\boldsymbol{X} = (X_1, \ldots, X_5)^T \in RV_+^5(2)$, we let $\boldsymbol{X}_t = (X_{1,t}, \ldots, X_{5,t})^T$ denote the random vector of the daily NO_2 level on day t and be iid copies of $\boldsymbol{X}$.

The goal is to test whether or not extreme NO_2 levels between each pair of stations exhibit partial tail correlation. The first step is to estimate the TPDM $\hat{\Sigma}_{\boldsymbol{X}}$. Let $\boldsymbol{x}_t$ denote the observed daily NO_2 level on day t. For each $i \neq j$, let $r_{t,ij} = \|(x_{t,i}, x_{t,j})\|$ and $(w_{t,i}, w_{t,j}) = (x_{t,i}, x_{t,j})/r_{t,ij}$. We let $\hat{\sigma}_{ij} = 2k^{-1}\sum_{i=1}^{n} w_{t,i}w_{t,j}\mathbb{I}[r_{t,ij} > r^*_{i,j}]$, where $k = \sum_{t=1}^{n}\mathbb{I}[r_{t,ij} > r^*_{ij}]$ is the number of exceedances. We set r^*_{ij} as the 0.95 quantile for radial components. Due to the pairwise estimation for the TPDM, the total mass 2 arises from the fact that each X_i has the unit scale after preprocessing. We can then calculate $\hat{\Sigma}^{-1}$, which is given in Table 1. The inverse TPDM includes small values close to zero. Our aim can now be described as trying to assess if each off-diagonal element is significantly different from 0.

Table 1: Inverse TPDM for all pairs of five stations

	1	2	3	4	5
1	2.10	-0.54	-0.19	-0.81	-0.23
2	-0.54	2.72	-1.14	-0.31	-0.58
3	-0.19	-1.14	2.28	-0.22	-0.38
4	-0.81	-0.31	-0.22	2.11	-0.47
5	-0.23	-0.58	-0.38	-0.47	2.01

For each $i \neq j$ for $i, j = 1, \ldots, 5$, let $\boldsymbol{X}_t^{(1)} = (X_{t,i}, X_{t,j})^T$ and $\boldsymbol{X}_t^{(2)} = (\boldsymbol{X}_t \setminus (X_{t,i}, X_{t,j}))^T$. Given the estimated TPDM $\hat{\Sigma}_{\boldsymbol{X}}$, we obtain $\hat{\boldsymbol{X}}_t = \hat{\boldsymbol{b}}^T \circ \boldsymbol{X}_t^{(2)} \in RV_+^2(2)$, where $\hat{\boldsymbol{b}} = \hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}$. After computing $\hat{\boldsymbol{X}}_t$ for all t, we take the top 5% of $\hat{\boldsymbol{X}}_t$ to find residual vectors $\boldsymbol{U}_t = t^{-1}(\boldsymbol{X}_t^{(1)}) - t^{-1}(\hat{\boldsymbol{X}}_t) = t^{-1}(\boldsymbol{X}_t^{(1)}) - \hat{\boldsymbol{b}}^T t^{-1}(\boldsymbol{X}_t^{(2)}) \in RV^2(2)$. Note that we suppress the index (i, j) in $\boldsymbol{U}_t$ for simplicity.

For each pair of $(X_i, X_j)^T$ given all other components, we estimate the off-diagonal element of the conditional TPDM $[\Sigma_{1|2}]_{12}$ and its variance. Let $r_{t,12} = \|(U_{t,1}, U_{t,2})\|$ and $(w_{t,1}, w_{t,2}) = (u_{t,1}, u_{t,2})/r_{t,12}$. We let $\hat{\sigma}^u_{12} = \tilde{m}^* k^{-1}\sum_{i=1}^{k} w_{t,1}w_{t,2}\mathbb{I}[r_{t,12} > r^*_{12}]$, where $k = \sum_{t=1}^{n}\mathbb{I}[r_{t,12} > r^*_{12}]$ and $\tilde{m}^*$ is an estimate for the total mass of $H_{\boldsymbol{U}}(\cdot)$. We choose r^*_{12} as the 0.98 quantile for radial components.

Under the null hypothesis that $\rho^E_{ij} = 0$, for each $i \neq j$, we calculate test statistics $t = \sqrt{k}(\hat{\sigma}^u_{12}/\hat{\tau}^u)$, where $\hat{\tau}^u$ is an estimate for $\tilde{m}\sqrt{Var(W_iW_j)}$. We employ the Tukey's exact procedure to adjust for multiple comparisons because the Tukey's exact procedure is well-suited for all pairwise comparisons where the number of exceedances is equal across all pairwise comparisons. We have the total number of observations $N = 103 \times 10 = 1030$ where each pairwise comparison has the equal number of threshold exccedances of 103 and there are 10 pairwise comparisons. The degrees of freedom is $df = N - 10 = 1020$. Having a critical value of $t_{crit} = 4.797$, we summarize test statistics in Table 2. If $|t| < 4.797$, then we fail to reject the null hypothesis that $\rho^E_{ij} = 0$.

Table 2: Test statistics for each pair of stations $i \neq j$ for $i, j = 1, \ldots, 5$

	1	2	3	4	5
1	-	1.69	1.69	2.37	9.89
2	1.69	-	6.18	7.83	3.27
3	1.69	6.18	–	2.42	4.50
4	2.37	7.83	2.42	-	5.31
5	9.89	3.27	4.50	5.31	-

We create an undirected graphical model for five stations given in Figure 3 by assuming partial tail correlation implies conditional independence. The extremal graph looks similar to a Markov chain. The thickness of lines is proportional to the test statistics, and describes the strength of the conditional relationship. The extremal graph from the partial tail correlation has 4 edges determined to be significant from the $\binom{5}{2} = 10$ possible edges.

7.2 Danube River Basin

We also employ the notion of partial tail correlation to investigate conditional relationships between the extremes of average daily river discharges in the upper Danube basin. The Danube is Europe's second largest river and the upper Danube extends from its source in Germany to Bratislava in Slovakia[2]. To assess flood risks caused by extreme river discharges, there are a number of gauging stations along the river and its tributaries. The main feature in the upper Danube basin is that there are physical flow connections among stations. This feature allows us to compare the estimated graphical structure to the known structure of the river network on the Danube. Figure 4 shows the river network in the upper Danube basin where the path $10 \to \cdots \to 1$ is the main channel and the 21 other locations are on tributaries.

[2]https://www.icpdr.org/main/danube-basin/river-basin

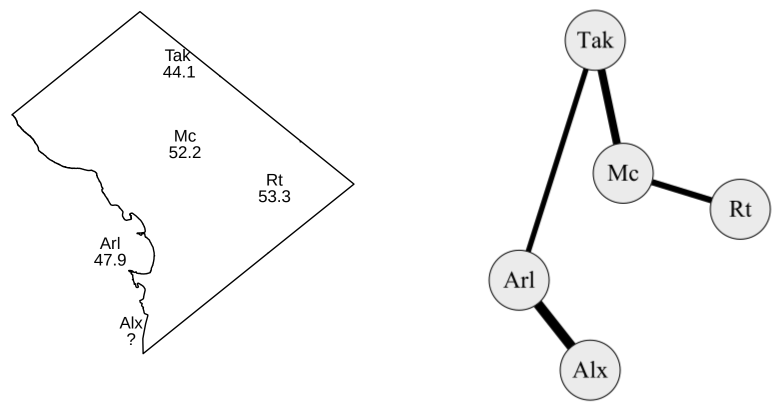

Figure 3: Left: An outline of Washington, D.C. with locations of five NO_2 monitors. Right: The extremal graph induced by partial tail tail correlation for five stations. The thickness of edges corresponds to absolute values of test statistics being greater than 4.797.

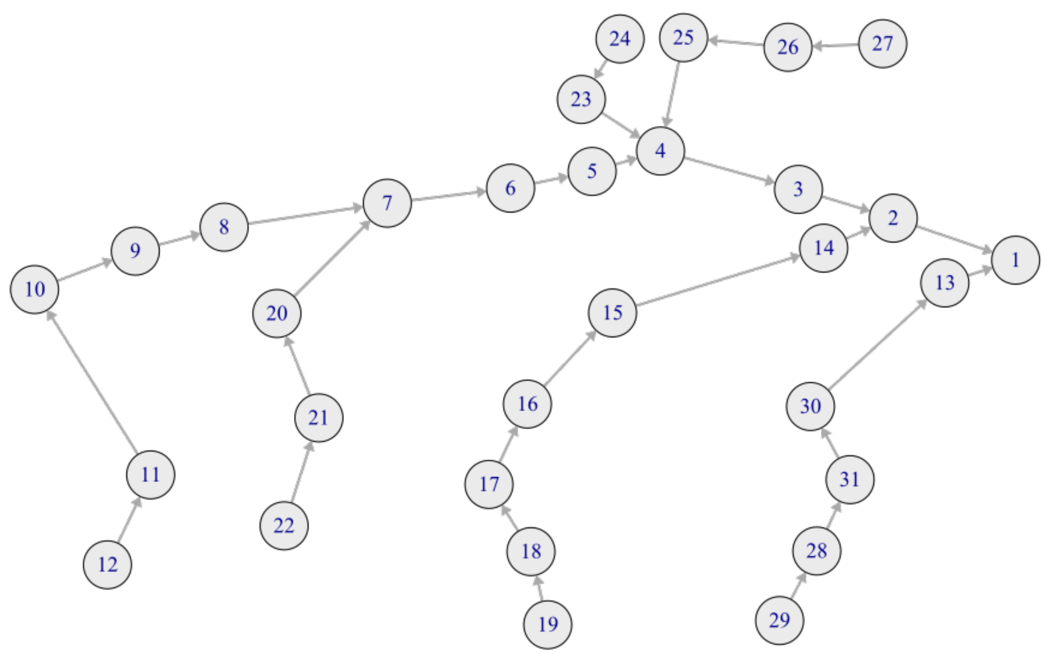

Figure 4: Physical flow connections in the upper Danube river basin

We analyze average daily river discharges from 31 gauging stations for 1960-2009. The data are available in the Bavarian Environmental Agency[3]. This data set has been widely used across multiple disciplines to assess flood risk. Asadi et al. [2015] used a spatial extremes model to fit data from these 31 stations. Engelke and Hitz [2020] fit an extremal undirected graphical model based on the Hüsler-Reiss model to this data.

We follow a similar preprocessing approach as Asadi et al. [2015] in order to compare results. Engelke and Hitz [2020] and Asadi et al. [2015] only considered June, July, and August because the main factor causing extreme flooding is extreme precipitation in these summer months. It results in $n = 50 \times 92 = 4600$ daily river discharges where all gauging stations have measurements. Focusing on the summer period helps remove seasonality. The overall trend in extreme river discharges on the Danube turns out to be insignificant by Katz et al. [2002]. Extreme discharges for each

[3] http://www.gkd.bayern.de

station occur in clusters because extreme discharges at downstream may occur a few days later from upstream stations. To remove temporal dependence, they set nonoverlapping timewindows of length $p = 9$ days and then take the largest value within each window, resulting in a declustered time series of $n = 428$ independent data from the original data. However, we decide to use the whole sample of size $n = 4600$ to get a large enough sample size by treating the data as independent samples.

We assume each $X_i \in RV_+^1(2)$ in our inner product space. Let $X_i^{(orig)}$ denote the random variable for average daily river discharges at the ith station for $i = 1, \ldots, 31$. For simplicity, we apply the empirical CDF to perform the marginal transformation $X_i = 1/\sqrt{(1 - \hat{F}_i(X_i^{(orig)}))} - \delta$ for each station so that X_i follows the shifted Pareto distribution. That is, each $X_i \in RV_+^1(2)$ and the shift $\delta = 0.9352$ is such that $E[t^{-1}(X_i)] = 0$. We assume $\boldsymbol{X} = (X_1, \ldots, X_{31})^T \in RV_+^{31}(2)$. We let $\boldsymbol{X}_t = (X_{1,t}, \ldots, X_{31,t})^T$ denote the random vector of the average daily river discharge on day t, which we treat as iid copies of $\boldsymbol{X}$.

We first investigate a sub-network for the stations on the main channel, $10 \to \cdots \to 1$. The physical flow connections look similar to a graphical model generated by an AR(1) model or the Markov chain. The goal is to test whether or not extreme discharges between each pair of stations exhibit partial tail correlation. To perform a hypothesis test for the partial tail correlation for each pair of stations, we first estimate the TPDM $\hat{\Sigma}_{\boldsymbol{X}}$. Let $\boldsymbol{x}_t$ denote the observed average daily discharge on day t. For each $i \neq j$, let $r_{t,ij} = \|(x_{t,i}, x_{t,j})\|$ and $(w_{t,i}, w_{t,j}) = (x_{t,i}, x_{t,j})/r_{t,ij}$. We let $\hat{\sigma}_{ij} = 2k^{-1} \sum_{i=1}^{k} w_{t,i} w_{t,j} \mathbb{I}[r_{t,ij} > r_{i,j}^*]$, where $k = \sum_{t=1}^{n} \mathbb{I}[r_{t,ij} > r_{ij}^*]$ is the number of exceedances. We set r_{ij}^* as the 0.95 quantile for radial components. We can then calculate $\hat{\Sigma}^{-1}$, which is given in Table 3. The goal is to assess if each off-diagonal element is significantly different from 0.

Table 3: Inverse TPDM for the main path.

	1	2	3	4	5	6	7	8	9	10
1	4.33	-5.41	4.62	-2.77	4.74	-4.32	-1.72	2.95	-1.41	-0.78
2	-5.41	29.24	-32.01	8.24	-4.45	3.00	2.17	-2.57	0.72	1.72
3	4.62	-32.01	46.69	-19.71	-5.91	7.00	-1.53	4.99	-4.34	0.31
4	-2.77	8.24	-19.71	39.78	-28.03	2.08	1.76	-4.24	1.15	1.57
5	4.74	-4.45	-5.91	-28.03	123.04	-101.71	11.22	6.83	0.78	-7.34
6	-4.32	3.00	7.00	2.08	-101.71	123.03	-29.01	-13.46	10.06	4.36
7	-1.72	2.17	-1.53	1.76	11.22	-29.01	25.30	-3.89	-2.96	-1.84
8	2.95	-2.57	4.99	-4.24	6.83	-13.46	-3.89	26.37	-22.79	6.03
9	-1.41	0.72	-4.34	1.15	0.78	10.06	-2.96	-22.79	39.47	-20.87
10	-0.78	1.72	0.31	1.57	-7.34	4.36	-1.84	6.03	-20.87	17.66

For each $i \neq j$ for $i, j = 1, \ldots, 10$, let $\boldsymbol{X}_t^{(1)} = (X_{t,i}, X_{t,j})^T$ and $\boldsymbol{X}_t^{(2)} = (\boldsymbol{X}_t \setminus (X_{t,i}, X_{t,j}))^T$. Given the estimated TPDM $\hat{\Sigma}_{\boldsymbol{X}}$, we obtain $\hat{\boldsymbol{X}}_t = \hat{\boldsymbol{b}}^T \circ \boldsymbol{X}_t^{(2)} \in RV_+^2(2)$, where $\hat{\boldsymbol{b}} = \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21}$ for all $t = 1, \ldots, 4600$. We only consider those for which $\hat{\boldsymbol{X}}_t$ exceeds the 0.98 quantile to find two dimensional residual vectors $\boldsymbol{U}_t = t^{-1}(\boldsymbol{X}_t^{(1)}) - t^{-1}(\hat{\boldsymbol{X}}_t) = t^{-1}(\boldsymbol{X}_t^{(1)}) - \hat{\boldsymbol{b}}^T t^{-1}(\boldsymbol{X}_t^{(2)}) \in RV^2(2)$. Note that we suppress the index (i, j) in $\boldsymbol{U}_t$ for simplicity.

Following the similar steps in NO_2 application, for each pair of $(X_i, X_j)^T$ given all other components, we estimate the off-diagonal element of the conditional TPDM $[\Sigma_{1|2}]_{12}$ and its variance. Let $r_{t,12} = \|(U_{t,1}, U_{t,2})\|$ and $(w_{t,1}, w_{t,2}) = (u_{t,1}, u_{t,2})/r_{t,12}$. We let $\hat{\sigma}_{12}^u = \tilde{m}^* k^{-1} \sum_{i=1}^{k} w_{t,1} w_{t,2} \mathbb{I}[r_{t,12} > r_{12}^*]$, where $k = \sum_{t=1}^{n} \mathbb{I}[r_{t,12} > r_{12}^*]$ and $\tilde{m}^*$ is an estimate for the total mass of $H_{\boldsymbol{U}}(\cdot)$. We choose r_{12}^* as the 0.98 quantile for radial components. For each pair of $(X_i, X_j)^T$ given all other components, we estimate the off-diagonal element of the conditional TPDM $[\Sigma_{1|2}]_{12}$ and its variance.

Under the null hypothesis that $\rho_{ij}^E = 0$, for each $i \neq j$, we calculate test statistics $t = \sqrt{k}(\hat{\sigma}_{12}^u / \hat{\tau}^u)$, where $\hat{\tau}^u$ is an estimate for $\tilde{m}\sqrt{Var(W_i W_j)}$. We employ the Tukey's exact procedure to adjust for multiple comparisons. We consider 10 nodes and 45 possible comparisons. The total number of observations is $N = 92 \times 45 = 4140$ where each pairwise comparison has the equal number of threshold exccedances of 92 and there are 45 pairwise comparisons. Hence, the degrees of freedom is $df = N - 45 = 4095$. Finding a critical value of $t_{crit} = 5.847$, we summarize test statistics in a matrix (4). If $|t| < 5.847$, then we fail to reject the null hypothesis that $\rho_{ij}^E = 0$.

Let $\mathcal{G} = (V, E)$ be an undirected graphic with a node set $V = \{1, \ldots, 10\}$ and its edge set E. Based on the test statistics in the table above, we create an undirected graphical model for the main path by assuming partial tail correlation implies conditional relationships in Figure 5. Each circle indicates a node. Extreme discharges at nearby stations tend to be

Table 4: Test statistics for each pair of stations $i \neq j$ for $i, j = 1, \ldots, 10$ in the main path.

	1	2	3	4	5	6	7	8	9	10
1	-	7.62	-3.46	3.12	-0.84	-0.35	4.77	-1.65	-0.18	1.29
2	7.62	-	21.40	-3.52	-1.02	0.43	-0.77	-1.24	-0.01	-0.52
3	-3.46	21.40	-	6.94	0.46	-0.53	-1.91	0.80	-0.06	0.81
4	3.12	-3.52	6.94	-	8.76	-1.66	0.85	3.51	-1.03	-1.73
5	-0.84	-1.02	0.46	8.76	-	16.93	-1.55	-0.88	0.78	-0.20
6	-0.35	0.43	-0.53	-1.66	16.93	-	8.92	-0.26	-0.80	-0.45
7	4.77	-0.77	-1.91	0.85	-1.55	8.92	-	2.82	2.00	0.06
8	-1.65	-1.24	0.80	3.51	-0.88	-0.26	2.82	-	9.62	-3.81
9	-0.18	-0.01	-0.06	-1.03	0.78	-0.80	2.00	9.62	-	15.21
10	1.29	-0.52	0.81	-1.73	-0.20	-0.45	0.06	-3.81	15.21	-

partially correlated in terms of tail behaviors. The thickness of lines is proportional to the test statistics, and describes the strength of the conditional relationship. Focusing on the main stream line, our extremal graph shows a resemblance to the physical flow connection. The graph from the partial tail correlation has 8 edges determined to be significant. The disconnection between a station 8 and 7 is not a big surprise because the actual distance between them is far away from each other.

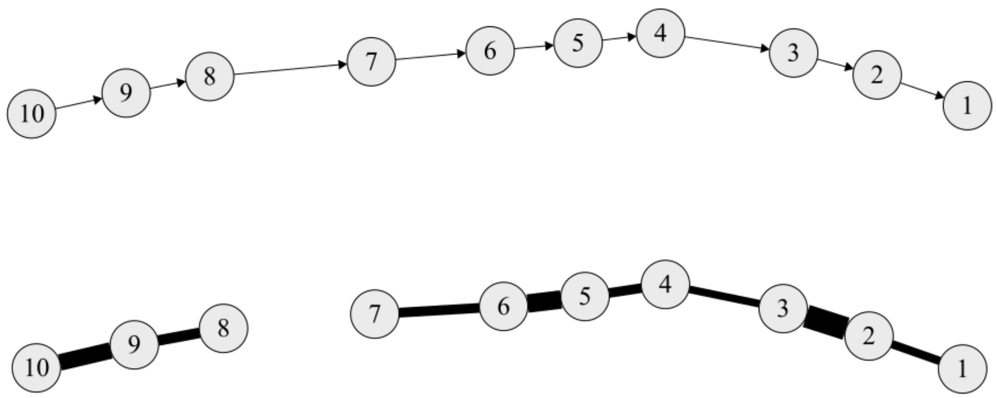

Figure 5: The known physical flow connections (above) versus the extremal graph induced by partial tail correlation (below) for the main path $10 \to \cdots \to 1$ in the upper Danube river basin. The thickness of edges corresponds to absolute values of test statistics being greater than 5.847.

We also investigate the whole river network in the upper Danube basin. Following the similar steps above, we standardize the off-diagonal element of the conditional TPDM for each pair of stations. We obtain the critical value of 7.189 via the Tukey's exact procedure. We then create an undirected graph on the whole river network. The extremal graph constructed from the partial tail correlation has 192 significant edges out of the 465 possible edges. The graph has a noteworthy reduction from all possible edges but is quite complicated to interpret.

The graph found by Engelke and Hitz [2020] was much more simple and closely resembled the physical flow network. However, their approach was much more model-based and used knowledge of the physical flow graph to perform stepwise model selection based on AIC. In recent work, Röttger et al. [2021] use a somewhat less model-based approach to fit a graphical model on this same Danube data and find a more connected network than their earlier estimate, but one which is still more simple than the one in Figure 5. We believe that including too many variables may introduce noise when estimating the TPDM.

8 Summary and Discussion

A vector space constructed from transformed linear combinations of independent regularly varying random variables provides a framework for linear prediction for extremes. We develop projection theorem as a natural way of defining partial tail correlation in the context of extremes. Similar to Gaussian cases, sparsity in the inverted TPDM can also be tied to the idea of the partial tail correlation. Without a distributional assumption, the notion of partial tail correlation can provide much less model-based approach for describing conditional relationships between variables at extreme levels.

Using the asymptotic normality result for the estimator of the conditional TPDM, we develop a hypothesis test for zero elements in the inverse extremal matrix. We observed the performance of conditional extremal relationships between variables in simulation study as well as in both applications: high NO_2 levels in Washington D.C. and extreme river discharges in the upper Danube river basin.

Especially, in the upper Daunbe river basin application, focusing on the main channel $10 \to \cdots \to 1$, one can explore further the oscillating pattern in the table of test statistics for each pair of stations. The study of this pattern may be tied to the idea of directed graphs. Furthermore, taking the spatial dependence into account can be another important factor when describing conditional relationships between variables at extreme levels.

Acknowledgements

We are aware that there has been independent and parallel work by Gong, Huser, and Opitz which also investigates partial tail correlation for extremes. Our understanding is that inference in their work is done from a perspective of model selection rather than hypothesis testing. This paper was completed during a Ph.D. at Colorado State University.

References

Havard Rue and Leonhard Held. *Gaussian Markov random fields: theory and applications*. Chapman and Hall/CRC, 2005.

Nadine Gissibl and Claudia Klüppelberg. Max-linear models on directed acyclic graphs. *Bernoulli*, 24(4A):2693–2720, 2018.

Sebastian Engelke and Adrien S Hitz. Graphical models for extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(4):871–932, 2020.

Jürg Hüsler and Rolf-Dieter Reiss. Maxima of normal random vectors: between independence and complete dependence. *Statistics & Probability Letters*, 7(4):283–286, 1989.

Jeongjin Lee and Daniel Cooley. Transformed-linear prediction for extremes. *arXiv preprint arXiv:2111.03754*, 2021.

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

Laurens De Haan and Ana Ferreira. *Extreme value theory: an introduction*. Springer Science & Business Media, 2007.

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

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

Jeongjin Lee. *Linear prediction and partial tail correlation for extremes*. PhD thesis, Colorado State University, 2022.

Hedegaard Anders Jessen and Thomas Mikosch. Regularly varying functions. *Publications de L'institut Mathematique*, 80(94):171–192, 2006.

Peter J Brockwell, Richard A Davis, and Stephen E Fienberg. *Time series: theory and methods: theory and methods*. Springer Science & Business Media, 1991.

Daren BH Cline. *Estimation and linear prediction for regression, autoregression and ARMA with infinite variance data*. PhD thesis, Colorado State University, 1983.

Sidney Resnick. The extremal dependence measure and asymptotic independence. 2004.

Peiman Asadi, Anthony C Davison, and Sebastian Engelke. Extremes on river networks. *The Annals of Applied Statistics*, 9(4):2023–2050, 2015.

Richard W Katz, Marc B Parlange, and Philippe Naveau. Statistics of extremes in hydrology. *Advances in water resources*, 25(8-12):1287–1304, 2002.

Frank Röttger, Sebastian Engelke, and Piotr Zwiernik. Total positivity in multivariate extremes, 2021. URL `https://arxiv.org/abs/2112.14727`.

A Projection Theorem

Theorem A.1. *(Projection theorem) Let $\mathcal{V}_A$ be the previously defined subspace of the Hilbert space $\mathcal{V}^q$ and $X \in \mathcal{V}^q$. Let $X_i = \sum_{j=1}^q a_{ij} \circ Z_j \in \mathcal{V}^q$, $i = 1, \dots, p$, and let $X = \sum_{j=1}^q a_j^* \circ Z_j \in \mathcal{V}^q$. Then*
i) $\hat{X} := P_{\mathcal{V}_A} X$ ($\hat{X}$ is the projection of X onto $\mathcal{V}_A$) has a unique element in $\mathcal{V}_A$ such that

$$||X \ominus \hat{X}||_{\mathcal{V}^q} = \inf_{Y \in \mathcal{V}_A} ||X \ominus Y||_{\mathcal{V}^q}, \text{ and}$$

ii) $\hat{X} \in \mathcal{V}_A$ such that $||X \ominus \hat{X}||_{\mathcal{V}^q} = \inf_{Y \in \mathcal{V}_A} ||X \ominus Y||_{\mathcal{V}^q}$ if and only if $\hat{X} \in \mathcal{V}_A$ and $(X \ominus \hat{X}) \in \mathcal{V}_A^{\perp}$.

Proof. i) Consider $X_i = \sum_{j=1}^q a_{ij} \circ Z_j$, $i = 1, \dots, p$, and $X = \sum_{j=1}^q a_j^* \circ Z_j$ in $\mathcal{V}^q$. For $\boldsymbol{X}_p = (X_1, \dots, X_p)^T$, consider $\boldsymbol{b}^T \circ \boldsymbol{X}_p \in \mathcal{V}_A$. $||X \ominus (\boldsymbol{b}^T \circ \boldsymbol{X}_p)||^2_{\mathcal{V}^q} = \sum_{j=1}^q (a_j^* - \boldsymbol{b}^T \boldsymbol{a}_{\cdot j})^2$ where $\boldsymbol{a}_{\cdot j}$ is the j^{th} column vector of $\boldsymbol{A}$. We assume Rank$(\boldsymbol{A}) = p$. Let $S_j = \{\boldsymbol{b} \in \mathbb{R}^p \text{ such that } \boldsymbol{b}^T \boldsymbol{a}_{\cdot j} = a^*\}$ and $f_j(\boldsymbol{b}) = (a_j^* - \boldsymbol{b}^T \boldsymbol{a}_{\cdot j})^2$. For $\boldsymbol{b} \notin S_j$, $\frac{\partial f_j(\boldsymbol{b})}{\partial \boldsymbol{b}} = 2\boldsymbol{a}_{\cdot j}[\boldsymbol{b}^T \boldsymbol{a}_{\cdot j} - a_j^*]$ and $\frac{\partial^2 f_j(\boldsymbol{b})}{\partial \boldsymbol{b} \partial \boldsymbol{b}^T} = 2\boldsymbol{a}_{\cdot j}\boldsymbol{a}_{\cdot j}^T(\boldsymbol{b}\boldsymbol{a}_{\cdot j} - a_j^*)$. As $\boldsymbol{a}_{\cdot j}\boldsymbol{a}_{\cdot j}^T$ is nonnegative definite, f_j is convex off of S_j. Since f_j is minimized on S_j, f_j is convex everywhere. Thus for $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ and any $w \in (0,1)$,

$$w f_j(\boldsymbol{b}_1) + (1-w) f_j(\boldsymbol{b}_2) \geq f_j(w\boldsymbol{b}_1 + (1-w)\boldsymbol{b}_2),$$

where equality above implies $\boldsymbol{b}_1^T \boldsymbol{a}_{\cdot j} = \boldsymbol{b}_2^T \boldsymbol{a}_{\cdot j}$. Equality does not hold for every j. $||X \ominus (\boldsymbol{b}^T \circ \boldsymbol{X}_p)||^2_{\mathcal{V}^q} = \sum_{j=1}^q f_j$ is strictly convex since $\boldsymbol{A}$ is full rank. $||X \ominus (\boldsymbol{b}^T \circ \boldsymbol{X}_p)||_{\mathcal{V}^q} \to \infty$ as $\max_{1 \leq j \leq p} |a_j^*| \to \infty$. Thus, $||X \ominus (\boldsymbol{b}^T \circ \boldsymbol{X}_p)||_{\mathcal{V}^q}$ must have a unique minimum.
ii) Suppose $\hat{X} \in \mathcal{V}_A$ and $(X \ominus \hat{X}) \in \mathcal{V}_A^{\perp}$. For any $Y \in \mathcal{V}_A$,

$$\begin{aligned} ||X \ominus Y||^2_{\mathcal{V}^q} &= \langle (X \ominus \hat{X}) \oplus (\hat{X} \ominus Y), (X \ominus \hat{X}) \oplus (\hat{X} \ominus Y) \rangle \\ &= ||X \ominus \hat{X}||^2_{\mathcal{V}^q} + ||\hat{X} \ominus Y||^2_{\mathcal{V}^q} \\ &\geq ||X \ominus \hat{X}||^2_{\mathcal{V}^q}, \end{aligned}$$

with equality iff $Y = \hat{X}$. Thus, $\hat{X}$ is such that $||X \ominus \hat{X}||_{\mathcal{V}^q} = \inf_{Y \in \mathcal{V}_A} ||X \ominus Y||_{\mathcal{V}^q}$.

Conversely if $\hat{X} \in \mathcal{V}_A$ and $(X \ominus \hat{X}) \notin \mathcal{V}_A$, then $\hat{X}$ is not the element of $\mathcal{V}_A$ closest to X since there exists

$$\tilde{X} = \hat{X} \oplus a \circ Y / ||Y||^2_{\mathcal{V}^q}$$

closer to X where Y is any element of $\mathcal{V}^q$ such that $\langle X \ominus \hat{X}, Y \rangle \neq 0$ and $a = \langle X \ominus \hat{X}, Y \rangle$.

$$\begin{aligned} ||X \ominus \tilde{X}||^2_{\mathcal{V}^q} &= \langle X \ominus \hat{X} \oplus \hat{X} \ominus \tilde{X}, X \ominus \hat{X} \oplus \hat{X} \ominus \tilde{X} \rangle \\ &= ||X \ominus \hat{X}||^2_{\mathcal{V}^q} + a^2 \circ \frac{1}{||Y||^2_{\mathcal{V}^q}} + 2\langle X \ominus \hat{X}, \hat{X} \ominus \tilde{X} \rangle \\ &= ||X \ominus \hat{X}||^2_{\mathcal{V}^q} - a^2 \circ \frac{1}{||Y||^2_{\mathcal{V}^q}} \\ &< ||X \ominus \hat{X}||^2_{\mathcal{V}^q}. \end{aligned}$$

□

B Property of Projection Mappings

Proposition B.1. *(Property of Projection Mappings) Let $P_{\mathcal{V}_A}$ be the projection mapping of $\mathcal{V}^q$ onto a subspace $\mathcal{V}_A$. Then,*
i) $P_{\mathcal{V}_A}(\alpha \circ X \oplus \beta \circ Y) = \alpha \circ P_{\mathcal{V}_A} X \oplus \beta \circ P_{\mathcal{V}_A} Y, \quad X, Y \in \mathcal{V}^q, \quad \alpha, \beta \in \mathbb{R}$.

[That is, the projection mapping $P_{\mathcal{V}_A}$ is a linear mapping.]
ii) For every $X \in \mathcal{V}^q$, there exist an element of $\mathcal{V}_A$ and an element of $\mathcal{V}_A^{\perp}$ such that

$$X = P_{\mathcal{V}_A} X \oplus (I - P_{\mathcal{V}_A})X$$

and this decomposition is unique.

Proof. i) $(\alpha \circ P_{\mathcal{V}_A} X) \oplus (\beta \circ P_{\mathcal{V}_A} Y) \in \mathcal{V}_A$ since $\mathcal{V}_A$ is a linear subspace of $\mathcal{V}^q$. In addition,

$$\alpha \circ X \oplus \beta \circ Y \ominus (\alpha \circ P_{\mathcal{V}_A} X \oplus \beta \circ P_{\mathcal{V}_A} Y) = \alpha \circ (X \ominus P_{\mathcal{V}_A} X) \oplus \beta \circ (Y \ominus P_{\mathcal{V}_A} Y) \in \mathcal{V}_A^{\perp}$$

since $\mathcal{V}_A^{\perp}$ is a linear subspace of $\mathcal{V}^q$. Thus, these two properties indicate $\alpha \circ P_{\mathcal{V}_A} X \oplus \beta \circ P_{\mathcal{V}_A} Y$ is the projection of $P_{\mathcal{V}_A}(\alpha \circ X \oplus \beta \circ Y)$. We note that this linear mapping is not necessarily true when $\alpha \neq 2$.
ii). To show uniqueness of decomposition, let $X = Y \oplus Z$, $Y \in \mathcal{V}_A$, $Z \in \mathcal{V}_A^{\perp}$ be another decomposition, then

$$Y \ominus P_{\mathcal{V}_A} X \oplus Z \ominus (I - P_{\mathcal{V}_A})X = 0.$$

By taking inner products of each side with $Y \ominus P_{\mathcal{V}_A}$, $||Y \ominus P_{\mathcal{V}_A} X||^2_{\mathcal{V}^q} = 0$ since $Z \ominus (I - P_{\mathcal{V}_A})X \in \mathcal{V}_A^{\perp}$. Hence $Y = P_{\mathcal{V}_A} X$ and $Z = (I - P_{\mathcal{V}_A})X$. □